
Economic Theory and Empirics: Selected Economic Papers

By Heng-fu Zou

Volume.3

May, 2012

Theory

- 1, Gong, Liutang & William Smith & Heng-fu Zou, 2007, Consumption and Risk with Hyperbolic Discounting, *Economics Letters*, 96(2), 153-160.
- 2, Gong, Liutang & William Smith & Heng-fu Zou, 2007, Asset Prices and Hyperbolic Discounting, *Annals of Economics and Finance*, 8, 397-414.
- 3, Luo Yulei & William Smith & Heng-fu Zou, 2009, The Spirit of Capitalism and Excess Smoothness, *Annals of Finance and Economics*, 10, 281-301.
- 4, Luo Yulei & William Smith & Heng-fu Zou, 2009, The Spirit of Capitalism and Excess Smoothness, *Journal of Money, Credit and Banking*, 41, 543-554.
- 5, Gong, Liutang & Xiaojun Zhao & Yunhong Yang & Heng-fu Zou, 2010, Stochastic Growth with Social-status Concern: The Existence of a Unique Stable Distribution, *Journal of Mathematical Economics*, 46, 505-518.
- 6, Luo Yulei & Liutang Gong & Heng-fu Zou, 2010, A Note on Entrepreneurial Risk, Capital Market Imperfections, and Heterogeneity, *Macroeconomic Dynamics*, 14, 269-284.
- 7, Gong Liutang & Hongyi Li & Dihai Wang & Heng-fu Zou, 2010, Health, Taxes and Growth, 11, 73-94.
- 8, Gong Liutang & Heng-fu Zou, 2011, Public Expenditure, Taxes, Federal Transfers, and Endogeneous Growth, *Journal of Public Economic Theory* 13, 973-991.
- 9, Gong Liutang & Yulei Luo & Heng-fu Zou, 2009, Social Status and the Term Structure of Interest Rates in Stochastic Production Economies, in *Economic Dynamics: Theory, Games and Empirical Studies*, Nova Science Publisher.
- 10, Wang Gaowang & Heng-fu Zou, 2011, Inflation Aversion and Macroeconomic Policy in a Perfect Foresight Monetary Model 28, 1802-1807.
- 11, Zou Heng-fu & Liutang Gong & Xinsheng Zeng, 2011, Inflation Aversion, *Annals of Finance and Economics* 12, 1-11.
- 12, Wang Gaowang & Heng-fu Zou, 2011, Inflation Aversion and the Optimal Inflation Tax 12, 13-30.

Empirics

- 13, Jin Jing & Heng-fu Zou, 2005, Fiscal Decentralization, Revenue and Expenditure Assignments and Growth in China, *Journal of Asian Economics* 16, 1047-1064.
- 14, Clarke George & Lixin Colin Xu & Heng-fu Zou, 2006, Finance and Income Inequality: What Do the Data Tell Us? *Southern Economic Journal* 72, 578-596.
- 15, Shen Chunli & Heng-fu Zou, 2008, China: Regional Disparities in Poverty Distribution, *Division of Labor & Transition Costs* 3, 17-56.
- 16, Wang Qian & Chunli Shen & Heng-fu Zou, 2009, Local Government Tax Effort in China: An Analysis of Provincial Tax Performance, *Region et Developpment* 29, 203-236.

Consumption and Risk with hyperbolic discounting

Liutang Gong^a, William Smith^{b,*}, Heng-fu Zou^{c,d}

^a *University of Beijing, China*

^b *Department of Economics, Fogelman, United States*

^c *University of Beijing, China and the World Bank, United States*

^d *World Bank, United States*

Received 18 June 2006; received in revised form 10 November 2006; accepted 20 December 2006

Available online 25 April 2007

Abstract

Hyperbolic discounting is not observationally equivalent to exponential discounting. It is always possible to calibrate an exponential model so that it predicts the same level of consumption as a hyperbolic model. However, the two models have radically different comparative statics.

© 2007 Elsevier B.V. All rights reserved.

Keywords: Consumption; Uncertainty; Hyperbolic discounting

JEL classification: D91; E21; G11

1. Introduction

The last decade has seen extensive research on dynamically inconsistent preferences [Ainslie and Haslam (1992), Laibson (1997a,b), Barro (1999), Bernheim et al. (1999), Gul and Pesendorfer (2001), Harris and Laibson (2001a,b, 2003), Krusell and Smith (2001)]. Only recently has there been any attempt to study the effects of dynamically inconsistent preferences on consumption and portfolio behavior under conditions of uncertainty. Palacios-Huerta (2003) has recently solved the savings problem for a version of Merton's (1969, 1971) classic model with constant relative risk aversion (CRRA) and hyperbolic discounting. He demonstrates that hyperbolic discounting raises consumption, so that a hyperbolic model

* Corresponding author. Tel.: +1 901 678 3675; fax: +1 901 678 5380.

E-mail address: wtsmith@memphis.edu (W. Smith).

is observationally equivalent to a model with exponential discounting, but with a higher discount rate. He infers that the canonical model remains “intact” after introducing hyperbolic discounting. This mirrors Barro’s (1999) argument that the neoclassical growth model remains “intact” after introducing a non-constant rate of time preference.

Our purpose in this paper is to make precise the sense in which hyperbolic discounting leaves the canonical model intact, and to explain the ways in which it does not. It is true that the *level* of consumption in a model with hyperbolic discounting is observationally equivalent to the level of consumption in a model of exponential discounting. However, hyperbolic discounting endogenizes the rate of time preference, making it depend in a non-linear way upon the expected growth in wealth. Thus the *comparative statics* of consumption under hyperbolic discounting are strikingly different from those under exponential discounting.

Consider how portfolio risk affects consumption. We demonstrate that the marginal effect of risk on consumption is always greater under hyperbolic discounting than under exponential discounting. Furthermore, hyperbolic discounting causes consumption to be a *concave* function of risk, while in the exponential benchmark the response is *linear*.

To understand these results it is useful to elaborate upon the intuitive story told by Harris and Laibson (2001a, p. 936). Suppose, for example, that relative risk aversion is less than one. Now imagine an increase in portfolio risk. In the exponential model, the investor responds by raising consumption. This is still true in the hyperbolic model. Now, however, the “current” self anticipates that the “future” self will be impatient and consume “too much.” He realizes that the future self will increase consumption even more in response to the increase in risk, so he attaches less value to future consumption at the margin. Therefore, the increase in risk lowers the current self’s discount factor on future consumption. This causes his current consumption to increase by more than in the exponential model. In other words, *hyperbolic discounting amplifies the marginal effect of risk on consumption*. However, this effect weakens as risk gets bigger, so that consumption increases at a decreasing rate as risk increases.

Harris and Laibson (2001b) and Luttmer and Mariotti (2003) also treat saving decisions in continuous-time with risky assets and hyperbolic discounting. Neither investigates the comparative statics of consumption.

2. Consumption and portfolio choice with hyperbolic discounting

We briefly recapitulate Palacios-Huerta’s (2003) model. The consumer maximizes expected lifetime utility over an infinite planning horizon. There is quasi-hyperbolic discounting: starting at time t the consumer’s discount function decays exponentially at the constant rate β until time $t+h$; just prior to time $t+h$ it drops discontinuously to a fraction $\delta \in (0,1]$ of its value, and then continues to decay at the rate β . In other words, the discount function is

$$\begin{aligned} e^{-\beta s}, & \quad t \leq s \leq t+h, \\ \delta e^{-\beta s}, & \quad t+h \leq s < \infty. \end{aligned} \quad (1)$$

Following Merton (1969, 1971), the consumer’s preferences are time-separable and the felicity function is CRRA. His expected lifetime utility is then

$$E_t U_t = E_t \int_t^{t+h} e^{-\beta s} \frac{c_s^{1-b}}{1-b} ds + \delta \int_{t+h}^{\infty} e^{-\beta s} \frac{c_s^{1-b}}{1-b} ds. \quad (2)$$

Intuitively, the current self makes decisions from time t to time $t+h$, when the next self takes charge. Setting $\delta=1$ in Eq. (2) recovers the exponential preferences used by Merton (1969, 1971).

The consumer can invest in two assets. The riskless asset pays a constant rate of return r , while the price P_t of the risky asset follows a geometric Brownian motion,

$$\frac{dP_t}{P_t} = \mu dt + \sigma dZ_t, \quad (3)$$

where Z_t is a Wiener process. If θ_t is the share of wealth W_t invested in the risky asset, then the consumer's budget constraint is

$$dW_t = \{[(1-\theta_t)r + \theta_t\mu]W_t - c_t\}dt + \sigma\theta_t W_t dZ_t. \quad (4)$$

The consumer chooses policies θ_t and c_t to maximize Eq. (2) subject to Eq. (4), and given initial wealth W_0 Palacios-Huerta (2003) solves this problem to arrive at the following optimal policies. The portfolio demand is exactly the same as in Merton (1969, 1971):

$$\theta_t^* = \frac{\mu - r}{b\sigma^2}. \quad (5)$$

Hyperbolic discounting has *no* effect on portfolio demands.

The consumption function is

$$c_t^* = \alpha_H W_t. \quad (6)$$

The marginal propensity to consume (MPC) α_H is determined implicitly by the equation

$$\alpha_H = \frac{\left[\beta + (1-\delta)\alpha_H e^{-\beta h} E_0(W_h/W_0)^{1-b} \right] - (1-b)[\mu_w - b\sigma_w^2/2]}{b}, \quad (7)$$

where $\mu_w = (1-\theta^*)r + \theta^*\mu$ and $\sigma_w^2 = \theta^{*2}\sigma^2$ are the optimal mean and variance of the rate of return to the portfolio. The subscript “H” denotes “hyperbolic,” to distinguish it from the exponential benchmark.¹ It is important to note that $E_0(W_h/W_0)^{1-b}$ is endogenous; its value will be given by Eq. (9) below.

Consider the properties of this consumption function. The second term in braces in Eq. (7) is the certainty equivalent rate of return to the portfolio. The response of consumption to changes in the certainty equivalent rate of return is governed by the magnitude of risk aversion, b .

The first term in braces is the “effective” rate of time preference. In the absence of hyperbolic discounting ($\delta=1$) the rate of time preference would reduce to β and we would recover the marginal propensity to consume in Merton (1969, 1971):

$$\alpha_M = \frac{\beta - (1-b)[\mu_w - b\sigma_w^2/2]}{b}. \quad (8)$$

¹ For small h , $\alpha_H > 0$ requires $b > 1-\delta$. The optimal policies must also satisfy a transversality condition. All derivations and technical details are in an appendix, available upon request.

The subscript “M” stands for “Merton.” Notice that in this case the MPC is a linear function of the certainty equivalent rate of return.

Comparing Eqs. (7) and (8), it is evident that hyperbolic discounting ($\delta < 1$) increases the effective rate of time preference from β to $\beta + (1 - \delta)\alpha e^{-\beta b} E_0(W_h/W_0)^{1-b}$. As explained by Harris and Laibson (2001a), the current self anticipates that the future self will consume too much and so attaches less value at the margin to future consumption. It follows that hyperbolic discounting raises consumption. Palacios-Huerta (2002) infers that hyperbolic discounting is observationally equivalent to exponential discounting, but with a higher discount rate.

But the marginal propensity to consume defined by Eq. (7) warrants closer scrutiny. Note that the rate of time preference with hyperbolic discounting depends upon the expected growth in wealth between period 0 and period h . Since wealth is log-normal, it is straightforward to calculate

$$E_0(W_h/W_0)^{1-b} = e^{(1-b)[\mu_w - \alpha_H - b\sigma_w^2/2]h}. \quad (9)$$

Changes in the mean and the variance of the rate of return, as well as changes in the MPC itself, alter the effective rate of time preference exponentially. This elucidates the non-linear way in which hyperbolic discounting affects time preference, and suggests that the comparative statics of consumption will be much richer than in the canonical model with exponential discounting.

3. Risk and consumption

How does uncertainty affect consumption in the presence of hyperbolic discounting? Since the portfolio demand is unaffected by discounting, we will abstract from the portfolio decision entirely in addressing this question: Henceforth, we assume there is no riskless asset. In this case, the marginal propensity to consume is determined implicitly by

$$\alpha_H = \frac{\left[\beta + (1 - \delta)\alpha_H e^{-\beta h + (1-b)[\mu - \alpha_H - b\sigma^2/2]h} \right] - (1-b)[\mu - b\sigma^2/2]}{b}. \quad (10)$$

The rate of time preference is an increasing function of the MPC.

To understand the comparative statics of the model with hyperbolic discounting, it is helpful to recall the MPC for the benchmark case with exponential discounting in Eq. (8). In that case consumption will increase or decrease with risk depending upon whether relative risk aversion is less than or greater than one,

$$\partial \alpha_M / \partial \sigma^2 = (1-b)/2. \quad (11)$$

The key thing to notice is that under exponential discounting consumption is a *linear* function of σ^2 .

Now consider the effects of a change in risk on the MPC in the general case with hyperbolic discounting. Differentiating Eq. (10) we find

$$\frac{\partial \alpha_H}{\partial \sigma^2} = (1-b) \frac{b}{2} \frac{1 - (1-\delta)h\alpha_H e^{-\beta h} E_0(W_h/W_0)^{1-b}}{b + [(1-b)\alpha_H h - 1](1-\delta)e^{-\beta h} E_0(W_h/W_0)^{1-b}}. \quad (12)$$

It is still true that the sign of the effect of risk on consumption hinges upon the degree of risk aversion. However, the magnitude of this effect is no longer constant, and depends upon the importance of

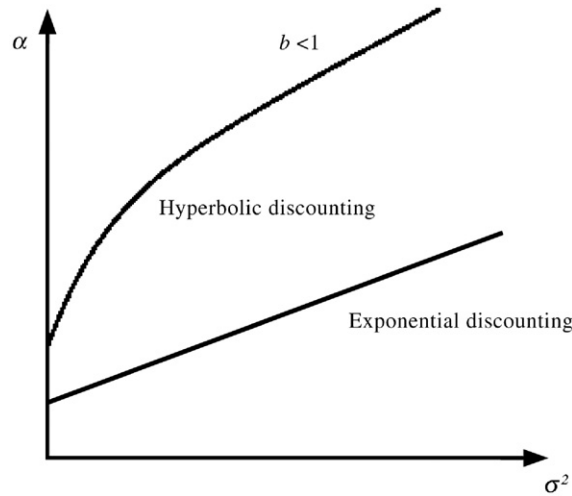


Fig. 1.

hyperbolic discounting (δ). Hyperbolic discounting disrupts the simple linear effect of risk on consumption with CRRA utility.

4. How does hyperbolic discounting alter the comparative statics of risk?

Proposition. *The absolute value of the marginal effect of risk on consumption is greater under hyperbolic discounting than under exponential discounting: $|\partial\alpha_H/\partial\sigma^2| > |\partial\alpha_M/\partial\sigma^2|$. Furthermore, if h is sufficiently small and $b > 1 - \delta$ consumption is a concave function of risk when discounting is hyperbolic: $\partial^2\alpha_H/\partial\sigma^2 < 0$.²*

In other words, hyperbolic discounting amplifies the effect of risk on consumption, relative to the exponential benchmark. Consumption still increases or decreases with risk depending upon the magnitude of relative risk aversion. Now, however, it increases or decreases at a decreasing rate, rather than at a constant rate.

Figs. 1 and 2 depict consumption as a function of risk for the cases where b is less than or greater than unity. The linear functions depict the exponential benchmarks. Notice that in either case consumption with hyperbolic discounting exceeds consumption under exponential discounting.

First suppose that $b < 1$ shown in Fig. 1. If discounting is hyperbolic, then consumption is an increasing, concave function of risk. The hyperbolic consumption function is always steeper than the exponential, so that the marginal effect of risk on consumption is greater under hyperbolic discounting than under exponential discounting. However, the marginal impact of risk on consumption decreases as risk increases. The intuition for this behavior is quite simple. Compare the change in the MPC under exponential discounting in Eq. (11) with the change in the MPC under hyperbolic discounting in Eq. (12). Recall that under exponential discounting the rate of time preference is just β while under hyperbolic discounting the “effective” rate of time preferences is $\beta + (1 - \delta)\alpha e^{-\beta b} E_0(W_h/W_0)^{1-b}$. With exponential discounting the rate of time preference is constant, so the MPC increases linearly with risk. Under hyperbolic discounting, however, the increase in the MPC feeds back to raise the effective rate of time

² The concavity of α_H holds for small h , but not for the case of instantaneous gratification, when $h=0$. In most applications, however, the consumer presumably has some ability to commit to decisions over short periods.

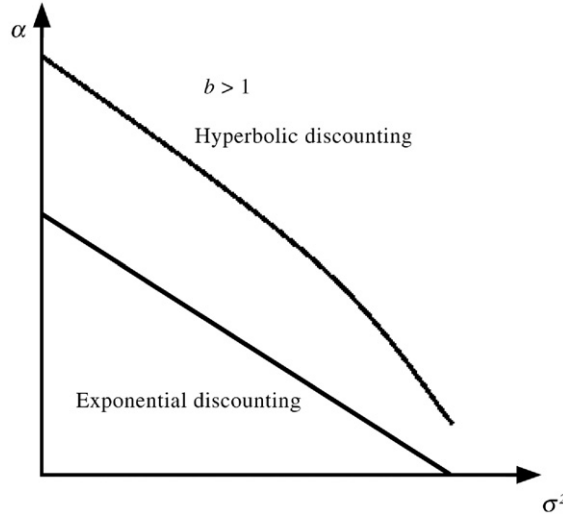


Fig. 2.

preference. This tends to magnify the increase in consumption caused by the increase in risk. However, the increase in risk also tends to lower the effective rate of time preference (for a given α_H) when $b < 1$ [see Eq. (9)]. This exerts a dampening effect on the MPC which increases as risk increases in magnitude.

Conversely, suppose that $b > 1$. In this case consumption is decreasing and concave in risk, as shown in Fig. 2. The slope of the hyperbolic curve is always more negative than for the exponential curve. Here, hyperbolic discounting amplifies the decrease in consumption caused by an increase in risk.

5. Conclusion

By endogenizing the rate of time preference, hyperbolic discounting amplifies the effect of changes in risk on consumption. This offers a broad warning about the extent to which “standard” models seem to remain “intact” in the face of hyperbolic discounting. The standard model is observationally equivalent to the hyperbolic model in the sense that the former can always be calibrated to match the consumption predicted by the latter. However, this does not imply that the comparative static predictions of the two models are the same.

Appendix A. Consumption and risk with hyperbolic discounting

A.1. The transversality condition and some essential inequalities

The optimal policies must satisfy the transversality condition (TVC)

$$\lim_{t \rightarrow \infty} E b e^{-\beta t} W_t^{1-b} = 0. \quad (\text{A.1})$$

As in Merton (1969, 1971) the feasibility condition $\alpha_H > 0$ is a sufficient condition for the TVC to be satisfied. The TVC in turn implies that $e^{-\beta h} E_0(W_h / W_0)^{1-b} < 1$. Thus, it also follows that $1 - (1 - \delta)h\alpha_H e^{-\beta h} E_0(W_h / W_0)^{1-b} > 0$.

Notice that for small h , $h\alpha_H < 1$. Notice also that [from Eq. (7) in the text] for small h it must be true that $b > 1 - \delta$ in order for $\alpha_H > 0$. Assuming that $b > 1 - \delta$ it then follows that, $b + [(1 - b)\alpha_H h - 1](1 - \delta)e^{-\beta h}E_0(W_h/W_0)^{1-b} > 0$ for sufficiently small h .

A.2. Existence and uniqueness of the MPC

Write the right-hand side of Eq. (8) as $\text{RHS}(\alpha)$. The TVC implies $\text{RHS}(0) > 0$. Furthermore, since $b + [(1 - b)\alpha h - 1](1 - \delta)e^{-\beta h}E_0(W_h/W_0)^{1-b} > 0$, it can be shown that $0 < d\text{RHS}/d\alpha < 1$. Therefore, RHS crosses the 45° line once.

A.3. Proof of the proposition

The first statement follows from comparing Eqs. (11) and (12) and using the fact that $1 > \alpha_H h$. The second statement follows from differentiating Eq. (12):

$$\frac{\partial^2 \alpha_H}{\partial \sigma^2} = (1 - b)b \frac{(1 - \delta)h e^{-\beta h}}{2} \Omega, \quad (\text{A.2})$$

where

$$\Omega = \frac{\left[(1 - \delta)e^{-\beta h}E_0(W_h/W_0)^{1-b} - 1 \right] E_0(W_h/W_0)^{1-b} h \frac{\partial \alpha_H}{\partial \sigma^2} - (1 - b)(1 - h\alpha) \left[\frac{\partial \alpha_H}{\partial \sigma^2} + \frac{b}{2} \right]}{\{b + [(1 - b)\alpha_H h - 1](1 - \delta)e^{-\beta h}E_0(W_h/W_0)^{1-b}\}^2}. \quad (\text{A.3})$$

Again, $1 > h\alpha_H$ for small h . The transversality condition implies that the first term in braces is negative.

Now consider the two cases mentioned in the Proposition. If $b < 1$ then $\partial \alpha_H / \partial \sigma^2 > 0$, so $\Omega < 0$. Therefore $\partial^2 \alpha_H / \partial \sigma^2 < 0$. If $b > 1$ then $\partial \alpha_H / \partial \sigma^2 < 0$. If the last term in braces is positive then $\Omega > 0$. It is easy to show, however, that this expression is positive if and only if $1 > h\alpha_H$. Therefore, if $b < 1$ then $\partial^2 \alpha_H / \partial \sigma^2 < 0$.

References

- Ainslie, G., Haslam, N., 1992. Hyperbolic discounting. In: Loewenstein, G., Elster, J. (Eds.), *Choice Over Time*. Russell Sage Foundation, New York, pp. 57–92.
- Barro, R., 1999. Ramsey meets laibson in the neoclassical growth model. *Quarterly Journal of Economics* 114, 1125–1152.
- Bernheim, D., Ray, D., Yeltekin, S., 1999. Self-Control, Saving, and the Low Asset Trap, Working Paper, Stanford University.
- Gul, F., Pesendorfer, W., 2001. Temptation and self-control. *Econometrica* 69, 1403–1435.
- Harris, C., Laibson, D., 2001a. Dynamic choices of hyperbolic consumers. *Econometrica* 69, 935–957.
- Harris, C., Laibson, D., 2001b. Instantaneous Gratification, Working Paper, Harvard University.
- Harris, C., Laibson, D., 2003. Hyperbolic discounting and consumption. In: Dewatripont, M., Hansen, L., Turnovsky, S. (Eds.), *Advances in Economics and Econometrics: Theory and Applications, Eight World Congress, vol. I*. Cambridge University Press, Cambridge.
- Krusell, P., Smith, A., 2001. Consumption and Savings Decisions with Quasi-Geometric Discounting, Working Paper, Carnegie Mellon University.
- Laibson, D., 1997a. Golden eggs and hyperbolic discounting. *Quarterly Journal of Economics* 112, 443–478.
- Laibson, D., 1997b. Hyperbolic Discount Functions and Time Preference Heterogeneity, Working Paper, Harvard University.

- Luttmer, E., Mariotti, T., 2003. Subjective discounting in an exchange economy. *Journal of Political Economy* 111, 959–989.
- Merton, R.C., 1969. Lifetime portfolio selection under uncertainty: the continuous time case. *Review of Economics and Statistics* 51, 247–257.
- Merton, R.C., 1971. Optimum consumption and portfolio rules in a continuous-time model. *Journal of Economic Theory* 3, 373–413.
- Palacios-Huerta, I., 2002. Consumption and Portfolio Rules Under Hyperbolic Discounting, Working Paper, Brown University.
- Palacios-Huerta, I., 2003. Consumption and Portfolio Rules Under Hyperbolic Discounting, Working Paper, Brown University.

Asset Prices and Hyperbolic Discounting

Liutang Gong

Guanghua School of Management, Peking University, Beijing, 100871, China
Institute for Advanced Study, Wuhan University, Wuhan, 430072, China

William Smith

Department of Economics, Fogelman College of Business & Economics
University of Memphis
E-mail: wtsmith@memphis.edu

and

Heng-fu Zou

Guanghua School of Management, Peking University
Institute for Advanced Study Wuhan University
The World Bank, Washington DC, 20433, USA

This paper explores the implications of hyperbolic discounting for asset prices and rates of return. Hyperbolic discounting has no effect on the equity premium. However, by making people less patient, causes stock prices to be lower, and interest rates higher, than with exponential discounting. In addition, hyperbolic discounting dampens the marginal effect of risk on stock prices, relative to the exponential case.

Key Words: Asset-Pricing; Hyperbolic Discounting.

JEL Classification Numbers: D91, E21, G11.

1. INTRODUCTION

Since Laibson's (1994, 1997a,b) path-breaking work, economists have probed deeply into the implications of dynamically inconsistent preferences [see Ainslie and Haslam (1992), Barro (1999), Bernheim, Ray, and Yeltekin (2000), Gul and Pesendorfer (2001), Harris and Laibson (2001a, b; 2003),

Krusell and Smith (2001) to mention just a few important papers]¹ Recently, one branch of the literature has focused on how dynamically inconsistent preferences affect consumption and portfolio behavior under conditions of uncertainty [Harris and Laibson (2001b), Luttmer and Mariotti (2003)]. In particular, Palacios-Huerta (2003) has adapted Merton's (1969, 1971) classic model of consumption and portfolio choice to incorporate hyperbolic discounting. This is a particularly appealing framework because it permits a clear picture of how hyperbolic discounting alters consumer behavior under uncertainty.

In Gong, Smith, and Zou (2006) we have employed Palacios-Huerta's model to explore the comparative statics of risk under hyperbolic discounting. With exponential discounting and constant-relative-risk-aversion (CRRA) utility consumption is a linear function of risk. With hyperbolic discounting, however, the rate of time preference becomes endogenous, so that risk affects consumption in a non-linear way. In particular, hyperbolic discounting amplifies the marginal effect of risk on consumption, relative to the exponential case. This means that it is not true — as often asserted — that hyperbolic discounting and exponential discounting are observationally equivalent. It is true that the level of consumption predicted by a hyperbolic model can be matched by imputing a higher rate if time preference to an exponential model. However, the two models offer radically different comparative static predictions.

In this paper we expand the model to investigate the implications of hyperbolic discounting for asset prices and rates of return. We incorporate the Palacios-Huerta model of consumption with hyperbolic discounting into a equilibrium asset-pricing model à la Lucas (1978). Hyperbolic discounting makes people less patient. This depresses savings and reduces the demand for stocks, so that stock prices fall and interest rates increase. Furthermore, hyperbolic discounting dampens the marginal effect of risk on stock prices, relative to the exponential case.

Two other papers have studied consumption behavior in continuous-time with risky assets and hyperbolic discounting. Harris and Laibson (2001b) work in continuous-time in order to avoid the “pathologies” that crop up in discrete-time models of with hyperbolic discounting.² They establish general existence results and prove that consumption is continuous and monotonic in wealth. Luttmer and Mariotti (2003) consider the continuous-time approximation of a discrete-time consumption/portfolio model with hyperbolic discounting. Like Palacios-Huerta (2002), they show that hyperbolic

¹Gul and Pesendorfer (2002) develop a model with dynamically consistent preferences.

²In discrete time, consumption may be discontinuous and non-monotonic in wealth and there may be multiple equilibria. See Laibson (1997b), Morris and Postlewaite (1997), O'Donoghue and Rabin (1999), Harris and Laibson (2001b), and Krusell and Smith (2000).

discounting affects consumption and the risk-free rate, but does not alter portfolio demands or excess returns. They do not investigate the comparative statics of consumption or of asset prices.

2. CONSUMPTION AND PORTFOLIO POLICIES

Following Palacios-Huerta (2002), imagine a consumer who has an infinite planning horizon and maximizes expected lifetime utility. He exhibits quasi-hyperbolic discounting, so that his discount function is

$$\begin{aligned} e^{-\theta s}, & \quad t \leq s \leq t+h, \\ \delta e^{-\theta s}, & \quad t+h \leq s < \infty. \end{aligned} \quad (1)$$

Beginning at time t the discount function decays exponentially at the constant rate θ until time $t+h$. At time $t+h$ it drops discontinuously by a fraction $\delta \in (0, 1]$; thereafter it continues to decay at the rate δ . This subsumes two important, special cases: If $\delta = 1$ we recover Merton's (1969, 1971) exponential discounting, while if $h \rightarrow 0$ there is "instantaneous gratification," proposed by Harris and Laibson (2001b).

The consumer has time-separable utility with constant relative risk aversion (CRRA). Expected lifetime utility is thus

$$E_t U_t = E_t \int_t^{t+h} e^{-\theta s} \frac{C_s^{1-\gamma}}{1-\gamma} ds + \delta \int_{t+h}^{\infty} e^{-\theta s} \frac{C_s^{1-\gamma}}{1-\gamma} ds, \quad (2)$$

where $\gamma > 0$ is the coefficient of relative risk aversion. Intuitively, Equation (2) says that the "current self" makes decisions from time t to time $t+h$, whereupon the "next self" starts to make the decisions.

There are two assets. A riskless asset pays a constant rate of return r . A risky asset has a price P_t that follows a geometric Brownian motion,

$$\frac{dP_t}{P_t} = \mu dt + \sigma dZ_t, \quad (3)$$

where Z_t is a Wiener process. Define λ_t as the share of wealth W_t invested in the risky asset. The budget constraint is

$$dW_t = \{[(1-\lambda_t)r + \lambda_t\mu]W_t - c_t\}dt + \sigma\lambda_t W_t dZ_t. \quad (4)$$

The consumer's problem is to choose policies λ_t and C_t to maximize Equation (2) subject to Equation (4), given initial wealth W_0 . Palacios-Huerta (2003) shows that the optimal policies for this problem are

$$\lambda_t^* = \frac{\mu - r}{\gamma\sigma^2}, \quad (5)$$

$$C_t^* = c_H W_t, \quad (6)$$

where the marginal propensity to consume (MPC) c_H is determined implicitly by the equation

$$c_H = \frac{[\theta + (1 - \delta)c_H e^{-\theta h} E_0(W_h/W_0)^{1-\gamma}] - (1 - \gamma)[\mu_w - \gamma\sigma_w^2/2]}{\gamma}, \quad (7)$$

and $\mu_w = (1 - \lambda^*)r + \lambda^*\mu$ and $\sigma_w^2 = \lambda^{*2}\sigma^2$ are the optimal mean and variance of the rate of return to the portfolio. The subscript “H” denotes “hyperbolic”.

The portfolio demand in Equation (5) is exactly the same as in Merton (1969, 1971). Therefore, as Palacios-Huerta (2002) emphasizes, hyperbolic discounting has no effect on portfolio demands.

To understand the consumption function in Equations (6) and (7) it is useful to consider the MPC for the exponential benchmark in Merton (1969, 1971):

$$c_M = \frac{\theta - (1 - \gamma)[\mu_w - \gamma\sigma_w^2/2]}{\gamma}. \quad (8)$$

The subscript “M” stands for “Merton”. The term in braces in Equation (8) is the certainty-equivalent rate of return to the portfolio. An increase in risk lowers the certainty-equivalent rate of return, which then increases or decreases consumption depending upon whether relative risk aversion γ , is less than or greater than one.³ The essential thing to note is that, in the presence of exponential discounting, consumption is a linear function of the constant rate of time preference and the certainty-equivalent rate of return.

Now compare Equations (7) and (8). It is clear on inspection that that hyperbolic discounting ($\delta < 1$) has the effect of increasing the rate of time preference from θ to $\theta + (1 - \delta)c_H e^{-\theta h} E_0(W_h/W_0)^{1-\gamma}$. Intuitively, the “current self” anticipates that the “next self” will consume too much and so attaches less value at the margin to future consumption. [Harris and Laibson (2001a)].⁴ Hyperbolic discounting raises consumption, relative to the exponential benchmark, by making people less patient.

³Since Weil (1989) we have known that it is really intertemporal substitution, rather than risk aversion, that governs the sign of the effect of risk on consumption. It would be straightforward to develop a version of this model with Generalized Isoelastic (GIE) preferences [Epstein (1987), Epstein and Zin (1989, 1991), Duffie and Epstein (1993a, b), Svensson (1989), Weil (1989)] in order to disentangle risk aversion from intertemporal substitution. However, doing so would not add much to the point here, and would distract attention from the time-separable benchmark used by Palacios-Huerta (2002). The reader should feel free to interpret the coefficient attached to the certainty-equivalent rate of return in Equations (8) and (9) as $1 - 1/\varepsilon$, where ε is the intertemporal elasticity of substitution for riskless consumption paths.

⁴Our effective rate of time preference corresponds to the effective discount factor in Harris and Laibson (2001a).

At first glance this would seem to suggest that hyperbolic discounting is observationally equivalent to exponential discounting: it is always possible to match the level of consumption predicted in a hyperbolic model by calibrating an exponential model to have a higher discount rate. Indeed, Palacios-Huerta (2003) asserts for this reason that the canonical model of consumption and portfolio choice remains “intact” after the introduction of hyperbolic discounting. This mirrors Barro’s (1999) argument that the neo-classical growth model remains “intact” after introducing a non-constant rate of time preference.

Gong, Smith, and Zou (2006) rebut this argument. Even if the two models can be calibrated to generate the same level of consumption, they still may make very different comparative static predictions. Notice that the rate of time preference with hyperbolic discounting [in Equation (7)] depends upon the expected growth in wealth between period 0 and period h . This is a manifestation of the general result of Harris and Laibson (2001b): the value function for a consumer with dynamically inconsistent, hyperbolic preferences is the same as to the value function for a consumer with dynamically consistent, exponential preferences and a wealth-dependent utility function. In other words, hyperbolic discounting induces an “indirect” form (i.e., through the value function) of the “spirit of capitalism” — the old notion [Weber (1958)] that people may derive utility from wealth itself, in addition to consumption — that has recently been used to explain asset prices [Bakshi and Chen (1996), Smith (2001), Gong and Zou (2002a)].⁵ In this literature, the level of wealth (or some other measure of status) yields utility. With hyperbolic discounting it is the growth of wealth that matters, rather than the level. This is similar in spirit to an idea originally espoused by Marshall (1979), and recently explored by Gootzeit, Schneider, and Smith (2002), that people derive utility from the act of saving, from the accumulation of wealth rather than the level of wealth.

Using the fact that wealth is log-normal, it is straightforward to calculate

$$E_0(W_h/W_0)^{1-\gamma} = e^{(1-\gamma)[\mu_w - c_H - \gamma\sigma_w^2/2]h}. \quad (9)$$

Changes in the mean and the variance of the rate of return, as well as changes in the MPC itself, alter the effective rate of time preference exponentially. In other words, the rate of time preference is endogenous. This will have profound implications for the comparative statics of the model.

Following Gong, Smith, and Zou (2006), consider how does uncertainty affects consumption in the presence of hyperbolic discounting. To simplify exposition, and to set the stage for the next section, we will assume that

⁵The spirit of capitalism has also been used to explain saving [Zou (1995)] and growth [Zou (1994), Smith (1999), Gong and Zou (2002a,b)].

$\lambda^* = 1$.⁶ In this case, the marginal propensity to consume is determined implicitly by

$$c_H = \frac{[\theta + (1 - \delta)c_H e^{-\theta h + (1 - \gamma)[\mu - c_H - \gamma\sigma^2/2]h}] - (1 - \gamma)[\mu - \gamma\sigma^2/2]}{\gamma}. \quad (10)$$

Given the transversality condition, there is a unique value of c_H that solves this equation.⁷ Note that the rate of time preference in Equation (10) is an increasing function of the MPC. This is similar to Harris and Laibson (2001a), where the current discount factor is a decreasing function of the future MPC. The MPC in Equation (10) is the fixed-point that captures the dependence of the current MPC on the future MPC.

Now consider how risk affects consumption. In the benchmark case with exponential discounting we have seen that consumption will increase or decrease linearly with σ^2 depending upon whether relative risk aversion is less than or greater than one,

$$\partial c_M / \partial \sigma^2 = (1 - \gamma) / 2. \quad (11)$$

In the general case with hyperbolic discounting we find

$$\frac{\partial c_H}{\partial \sigma^2} = (1 - \gamma) \frac{\gamma}{2} \frac{1 - (1 - \delta)h c_H e^{-\theta h} E_0(W_h/W_0)^{1-\gamma}}{\gamma + [(1 - \gamma)c_H h - 1](1 - \delta)e^{-\theta h} E_0(W_h/W_0)^{1-\gamma}}. \quad (12)$$

The direction of the effect of risk on consumption still depends upon the magnitude of relative risk aversion. However, with hyperbolic discounting risk no longer has a simple linear effect of risk on consumption.

In Gong, Smith, and Zou (2006) we demonstrate

PROPOSITION 1. *The absolute value of the marginal effect of risk on consumption is greater under hyperbolic discounting than under exponential discounting: $|\partial c_H / \partial \sigma^2| > |\partial c_M / \partial \sigma^2|$. Furthermore, if h is sufficiently small and $b > 1 - \delta$, consumption is a concave function of risk when discounting is hyperbolic: $\partial^2 c_H / \partial \sigma^2 < 0$.⁸*

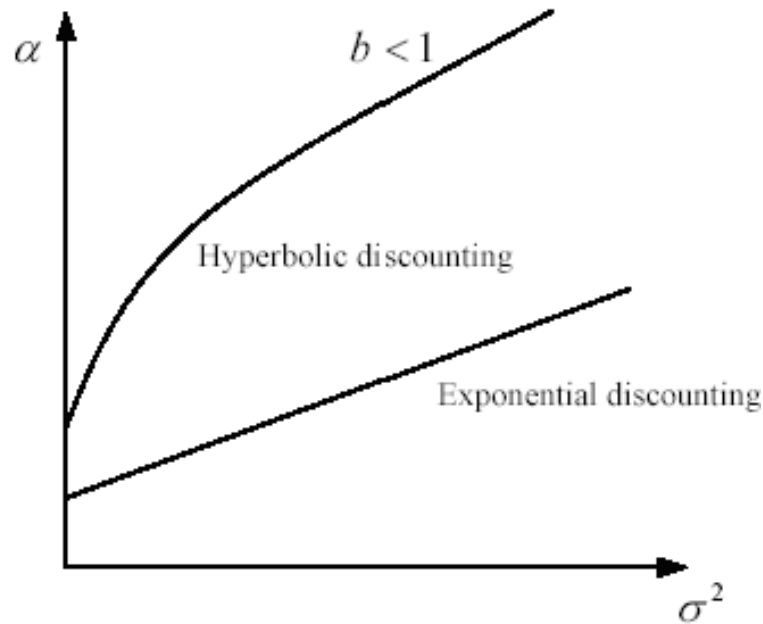
Intuitively, consumption still increases or decreases with risk depending upon the magnitude of relative risk aversion. However, hyperbolic discounting amplifies the effect of risk on consumption, relative to the exponential

⁶In an equilibrium where the riskless asset is in zero net supply $\lambda_t = 1$.

⁷Write the right-hand side of Equation (8) as $RHS(c_H)$. The TVC implies $RHS(0) > 0$. Furthermore, since $\gamma + [(1 - \gamma)c_H h - 1](1 - \delta)E_0(W_h/W_0)^{1-\gamma} > 0$, it can be shown that $0 < dRHS/dc_H < 1$. Therefore, RHS crosses the 45° line once.

⁸Although is concave for small, nonzero it is not when there is instantaneous gratification. It is plausible to think that consumers usually are able to commit to decisions over short periods.

FIG. 1.

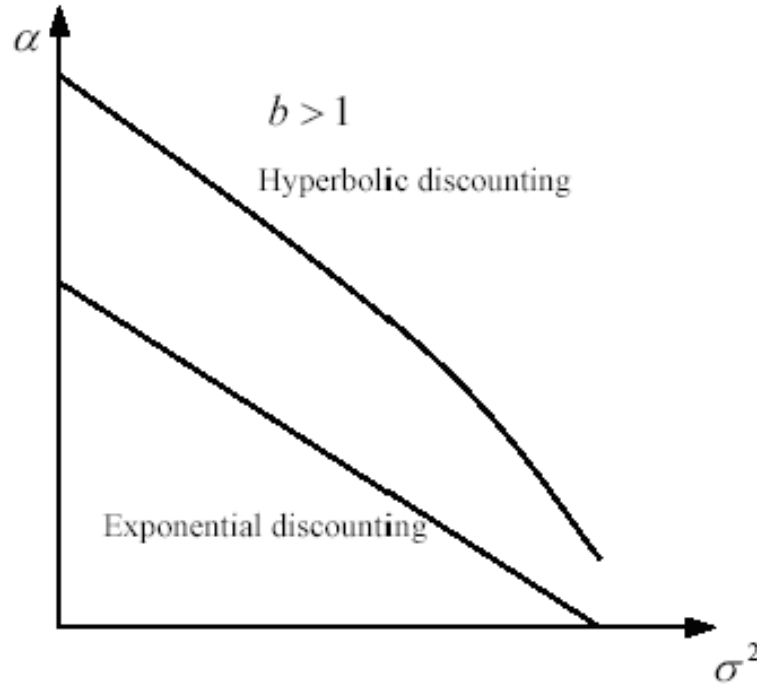


benchmark. Moreover, the effect is no longer linear: consumption increases (or decreases, depending upon γ) at a decreasing rate, as risk increases.

In Figures I and II we show consumption as a function of risk for the cases where γ is less than or greater than unity, respectively. The linear functions depict the exponential benchmarks.

Consider the case when $\gamma < 1$, in Figure I. If discounting is hyperbolic, then consumption is an increasing, concave function of risk, while if discounting is exponential consumption is an increasing linear function of risk.. The hyperbolic consumption function is always steeper than the exponential consumption function. This means that the marginal effect of risk on consumption is greater under hyperbolic discounting than under exponential discounting. However, because the hyperbolic consumption function is concave, the marginal impact of risk on consumption decreases as risk increases. Why, intuitively, is this happening? Under exponential discounting the rate of time preference is just θ ; since the rate of time preference is exogenous the MPC increases linearly with risk. Under hyperbolic discounting, however, the “effective” rate of time preferences is $\theta + (1 - \delta)c_H e^{-\theta b} E_0(W_h/W_0)^{1-\gamma}$. Because the rate of time preference is

FIG. 2.



endogenous, the increase in the MPC feeds back to raise the effective rate of time preference. This magnifies the increase in consumption caused by the increase in risk. However, the increase in risk also has a direct effect on the effective rate of time preference: for a given c_H an increase in risk lowers the rate of time preference when $\gamma < 1$ [see Equation (9)]. This exerts a countervailing effect on the MPC, the magnitude of which increases as risk increases.

Conversely, consider the case where $\gamma > 1$, in Figure II. Now consumption decreases with in risk: the relationship is again linear in with exponential discounting and concave with hyperbolic discounting. The slope of the function is always more negative in the hyperbolic than in the exponential case. Hence, hyperbolic discounting amplifies the decline in consumption associated with an increase in risk.

3. IMPLICATIONS FOR ASSET PRICES

To develop the implications of hyperbolic discounting for asset pricing, consider the following Lucas (1978) “tree” model. A tree yields “fruit” D_t

(dividends) according to the geometric Brownian motion:

$$\frac{dD_t}{D_t} = \nu dt + \sigma_D dZ_t. \quad (13)$$

Investors can buy shares in the tree (a stock) at the price (ex-dividend) P_t . The supply of shares is inelastic and normalized in size to one. Using the notation in Equation (3), the cum dividend rate of return is then

$$\frac{dP_t}{P_t} + \frac{D_t}{P_t} = \mu dt + \sigma_D dZ_t, \quad (14)$$

where $\mu = \pi + D_t/P_t$ and π is expected capital gains. In equilibrium D_t/P_t will be constant, so that the expected rate of return will also be a constant.

An equilibrium consists of a pricing function $P_t = f(D_t)$ and a risk-free interest rate r such that for $t \in [0, \infty)$

- (1) the representative consumer obeys the optimal policies in Equations (5), (6), and (7),
- (2) all dividends are consumed, so that $C_t = D_t$ and
- (3) the riskless asset is in zero net supply, $\lambda_t = 1$.

In Appendix B we show

PROPOSITION 2. *The equilibrium price function and interest rate are*

$$P_t = A_H D_t, \quad (15)$$

$$r = \nu + 1/A_H - \gamma \sigma_D^2, \quad (16)$$

where

$$A_H = \frac{1 - (1 - \delta)e^{-\theta h + (1 - \gamma)(\nu - \gamma \frac{\sigma_D^2}{2})}}{\theta - (1 - \gamma)(\nu - \gamma \frac{\sigma_D^2}{2})} \quad (17)$$

The subscript “ H ” again denotes “hyperbolic”. Notice that the stock price is proportional to dividends, so that capital gains is equal to the growth rate of dividends; that is, $dP_t/P_t = dD_t/D_t$, so that $\pi = \nu$ and $\sigma^2 = \sigma_D^2$.⁹

Sketch of Proof:

⁹We assume that the denominator in Equation (17) is positive. This is a necessary and sufficient condition for the TVC to be satisfied in the exponential version of the model, discussed below.

In the appendix we demonstrate that the pricing function must satisfy the following, non-linear, second-order differential equation:

$$1 - (1 - \delta)e^{-\theta h + (1-\gamma)(\left[\frac{f'(D_t)D_t}{f(D_t)} + \frac{f''(D_t)D_t}{f(D_t)}\right]\nu - \gamma\left[\frac{f'(D_t)D_t}{f(D_t)}\right]^2 \frac{\sigma_D^2}{2})h} \\ = f \left[\theta - (1 - \gamma) \left(\left[\frac{f'(D_t)D_t}{f(D_t)} + \frac{f''(D_t)D_t}{f(D_t)} \right] \nu - \gamma \left[\frac{f'(D_t)D_t}{f(D_t)} \right]^2 \frac{\sigma_D^2}{2} \right) \right] \quad (18)$$

The solution to this equation is given by equations (15) and (17). The equilibrium interest rate in Equation (16) then follows from the proportionality of the asset price to dividends and the fact that $\lambda_t = 1$ in equilibrium.

How does hyperbolic discounting affect asset prices and rates of return? Notice first that in equilibrium all wealth is invested in the stock, so that $\lambda_t = 1$. Using the portfolio demand in Equation (5) and the fact that $\sigma^2 = \sigma_D^2$ it follows that

$$\mu - r = \gamma \sigma_D^2, \quad (19)$$

where $\mu = \nu + 1/A_H$. This implies

PROPOSITION 3. *Hyperbolic discounting has no effect on the equity premium.*

Hyperbolic discounting is of no use to explaining the equity premium paradox, for the simple reason that it does influence portfolio demands.

However, hyperbolic discounting does affect the levels of stock prices and interest rates. To see this, it is useful to consider the exponential benchmark as a special case. When $h = 0$ the equilibrium stock price in equations (15) and (17) reduces to $P_t = A_M D_t$, where

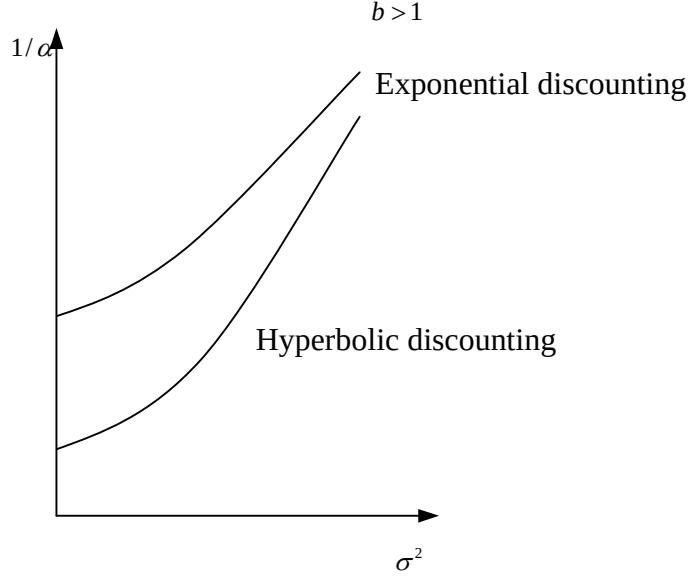
$$A_M = \frac{1}{\theta - (1 - \gamma)(\nu - \gamma \frac{\sigma_D^2}{2})} \quad (20)$$

This is the equilibrium stock price that would emerge if the consumption/portfolio model in Merton (1969, 1971) were embedded in a Lucas (1978) equilibrium model, so “ M ” is again a mnemonic for “Merton”. Comparing this to the consumption in Equation (8), it is evident that the stock price is inversely proportional to the MPC.

Now compare Equation (17) and (19). As suggested by Palacios-Huerta (2003), hyperbolic discounting will lower the level of the stock price by raising the discount rate. From Equation (16) this also increases the interest rate, by increasing the dividend/price ratio. Thus.

PROPOSITION 4. *Hyperbolic discounting lowers stock prices and raises the risk-free rate.*

FIG. 3.



Now consider how risk affects the stock price. In the exponential model [Equation (19)] it is immediate that

$$\frac{\partial A_M}{\partial \sigma_D^2} = -(1-\gamma) \frac{\gamma}{2} \frac{1}{\left[\theta - (1-\gamma) \left(\nu - \gamma \frac{\sigma_D^2}{2} \right) \right]^2}, \quad (21)$$

$$\frac{\partial^2 A_M}{\partial \sigma_D^2} = (1-\gamma)^2 \frac{\gamma^2}{2} \frac{1}{\left[\theta - (1-\gamma) \left(\nu - \gamma \frac{\sigma_D^2}{2} \right) \right]^3} > 0. \quad (22)$$

In the canonical model an increase in uncertainty about dividend growth will lower the stock price if $\gamma < 1$, and raise it if $\gamma > 1$. Furthermore, from Equation (17), the stock price will be convex in risk. The intuition is straightforward. Suppose that $\gamma > 1$. An increase in risk will then lower the MPC. Since people save more, the demand for the stock increases and its price rises. The stock price increases at an increasing rate because it is inversely proportional to the MPC.

What happens when there is exponential discounting? In Appendix C we prove

PROPOSITION 5. *The absolute value of the marginal effect of risk on the price of the risky asset is greater under hyperbolic discounting than under*

exponential discounting: $|\partial A_H / \partial \sigma_D^2| < |\partial A_M / \partial \sigma_D^2|$. Furthermore, if h is sufficiently small then the asset price is a convex function of risk when discounting is hyperbolic: $\partial^2 A_H / \partial \sigma_D^2{}^2 > 0$.

In other words, hyperbolic discounting dampens the marginal effect of risk on stock prices, relative to the effect predicted by the exponential model. Why does this happen? Consider again the empirically plausible case where $\gamma > 1$. This is depicted in Figure III. In the exponential model an increase in risk lowers consumption [Equation (11)]. Since people are saving more, the demand for the risky asset increases, and with it the price of the risky asset [Equation (20)]. This effect also occurs in the hyperbolic model. However, in the hyperbolic model the increase in risk also raises the rate of time preference by changing the expected growth of wealth. Since people are less patient, savings falls by more than in the exponential model, causing the price of the asset to decrease relative to the increase in the exponential model.

4. CONCLUSION

By endogenizing the rate of time preference, hyperbolic discounting introduces a non-linearity into the consumption/portfolio decision. We have shown [Gong, Smith, and Zou (2006)] that this causes the comparative static predictions of the hyperbolic model to differ radically from the exponential model. Hyperbolic discounting amplifies the effect of changes in risk on consumption.

In this paper we have explored the implications of this non-linearity for asset prices and rates of return. Hyperbolic discounting does not affect the equity premium. However, it does alter the way in which the level of stock prices and interest rates are affected by risk. Hyperbolic discounting induces people to save less than in the exponential case, lowering the demand for stocks. This lowers stock prices and raises the risk-free rate. In addition, hyperbolic discounting reduces the marginal effect of risk on stock prices, relative to the exponential case.

The non-linear comparative statics induced by hyperbolic discounting should also have interesting implications for macroeconomic policy. Gong, Smith, Turnovsky, and Zou (2006) incorporate hyperbolic discounting into a model of fiscal policy in a stochastic growing economy. In the presence of hyperbolic discounting taxes on the stochastic components of capital and wage income have magnified effects on growth rates and welfare, relative to the benchmark exponential model.

APPENDIX A

Derivation of Proposition 1.

The transversality condition is

$$\lim_{t \rightarrow \infty} E b e^{-\beta t} W_t^{1-\gamma} = 0. \quad (\text{A.1})$$

As in Merton (1969, 1971) feasibility $c_H > 0$ is necessary and sufficient for the TVC to be satisfied. If the TVC is satisfied then

$$e^{-\beta h} E_0(W_h/W_0)^{1-\gamma} < 1. \quad (\text{A.2})$$

Therefore, since $\delta \leq 1$, it must also be true that

$$1 - (1 - \delta) h c_H e^{-\beta h} E_0(W_h/W_0)^{1-\gamma} > 0 \quad (\text{A.3})$$

Inequalities (A.2) and (A.3) will be important in the ensuing comparative statics.

For small h , $h c_H < 1$. Equation (7) in the text implies that for small h it must also be true that $\gamma > 1 - \delta$ in order for $c_H > 0$. Given $\gamma > 1 - \delta$ it then follows that, $\gamma + [(1 - \gamma) c_H h - 1](1 - \delta) e^{-\beta h} E_0(W_h/W_0)^{1-\gamma} > 0$ for sufficiently small h .

The first statement follows from comparing Equations (11) and (12) and using the fact that $1 > c_H h$.

The second statement follows from differentiating Equation (12):

$$\frac{\partial^2 c_H}{\partial \sigma^2} = (1 - \gamma) \gamma \frac{(1 - \delta) h e^{-\beta h}}{2} \Omega, \quad (\text{A.4})$$

where

$$\Omega = \frac{[(1 - \delta) e^{-\beta h} E_0(W_h/W_0)^{1-\gamma} - 1] E_0(W_h/W_0)^{1-\gamma} h \frac{\partial c_H}{\partial \sigma^2} - (1 - \gamma)(1 - h c_H) [\frac{\partial c_H}{\partial \sigma^2} + \frac{\gamma}{2}]}{\{\gamma + [(1 - \gamma) c_H h - 1](1 - \delta) e^{-\beta h} E_0(W_h/W_0)^{1-\gamma}\}^2}. \quad (\text{A.5})$$

Again, $1 > h c_H$ for small h . We have seen that the transversality condition implies that the first term in braces is negative.

Consider the two cases mentioned in the proposition. On the one hand, if $\gamma < 1$ then $\partial c_H / \partial \sigma^2 > 0$, so $\Omega < 0$. Therefore $\partial^2 c_H / \partial \sigma^2 < 0$. On the other hand, if $\gamma > 1$ then $\partial c_H / \partial \sigma^2 < 0$. If the last term in braces is positive then $\Omega > 0$. It can be shown that this expression is positive if and only if $1 > h c_H$. Thus if $\gamma < 1$ then $\partial^2 c_H / \partial \sigma^2 < 0$.

APPENDIX B

Derivation of Proposition 2

The derivation is similar to that in Smith (2001). First, since all dividends are consumed, it follows that $D_t = c_H^* W_t$, where c_H^* denotes the equilibrium value of the MPC. However $W_t = P_t$ because there is one share of stock and the riskless asset is in zero net supply. Therefore

$$D_t = c_H^* P_t. \quad (\text{B.1})$$

Now evaluate the MPC in Equation (7) at $\lambda_t = 1$. This yields Equation (10), which we report here for convenience

$$c_H^* = \frac{|\theta + (1 - \delta)c_H^* e^{-\theta h + (1 - \gamma)[\mu - c_H^* - \gamma\sigma^2/2]h} - (1 - \gamma)|\mu - \gamma\sigma^2/2|}{\gamma}. \quad (\text{B.2})$$

Apply Ito's lemma to the function $f(D_t)$:

$$\frac{dP_t}{P_t} = \left[\frac{f'(D_t)}{f(D_t)} + \frac{f''(D_t)}{f(D_t)} \right] \nu D_t dt + \frac{f'(D_t)}{f(D_t)} \sigma_D D_t dZ_t. \quad (\text{B.3})$$

This implies that the mean and variance of capital gains are

$$\pi = \left[\frac{f'(D_t)D_t}{f(D_t)} + \frac{f''(D_t)D_t}{f(D_t)} \right] \nu \quad (\text{B.4})$$

$$\sigma^2 = \left[\frac{f'(D_t)D_t}{f(D_t)} \right]^2 \sigma_D^2. \quad (\text{B.5})$$

Consider the term $\mu - c_H^*$ in the exponential function in Equation (B.2): By definition $\mu = \pi + D_t/P_t$. In equilibrium, however, $D_t/P_t = c_H^*$. Therefore $\mu - c_H^* = \pi$. Using this fact along with Equations (B.4) and (B.5) yields

$$\begin{aligned} & 1 - (1 - \delta)e^{-\theta h + (1 - \gamma)[\frac{f'(D_t)D_t}{f(D_t)} + \frac{f''(D_t)D_t}{f(D_t)}]\nu - \gamma[\frac{f'(D_t)D_t}{f(D_t)}]^2 \frac{\sigma_D^2}{2}]h} \\ &= f \left[\theta - (1 - \gamma) \left(\left[\frac{f'(D_t)D_t}{f(D_t)} + \frac{f''(D_t)D_t}{f(D_t)} \right] \nu - \gamma \left[\frac{f'(D_t)D_t}{f(D_t)} \right]^2 \frac{\sigma_D^2}{2} \right) \right] \end{aligned} \quad (\text{B.6})$$

This is equation (15) in the text.

Conjecture that the equilibrium price is proportional to dividends:¹⁰

$$P_t = A_H D_t. \quad (\text{B.7})$$

¹⁰We ignore bubble solutions.

It follows that $\pi = \nu$ and $\sigma^2 = \sigma_D^2$. Equation (B.6) then reduces to

$$1 - (1 - \delta)e^{-\theta h + (1 - \gamma)(\nu - \gamma\sigma_D^2)h} = A_H \left[\theta - (1 - \gamma) \left(\nu - \gamma \frac{\sigma_D^2}{2} \right) \right] \quad (\text{B.8})$$

Solving for yields

$$A_H = \frac{1 - (1 - \delta)e^{-\theta h + (1 - \gamma)(\nu - \gamma\sigma_D^2)h}}{\theta - (1 - \gamma) \left(\nu - \gamma \frac{\sigma_D^2}{2} \right)}. \quad (\text{B.9})$$

This determines the equilibrium pricing function. To find the equilibrium interest rate, note that since $\lambda_t = 1$ in equilibrium, then $\mu - r = \gamma\sigma^2$. However, we have seen that $\mu = \pi + D_t/P_t$, $\pi = \nu$, and $\sigma^2 = \sigma_D^2$. Therefore $r = \nu + 1/A_H - \gamma\sigma_D^2$.

APPENDIX C

Derivation of Proposition 4

To simplify notation, define $x = \theta - (1 - \gamma)(\nu - \sigma_D^2/2)$. The price-dividend ratio in Equation (B.9), or in Equation (17) in the text, can then be written as

$$A_H = \frac{1 - (1 - \delta)e^{-xh}}{x}. \quad (\text{C.1})$$

Similarly, the price-dividend ratio for the exponential model [Equation (19) in the text] is simply

$$A_M = \frac{1}{x}. \quad (\text{C.2})$$

Equations (20) and (21) can now be expressed as $\partial A_m / \partial \sigma_D^2 = -\frac{x_{\sigma^2}}{x^2}$ and $\partial A_m^2 / \partial \sigma_D^2 = x_{\sigma^2}^2 / x^3$, where $x_{\sigma^2} = \partial x / \partial \sigma_D^2 = \gamma(1 - \gamma)/2$.

Now consider the marginal effect of risk on the asset price:

$$\frac{\partial A_H}{\partial \sigma_D^2} = \frac{\partial A_M}{\partial \sigma_D^2} [1 - (1 - \delta)e^{-xh}(1 + xh)] \quad (\text{C.3})$$

To sign this expression, substitute A_H in Equation (B.9) into inequality (A.3). This implies that the expression in brackets in Equation (C.3) is unambiguously positive. Therefore, $\partial A_H / \partial \sigma_D^2 \geq 0$ as $\gamma \geq 1$. Equation (C.3) also implies that $\left| \frac{\partial A_H}{\partial \sigma_D^2} \right| < \left| \frac{\partial A_M}{\partial \sigma_D^2} \right|$.

With a bit of tedious algebra it can be shown that

$$\frac{\partial A_H^2}{\partial \sigma_D^2} = \frac{\partial A_M^2}{\partial \sigma_D^2} \left[1 - \frac{1}{2}(1 - \delta)e^{-xh}(1 + xh + x^2h^2) \right]. \quad (\text{C.4})$$

To sign this expression, recall that $1 > (1 - \delta)e^{-xh}(1 + xh)$. Now consider the quadratic expression in Equation (C.4). It is straightforward to show that¹¹

$$1 + xh > \frac{1 + xh + x^2h^2}{2}. \quad (\text{C.5})$$

It follows that the expression in brackets is positive for small h . Since the exponential price function is convex in risk, the hyperbolic price function must also. That is, $\frac{\partial A_H^2}{\partial \sigma_D^2} > 0$.

REFERENCES

- Ainslie, G. and N. Haslam, 1992, Hyperbolic Discounting, in G. Loewenstein and J. Elster, eds., *Choice Over Time*, (Russell Sage Foundation, New York) 57-92.
- Bakshi, G. and Z. Chen, 1996, The Spirit of Capitalism and Stock Market Prices. *American Economic Review* **86**, 133-157.
- Barro, R., 1999, Ramsey Meets Laibson in the Neoclassical Growth Model. *Quarterly Journal of Economics* **114**, 1125-1152.
- Bernheim, D., D. Ray, and S. Yeltekin, 1999, Self-Control, Saving, and the Low Asset Trap, Working Paper, Stanford University.
- Duffie, D. and L. Epstein, 1993a, Asset Pricing with Stochastic Differential Utility. *Review of Financial Studies* **5**, 411-436.
- Duffie, D. and L. Epstein, 1993b, Stochastic Differential Utility. *Econometrica* **60**, 353-394.
- Epstein, L., 1987, Risk Aversion and Asset Prices. *Journal of Monetary Economics* **22**, 179-192.
- Epstein, L. and S. Zin, 1989, Substitution, Risk Aversion, and the Temporal Behavior of Consumption and Asset Returns: A Theoretical Framework. *Econometrica* **57**, 937-969.
- Epstein, L. and S. Zin, 1991, Substitution, Risk Aversion, and the Temporal Behavior of Consumption and Asset Returns: An Empirical Analysis. *Journal of Political Economy* **99**, 263-286.
- Gong, L., W. Smith, and H. Zou, 2006, Consumption and Risk with Hyperbolic Discounting, forthcoming, *Economics Letters*.
- Gong, L., W. Smith, S. Turnovsky, and H. Zou., 2006, Fiscal Policies in a Stochastic Growth Model with Hyperbolic Discounting, Working Paper, Beijing University.
- Gong, L. and H. Zou, 2002, Direct Preferences for Wealth, the Risk Premium Puzzle, Growth, and Policy Effectiveness. *Journal of Economic Dynamics and Control* **26**, 243-270.
- Gong, L. and H. Zou, 2002, Money, Social Status, and Capital Accumulation in a Cash-in-Advance Model. *Journal of Money Credit and Banking*, **33**, 284-293.

¹¹Proof: If this inequality holds then $1 + xh > x^2h^2$. When plotted as a function of h the left hand side of this inequality has a positive vertical intercept of 1 and a positive slope of x . The right hand side is equal to zero when $h = 0$ and at the rate $2x^2h$. Therefore for sufficiently small h the left hand side is greater than the right hand side.

- Gootzeit, M., J. Schneider, and W. Smith, 2002, Marshallian Recursive Preferences and Growth. *Journal of Economic Behavior and Organization* **49**, 381-404.
- Gul, F. and W. Pesendorfer, 2001, Temptation and Self-Control. *Econometrica* **69**, 1403-1435.
- Gul, F. and W. Pesendorfer, 2002, Self-Control and the Theory of Consumption, Working Paper, Princeton University.
- Harris, C. and D. Laibson, 2001a, Dynamic Choices of Hyperbolic Consumers. *Econometrica* **69**, 935-957.
- Harris, C. and D. Laibson, 2001b, Instantaneous Gratification, Working Paper, Harvard University.
- Harris, C. and D. Laibson, 2003, Hyperbolic Discounting and Consumption, in M. Dewatripont, L. Hansen, and S. Turnovsky, eds., *Advances in Economics and Econometrics: Theory and Applications, Eight World Congress, Vol. I*, (Cambridge University Press, Cambridge).
- Krusell, P. and A. Smith, 2001, Consumption and Savings Decisions with Quasi-Geometric Discounting, Working Paper, Carnegie Mellon University.
- Laibson, D., 1994, Self Control and Savings, Ph.D. Dissertation, Massachusetts Institute of Technology.
- Laibson, D., 1997a, Golden Eggs and Hyperbolic Discounting. *Quarterly Journal of Economics* **112**, 443-478.
- Laibson, D., 1997b, Hyperbolic Discount Functions and Time Preference Heterogeneity, Working Paper, Harvard University.
- Lucas, R., 1978, Asset Prices in an Exchange Economy. *Econometrica* **46**, 1429-1445.
- Luttmer, E. and T. Mariotti, 2002, Subjective Discounting in an Exchange Economy. *Journal of Political Economy* **111**, 959-989.
- Marshall, A., 1979, *Principles of Economics*, 8th edition (MacMillan, London).
- Merton, R.C., 1969, Lifetime Portfolio Selection under Uncertainty: The Continuous Time Case. *Review of Economics and Statistics* **51**, 247-257.
- Merton, R.C., 1971, Optimum Consumption and Portfolio Rules in a Continuous-Time Model. *Journal of Economic Theory* **3**, 373-413.
- Morris, S. and A. Postlewaite, 1997, Observational Implications on Nonexponential Discounting, Working Paper.
- O'Donoghue, T. and M. Rabin, 1999, Doing It Now or Later. *American Economic Review* **89**, 103-124.
- Palacios-Huerta, I., 2002, Consumption and Portfolio Rules Under Hyperbolic Discounting, Working Paper, Brown University.
- Smith, W., 1999, Risk, the Spirit of Capitalism, and Long Term Growth: the Implications of a Preference for Capital. *Journal of Macroeconomics* **21**, 241-261.
- Smith, W., 2001, How Does the Spirit of Capitalism Affect Stock Market Prices. *Review of Financial Studies* **14**, 1215-1232.
- Svensson, L., 1989, Portfolio Choice with Non-expected Utility in Continuous Time. *Economics Letters* **30**, 313-317.
- Turnovsky, S., 2000a, *Methods of Macroeconomic Dynamics*, 2nd ed., MIT Press.
- Turnovsky, S., 2000b, Government Policy in a Stochastic Growth Model with Elastic Labor Supply. *Journal of Public Economic Theory* **2**, 389-433.

Turnovsky, S. 2000c, Fiscal Policy, Elastic Labor Supply, and Endogenous Growth. *Journal of Monetary Economics* **45**, 185-210.

Weber, M., 1958, The Protestant Ethic and the Spirit of Capitalism (Charles Scribner's Sons, New York).

Weil, P., 1989, Non-expected Utility in Macroeconomics. *Quarterly Journal of Economics* **105**, 29-42.

Zou, H., 1994, The 'Spirit of Capitalism' and Long Run Growth. *European Journal of Political Economy* **10**, 279-293.

Zou, H., 1995, The Spirit of Capitalism and Savings Behavior. *Journal of Economic Behavior and Organization* **28**, 131-143.

October 2006

The Spirit of Capitalism, Precautionary Savings, and Consumption

by

Yulei Luo,^{*} William T. Smith,^{**} Heng-fu Zou^{***}

Recent research has shown that the “spirit of capitalism” – a preference for wealth itself, in addition to consumption – has important implications for growth and asset pricing. This paper explores how the spirit of capitalism affects saving and consumption behavior. We demonstrate that the spirit of capitalism may reduce the importance of precautionary savings. It can also explain the excess sensitivity puzzle: the spirit of capitalism causes dramatic deviations from a random walk. It may also offer a partial explanation of the excess smoothness puzzle.

Key words: The Spirit of Capitalism, Precautionary Savings, The Excess Sensitivity and

Smoothness of Consumption

JEL classification: D11, E21

^{*}The University of Hong Kong, ^{**}University of Memphis, ^{***}World Bank, Peking University, and Wuhan University. Address correspondence to W.T. Smith, Department of Economics, Fogelman College of Business & Economics, University of Memphis, Memphis, TN 38152. Phone: (901) 678-3675 e-mail: wtsmith@memphis.edu

The great sociologist Weber (1948) viewed the desire to accumulate wealth as an end in itself as the defining characteristic of capitalist societies. Recently, this notion of the "spirit of capitalism" has been used to address a range of issues, including savings [Zou (1995)], growth [Gong and Zou (2001, 2002), Yang and Zou (2001), Zou (1994), Smith (1999)], asset pricing [Bakshi and Chen (1996), Smith (2001), Yang and Zou (2001)] and the distribution of wealth [Luo and Young (2002)]. In this paper we extend the inquiry to explore how the spirit of capitalism affects precautionary savings and the dynamics of consumption.

Section I of the paper develops a very general model of precautionary savings in the presence of the capitalist spirit. As in most models of precautionary savings, we consider a consumer who knows his current wage but is uncertain about his future wage income.¹ Following Zou (1994) we model of spirit of capitalism by allowing the consumer to derive utility from wealth itself, in addition to consumption. We use a very general utility function, and assume only that income is a discrete-time diffusion process. Using the methods of Grossman and Shiller (1982), we derive an *exact* expression for the expected growth of consumption in the continuous-time limit. This is of some interest in itself, since most of the literature on consumption relies upon approximations [Baxter and Jermann (1999), Lettau and Ludvigson (2001), Gourinchas and Parker (2001)], or restrictive assumptions about preferences [constant absolute risk aversion (CARA) as in [Blanchard and Mankiw (1988), Hall (1988), Caballero (1990), Alessie and Lusardi (1997), and Smith (1998,2002)], or the income process [for example, log-normality in Hansen and Singleton (1983) and Carroll (1992)]. We demonstrate that the spirit of capitalism causes the random walk hypothesis [Hall (1978, 1988)] to fail: (nonstochastic) changes in wealth can be used to predict changes in consumption if there is a spirit of capitalism.

¹ See excellent survey by Carroll (2000). Other important contributions are [Blanchard and Mankiw (1988), Caballero (1990, 1991), Dynan (1993), Carroll (1992), Gourinchas and Parker (1999), Irvine and Yang (1995), Kimball and Mankiw (1989), Skinner (1988), Weil (1993), Zeldes (1993)].

To shed more light on how the spirit of capitalism affects savings and consumption Section II develops a simple model that allows a closed form solution: consumers have constant absolute risk aversion (CARA) preferences over consumption and wealth, and income is AR(1). This model has interesting implications for both precautionary savings and consumption dynamics. First, the capitalist spirit usually (unless income is stationary and converges to its ergodic distribution very rapidly) reduces the precautionary savings premium.

Second, the spirit of capitalism can explain the *excess sensitivity* of consumption to anticipated changes in income [Flavin (1981), Campbell and Mankiw (1989), Deaton (1992)]: When there is a spirit of capitalism, the growth of consumption can be predicted by expected changes in income. This is consistent with the empirical results of Campbell and Mankiw (1989), who regress consumption growth on income growth. The coefficient attached to income growth in their regression is large and statistically significant. They interpret this coefficient as the proportion of “rule-of-thumb” consumers” in the economy. Our model suggests that this coefficient can also be interpreted as a measure of the strength of the spirit of capitalism.

Third, the spirit of capitalism may offer a partial explanation of the excess sensitivity of consumption growth to unanticipated changes in income. Friedman's famous (1957) permanent income hypothesis suggests that consumption should be smoother than income. However, Campbell and Deaton (1989) and Deaton (1992) argue convincingly that income is non-stationary. In this case, the permanent income hypothesis predicts that innovations to income should be associated with larger innovations in consumption. This is the *excess smoothness* paradox: if income is nonstationary, observed consumption growth is much too smooth, relative to the consumption path predicted by the permanent income hypothesis. In our model, the spirit of capitalism mitigates the effect of an income innovation on consumption growth as long as income is non-stationary.

A plausible calibration of the model suggests that by itself the spirit of capitalism can only partially explain excess smoothness. In Section III, therefore, we incorporate two new features that heighten the ability of the spirit of capitalism to account for excess smoothness: the first is a more realistic income process; the second is uncertainty about the interest rate.

Section V offers some concluding thoughts.

I. The Model

Suppose that time is divided into discrete intervals of length Δt (we will be considering the limit as $\Delta t \rightarrow 0$). Imagine a consumer with an infinite planning horizon and a constant rate of time preference $\theta > 0$. He maximizes the lifetime expected utility of time-separable preferences defined over consumption c_t and wealth w_t :

$$E_0 \sum_{t=0}^{\infty} e^{-\theta t} U(c_t, w_t) \Delta t. \quad (1)$$

We assume that $U(c_t, w_t)$ is twice continuously differentiable in c_t and w_t , and that

$$U_c > 0, U_w > 0, U_{cc} < 0, \text{ and } U_{ww} < 0.$$

The consumer can borrow and lend at the riskless rate of return r , and receives income (or more generally, non-asset income) of y_t in each period. His budget constraint is therefore

$$w_{t+\Delta t} = (w_t + y_t \Delta t - c_t \Delta t)(1 + r \Delta t) \quad (2)$$

In keeping with most of the literature on precautionary savings [Carroll (2001)], we assume that the consumer knows his income in each period, but is uncertain about its future evolution. To use Merton's (1975) terminology, there is "future" uncertainty, but not current uncertainty.² We employ a very general income process, assuming only that it is a discrete-time diffusion

$$\Delta y_t = y_{t+\Delta t} - y_t = \mu_{y,t} \Delta t + \sigma_{y,t} \Delta z_t, \quad (3)$$

² Gourinchas and Parker (1999) and Turnovsky and Smith (2006) also allow for current uncertainty.

where Δz_t is the increment to the Wiener process z_t . The conditional expectation and standard deviation of the growth of income, $\mu_{y,t}$ and $\sigma_{y,t}$, may be time-varying.

The consumer maximizes the expected lifetime utility in Equation (1), subject to the budget constraint in Equation (2), given initial wealth w_0 and the income process in Equation (3). The Euler equation for this problem is

$$1 = e^{-\theta\Delta t} E_t \left[\frac{U_c(c_{t+\Delta t}, w_{t+\Delta t})}{U_c(c_t, w_t)} (1 + r\Delta t) + \frac{U_w(c_{t+\Delta t}, w_{t+\Delta t})}{U_c(c_t, w_t)} \right]. \quad (4)$$

Using the limiting arguments of Grossman and Shiller (1982), we can use this first-order condition to infer the stochastic process for the optimal consumption path. Define $\sigma_{c,t}$ as the (possibly time-varying) instantaneous variance of the growth of consumption. In Appendix A we prove

Proposition 1. *In the continuous-time limit, as $\Delta t \rightarrow 0$, the expected growth of consumption is*

$$E_t dc_t = \left[-\frac{U_c(c_t, w_t)}{U_{cc}(c_t, w_t)} (\theta - r) - \frac{U_{cw}(c_t, w_t)}{U_{cc}(c_t, w_t)} \frac{dw_t}{dt} - \frac{U_{ccc}(c_t, w_t)}{U_{cc}(c_t, w_t)} \frac{\sigma_{c,t}^2}{2} \right] dt. \quad (5)$$

The first term in Equation (5) is the familiar continuous-time Euler equation in nonstochastic models: the consumption profile in non-stochastic models depends upon the difference between the interest rate and the rate of time preference. The last term is a precautionary savings premium [Carroll (1992), Carroll and Kimball (1997)] that changes the slope of the consumption profile. The innovations here are two-fold.

First, if there is a spirit of capitalism ($U_{cw} \neq 0$) then the expected growth of consumption depends upon the (instantaneously non-stochastic) growth in wealth. In other words, growth in wealth can be used to predict growth in consumption. Thus, *the spirit of capitalism causes the random walk hypothesis to fail.*

Second, expressions like Equation (5) are common in the consumption literature [for example Carroll (1992), Baxter and Jermann (1999), Campbell (1994), Lettau and Ludvigson

(2001)] as *log-linear approximations*. Log-linear Euler equations are very popular in solving stochastic general equilibrium models in macroeconomics. Other papers arrive at closed form solutions, but impose restrictive assumptions about either preferences or the income process. Here, however, this relationship holds *exactly* in the continuous-time limit, for a very general income process and a very general class of preferences.

II. CARA Preferences

In order to shed more light on how the spirit of capitalism affects the precautionary premium and dynamics of consumption it will be useful to consider an example that permits a closed-form solution. To this end, suppose that time is continuous and that the felicity function is of the following form:

$$-\frac{1}{a} E_0 \int_0^{\infty} e^{-\theta t - ac_t - bw_t} dt. \quad (6)$$

The parameter a is the coefficient of absolute risk aversion with respect to consumption. The parameter $b \geq 0$ measures the strength of the spirit of capitalism. If $b = 0$ there is no spirit of capitalism, we recover the familiar case of constant absolute risk aversion (CARA) defined over consumption alone.

In continuous-time, the consumer's budget constraint is

$$dw_t = (rw_t + y_t - c_t)dt. \quad (7)$$

Income is a continuous-time, first-order autoregressive (Ornstein-Uhlenbeck) process

$$dy_t = \rho \left(\frac{\mu}{\rho} - y_t \right) dt + \sigma dz_t. \quad (8)$$

The steady-state mean of income is $\bar{y} = \mu / \rho$. The parameter ρ governs the speed of convergence (or divergence) from the steady state. If $\rho > 0$ the process is stationary; deviations of income from the steady state are temporary. If $\rho < 0$ the process is non-stationary and innovations to income are "super-permanent." This last case catches the flavor of Campbell

and Deaton's (1992) argument that income is non-stationary.³ We will introduce a more realistic income process later, in Section III.A.

The consumer chooses a consumption policy to maximize expected lifetime utility in Equation (6), subject to the budget constraint in Equation (7), and given the income process in Equation (8). If $b = \rho = 0$ this reduces to the canonical precautionary savings problem in Blanchard and Mankiw (1988), and Hall (1988). If $b = 0$ but $\rho > 0$ it becomes a continuous-time version of the precautionary savings problem with autoregressive income of Caballero (1990).

We demonstrate in Appendix B that the solution to this problem is the consumption function

$$c(w_t, y_t) = \Omega(w_t, y_t) - \Gamma \frac{\sigma^2}{2} \quad (9)$$

where

$$\Omega(w_t, y_t) = \frac{\theta - r - b/a}{ar + b} + \frac{1}{r + b/a + \rho} \mu + \frac{r + b/a}{r + b/a + \rho} y_t + r w_t \quad (10)$$

and

$$\Gamma = a \frac{r + b/a}{(r + b/a + \rho)^2} \quad (11)$$

This solution shares features common to all CARA models of precautionary savings [Blanchard and Mankiw (1988), Hall (1988), Caballero (1990), Alessie and Lusardi (1997), Smith (1998, 2002)]. First, consumption decomposes into two parts, certainty-equivalent

³ Campbell and Deaton (1992) argument argue that income is a non-stationary *second-order* process. It is not possible to derive a closed-form solution to the precautionary savings problem with a second-order process, so we have focused on the first-order process in Equation (3). This is tractable, yet allows a form of non-stationarity.

consumption $\Omega(w_t, y_t)$ and a risk adjustment $-\Gamma \frac{\sigma^2}{2}$.⁴ Second, certainty-equivalent

consumption is a linear function of financial wealth w_t and wage income y_t .

Notice that this solution is identical to the consumption function without the spirit of capitalism but with an interest rate of $r + b/a$ rather than r . In other words, the spirit of capitalism has the effect of raising the interest rate: r is the market rate, while $r + b/a$ is the effective “psychological” rate at which the consumer discounts wage income.

It is also illuminating to express the consumption function in terms of human wealth. Following the literature, we can define human wealth as the expected present value of future labor income discounted at the appropriate interest rate. In the presence of the spirit of capitalism the effective rate of interest is $r + b/a$. Therefore human wealth is

$$h_t = E_t \int_t^\infty e^{-(r+b/a)(s-t)} y_s ds. \quad (12)$$

As shown in Cox, Ingersoll, and Ross (1985), straightforward calculations imply that⁵

$$h_t = \frac{1}{r + b/a + \rho} \left(y_t + \frac{\mu}{r + b/a} \right) \quad (13)$$

⁴ It is “certainty-equivalent” in the sense that it is the consumption predicted by a non-stochastic model with CARA utility. “Certainty-equivalent” is often used to describe linear-quadratic preferences, which do not generate a precautionary savings premium.

⁵ We use the fact that

$$E_t[y_{i,s}] = y_{i,t} e^{-\rho_i(s-t)} + \mu_i (1 - e^{-\rho_i(s-t)}) \text{ for } i = 1, 2.$$

If $r + b/a + \rho < 0$ the integral in Equation (12) diverges, so that human wealth is undefined.

Henceforth we therefore assume that $r + b/a + \rho > 0$. In other words, income cannot be “too” nonstationary.

Using Equation (13), the consumption function in equations (9) – (11) can then be rewritten as

$$c(w_t, y_t) = \frac{1}{a} \frac{\theta - r - b/a}{r + b/a} + (r + b/a)h_t + rw_t - \Gamma \frac{\sigma^2}{2} \quad (14)$$

Consumption is linear in financial wealth and human wealth.

We will now explore how the spirit of capitalism (captured by the parameter “ b ”) affects the precautionary premium and the time-series properties of consumption.

II.A The Precautionary Savings Premium

Consider first the precautionary savings premium. In the absence of the capitalist spirit ($b = 0$), the precautionary savings premium is simply

$$P_{b=0} = \frac{r}{(r + \rho)^2} a \frac{\sigma^2}{2}. \quad (15)$$

Note for future reference that $\partial P_{b=0} / \partial r \leq 0$ as $r \geq \rho$. That is, an increase in the interest rate will decrease or increase the premium depending upon whether $r > \rho$ is positive or negative. Nonstationary income ($\rho < 0$) is a sufficient condition for an increase in the interest rate to reduce the precautionary premium.

Things are quite different if there is a capitalist spirit ($b > 0$). In this case the premium is

$$P_{b>0} = a \frac{r + b/a}{(r + b/a + \rho)^2} \frac{\sigma^2}{2} \quad (16)$$

Recall that the “effective” interest rate in the presence of the spirit of capitalism is $r + b/a$ rather than r . It follows that the spirit of capitalism may decrease or increase the precautionary

premium depending upon whether the effective rate of interest exceeds the rate of time preference: It follows that the spirit of capitalism may decrease or increase the precautionary premium depending upon whether the effective rate of interest exceeds the rate of time preference: That is, $\partial P_{b>0} / \partial b \leq 0$ as $r + b/a \geq \rho$. If income is non-stationary ($\rho < 0$) then the premium always decreases with the spirit of capitalism. If income is stationary ($\rho > 0$), and if $r + b/a \geq \rho$, then the premium still decreases with the strength of the spirit of capitalism. However, if income is stationary and $r + b/a < \rho$ then the premium will actually decrease with the spirit of capitalism.⁶ We now have

Proposition 2. *An increase in the spirit of capitalism always lowers the precautionary premium if income is non-stationary. If income is stationary, then the precautionary premium initially increases with the spirit of capitalism, and then falls.*

II.B Excess Sensitivity and Excess Smoothness

The spirit of capitalism has important implications for the two fundamental puzzles of consumption dynamics, excess sensitivity and excess smoothness. To see why, consider first the case where there is no spirit of capitalism. Using Equations (9), (10) and (11) and setting $b = 0$, the growth of consumption is simply

$$dc_t = r \left[\frac{r - \theta}{ar} + a \frac{r}{(r + \rho)^2} \frac{\sigma^2}{2} \right] dt + \frac{r}{r + \rho} \sigma dz_t \quad (17)$$

Note two properties of consumption growth in this benchmark case. First, it reflects Hall's (1978, 1988) classic result that that consumption should be a random walk under rational expectations. As shown by Caballero (1990), Hall's conclusion is not affected by the persistence of income: Current and lagged consumption and income cannot help predict the growth of consumption. In fact, it has been documented again and again [Flavin (1981),

⁶ It is straightforward to show that when

Campbell and Mankiw (1989), and Deaton (1992), to mention just some classic references] that changes in income predict changes in consumption. This is the *excess sensitivity* puzzle.

Second, the innovation to consumption is equal the annuity value of the innovation to income [also a result due to Caballero (1990)]. This implies that if income is stationary ($\rho > 0$) the variance of consumption growth is less than the variance of income growth, as one would expect from the consumption smoothing suggested by the permanent income hypothesis. If income is non-stationary ($\rho < 0$) however, the variance of consumption growth *exceeds* the variance of income growth. This leads to the *excess smoothness puzzle*, or Deaton (1992) paradox: if income is non-stationary, observed consumption growth is actually too smooth relative to what the permanent income hypothesis predicts.

How does the spirit of capitalism ($b > 0$) alter these predictions? The expected growth of consumption is now

$$E_t dc_t = r \left[\frac{r + b/a - \theta}{ar} + a \frac{r + b/a}{(r + b/a + \rho)^2} \frac{\sigma^2}{2} \right] - \frac{b}{a} dw_t. \quad (18)$$

If there is a capitalist spirit, so that $b > 0$, the expected change in consumption can be predicted by the growth in wealth. This is a special case of Proposition 1, Equation (4) in the general model in Section 1.

To develop further insights about how the spirit of capitalism affects consumption dynamics, it is useful to use the budget constraint in Equation (7) to rewrite consumption growth as

$$dc_t = r \left[\frac{1}{a} \frac{r - \theta + b/a}{r + b/a} + \Gamma \frac{\sigma^2}{2} \right] dt + \frac{b/a}{r + b/a + a} E_t dy_t + \frac{r + b/a}{r + b/a + \rho} \sigma dz_t \quad (19)$$

Consumption is no longer a random walk when there is a capitalist spirit: the anticipated growth of wage income can be used to predict changes in consumption when $b > 0$. We therefore have

Proposition 3. *The spirit of capitalism can explain the excess sensitivity of consumption to income.*

This suggests an alternative interpretation of the empirical results of Campbell and Mankiw (1989). They regress consumption growth on income growth. The estimated coefficient attached to income growth in this regression is large and statistically significant. They interpret this number as the proportion of “rule-of-thumb” consumers” in the economy. Proposition 3 implies that this coefficient can also be viewed as a measure of the strength of the spirit of capitalism.

Finally, consider the excess smoothness puzzle. In the absence of the spirit of capitalism, when $b = 0$, the growth of consumption is given by Equation (17). Hence, the standard deviation of consumption growth is

$$std(dc_t) = \frac{r}{r + \rho} \sigma. \quad (20)$$

From Equation (17), however, the standard deviation of consumption growth in the presence of the spirit of capitalism is

$$std(dc_t) = \frac{r + b/a}{r + b/a + \rho} \sigma. \quad (21)$$

To show how the spirit of capitalism affects excess smoothness, we define the excess smoothness ratio as⁷

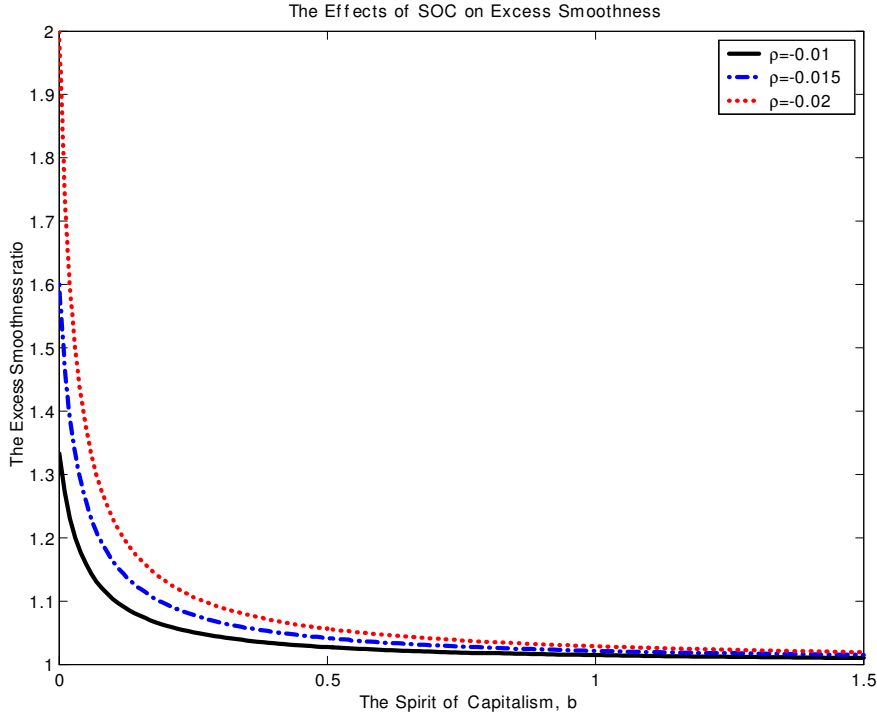
$$\lambda = \frac{std(dc_t)}{std(dy_t)} = \frac{r + b/a}{r + b/a + \rho} \quad (22)$$

Since absolute risk aversion $a > 0$, it follows immediately that $\lambda > 1$ as $\rho > 0$. Indeed, $\partial \lambda / \partial b > 0$ as $\rho > 0$. This implies that the spirit of capitalism mitigates the volatility of consumption growth when income is non-stationary. Therefore,

⁷ Note that $std(dy_t) = \sigma^2$.

Proposition 4. *The spirit of capitalism can explain the excess smoothness puzzle, by reducing the volatility of consumption growth when income is non-stationary.*

The following figure plots the relationship between the excess smoothness ratio and the spirit of capitalism (b) when labor income is non-stationary. It is obvious from this figure that the spirit of capitalism can reduce the ratio of the standard deviation of consumption growth to the standard deviation of labor income growth, that is, the excess smoothness ratio. As b increases, the ratio converges to 1. In the US aggregate data, the ratio is close to .58. Hence, the spirit of capitalism itself cannot resolve the excess smoothness puzzle in this simple model. In the next Section, we will show how the spirit of capitalism can help resolve this puzzle in two, more realistic, setups.



III. Extensions

In this section we enrich the basic model by incorporating two new features. First we introduce a more realistic income process. Second, we allow for interest rate risk. Both extensions enhance the ability of the spirit of capitalism to explain the excess smoothness puzzle.

III.A Extension 1: A More Realistic Labor Income Process

In this section, we consider a more realistic labor income process. The empirical literature often specifies labor income as a sum of two distinct components: one is a permanent (or very persistent) process, for example, a unit-root process, and the other is a transitory process, for example, a white noise process.⁸ Here we specify labor income as

$$y_t = y_{1,t} + y_{2,t}. \quad (23)$$

Thus

$$dy_t = dy_{1,t} + dy_{2,t} \quad (24)$$

where we assume

$$\begin{aligned} dy_{1,t} &= \rho_1 \left(\frac{\mu_1}{\rho_1} - y_{1,t} \right) dt + \sigma_1 dz_{1,t} \\ dy_{2,t} &= \rho_2 \left(\frac{\mu_2}{\rho_2} - y_{2,t} \right) dt + \rho_{12} \sigma_2 dz_{1,t} + \sqrt{1 - \rho_{12}^2} \sigma_2 dz_{2,t} \end{aligned} \quad (25)$$

and $z = (z_1, z_2)'$ is a standard Brownian motion in R^2 . ρ_{12} is the instantaneous correlation coefficient between the two labor income components, and the parameters ρ_1 and ρ_2 measure the persistence of the two individual components of labor income, respectively. Without loss of generality, we assume that $y_{1,t}$ is more persistent than $y_{2,t}$, that is, $\rho_1 < \rho_2$.

Following the same procedure used in the benchmark model, we can derive the consumption function as follows

$$c_t = \Omega(w_t, y_{1,t}, y_{2,t}) - P_{b>0} \quad (26)$$

where

⁸ See Pischke (1995).

$$\begin{aligned}\Omega(w_t, y_{1,t}, y_{2,t}) = & rw_t + \frac{r+b/a}{r+b/a+\rho_1} y_{1,t} + \frac{r+b/a}{r+b/a+\rho_2} y_{2,t} \\ & + \frac{1}{r+b/a+\rho_1} \mu_1 + \frac{1}{r+b/a+\rho_2} \mu_2 + \frac{1}{a} \frac{\theta-r-b/a}{r+b/a}\end{aligned}\quad (27)$$

and

$$P_{b>0} = a \frac{r+b/a}{(r+b/a+\rho_1)^2} \frac{\sigma_1^2}{2} + a \frac{r+b/a}{(r+b/a+\rho_2)^2} \frac{\sigma_2^2}{2} + 2a\rho_{12} \frac{r+b/a}{(r+b/a+\rho_1)(r+b/a+\rho_2)} \frac{\sigma_1\sigma_2}{2}\quad (28)$$

The appropriate measure of human wealth in this case turns out to be⁹

$$h_t = \frac{1}{r+b/a+\rho_1} \left(y_{1,t} + \frac{\mu_1}{r+b/a} \right) + \frac{1}{r+b/a+\rho_2} \left(y_{2,t} + \frac{\mu_2}{r+b/a} \right)\quad (29)$$

This allows the consumption function to be expressed as

$$c_t = \frac{1}{a} \frac{\theta-r-b/a}{r+b/a} + (r+b/a)h_t + rw_t - P_{b>0}\quad (30)$$

This is the same as Equation (14) in the simple model, except for two things: first, the more complicated expression in Equation (29) has been substituted Equation (13) for human wealth; second, the precautionary premium in Equation (28) now involves the variances and covariances of the two shocks, as well as their autoregressive parameters.

Implications for Precautionary Savings

Based on Equation (28), it is straightforward to show that (i) the higher the coefficient for absolute risk aversion (a), the larger the precautionary premium, (ii) more persistent (lower ρ_i) income process or more volatile (σ_i^2) income shock induces larger premium, *ceteris paribus*. Furthermore, an instantaneous positive (negative) correlation ρ_{12} increases (decreases) the total exposure of labor income risk and thus increase the precautionary savings.

⁹ In this case we assume $r+b/a+\rho_i > 0, i=1,2$ to ensure that human wealth is well-defined.

As shown in the benchmark case, the first two terms in Equation (28) mean that the spirit of capitalism lowers the precautionary premium if $\rho_i < 0$ or if the speed of convergence of income is not too rapid ($r + b/a \geq \rho_i$). If the speed of convergence is very rapid ($r + b/a < \rho_i$), then the premium initially increases with the spirit of capitalism and then falls. The third term in Equation (28) means that if the two components in income are positively correlated ($\rho_{12} > 0$), the spirit of capitalism has the same effects on precautionary savings as in the first two terms, while the spirit of capitalism has reversed effects on precautionary savings if ($\rho_{12} < 0$).

Implications for Excess Sensitivity and Excess Smoothness

We can now derive the expression for consumption growth as follows

$$\begin{aligned} dc_t &= rdw_t + a \frac{r+b/a}{r+b/a+\rho_1} dy_{1,t} + \frac{r+b/a}{r+b/a+\rho_2} dy_{2,t} \\ &= r \left[-\frac{1}{a} \frac{\theta - r - b/a}{r+b/a} + P_{b>0} \right] dt + \left[\frac{b/a}{r+b/a+\rho_1} y_{1,t} + \frac{b/a}{r+b/a+a\rho_2} y_{2,t} \right] dt \\ &\quad + \left[\frac{r+b/a}{r+b/a+\rho_1} \sigma_1 dz_{1,t} + \frac{r+b/a}{r+b/a+\rho_2} (\rho_{12}\sigma_2 dz_{1,t} + \sqrt{1-\rho_{12}^2} \sigma_2 dz_{2,t}) \right] \end{aligned} \quad (31)$$

Hence, the spirit of capitalism can explain the excess sensitivity of consumption to income, that is, consumption growth can be predicted by expected income growth. Furthermore, the spirit of capitalism can also mitigate the excess smoothness puzzle. To see this, note that in this case

$$\begin{aligned} std(dy_t) &= std(dy_{1,t} + dy_{2,t}) \\ &= \sqrt{\sigma_1^2 + \sigma_2^2 + 2\rho_{12}\sigma_1\sigma_2} \end{aligned} \quad (32)$$

and

$$std(dc_t) = \sqrt{\left(\frac{r+b/a}{r+b/a+\rho_1} \sigma_1 + \frac{r+b/a}{r+b/a+\rho_2} \rho_{12}\sigma_2 \right)^2 + (1-\rho_{12}^2)\sigma_2^2}. \quad (33)$$

We can then write the excess smoothness ratio as follows

$$\lambda = \sqrt{\frac{\left(\frac{r+b/a}{r+b/a+\rho_1} \sigma_1 + \frac{r+b/a}{r+b/a+\rho_2} \rho_{12}\sigma_2 \right)^2 + (1-\rho_{12}^2)\sigma_2^2}{\sigma_1^2 + \sigma_2^2 + 2\rho_{12}\sigma_1\sigma_2}} \quad (34)$$

The properties of the income process affect excess smoothness of consumption in complicated ways. For example, if the two components have the same persistence ($\rho_1 = \rho_2$),

$$\lambda = a \frac{r+b/a}{r+b/a+\rho_1}. \quad \text{This means that, as in the benchmark case, the spirit of capitalism does}$$

reduce the smoothness of consumption growth, but it cannot eliminate the excess smoothness

puzzle because $\lambda \geq 1$ when $\rho_1 < 0$. However, if the two components have different

degrees of persistence (without loss of generality, we suppose that $\rho_1 < \rho_2$), λ could be

less than 1 if the spirit of capitalism is strong.

We can see this from a simple numerical example. For simplicity, assume that the two components are perfectly correlated ($\rho_{12} = 1$). The excess smoothness ratio becomes

$$\lambda = \frac{r+b/a}{r+b/a+\rho_1} \frac{\sigma_1 + \frac{r+b/a+\rho_1}{r+b/a+\rho_2} \sigma_2}{\sigma_1 + \sigma_2}. \quad (35)$$

Note that since $\rho_1 < \rho_2$, $\frac{\sigma_1 + \frac{r+b/a+\rho_1}{r+b/a+\rho_2} \sigma_2}{\sigma_1 + \sigma_2} < 1$. In the absence of a spirit of capitalism

$$\lambda = \frac{r}{r+\rho_1} \frac{\sigma_1 + \frac{r+\rho_1}{r+\rho_2} \sigma_2}{\sigma_1 + \sigma_2}. \quad (36)$$

Setting $a = 1$, $r = 0.04$, $\rho_1 = -0.02$, and $\rho_2 = 0$, we get $\lambda > 1$ for any positive

values of σ_1 and σ_2 . Introducing the spirit of capitalism, λ becomes less than 1

for some plausible values of σ_1 and σ_2 . For example, when $\sigma_1 = 0.03$ and

$\sigma_2 = 0.02$, without the spirit of capitalism, $\lambda = 1.6 > 1$, while $\lambda = 0.82 < 1$ in the

presence of soc. In sum, in this model where there are two distinct components in income

process, λ converges to $\frac{\sigma_1 + \frac{r+\rho_1}{r+\rho_2} \sigma_2}{\sigma_1 + \sigma_2} < 1$ when the spirit of capitalism is strong enough.

Hence, in some cases with our more realistic income process, the spirit of capitalism could help resolve the excess smoothness puzzle.

III.B Extension 2: Interest Rate Risk

So far we have assumed consumers only face labor income risk and they smooth their consumption over time by borrowing or lending at a constant risk-free interest rate. However, in reality, consumers also face substantial risk for holding financial wealth that would largely affect their optimal consumption and saving decisions. In this section, we will explore the implications of soc for precautionary saving and consumption dynamics in the model with both labor income risk and interest rate risk. In this case, the consumer's budget constraint becomes

$$dw_t = (rw_t + y_t - c_t)dt + d\omega_t \quad (37)$$

where ω_t is a Brownian motion with $E[d\omega] = 0$ and $\text{var}[d\omega] = \sigma^2$, and summarizes interest rate risk. Further, interest rate risk is instantaneously correlated with labor income risk, that is, $\rho_{wy} \neq 0$.

In Appendix C we show that the solution to this problem is:

$$c_t = \Omega(w_t, y_t) - P_{b>0}, \quad (38)$$

where

$$\Omega(w_t, y_t) = \frac{1}{a} \frac{\theta - r - b/a}{r + b/a} + \frac{r + b/a}{r + b/a + \rho} y_t + rw_t + \frac{1}{r + b/a + \rho} \mu \quad (39)$$

and the precautionary saving premium is

$$P_{b>0} = a \frac{r + b/a}{(r + b/a + \rho)^2} \frac{\sigma^2}{2} + a^2 (r + b/a) \frac{\sigma^2}{2} + a^2 \frac{r + b/a}{r + b/a + \rho} \rho_{wy} \sigma \sigma. \quad (40)$$

Implications for Precautionary Savings

Without the spirit of capitalism, the precautionary saving premium becomes

$$P_{b=0} = \frac{ar}{(r + \rho)^2} \frac{\sigma^2}{2} + a^2 r \frac{\sigma^2}{2} + \frac{a^2 r}{r + \rho} \rho_{wy} \sigma \sigma. \quad (41)$$

The second term means that the second risk increases the precautionary premium because the consumers face more risk than in our benchmark model, and the third term implies that the

correlation between the two risks could affect the precautionary savings premium. If the two risks are positively correlated, that is, $\rho_{wy} > 0$, mean reversion ($\rho > 0$) tends to reduce precautionary savings, while non-stationarity ($\rho < 0$) tends to increase precautionary savings when $\rho < r$. However, if the two risks are negatively correlated, the effects of ρ on precautionary savings would be reversed.

For the case with the spirit of capitalism, equation (39) implies that the additional financial risk contributes to the precautionary savings premium by the following two terms

$$a^2(r + b/a)\frac{\sigma^2}{2} + a^2 \frac{r + b/a}{r + b/a + \rho} \rho_{wy} \sigma \sigma. \quad (42)$$

The first term implies that a more volatile financial risk induces larger precautionary saving demand. The second term means that a positive (negative) correlation between the two risks increases (decreases) the total risk exposure of income risk and thus induces a larger (smaller) precautionary premium. It is obvious that the spirit of capitalism increases the impacts of σ^2 and reduces the impacts of ρ_{wy} on the precautionary premium.

Implications for Excess Sensitivity and Excess Smoothness

Setting $b = 0$, the growth of consumption is then

$$dc_t = r \left[\frac{r - \theta}{ar} + P_{b=0} \right] dt + \frac{r}{r + \rho} \sigma dz_t + rd\omega_t, \quad (43)$$

while with the spirit of capitalism, it becomes

$$dc_t = r \left[\frac{1}{a} \frac{r - \theta + b/a}{r + b/a} + P_{b>0} \right] dt + \frac{b/a}{r + b/a + \rho} E_t dy_t + \frac{r + b/a}{r + b/a + \rho} \sigma dz_t + rd\omega_t. \quad (44)$$

Hence, in the first case, the excess smoothness ratio is

$$\begin{aligned}\lambda &= \frac{std(dc_t)}{std(dy_t)} \\ &= \sqrt{\left(\frac{r}{r+\rho}\right)^2 + r^2\left(\frac{\varpi}{\sigma}\right)^2 + 2\rho_{wy} \frac{r^2}{r+\rho} \frac{\varpi}{\sigma}};\end{aligned}\tag{45}$$

and in the second case, the ratio is

$$\lambda = \sqrt{\left(\frac{r+b/a}{r+b/a+\rho}\right)^2 + r^2\left(\frac{\varpi}{\sigma}\right)^2 + 2\rho_{wy} \frac{(r+b/a)r}{r+b/a+\rho} \frac{\varpi}{\sigma}}.\tag{46}$$

Note that since $\rho_{wy} \in [-1,1]$, $\lambda \in \left[\frac{r+b/a}{r+b/a+\rho} - r\frac{\varpi}{\sigma}, \frac{r+b/a}{r+b/a+\rho} + r\frac{\varpi}{\sigma}\right]$. Hence, when

$\rho < 0$, both negative correlation between labor income risk and interest rate risk and the spirit of capitalism reduce the excess smoothness ratio. As documented in Campbell and Viceira (2002), in the US data (CRSP data on the NYSE value-weighted stock return relative to the Treasury bill rate), the correlations between labor income and stock returns are positive for all education groups (0.328 for the group with no high school education, 0.371 for the group with high school education, and 0.516 for the group with college education). However, the stock returns used in their study are not equivalent to the stochastic process for the interest rate used in this paper. We haven't modeled the stochastic process for the interest rate explicitly, so it is difficult to find the empirical counterpart of this process. Theoretically, the correlation could be any value between -1 and 1. Therefore, for given r and a , incorporating soc could lower the ratio λ to a value less than 1 in the presence of interest rate risk and thus resolve the excess smoothness puzzle.

IV. Conclusion

Ever since Weber (1948), the spirit of capitalism has been recognized by sociologists as an essential aspect of modern, capitalist economies. It has only been over the last decade, however, that – as part of a broader effort to address problems of envy and “keeping-up-with the Joneses” – economists have formalized this notion. The modern economic literature on the

spirit of capitalism has investigated its implications for growth and asset-pricing. Conspicuously absent from this literature has been how the spirit of capitalism affects consumption and savings behavior under uncertainty. In this paper we fill this gap by incorporating the spirit of capitalism into a model of precautionary savings.

Our basic model with a simple, AR1 income process suggests that the capitalism spirit may increase or decrease the precautionary premium, depending upon the degree of non-stationarity of income. It also shows that the spirit of capitalism provides a simple explanation excess smoothness: it generates dramatic deviations from the random walk hypothesis. By itself, the spirit of capitalism does not provide a plausible explanation for excess smoothness. However, we show in richer models with either more complicated income processes or interest rate risk, the spirit of capitalism may also resolve the excess smoothness puzzle as well.

Appendix A:

Derivation of Proposition 1

The derivation employs the limiting argument pioneered by Grossman and Shiller (1982), and more recently employed by Bakshi and Chen (1996a, 1996b) and Smith (2001).

First, assume that the optimal consumption policy is a discrete-time diffusion:

$$\Delta c_t = c_{t+\Delta t} - c_t = \mu_{c,t} \Delta t + \sigma_{c,t} \Delta z_{c,t} \quad (\text{A.1})$$

For convenience, express the (instantaneously non-stochastic) growth in wealth using analogous notation:

$$\Delta w_t = \mu_{w,t} \Delta t. \quad (\text{A.2})$$

Now consider the Euler equation in Equation (4). Take a second-order Taylor series of the right-hand side around $\Delta t = 0$, c_t , and w_t . Using Equations (A.1) and (A.2), this leads to

$$0 = U_c(c_t, w_t)(r - \theta) + U_w(c_t, w_t)U_{cc}(c_t, w_t)\mu_{c,t}\Delta t + U_{cw}(c_t, w_t)\mu_{w,t}\Delta t + U_{ccc}(c_t, w_t)\frac{E_t \Delta c_t^2}{2}. \quad (\text{A.3})$$

Take the limit as $\Delta t \rightarrow 0$. Use Equation (A.1) and the Ito multiplication rule to write

$E_t \Delta c_t^2 = \sigma_{c,t}^2 \Delta t$. Dividing by Δt and rearranging leads to Equation (5) in the text.

Appendix B:

Derivation of the Consumption Function in the CARA Example

The derivation is a straightforward application of the methods in Merton (1971).

Define the value function as $J(w_t, y_t)$. The Bellman equation for this problem is then

$$0 = \max_c -\frac{e^{-ac_t-bw_t}}{a} - \theta J + J_w[rw_t + y_t - c_t] + J_y \rho \left(\frac{\mu}{\rho} - y_t \right) + J_{yy} \frac{\sigma^2}{2}. \quad (\text{B.1})$$

Performing the indicated optimization yields the first-order condition

$$e^{-ac_t-bw_t} = J_w. \quad (\text{B.2})$$

Substitute Equation (B.2) back into Equation (B.1) to arrive at the partial differential equation

$$0 = -\frac{J_w}{a} - \theta J + J_w \left(rw_t + y_t + \frac{\ln J_w + bw_t}{a} \right) + J_y \rho \left(\frac{\mu}{\rho} - y_t \right) + J_{yy} \frac{\sigma^2}{2}. \quad (\text{B.3})$$

Conjecture that the value function is of the form

$$J(w_t, y_t) = -\frac{e^{-\alpha_0 - \alpha_1 w_t - \alpha_2 y_t}}{\alpha_1}, \quad (\text{B.4})$$

where α_0 , α_1 , and α_2 are constants to be determined. Using this conjecture, Equation (B.3)

reduces to

$$0 = -\frac{1}{a} + \frac{\theta}{\alpha_1} + rw_t + y_t - \frac{\alpha_0 + (\alpha_1 - b)w_t + \alpha_2 y_t}{a} + \frac{\alpha_2}{\alpha_1} \rho \left(\frac{\mu}{\rho} - y_t \right) + \frac{\alpha_2^2}{\alpha_1} \frac{\sigma^2}{2}. \quad (\text{B.5})$$

Collecting terms, the constants turn out to be

$$\alpha_1 = ra + b \quad (\text{B.6})$$

$$\alpha_2 = \frac{a\alpha_1}{\alpha_1 + a\rho} \quad (\text{B.7})$$

$$\alpha_0 = \frac{\theta a}{ra + b} - 1 + \frac{a^2}{a(r + \rho) + b} \mu - \frac{ra + b}{[a(r + \rho) + b]^2} a^3 \frac{\sigma^2}{2}. \quad (\text{B.8})$$

Substituting these back into the first-order condition (B.2) yields the consumption function in Equations (9), (10), and (11) of the text.

The value function must also satisfy the transversality condition

$$\lim_{t \rightarrow \infty} E e^{-\theta t} J(w_t, y_t) = 0. \quad (\text{B.9})$$

Some tedious algebra reveals that a sufficient condition for this to be satisfied is that the effective rate of interest be positive, $r + b/a > 0$.

Appendix C:

Derivation of the Consumption Function with Interest Rate Risk

The derivation is similar to that in appendix B. The Bellman equation for this problem is

$$\begin{aligned} 0 = \max_{c_t} & \left[-\frac{\exp(-ac_t - bw_t)}{a} - \theta J + J_w(rw_t + y_t - c_t) + J_{ww} \frac{\varpi^2}{2} \right. \\ & \left. + J_y \rho \left(\frac{\mu}{\rho} - y_t \right) + J_{yy} \frac{\sigma^2}{2} \right]. \end{aligned} \quad (\text{C.1})$$

The first order condition implies that

$$\exp(-ac_t - bw_t) = J_w \quad (\text{C.2})$$

Substituting it back into Equation (C.1) yields

$$\begin{aligned} 0 = & -\frac{J_w}{a} - \theta J + J_w(rw_t + y_t + \frac{\ln J_w + bw_t}{a}) + J_{ww} \frac{\varpi^2}{2} \\ & + J_y \rho \left(\frac{\mu}{\rho} - y_t \right) + J_{yy} \frac{\sigma^2}{2} + J_{yw} \rho_{wy} \sigma \varpi. \end{aligned} \quad (\text{C.3})$$

Guess that the value function takes the form

$$J(w_t, y_t) = -\frac{\exp(-\alpha_0 - \alpha_1 w_t - \alpha_2 y_t)}{\alpha_1}, \quad (\text{C.4})$$

where α_0 , α_1 , and α_2 are undetermined coefficients. Using this conjecture, we have

$$\begin{aligned}
J_w &= \exp(-\alpha_0 - \alpha_1 w_t - \alpha_2 y_t), \\
J_{ww} &= -\alpha_1 J_w, \\
J_y &= \frac{\alpha_2}{\alpha_1} J_w, \\
J_{yy} &= -\frac{\alpha_2^2}{\alpha_1} J_w, \\
J_{wy} &= -\alpha_2 J_w, \\
J &= -\frac{1}{\alpha_1} J_w.
\end{aligned} \tag{C.5}$$

and

$$\begin{aligned}
0 &= -\frac{1}{a} + \frac{\theta}{\alpha_1} + [rw_t + y_t + \frac{(-\alpha_0 - \alpha_1 w_t - \alpha_2 y_t) + bw_t}{a}] - \alpha_1 \frac{\varpi^2}{2} \\
&\quad + \frac{\alpha_2}{\alpha_1} \rho \left(\frac{\mu}{\rho} - y_t \right) - \frac{\alpha_2^2}{\alpha_1} \frac{\sigma^2}{2} - \alpha_2 \rho_{wy} \sigma \varpi.
\end{aligned} \tag{C.6}$$

Collecting terms yields

$$\begin{aligned}
0 &= \left[-\frac{1}{a} + \frac{\theta}{\alpha_1} - \frac{\alpha_0}{a} - \alpha_1 \frac{\varpi^2}{2} + \frac{\alpha_2}{\alpha_1} \mu - \frac{\alpha_2^2}{\alpha_1} \frac{\sigma^2}{2} - \alpha_2 \rho_{wy} \sigma \varpi \right] \\
&\quad + \left[r + \frac{b - \alpha_1}{a} \right] w_t \\
&\quad + \left[1 - \frac{\alpha_2}{a} - \frac{\alpha_2}{\alpha_1} \rho \right] y_t.
\end{aligned} \tag{C.7}$$

Matching the terms yields

$$\begin{aligned}
\alpha_1 &= ra + b \\
\alpha_2 &= \frac{a\alpha_1}{\alpha_1 + \rho a} \\
a_0 &= -1 + a \frac{\theta}{\alpha_1} - a \alpha_1 \frac{\varpi^2}{2} + a \frac{\alpha_2}{\alpha_1} \mu - a \frac{\alpha_2^2}{\alpha_1} \frac{\sigma^2}{2} - a \alpha_2 \rho_{wy} \sigma \varpi \\
&= -1 + \frac{a\theta}{ra + b} + \frac{a^2}{(r + \rho)a + b} \mu - a^3 \frac{ra + b}{[(r + \rho)a + b]^2} \frac{\sigma^2}{2} \\
&\quad - a(ra + b) \frac{\varpi^2}{2} - a^2 \frac{ra + b}{(r + \rho)a + b} \rho_{wy} \sigma \varpi.
\end{aligned} \tag{C.8}$$

Hence, we have

$$\begin{aligned}
c_t &= -\frac{\ln J_w + bw_t}{a} \\
&= \frac{\alpha_0 + (\alpha_1 - b)w_t + \alpha_2 y_t}{a} \\
&= \Omega(w_t, y_t) - P_{b>0}
\end{aligned} \tag{C.9}$$

where $\Omega(w_t, y_t)$ and $P_{b>0}$ are defined in Equations (39) and (40) in the text.

References

- Alessie, Rob, and Annamaria Lusardi. (1997). "Consumption, Saving, and Habit Formation." *Economics Letters* 55: 103-108.
- Bakshi, Gurdip S., and Zhiwu Chen. (1996a). "Inflation, Asset Prices, and the Term Structure of Interest Rates in a Monetary Economy." *Review of Financial Studies* 9: 241-275.
- Bakshi, Gurdip S., and Zhiwu Chen. (1996b). "The Spirit of Capitalism and Stock-Market Prices." *American Economic Review* 86: 133 – 157.
- Baxter, Marianne., Urban Jermann. (1999). "Household Production and the Excess Sensitivity of Consumption to Current Income." *American Economic Review* 89: 902 - 920.
- Blanchard, Olivier J., and Stanley Fischer. (1989). *Lectures on Macroeconomics*. Cambridge, MA.: MIT Press., 185-216.
- Blanchard, Olivier., and Gregory N. Mankiw, (1988). "Consumption: Beyond Certainty Equivalence." *American Economic Association Papers and Proceedings* 173-177.
- Caballero, Ricardo.J., (1990). "Consumption puzzles and Precautionary Savings." *Journal of Monetary Economics* 25: 113 - 136.
- Caballero, Ricardo.J., (1991). "Earnings Uncertainty and Aggregate Wealth." *American Economic Review* September: 859 - 871.
- Campbell, John (1994). "Inspecting the Mechanism: An Analytical Approach to the Stochastic Growth Model." *Journal of Monetary Economics* 33: 463-506, 1994.
- Campbell, John., and Angus Deaton. (1989). "Why Is Consumption So Smooth?" *Review of Economic Studies* 56: 205-251.
- Campbell, John., and Gregory Mankiw. (1989). "Consumption, Income, and Interest Rates: Reinterpreting the time series evidence." *NBER Macroeconomics Annual 1989*, Cambridge, Mass. MIT Press, 185-216.
- Carroll, Christopher. (1992). "The Buffer Stock Theory of Saving: Some Macroeconomic Evidence." *Brookings Papers on Economic Activity* 2: 61-156.
- Carroll, Christopher. (2001). "A Theory of the Consumption Function, With and Without Liquidity Constraints." *Journal of Economic Literature* 15: 23-46.
- Carroll, Christopher, and Miles Kimball. (1997). "On the Concavity of the Consumption Function." *Econometrica*, 64(4), 981-992.
- Cox, John C., John.E Ingersoll. Jr., and Stephen.A. Ross (1985). "A theory of the term structure of interest rates." *Econometrica*, 53, 385-407.
- Deaton, Angus. (1992). *Understanding Consumption*. Oxford: Oxford University Press.

- Debreu, G. (1976). "Least Concave Utility Functions." *Journal of Mathematical Economics* 8: 337-360.
- Dynan, Karen. (1993). "How Prudent are Consumers?" *Journal of Political Economy* 101: 1104-1113.
- Flavin, Marjorie. (1981). "The Adjustment of Consumption to Changing Expectations about Future Income." *Journal of Political Economy* 89: 974-1009.
- Friedman, Milton. (1957). *A Theory of the Consumption Function*. Chicago: Princeton: Princeton University Press.
- Gong, Liutang, and Heng-fu Zou. (2001). "Money, Social Status, and Capital Accumulation in a Cash-in-Advance Model." *Journal of Money Credit and Banking* 33: 284-293.
- Gong, Liutang, and Heng-fu Zou. (2002). "Direct Preferences for Wealth, the Risk Premium Puzzle, Growth, and Policy Effectiveness." *Journal of Economic Dynamics and Control* 26: 271-302.
- Gourinchas, Pierr.O. and Jonathan Parker. (1999). "Consumption over the Life Cycle." NBER Working Paper No. 7271.
- Grossman, Sanford.J. and Robert. J. Shiller. (1982). "Consumption Correlatedness and Risk Measurement in Economies with Non-Traded Assets and Heterogeneous Information." *Journal of financial Economics* 10: 195-210.
- Hall, Robert. (1978). "Stochastic Implications of the Lifecycle-Permanent Income Hypothesis: Theory and Evidence." *Journal of Political Economy* December: 971-987.
- Hall, Robert. (1988). "Intertemporal Substitution in Consumption." *Journal of Political Economy* April: 339-357.
- Hansen, Lars.P., and Kenneth Singleton. (1983). "Stochastic Consumption, Risk Aversion, and the Temporal Behavior of Asset Returns." *Journal of Political Economy* 91: 249 - 265.
- Irvine, Ian., and Wang, Susheng. (1995). "Earnings Uncertainty and Aggregate Wealth: Comment." *American Economic Review* 84: 1463 - 1469.
- Kihlstrom, Richard, and Leonard J. Mirman. (1974). "Risk Aversion with Many Commodities." *Journal of Economic Theory* 8: 361-368.
- Kihlstrom, Richard, and Leonard J. Mirman. (1981). "Constant, Increasing, and Decreasing Risk Aversion with Many Commodities." *Review of Economic Studies* 48: 271-280.
- Kimball, Miles. and Gregory N. Mankiw. (1989). "Precautionary Savings and the Timing of Taxes." *Journal of Political Economy* August, 863 - 879.
- Lettau, Martin and Sydney Ludvigson (2001). "Consumption, Aggregate Wealth and Expected Stock Returns" *Journal of Finance*, LVI (3), 815-849.

- Luo, Yulei, and Eric. R. Young. (2002). "The Wealth Distribution and the Demand for Status." Manuscript, Princeton University
- Merton, Robert. (1971). "Optimal Consumption and Portfolio Policies in a Continuous-Time Model." *Journal of Economic Theory* 3: 373- 413.
- Merton, Robert. (1975). "An Asymptotic Theory of Growth Under Uncertainty." *Review of Economic Studies* 42: 375-393.
- Skinner, Jonathan. (1988). "Risky Income, life Cycle Consumption, and Precautionary Savings." *Journal of Monetary Economics* September: 237-256.
- Smith, Williams. (1998). "Birth, Death, and Consumption: Overlapping Generations and the Random Walk Hypothesis." *International Economic Journal* 12: 105-116.
- Smith, Williams (1999). "Risk, the Spirit of Capitalism, and Long-Term Growth" *Journal of Macroeconomics*, 21, 241-262.
- Smith, Williams. (2001). "How Does the Spirit of Capitalism Affect Stock-Market Prices?" *Review of Financial Studies* 14: 1215-1232.
- Smith, Williams (2002). "Consumption and Saving with Habit Formation and Durability." *Economics Letters*.75: 369-375.
- Weber, Marx. (1948). *The Protestant Ethic and the Spirit of Capitalism*, translated by Talcott Parsons, George Allen & Unwin, London.
- Weil, Philippe. (1993). "Precautionary Savings and the Permanent Income Hypothesis." *Review of Economic Studies* 60: 367-383.
- West, Kenneth. (1988). "The Insensitivity of Consumption to News about Income." *Journal of Monetary Economics* 21: 17-33.
- Yang, Yunhong. and Heng-fu Zou. (2001). "Social Status, Non-expected Utility, Asset Pricing, and Growth." Working Paper, Peking University.
- Zeldes, Stephen. P. (1989). "Optimal Consumption with Stochastic Income: Deviations from Certainty Equivalence." *Quarterly Journal of Economics* May: 275 - 298.
- Zou, Heng-fu. (1994). "'The Spirit of Capitalism' and Long-Run Growth." *European Journal of Political Economy* 10: 279-293.
- Zou, Heng-fu. (1995). "The Spirit of Capitalism and Savings Behavior." *Journal of Economic Behavior and Organization*, 28, 131-143.

A Note on Entrepreneurial Risk, Capital Market Imperfections, and Heterogeneity*

Yulei Luo[†]

Liutang Gong

University of Hong Kong

Peking University

Heng-fu Zou

The World Bank, CUFU, Shenzhen University, and Wuhan University

May 5, 2009

Abstract

Empirical evidence shows that entrepreneurs hold a large fraction of wealth, have higher saving rates than workers, and face substantial uninsurable entrepreneurial and investment risks. This paper constructs a heterogeneous-agent general equilibrium model with uninsurable entrepreneurial risk and capital market imperfections to explore the implications of uninsurable entrepreneurial risk for wealth distribution and aggregate activity in an incomplete market economy. It is shown that entrepreneurial risk can substantially affect both the wealth distribution and the macroeconomy.

JEL Classification Numbers: *C68, E21, E24, E32.*

Keywords: Wealth Distribution, Idiosyncratic Entrepreneurial Risk, Incomplete Capital Markets

*An earlier version of this paper was circulated under the title: “Uninsurable Entrepreneurial Risk, Capital Market Imperfections, and Heterogeneity in the macroeconomy.” We thank Chris Sims, Pierre-Olivier Gourinchas, Per Krusell, and Eric Young, as well as seminar participants at Princeton University and Memphis University for helpful suggestions and comments. Gong thanks the National Science Foundation for Distinguished Young Scholars (#70725006) for financial support. The Fortran code solving the benchmark model is available from the corresponding author by request.

[†]Address corresponding to: Dr. Yulei Luo, School of Economics and Finance, The University of Hong Kong, Hong Kong; email: yluo@econ.hku.hk.

1. Introduction

What role do uninsurable entrepreneurial risk and capital market imperfections play in shaping the wealth distribution in the economy? What are the impacts of financial development on the level and volatility of aggregate output and capital? In this paper, we construct a heterogeneous-agent dynamic stochastic general equilibrium model to address these important macroeconomic questions quantitatively. The main features of this model are that occupational choice is endogenous, capital markets are incomplete in the sense that idiosyncratic risks cannot be fully insured, and contracts between borrowers and lenders are imperfectly enforceable.

The literature has typically found that simple models based on standard and identical preferences and on uninsurable shocks to labor income cannot account for the observed U.S. Gini coefficient of 0.803 on wealth.¹ For example, Aiyagari (1994) finds considerably less wealth concentration in a model with only idiosyncratic labor earnings uncertainty; Krusell and Smith (1998) find the same in models with both aggregate and idiosyncratic shocks. Among the infinite horizon models that try to reproduce the wealth distribution in the U.S. data, two kinds of models have performed well. One is the stochastic- β model by Krusell and Smith, which achieves an improved fit in the upper tail of the wealth distribution by assuming that individual discount factors are idiosyncratic. The second is the model of Castañeda, et al. (2003), which assumes extremely volatile uninsurable idiosyncratic shocks to labor income. Both models achieve their better fits by introducing individual-specific exogenous disturbances, which arguably weakens the models as explanations of the wealth distribution.² The reason that these models do poorly in explaining the facts is that the only motive to save is precautionary: in order to smooth consumption, agents build a buffer stock of wealth. However, as discussed in the literature on precautionary savings, once the buffer has reached a certain level, the incentive to save becomes weak. The introduction of life cycle features, as in Huggett (1996), increases the concentration

¹This number is taken from Budría et al. (2001); the data used are from the Survey of Consumer Finances (SCF), 1998 wave. Previous estimates using different waves of the SCF are quite similar to this number.

²For a detailed discussion of the successes and failures of these models in accounting for wealth inequality, see a survey by Quadrini and Ríos-Rull (1997).

of wealth as measured by the Gini index. However, the life-cycle model generates this higher concentration of wealth by increasing the proportion of households with zero or negative wealth, rather than by generating a higher concentration at the top of the distribution. Therefore, there must be other mechanisms inducing some agents to accumulate and maintain very high levels of wealth.

Entrepreneurship has been recently used to study household savings, the distribution of wealth, and social mobility; see Quadrani (2001), Fernández-Villaverde, et al. (2003; henceforth FGC), Gentry and Hubbard (2004), and Cagetti and De Nardi (2006). In the data, entrepreneurs are a small fraction of the population, but have a high saving rate and hold a large share of total wealth. For instance, in the 1989 SCF entrepreneurs are 8.7% of the sample, but hold 39% of total net worth. Both Quadrini (2001) and Gentry and Hubbard (2004) document that the large wealth holdings of entrepreneurs are due not only to the fact that entrepreneurs earn more income, but also to their saving a larger fraction of their income than non entrepreneurs. Evans and Jovanovic (1989) is another influential work. They show that wealthier people are more inclined to become entrepreneurs because of liquidity constraints: capital is essential for starting a firm, and liquidity constraints tend to exclude those with insufficient funds at their disposal. It will be shown that our quantitative results can also confirm their empirical result.

The model in this paper is constructed along the line of heterogeneous-agent models originally developed by Aiyagari (1994) and is closely related to that studied in FGC (2003). In FGC (2003), the interest and wage rates are set exogenously in solving individuals' optimization problems; in other words, their model cannot generate the *equilibrium* interest and wage rates. In contrast, in this paper, we assume that there are two production sectors (the corporate sector and the entrepreneur sector) and the interest and wage rates can be determined by the production of the corporate sector in equilibrium. This novel feature, together with other features such as uninsurable idiosyncratic risks, occupational choices, and capital market imperfections, makes our model more difficult to solve because the equilibrium factor prices now depend on both aggregate capital stock and aggregate labor employment in the corporate sector, which are not

simple functions of a known moment of the distribution, as they depend on the current optimal decisions of all entrepreneurs and workers. Therefore, we need to add some extra steps to guarantee that all markets are clear and all households know the current factor prices before they make decisions.

After calibrating and solving our benchmark model, we find that the model can generate the wealth distribution observed in the U.S. data. Furthermore, we show that due to uninsurable entrepreneurial risks and capital market imperfections, agents choose to save more to undertake entrepreneurial activity. Finally, we find that the economy with more volatile entrepreneurial risks generates greater wealth inequality. The paper is organized as follows. In Section 2, we characterize the model economy. In Section 3, we characterize households' optimization problems and define the recursive competitive equilibrium. In Section 4, we set the parameter values and present main findings. Section 5 concludes.

2. The Model Economy

The model economy is populated by a continuum of infinitely lived households measured by 1. In each period, every household makes a decision to establish or run its own business (be an entrepreneur) or to be a worker who supplies his or her labor to the competitive labor market. There are three sectors in the model: the household sector, the production sectors (the corporate and noncorporate/entrepreneurial sectors), and the financial intermediation sector. The workers face partially uninsurable labor income risk, and the entrepreneurs face idiosyncratic uninsurable entrepreneurial risk. There is one final good that can be used either for consumption or for capital services. The timing of the economy is as follows: (i) At the beginning of each period, different idiosyncratic shocks are realized; (ii) then, the households will produce according to their occupational choices made in the previous period to be entrepreneurs or workers; (iii) next, depending on their present shocks and their access to the credit market, they will decide if they want to become entrepreneurs or workers in the next period by comparing their conditional expectations of the next period's value functions; (iv) finally, after production, households decide

how much to save and consume. And all markets clear.

2.1. The Household Sector

Preferences. We consider a model economy with a continuum (with the measure of 1) of ex ante identical, infinitely lived households. Households have standard preferences over consumption and leisure and maximize the expected lifetime utility as follows:

$$E_0 \left[\sum_{t=0}^{\infty} \beta^t u(c_t) \right], \quad (2.1)$$

where c_t is current consumption, $u(c_t) = \frac{c_t^{1-\sigma}-1}{1-\sigma}$ is the standard CRRA utility function (note that when $\sigma = 1$, $u(c_t) = \log c_t$), and $\beta \in (0, 1)$ is the discount factor.

Households' economic activities. Households are assumed to engage in two economic activities: production according to their occupational choices and wealth accumulation. Here we assume that a fixed cost to be an entrepreneur (measured in utility units) is a fixed number, which will be calibrated later, and occupational choices cannot be reverted in the same period.

Idiosyncratic risk to labor productivity and entrepreneurial skills. Following Aiyagari (1994), we also assume the workers face idiosyncratic shocks to labor efficiency. Each agent is endowed with one unit of time. This unit of time has stochastic productivity as labor input, ε ; it can take a value from $\Omega_w = \{\varepsilon_1, \dots, \varepsilon_n\}$, where $\varepsilon_1 < \dots < \varepsilon_n$. When $\varepsilon = \varepsilon_n$, we think of the agent as having the highest labor productivity, and when $\varepsilon = \varepsilon_1$, we think of him or her as having the lowest labor productivity. The non zero labor services can be used in his or her own business (to be an entrepreneur) or supplied to the labor market at the competitive wage rate. Following Quadrini (2001), we also assume that labor has the same productivity in both activities and supplies all the services of labor in the market.

For comparison with the shocks to labor efficiency, which takes on additive form, we assume entrepreneurs also face idiosyncratic entrepreneurial risks that take a multiplicative form and are drawn randomly from the set, $\Omega_e = \{\theta_1, \dots, \theta_N\}$, and similarly we assume that $\theta_1 < \dots < \theta_N$.

Following FGC (2003), the distribution of these two idiosyncratic shocks depends on the agent's past shocks as well as on occupational choices in the last period: If the agent was an entrepreneur (a worker) in the last period $t - 1$ and remains an entrepreneur (a worker) in the current period t , he or she will draw an entrepreneurial shock (a labor productivity shock) from a first-order Markov chain $P(\theta_t|\theta_{t-1})$ ($Q(\varepsilon_t|\varepsilon_{t-1})$) defined on Ω_e (Ω_w); if the agent was an entrepreneur (a worker) in the last period $t - 1$ and wants to be a worker (an entrepreneur) in the current period t , he will draw a labor productivity shock (an entrepreneurial shock) from a probability distribution $\tilde{Q}(\varepsilon_t)$ ($\tilde{P}(\theta_t)$) defined on Ω_w (Ω_e).

2.2. The Production Sectors

As in Quadrini (2001) and Cagetti and De Nardi (2006), we have two production sectors: the corporate sector, composed of large firms and corporations, and the noncorporate sector, composed of entrepreneurs. The two sectors differ in their production technologies. Suppose that entrepreneurship is formed by running business projects, and entrepreneurs face uninsurable entrepreneurial risks and financial constraints. The first factor causes the whole household wealth to be invested in the business, and the second one makes the demand for capital of these small firms closely dependent on the net worth of the owners.

The noncorporate sector/entrepreneur sector

The production function in the entrepreneur sector is

$$y = f(\theta, k, l) = \theta k^\mu l^\omega, \quad 0 < \mu + \omega < 1, \quad (2.2)$$

where θ is the entrepreneurial ability/productivity, i.e., the capacity to invest capital productively, k the individual entrepreneurial capital, and l the labor input. Entrepreneurs can borrow and invest capital in a technology whose return depends on their own entrepreneurial ability. That is, those with higher ability levels have higher average and marginal returns from investing.

The corporate sector

The corporate sector is populated by large firms with a standard Cobb-Douglas technology,

$$Y_c = F(K_c, L_c) = K_c^\alpha L_c^{1-\alpha},$$

where K_c and L_c are aggregate corporate capital and labor, respectively. In equilibrium, the interest rate and the wage rate are given by the marginal products of each factor,

$$W(K_c, L_c) = (1 - \alpha) (K_c/L_c)^\alpha \text{ and } R(K_c, L_c) = \alpha (K_c/L_c)^{\alpha-1} - \delta_c, \quad (2.3)$$

respectively, where δ_c is the rate of depreciation in the corporate sector.

2.3. The Financial Intermediation Sector

The financial intermediation sector in this model can collect deposits from households by paying the interest rate R and make loans to either entrepreneurs asking for funds or the corporate sector. The lending is based on a constant-returns-to-scale technology with a proportional cost per unit of funds lent. Competition among banks makes (i) intermediation profits zero, (ii) the lending rates equal to R for loans to the corporate sector, and (iii) $R^e = R + \eta$ for loans to entrepreneurs; here η is the proportional cost per unit of funds faced by entrepreneurs. Based on the data about household borrowing and lending to banks and other intermediation sectors in Quadrani (2001), the lending rate could be set around (0.035, 0.055).

2.4. Demand for Capital and Business Profits

As a result of borrowing constraints, firms cannot operate at the level that maximizes their profits. Because household asset holdings, used as collateral, determine the tightness of these constraints, the demand for capital and labor will depend both on shocks and on the level of asset holdings. In case the borrower does not repay the loan with the interest, i.e., in the case of bankruptcy, the bank gets a share $0 < \kappa < 1$ of the profits of the firm. This amount can be regarded as the quantity that the bank will get if it uses the legal system to enforce the contract.

In addition, the bank cannot seize the household's assets. Finally, a default decision today does not have a reputational consequence in the future. Therefore, the bank will only lend an amount such that the firm does not have any incentive to default, and this amount may not be the one needed by the entrepreneur to operate the firm at an optimal level.

At the beginning of the current period, after observing the shocks, the entrepreneur decides his demand for inputs to maximize his profits:

$$\pi(\theta, a) = \max_{\{k_t, b_t, l_t\}} \{\theta k^\mu l^\omega - Wl - (R^e + \delta_e)b\} \quad (2.4)$$

$$\text{s.t.: } k_t \leq a_t + b_t, \quad (2.5)$$

$$\pi(\theta, a) \geq (1 - \kappa)\pi(\theta, a) + (1 + R^e)b, \quad (2.6)$$

where a_t is asset holding, b_t is the quantity borrowed from banks, k_t is the demand for capital, and δ_e is the depreciation rate in the entrepreneurial sector. The second equation above is the incentive compatibility constraint, which implies that the total profit an entrepreneur needs is higher than the entrepreneur's income if he defaults. Thus, we cannot observe any default in equilibrium. The first term on the right-hand side of that equation is the profit that the household keeps for itself, and the second term is the amount of payments to the financial intermediary because of default. Using the same procedure as in FGC (2003),³ we can solve the above problem and derive the demand for inputs as well as the profit function: $k = k_e(\theta, a)$, $l = l_e(\theta, a)$, and $p = \pi(\theta, a)$.

3. Household Optimization Problems and Steady State Equilibrium

In this section, we first present households' optimization problems and then define a steady state equilibrium for the economy. The optimal occupational choice and decision problem for a

³We thank Fernandez-Villaverde for sharing with us his Matlab code for solving the demand functions.

worker can be characterized by the stochastic Bellman equation

$$\begin{aligned} v_w(a, \varepsilon; \Gamma) &= \max_{c, a'} \left\{ u(c) + \beta \max \left[\int v_w(a', \varepsilon'; \Gamma') Q(\varepsilon, d\varepsilon'), \int v_e(a', \theta'; \Gamma') \tilde{P}(d\theta') - \psi \right] \right\} \\ \text{s.t.} \quad &: c + a' = [1 + R(K_c, L_c)]a + W(K_c, L_c)(1 - \tilde{h})\varepsilon \end{aligned}$$

and $a' \geq \bar{a}$, where Γ denotes the current distribution of agents over asset holdings and idiosyncratic shocks. The worker makes occupational choices by comparing the conditional expectations of two value functions for being a worker and being an entrepreneur in the next period. We denote the worker's decision rules for consumption and asset holdings as $c = c_w(a, \varepsilon)$ and $a' = a_w(a, \varepsilon)$, respectively.

Similarly, the optimization problem of an entrepreneur can be characterized by

$$\begin{aligned} v_e(a, \theta; \Gamma) &= \max_{c, a'} \left\{ u(c) + \beta \max \left[\int v_w(a', \varepsilon'; \Gamma') \tilde{Q}(d\varepsilon'), \int v_e(a', \theta'; \Gamma') P(\theta, d\theta') - \psi \right] \right\} \\ \text{s.t.} \quad &: c + a' = \pi(\theta, a) + [1 + R(K_c, L_c)]a + W(K_c, L_c)(1 - \tilde{h}) \end{aligned} \quad (3.2)$$

and $a' \geq \bar{a}$, where we denote the entrepreneur's decision rules for consumption and asset holdings as $c = c_e(a, \theta; \Gamma)$ and $a' = a_e(a, \theta; \Gamma)$, respectively. An entrepreneur will remain an entrepreneur in the next period if $\int v_w(a', \varepsilon'; \Gamma') \tilde{Q}(d\varepsilon') < \int v_e(a', \theta'; \Gamma') P(\theta, d\theta') - \psi$; otherwise, he will choose to be a worker. Similarly, a worker will remain a worker in the next period if $\int v_w(a', \varepsilon'; \Gamma') Q(\varepsilon, d\varepsilon') > \int v_e(a', \theta'; \Gamma') \tilde{P}(d\theta') - \psi$. Define $\chi_i(a, \varepsilon, \theta; \Gamma)$ as the decision rules governing whether an agent stays in the same occupation; we can use the following indicator function to specify occupational choices:

$$\chi_i(a, \varepsilon, \theta; \Gamma) = \begin{cases} 0, & \text{if he stays in the same occupation.} \\ 1, & \text{otherwise,} \end{cases} \quad (3.3)$$

where $i = w, e$.

Definition A recursive competitive equilibrium for the steady state economy is a set of decisions rules, $c_w, c_e, a_w, a_e, k, l, K_c$, and L_c , and a set of value functions, v_w and v_e , the pricing functions, R and W , and a law of motion for the measure of agents, H , such that:

- (1) The decision rules, c_w and a_w , and value function v_w , solve problem (3.1), given the functions v_e, R, W , and H .
- (2) The decision rules, a_e, a_e, k, l , and value function v_e , solve problem (3.2), given the functions v_w, R, W , and H .
- (3) The occupational decision rule, χ_i , is determined by (3.3), given v_w and v_e .
- (4) R and W are competitive, i.e., they are equal to the marginal productivity of capital and labor (net of depreciation) in the corporate sector.
- (5) The firms' decision rules, k and l , solve problem (2.4).
- (6) Prices are such that capital and labor markets clear:⁴

$$\int k d\Gamma_e + K_c = \int a d\Gamma \text{ and } \int l d\Gamma_e + L_c = \int (1 - \tilde{h}) \varepsilon d\Gamma_w. \quad (3.4)$$

- (7) The law of motion for the distribution is consistent with individual optimal behavior, and it is invariant.

Although we cannot guarantee theoretically the existence and uniqueness of the equilibrium described above because of the nonconvexity problem in the household problem, practically, the existence and uniqueness of the equilibrium does hold in this model because the computational

⁴Note that the left-hand side of the capital market (labor market) equation is the aggregate capital (labor) demand in the entrepreneur sector and the corporate sector, and the right-hand side is the aggregate capital (labor) supply from all agents. We also assume here that entrepreneurs do not use their own labor in production activity.

evidence shows that the value functions in this model are strictly concave for all reasonable parameter choices. This conclusion, together with the assumptions for stochastic shocks, can guarantee the existence of the unique invariant measure.

4. Main Findings

4.1. Parameterization

The quantitative properties of the model's competitive equilibrium cannot be established analytically, and they need to be studied using numerical methods. Computing the recursive competitive equilibrium involves three steps. First, we need to impose restrictions on the functional forms. Second, we select as many parameters as possible either by matching long-run properties of the model economy to the U.S. data or by using previous empirical evidence. In the last step, we need to develop a numerical algorithm to solve the competitive equilibrium up to an arbitrarily small error. Our computational algorithm is a combination of the ones used in FGC (2003), Young (2006), and Luo and Young (in press).⁵ The model period is set to one year, which is standard in the literature. The discount factor, β , is set together with the share of capital in the entrepreneurial sector, μ , so that the capital output ratio of the whole economy in the steady state is equal to 2.5. We set $\alpha = 0.36$, which is the standard choice in the literature, and $\mu = 0.36$ and $\omega = 0.52$ so that $\mu + \omega = 0.88 < 1$.⁶ The implicit degree of decreasing returns to scale (12%) generates a portion of income earned by entrepreneurs that matches the PSID data. In our benchmark exercise, we set $\phi = 0$. We also choose the parameter ψ , which governs the amount of effort to be an entrepreneur, to be 0.5. In this way, the number of entrepreneurs is around 8.6%, which matches the number in the U.S. data (SCF and PSID). Table 1 summarizes our parameter choices for the baseline model.

Next, to parameterize the stochastic idiosyncratic labor productivity, we follow Storesletten,

⁵It is available from the corresponding author by request.

⁶Note that the production function in the corporate sector is the standard Cobb-Douglas one, whereas the one in the entrepreneurial sector is strictly decreasing returns to scale.

et al. (2007). They argue that the specification of labor income for an individual household must allow for persistent and transitory components. Based on their empirical work from the PSID data, we specify $\log(y_i)$ to be

$$\log(y_i) = \omega_i + \epsilon_i, \quad (4.1)$$

$$\omega'_i = \rho\omega_i + v'_i, \quad (4.2)$$

where $\epsilon_i \sim N(0, \sigma_\epsilon^2)$ is the transitory component and ω_i is the persistent component. The innovation term associated with ω_i is assumed to be distributed as $N(0, \sigma_v^2)$. They estimate $\rho = 0.935$, $\sigma_\epsilon^2 = 0.01$, and $\sigma_v^2 = 0.061$. The unconditional variance of $\log(y_i)$ is then $\text{var}[\log(y_i)] = \frac{\sigma_\epsilon^2}{1-\rho^2} + \sigma_v^2 = 0.14051$. This process attributes about half the unconditional variance to the persistent component and half to the transitory component. We then approximate this process with a three-state Markov chain, which is characterized in Table 2.

For the points in θ , we follow Quadrini (2001) and FGC (2003). Conditional on θ_1 , θ_2 and θ_3 are set to obtain the demand for capital in the medium shock to be 10 times larger than that in the low shock and for firms in the high shock to be 100 times larger. Finally, θ_1 is set to make the ratio of entrepreneurial wealth to total wealth in the economy match around 0.4. In Table 3, we choose the diagonal elements in Q to match the empirical exit and entry rates from entrepreneurship, although we cannot reach the high exit rate as reported in the data, which is around 24%. In the nondiagonal parts, for the low shocks, we divide the rest of the probability into two equal numbers, 0.16 each; for the medium shock, we set them to capture the growth of the firms; for the high shocks, we just assume they drop to the medium level of 10%. The choice of $q = \begin{bmatrix} 0.6 & 0.3 & 0.1 \end{bmatrix}$ is motivated by Quadrini (2001), in which he chooses three bins that assign 60% of entrepreneurs to small projects, 30% to middle-sized projects, and 10% to large projects, respectively.

4.2. Baseline Results

As in FGC (2003), our model with endogenously determined interest and wage rates also predicts that the firms are always constrained in their financial decisions in the range of asset holdings where the measure of households is positive. It is obvious that entrepreneurs can borrow more from outside and operate more profitably if their own asset holdings are high. Figure 1 shows the constrained demand for capital. The difference between the demand for capital and the 45° line is the amount they can borrow from banks. It clearly shows that the higher the wealth holdings of the entrepreneurs, the higher the capital they demand. That is, they will run larger projects. Figure 2 shows that the level of profits increases with asset holdings until the firm can operate at the optimum level.

The value functions for workers and entrepreneurs are plotted in Figures 3 and 4, respectively. Theoretically, it is obvious that occupational choices of the households may make the value functions nonconcave, because an individual cannot be half an entrepreneur and half a worker. However, these two figures show that the value functions for workers and entrepreneurs are both strictly concave. This finding is robust for any set of our parameter choices, and is consistent with that obtained in Gomes, et al. (2001) and FGC (2003): Theoretical departures from concavity are not a serious problem.

Another important question in the literature about entrepreneurship is who prefers to be an entrepreneur and run his own projects. As in FGC (2003), we also find that those households with asset holdings in the middle of the distribution are most likely to become entrepreneurs, because they can borrow enough capital and run more profitable projects, whereas households with low asset holdings prefer to be workers because they cannot borrow enough capital due to the borrowing constraints. In contrast, for households with high levels of asset holdings, running their own businesses is not very attractive because they can earn enough interest income from their own assets. Furthermore, our model also predicts that the percentage of workers is much larger than that of entrepreneurs, and most workers hold assets less than 5, whereas most

entrepreneurs hold asset less than 20. That is, uninsurable idiosyncratic risks and capital market imperfections make households that want to become entrepreneurs accumulate large wealth.

4.3. Implications of Entrepreneurial Risk on Wealth Inequality and Aggregate Activity

Figure 5 clearly shows how our model can generate a skewed wealth distribution. Compared with the Aiyagari model, our benchmark model can fit the data much better. In our model, the top 1%, 5%, and 10% of agents hold 22%, 49%, and 63% of wealth, respectively; these numbers are quite similar to those reported from the US data (26%, 47%, and 60%, respectively). As we discussed above, it is difficult to estimate or calibrate the stochastic process of entrepreneurial risk because of the lack of good micro data. Therefore, to examine the effects of entrepreneurial risk on both the wealth distribution and aggregate activity, we assume that the entrepreneurial risk follows an $AR(1)$ process. Hence, in this section, we just study how the changes in the volatility of entrepreneurial risk affect the economy instead of trying to match the data perfectly. Specifically, we assume that the entrepreneurial shock follows an $AR(1)$ process,

$$\theta'_i - \bar{\theta} = \rho_e (\theta_i - \bar{\theta}) + \zeta_i, \quad (4.3)$$

where $\bar{\theta} = 1.55$, $\rho_e = 0.65$, and $\zeta_i \sim N(0, \sigma_\zeta^2)$.⁷ We then approximate this process with a three-state Markov chain as we did before.

Table 4 and Figure 6 provide a summary of the effects of the entrepreneurial risks on aggregate quantities and wealth inequality. The table shows that in the economy with uninsurable entrepreneurial risks, both entrepreneurial capital and aggregate capital increase with the volatility of the entrepreneurial shock because of the precautionary savings motive and borrowing constraints. This mechanism of wealth accumulation is similar to the one in the Bewley-Aiyagari economies. In addition, the equilibrium interest rate is decreasing with σ_ζ^2 as the entrepreneur-

⁷Changing the values of $\bar{\theta}$ and ρ_e does not change our main results reported in Table 4.

ial risk also increases the ratio K_c/L_c . It is shown in a one-sector growth model of Angeletos and Calvet (2006) that idiosyncratic production shocks introduce a risk premium on private capital and reduce the demand for investment. Our model tells another story. In our setup, although the introduction of entrepreneurial risks reduces the demand for the investment in the entrepreneur sector because of a risk premium on private capital, the effect of the precautionary savings motive dominates. Therefore, the economy with higher volatility may be characterized by higher entrepreneurial capital due to the net effect of precautionary savings, a risk premium of private capital, a lower risk-free rate, and higher aggregate capital due to the reallocation of capital and labor in the two sectors. Table 4 also shows that the fraction of entrepreneurs in the economy is increasing with σ_ζ^2 . The intuition is simple: entrepreneurs with higher wealth levels can borrow more funds and run more profitable projects; in this case choosing to be an entrepreneur becomes more attractive. Furthermore, the economy with high volatility of entrepreneurial risks will generate greater wealth inequalities. Figure 6 plots the Lorenz curves for different volatilities. Our model's prediction is very intuitive: for the economy with high volatility, the entrepreneurs hit by a sequence of good shocks will become wealthier and those hit by a sequence of bad shocks will keep losing and then have to close their own businesses and become workers.

4.4. Effects of Imperfect Enforcements

In this section, we examine the impacts of contract enforcements on the economy by adjusting the appropriability factor κ . This factor may also be a measure of the degree of financial development in the equilibrium. Based on the number reported by Moddy's investors' service, we set $\kappa = 0.4, 0.6$, and 0.8 , respectively. All other parameters are the same as those used in the baseline model. Table 5 summarizes the main effects of κ , i.e., the tightness of the borrowing constraints, on aggregate quantities and cross-sectional properties. This table shows that, when κ increases, i.e., the borrowing constraints become tight, both K and K_e go up. The intuition behind this result is simple: in the economy with a high appropriability factor, in equilibrium,

where no default is observed, entrepreneurs could borrow more capital and employ more labor, and then run more profitable projects and produce more output in the entrepreneurial sector. Furthermore, more workers with higher levels of asset holdings would choose to be entrepreneurs. This effect can slightly reduce the wealth inequalities in the economy because more people become entrepreneurs, and more entrepreneurs become even richer.

5. Conclusions

This paper presents and solves a heterogeneous-agent general equilibrium model with occupational choices, uninsurable idiosyncratic labor and entrepreneurial risks, and incomplete markets including both the absence of a state contingent market for idiosyncratic risks and credit market imperfections. We demonstrate in this model that introducing entrepreneurial risks and capital market imperfections can substantially increase the wealth inequalities and thus provide a better match with the U.S. data. We also demonstrate that uninsurable entrepreneurial risk can increase aggregate entrepreneurial capital stock because of precautionary motives and borrowing constraints.

References

- [1] Aiyagari, S. Rao (1994) Uninsured idiosyncratic risk and aggregate saving. *Quarterly Journal of Economics* 109, 659–684.
- [2] Angeletos, George-Marios and Laurent-Emmanuel Calvet (2006) Idiosyncratic production risk, growth, and the business cycle. *Journal of Monetary Economics* 53 (6), 1095–1115.
- [3] Budría Rodríguez, Santiago, Javier Díaz-Giménez, Vincenzo Quadrini, and Jose-Victor Ríos-Rull (2001) Updated facts on the U.S. distributions of earnings, income, and wealth. Federal Reserve Bank of Minneapolis Quarterly Review 26, 2–35.
- [4] Castañeda, Ana, Javier Díaz-Giménez, and Jose-Victor Ríos-Rull (2003) Accounting for earnings and wealth inequality. *Journal of Political Economy* 111(4), 818–857.

- [5] Cagetti, Marco and Mariacristina De Nardi (2006) Entrepreneurship, frictions and wealth. *Journal of Political Economy* 114 (5), 835-870.
- [6] Evans, David S. and Boyan Jovanovic (1989) An estimated model of entrepreneurial choice under liquidity constraints. *Journal of Political Economy* 97, 808-827.
- [7] Fernandez-Villaverde Jesus, Jose Enrique Galdon-Sanchez, and Luis Carranza (2003) Entrepreneurship, financial intermediation and aggregate activity. Working Paper, University of Pennsylvania.
- [8] Gentry, William and Robert Hubbard (2004) Entrepreneurship and household savings. *Advances in Economic Analysis & Policy* 4(1).
- [9] Gomes, Joao, Jeremy Greenwood, and Sergio Rebelo (2001) Equilibrium unemployment. *Journal of Monetary Economics* 48, 109-152.
- [10] Huggett, Mark (1996) Wealth distribution in life cycle economies. *Journal of Monetary Economics* 38 (3), 469-494.
- [11] Krusell, Per and Anthony. A. Smith, Jr. (1998) Income and wealth heterogeneity in the macroeconomy. *Journal of Political Economy* 106, 867-896.
- [12] Luo, Yulei and Eric R. Young (in press) The Wealth distribution and the demand for status. *Macroeconomic Dynamics*.
- [13] Quadrini, Vincenzo and Jose-Victor Ríos-Rull (1997) Understanding the U.S. distribution of wealth. *Federal Reserve Bank of Minneapolis Quarterly Review* 21, 675-687.
- [14] Quadrini, Vincenzo (2001) Entrepreneurship, saving and social mobility. *Review of Economic Dynamics* 3, 1-40.
- [15] Storesletten, Kjetil, Christopher I. Telmer, and Amir Yaron (2007) Asset pricing with idiosyncratic risk and overlapping generations. *Review of Economic Dynamics* 10 (4), 519-548.

- [16] Young, Eric R. (2006) Approximate aggregation. Working Paper, University of Virginia.

Table 1. Parameter choices

β	σ	α	δ	μ	ω	η	ψ
0.9	1	0.36	0.06	0.36	0.52	0	0.5

Table 2. Idiosyncratic shocks to labor productivity

$\varepsilon = \begin{bmatrix} 0.57 & 0.93 & 1.51 \end{bmatrix}, P = \begin{bmatrix} 0.75 & 0.24 & 0.01 \\ 0.19 & 0.62 & 0.19 \\ 0.01 & 0.24 & 0.75 \end{bmatrix}, p = \begin{bmatrix} 0.31 & 0.38 & 0.31 \end{bmatrix}$
--

Table 3. Idiosyncratic shocks to entrepreneurial activity

$\theta = \begin{bmatrix} 1 & 1.26 & 1.68 \end{bmatrix}, Q = \begin{bmatrix} 0.68 & 0.16 & 0.16 \\ 0.1 & 0.7 & 0.2 \\ 0.0 & 0.1 & 0.9 \end{bmatrix}, q = \begin{bmatrix} 0.6 & 0.3 & 0.1 \end{bmatrix}$

Table 4. The aggregate effects of uninsurable entrepreneurial risks

σ_ζ	K	K_e	K_c/L_c	Y_e	Y_c	# of entrep.	R	W
0.14	1.380	0.176	4.225	0.117	0.482	6.0%	0.087	1.075
0.16	1.554	0.304	5.169	0.207	0.441	9.6%	0.069	1.157
0.18	1.688	0.642	10.684	0.454	0.233	17.8%	0.021	1.511

Table 5. The aggregate effects of contract enforcements

κ	K	K_e	K_c/L_c	Y_e	Y_c	# of entrep.	R	Gini
0.4	1.912	0.820	7.690	0.422	0.300	8.23%	0.039	0.638
0.6	1.961	0.932	8.366	0.473	0.268	9.53%	0.035	0.635
0.8	1.973	1.023	8.716	0.512	0.240	10.3%	0.033	0.627

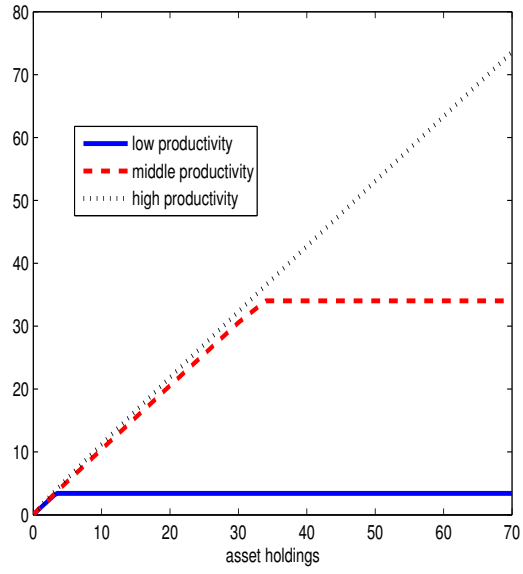


Figure 1. The demand for capital.

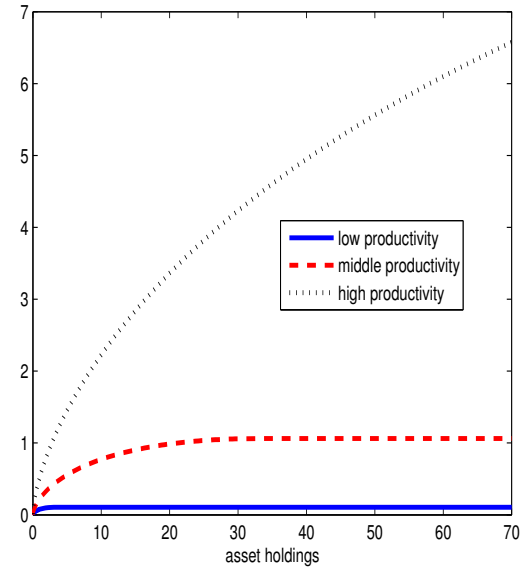


Figure 2. The profits.

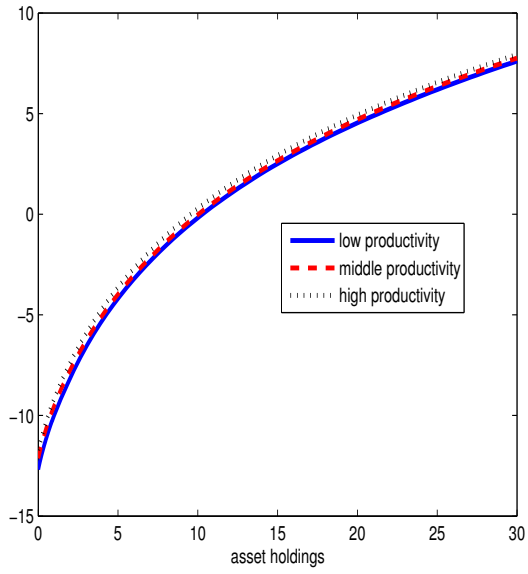


Figure 3. Worker's value function.

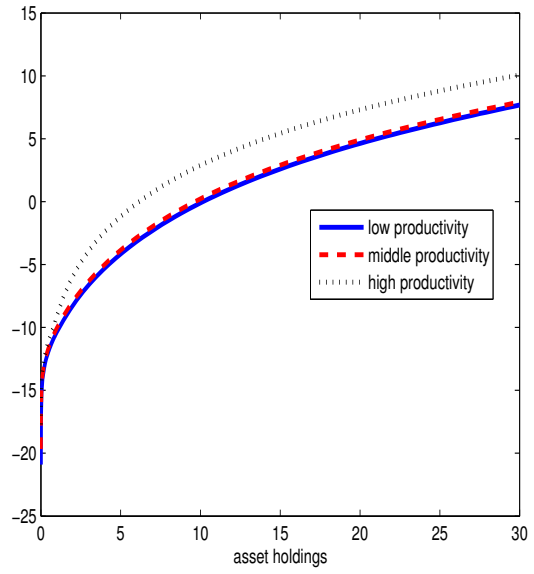


Figure 4. Entrepreneur's value function.

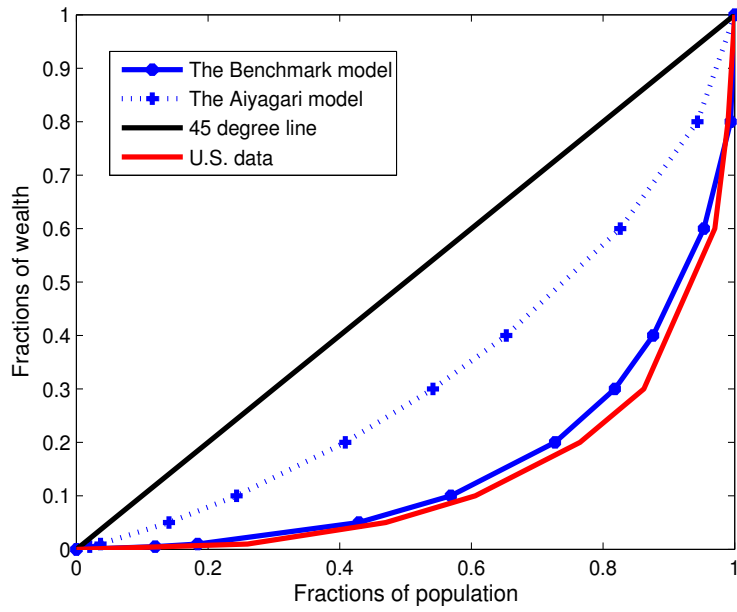


Figure 5. The Lorenz curves for wealth.

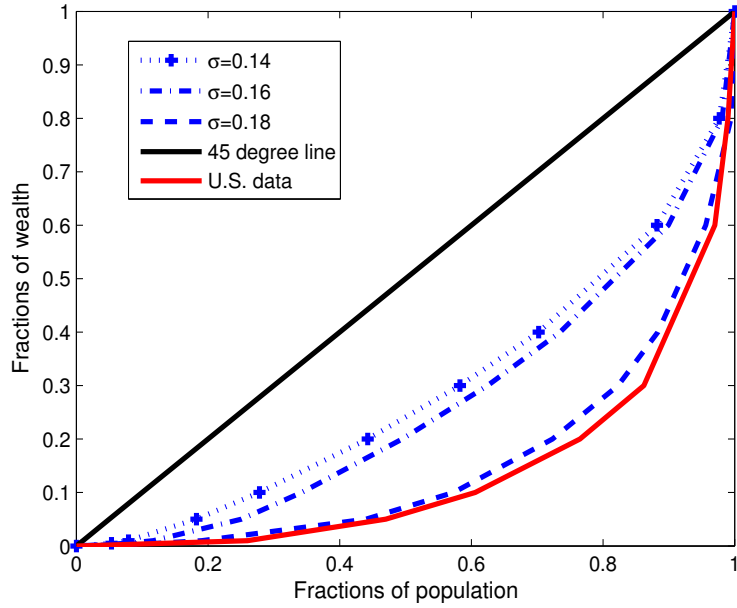


Figure 6. The Lorenz curves for wealth.

Health, Taxes, and Growth

Liutang Gong

Guanghua School of Management, Peking University, China
E-mail: ltgong@gsm.pku.edu.cn

Hongyi Li

BA Faculty, Chinese University of Hong Kong, Hong Kong
E-mail: hongyi@baf.msmail.cuhk.edu.hk

Dihai Wang*

School of Economics, Fudan University, China
E-mail: wangdihai@gmail.com

Heng-fu Zou

CEMA, Central University, China
School of Advanced Study (SAS), Shenzhen University, China
IAS, Wuhan University, China
Guanghua School of Management, Peking University, China
E-mail: zouhengfu@gmail.com

This paper studies capital accumulation and consumption in the traditional Ramsey model under an exogenous growth framework. The model has three important features: (1) treating health as a simple function of consumption, which enable the study of health and growth in an aggregate macroeconomic model; (2) the existence of multiple equilibria of capital stock, health, and consumption, which is more consistent with the real world situation - rich countries may end up with high capital, better health, and higher consumption than poor countries; (3) the fundamental proposition of a consumption tax instead of capital taxation from the traditional growth model does not hold anymore in our model. As long as consumption goods contribute to health formation, the issue of a consumption tax versus an income (or capital) tax should be re-examined.

Key Words: Health; Capital accumulation; Taxation.

JEL Classification Numbers: H0, I1, O3, O4.

* Corresponding author: Dihai Wang, Room 615, School of Economics, Fudan University, No.220, Handan Road, Shanghai, China, Zip Code: 200433

1. INTRODUCTION

The relationship between health and economic growth has long attracted researchers as well as practitioners from many disciplines including economics, sociology, physiology, etc. There have already been a large number of evidences indicating that health is positively related to economic growth, but it is still not clear how health interacts with economic growth. Moreover, existing studies are largely limited to examining the empirical relationship between health, measured by health expenditure, intensity of health care, and life expectancy, and economic growth. Recently there have been some theoretical studies analyzing this causality relationship between health and economic growth (Ehrlich and Lui, 1991; Barro, 1996; Zon and Muysken, 2001, 2003; Morand, 2004; etc.). However, since the interaction mechanism between health and growth is quite complicated, these theoretical studies only analyzing a portion of the whole interaction mechanism between health and growth. To our knowledge, no study has investigated how income and nutrition improvement influence on growth, which is another important channel through which health affects economic growth as indicated by Fogel (1994a, 1994b, 2002) who argued that the combined effort of the increases in the dietary energy available for work, and of the increased human efficiency in transforming dietary energy into work output, appears to account for about 50 percent of the British economic growth since 1790 (Fogel, 1994a, p.388). In this paper, following Fogel's research, we intend to explore the interaction between health and growth through the channel that increases in income and nutrition improve the health capital accumulation and hence raise the labor productivity. Furthermore, we also want to study when consumption affects health capital and hence labor productivity, whether the accumulation of health capital will lead to endogenous economic growth or it is just a by-product of economic growth.

Using an extended Ramsey (1928) model, we assume that consumption not only increases agents' utility but also improves agents' health. Under this assumption, we study the relationship among consumption, health capital and physical capital accumulation, and discuss the effect of health on economic growth. We find that health capital is not the motivation but the by-product of economic growth, which is consistent with Boumol (1967) and Zon and Muysken (2001, 2003). We also find that health capital accumulation is able to magnify economic growth driven by exogenous technology, which is consistent with Fogel's results (Fogel, 1994a, 1994b, 2002). Moreover, in the case of a special product function, we also find the existence of multiple equilibria of capital stock, health, and consumption, which is highly relevant to the real world situation that rich countries may end up with higher capital accumulation, better health, and higher consumption than the poor countries. This result helps to understand the

polarization between the developing countries and the developed countries in the real world. Finally, we also reconsider the effects of consumption tax and capital tax. We find that the fundamental proposition of a consumption tax instead of a capital tax contributes to growth in the traditional growth models does not hold anymore: Once the consumption goods contribute to health formation, the issue of a consumption tax versus an income (or capital) tax should be re-examined. It is necessary to point out that the consumption here denotes the categories of commodity which are able to benefit health improvement.

There are increasing theoretical and empirical investigations on the effect of health on economic growth. The empirical studies can be divided into three categories (Jamison, et al., 2004). The first category comprises the historical case studies that may be more or less quantitative (Fogel, 1994a, 1994b, 2002; Strauss and Thomas, 1998; Sohn, 2000). As stated above, these studies all concluded that nutritional improvement is the main force that enhances health human capital improvement and hence economic growth in the long term. The second category is characterized by many “micro” studies which involve either household surveys that include one or more measures of health status along with other extensive information, or the assessment of the impact of specific diseases. Strauss and Thomas (1998) provided a major review (extensively updated by Thomas and Frankenberg, 2002), and Savedoff and Schultz (2000) surveyed methods used in the household studies and summarized findings of recent analyses from five Latin American countries. Recent studies include Liu et al (2008) on China and Laxminarayan (2004) on Vietnam. This literature confirms that health is positively associated with productivity on the micro level, which is consistent with our assumption that health human capital constitutes a type of production factor. The third category focuses on the relationship between health and economic growth from a macroeconomic perspective. These studies mainly rely on cross-national data to assess the impact of health at the national level, measured in life expectancy, adult survival rates, adult mortality rates or other indexes, on income growth rates and most confirmed that health is positively related to growth (Hicks, 1979; Wheeler, 1980; Barro, 1996; Sachs & Warner, 1997; Bloom and Williamson, 1998; Arora, 2001; Bloom et al., 2004; McDonald and Roberts, 2006; Lorentzen, et al., 2008). On the microeconomic and macroeconomic contribution of health to economic growth and development, Shurcke, et al. (2006) reviewed recent evidence.

The theoretical studies on the relationship between health and growth did not appear until about 20 years ago. Early theoretical studies on this issue mainly focused on the provision of health services from a microeconomic demand perspective and did not analyze the effect of health in the form of human capital promotes economic growth (Grossman, 1972;

Muurinen, 1982; Forster, 1989; Ehrilch and Lui, 1991; Johansson & Lofgren, 1995; Mertz, 1997). Barro (1996) is the first study to propose a theoretical framework to analyze the macroeconomic effects of health as one of the most important components of human capital on economic growth. In a three-sector neoclassical growth model considering simultaneous both health and education human capitals, Barro analyzed the effects of health human capital on education and physical capital and the interaction between these three forms of capitals, and further discussed the effects of public policy of health services as a publicly subsidized private good and as a public good. Muysken, et al. (1999) also investigated the growth implications of endogenous health on steady-state growth and the transitional dynamics in a standard neo-classical growth framework.

Extending the Lucas (1988) endogenous growth model to include health investment and take into account that health services can provide utility, Zon and Muysken (2001, 2003) discussed the macroeconomic effects of health investment on economic growth. Compared to Barro (1996), besides the effect of health on labor productivity, Zon and Muysken (2001, 2003) considered three other channels through which health influences economic growth: 1) better health helps the accumulation of education human capital; 2) health services increase an agent's utility; and 3) health improvement increases longevity and hence leads to an aging population. While the first two effects of health on labor productivity and on education human capital accumulation tend to facilitate economic growth, the last two effects suggest that health investment may exceed the optimal level at which the marginal contribution of health investment to growth equals the marginal cost. This may crowd out resources which could have been used for physical capital investment. Therefore, in such a situation, health investment may impede the progress of economic growth. By introducing the effects of skill-driven technological change (henceforth SDTC) into the Zon and Muysken (2001, 2003) framework, Hosoya (2002, 2003) further investigated the relationships among economic growth, average health level, labor allocation, and longevity of the population in an endogenous growth model that integrates SDTC and human capital accumulation through formal schooling with health human capital accumulation. In addition, through integrating the accumulation of human capital, innovation in medical technology, health and longevity into a four-sector (education, consumption goods, R&D sector devoted to health research, and health goods) endogenous growth model with "keeping up with the Jones" preferences and an altruism utility function, Sanso and Asia (2006) also studied the bidirectional interaction between health and economic growth. They concluded that health, by influencing longevity, may become a source of endogenous growth.

In order to explain the real-world situation that rich countries may end up with higher capital, better health, and higher consumption than poor countries, the existence of multiple steady states and the poverty trap are also important issues in the literature on the relationship between health and economic development. Chakraborty (2004) and Bunzel and Qiao (2005) introduced endogenous mortality risk into a two-period overlapping generations model to study the effect of health (measured in mortality) on economic growth and confirmed the existence of multiple steady states. Hemmi, et al. (2007) studied the interaction between decisions on financing after-retirement health shocks and precautionary saving motives, and demonstrated that, at low levels of income, individuals choose not to save to finance the cost of after-retirement health shocks. However, once individuals become sufficiently rich, they do choose to save to finance the cost of these shocks. Therefore, this change in the individual saving behavior may also give rise to multiple steady state equilibria and result in the poverty trap.

Compared with the above literature, this paper has two important contributions to the existing literature: first, we analyze the effects of health improvement derived from increasing consumption and nutrition intake on the long-run economic growth, which have been ignored by all the previous studies; second, we build on the existing literature and discuss the effects of fiscal policies on the long-run capital stock and consumption level with health capital stock included as a variable.

This paper is organized as follows. Section 2 presents a theoretical model with health generated by consumption. Section 3 develops the accumulation of physical capital and health capital in an exogenous growth model. Multiple equilibria have been found in this framework. Section 4 studies the effects of the income tax and the consumption tax on the long-run consumption level and capital stock. Section 5 concludes with a discussion on the implications of these results and the future research.

2. BASIC MODEL

Consider an intertemporal model with the representative agent choosing his consumption path, c , and his capital accumulation path, k , to maximize his discounted utility, namely

$$\max \int_0^{\infty} u(c) e^{-\beta t} dt \quad (1)$$

subject to

$$\dot{k} = y - c - \delta k \quad (2)$$

where a dot over a variable denotes the derivative of the variable with respect to time, y denotes the agent's income, and $k(0) = k_0$ the initial capital stock. The discount rate β ($0 < \beta < 1$) is a given constant. The instantaneous utility function is defined as $u(c)$. It is assumed that the marginal utility of consumption is positive, but diminishing, i.e. $u'(\cdot) > 0$ and $u''(\cdot) < 0$.

One of the main channels through which health influences the economic growth lies in the production function in which an increase in health can improve the labor productivity. In this paper, the production function is assumed as follows:

$$y = f(k, hl), \quad (3)$$

where l and h denote labor supply and health capital respectively. Compared with the normal neoclassical production function, the uniqueness of the above production function lies in the health capital entering into the product function. In fact, the existing literature points out several channels through which better health will raise the productivity and output. Most directly, healthier workers have more energy and robustness and are able to work harder and for a longer time. People with healthier body are less likely to be caught by disease and have lower chance to be absent from work. The fact that labor productivity is positively associated with health has been confirmed both in empirical micro- and macro-economic researches, especially in low-income settings (Strauss and Thomas, 1998; Bloom, et al, 2004; etc.). In addition, there are some indirect channels through which health influences productivity. For instance, improvement in health raises the incentive to acquire more schooling, since investment in schooling can be amortized over a longer working life. Healthier students also have lower absenteeism and higher cognitive function, and thus receive a better education for a given level of schooling (Howitt, 2005; Kalemli-Ozcan, et al., 2000; Weil, 2007; etc.). All these factors lead to healthier people with higher productivity. Therefore, it is very rational and natural for the health variable to enter the production function, just as Barro (1996), Issa (2003), Hosoya (2002, 2003), Muysken, et al. (1999), Zon and Muysken (2001, 2003), Weil (2007) did. Furthermore, just as what Fogel (2002, p.24) observed, the contribution of nutrition and health to economic growth may be thought of as labor-enhancing technological changes. In Zon and Muysken (2001, p. xiii), they also considered the contribution of health to production ability as Harrod-neutral technical change. In addition, we assume that

$$f_h > 0, \quad f_k > 0, \quad f_{hh} < 0, \quad f_{kk} < 0, \quad f_{kk}f_{hh} > f_{hk}^2 \quad (4)$$

which implies that the marginal productivity of physical capital and health capital are positive but diminishing, and the production function is convex in h and k .

The second main aspect of the interaction mechanism between health and economic growth in our paper lies in the effect of income on the health through consumption and nutrition improvement. As most economists observed, there are at least three main ways to improve an individual's health. First, sufficient nutrition is indispensable to keep a healthy body. Fogel (1994a, 1994b, 2002) and Strauss and Thomas (1998) indicated that, measured in life expectation or in height, an increase in nutrition is the main factor to improve the population's health in the long run in many countries, including Britain, France, United states, Vietnam and others. For the case of the underdeveloped periods of developed countries or the presently low- and middle-income countries, the main approach to improve health is still to increase nutrition and calorie intakes which are mainly embodied in food consumption. The second approach to improve health is health investment (Grossman, 1972; Strauss and Thomas, 1998; Zon and Muysken, 2001, 2003). By Grossman (1972), the health investment includes the own time of consumers and market goods such as medical care, diet, exercise, recreation, housing, which are obvious included in total consumption. Moreover, the health investment may also include an individual's medical cure activities when he/she is caught by some diseases or infections, in that these actions can shorten ill health time and/or avoid incidental death caused by illness (Zon and Muysken, 2003). The third way of health improvement may be related to an individual's knowledge on health protection and life behavior. Since the goal of this paper is to study the relationship between the health and the long term growth, we mainly focus on health derived from improvement in nutrition and consumption. In the long term, just as Fogel (1994a, 1994b, 2002) and Strauss and Thomas (1998) indicated, income and hence total consumption is the main force that promote health improvement. To this end, we assume that health is determined mainly by an agent's consumption, and people with more consumption will be much healthier, though other factors are also crucial factors to determine the health level. Therefore, we assume that the health generation function is given below¹

$$h = h(c) \tag{5}$$

¹Note that in equation (5), health is considered as a flow variable rather than a stock variable and hence no depreciation is allowed as well. However, even if in the case that health is a stock variable and there exists health capital depreciation, the general conclusion of the paper is not affected.

We assume that the marginal health productivity of consumption is non-negative and non-increasing:

$$h'(c) \geq 0, \quad h''(c) \leq 0 \quad (6)$$

The assumption of nondecreasing $h(c)$ implies that, with the increase of consumption, the health capital h will at least not decrease. Alternatively, we can assume that $h(c)$ is not a monotonic function. For example, there exists a consumption level, $\bar{c} > 0$, such that $h(c)$ increases when consumption is less than $\bar{c}$; and the function $h(c)$ is kept constant when consumption is larger than $\bar{c}$. That is to say, we have $h'(c) \geq 0$, when $c < \bar{c}$; and $h'(c) \leq 0$, otherwise. We will discuss this kind of health generation function in section 3.3.

In order to solve the consumer's optimization problem, we define the Hamiltonian associated with the optimization problem

$$H = u(c) + \lambda[f(k, h(c)) - c - \delta k] \quad (7)$$

where λ is the co-state variable representing the marginal utility of physical capital investment measured in utility. By the Pontryagin's Principle, we obtain the first-order conditions

$$\lambda = u'(c) + \lambda f_h(k, h(c))h'(c) \quad (8)$$

$$\dot{\lambda} = \lambda[\beta + \delta - f_k(k, h(c))] \quad (9)$$

and the transversality condition $\lim_{t \rightarrow \infty} \lambda k e^{-\beta t} = 0$.

PROPOSITION 1. *Under the above assumptions on the utility function, production function and health generation function, if and only if a pair of real number, $(c(t), k(t))$, satisfies*

$$1 > f_h(k, h(c))h'(c) \quad (10)$$

then the pair $(c(t), k(t))$ satisfying equations (6), (8), (9) and the transversality condition which maximizes the objective function arrives.

Proof. (See appendix A) ■

Equation (8) indicates that the marginal value of physical capital investment equals the marginal value of consumption, which is the sum of the marginal utility of consumption and the marginal contribution of consumption to production. From equation (8), we can express λ as a function of consumption and capital stock, $\lambda(c, k)$.

$$\lambda = \frac{u'(c)}{1 - f_h(k, h(c))h'(c)} \quad (11)$$

In equation (11), $f_h(k, h(c))h'(c)$ denotes the increase in production brought by increasing the unit consumption through increasing health capital and hence improving productivity, and $1 - f_h(k, h(c))h'(c)$ denotes the cost of increasing the unit consumption measured in consumption goods. Hence, the right side of equation (11) represents the marginal value of increasing the unit consumption or the marginal cost of increasing the unit investment measured in utility. The left side of (11) represents the marginal value of investment. Therefore, equation (11) implies that the agent divides his/her income between investment and consumption subject to that the marginal value of investment equals the marginal cost. Compared with the standard Ramsey model, the uniqueness of this consumption optimal condition is that there is an additional term $f_h(k, h(c))h'(c)$ in the denominator of the right side in equation (11). If consumption has no effect on health, i.e., $h'(c) = 0$, then equation (11) is the same as in the standard Ramsey model.

From equation (11), we know why the condition of $1 > f_h(k, h(c))h'(c)$ should be satisfied if an agent's investment is optimal. Given any positive investment, as we can see from equation (11), if $1 \leq f_h(k, h(c))h'(c)$, then the marginal value of investment measured in utility will be negative or zero. Since the marginal utility of consumption, $u'(c)$, is definitely positive, a decrease in investment or/and an increase in consumption always increases the utility. Therefore, if $1 \leq f_h(k, h(c))h'(c)$, the agent who maximizes his/her lifetime utility will keep increasing his/her consumption and decreasing his/her investment till the marginal value of investment becomes positive and equals the marginal cost of investment.

Differentiating equation (11) with respect to c and k respectively, we are able to obtain the following short-run effects of consumption and capital stock on the marginal value of capital:

$$\lambda_c = \frac{u_{cc}[1 - f_h(k, h(c))h'(c)] + u_c[f_{hh}(k, h(c))(h'(c))^2 + f_h(k, h(c))h''(c)]}{[1 - f_h(k, h(c))h'(c)]^2} < 0 \quad (12)$$

$$\lambda_k = \frac{u_c f_{hk}(k, h(c))h'(c)}{[1 - f_h(k, h(c))h'(c)]^2} > 0 \quad (13)$$

From equations (12) and (13), it is clear that when consumption increases, the marginal value of investment will decrease, which is the same as the standard Ramsey model. The difference between our model and the standard Ramsey model is that the marginal value of investment decrease more in our model than in the standard Ramsey model, which results from the decreasing marginal health productivity of consumption ($u_c f_{hh}(k, h(c))(h'(c))^2$) and the decreasing marginal productivity of health ($u_c f_h(k, h(c))h''(c)$). However, when capital stock increases, the marginal value of investment

will increase, which is constant in the standard Ramsey model. The reason for this result is that in a standard Ramsey model, the marginal cost of investment, $u'(c)$, has no correlation with the capital stock which results in the marginal value of the optimal investment, which also equals to $u'(c)$, has no correlation with the capital stock. However, in our model, the marginal cost of investment, $u'(c)/[1 - f_h(k, h(c))h'(c)]$, is determined not only by consumption but also by capital stock. Therefore, when capital stock increases, the marginal productivity of capital will also increase, and hence the decrease in production brought by increasing the unit consumption will decrease. Consequently, with capital stock increasing, the marginal value of the optimal consumption or/and the marginal cost of the optimal investment will increase, which results in the increasing marginal value of the optimal investment, λ .

By equations (5), (8), (9) and (11), we derive the dynamic equation of consumption as follows

$$\dot{c} = -\frac{\lambda}{\lambda_c}[f_k(k, h(c)) - \delta - \beta] - \frac{\lambda_k}{\lambda_c}[f(k, h(c)) - c - \delta k] \quad (14)$$

Equations (2) and (14) determine the accumulation paths for capital stock and consumption. In the following sections, we analyze the dynamic behavior of the consumption, capital accumulation, and hence health accumulation.

3. DYNAMICS OF CAPITAL ACCUMULATION AND CONSUMPTION

By equations (2) and (14), the consumption and the capital stock approach the steady-state value when $\dot{c} = \dot{k} = 0$. It can be characterized as

$$f(k, h(c)) - c - \delta k = 0 \quad (15)$$

$$f_k(k, h(c)) - \delta - \beta = 0 \quad (16)$$

Under the assumption of the neoclassical production function, the existence of a steady state is obvious. But we cannot guarantee its uniqueness. We will give examples for the existence of unique steady state and multiple steady states. In Appendix B, we study the stability of the steady state. The saddle-point stability requires that

$$\Lambda \equiv \beta h'(c)f_{kh}(k, h(c)) + [1 - f_h(k, h(c))h'(c)]f_{kk}(k, h(c)) < 0 \quad (17)$$

In generally, we cannot determine the stability and the uniqueness of the steady state, we will present some examples to analyze it.

3.1. Unique Steady State

Consider the following special forms of the utility function, the output production function and the health generation function

$$u(c) = \frac{c^{1-\sigma} - 1}{1-\sigma}, \quad f(k, h) = Ak^\alpha h^{1-\alpha}, \quad h(c) = \xi c + \eta c^\theta \quad (18)$$

where A denotes technology level, α the elasticity of capital with respect to output, σ the intertemporal substitute elasticity. θ , ξ and η are parameters in the health generation function. All these parameters are assumed to be positive constants. By equations (8), (15) and (16), the steady state satisfies

$$c^{-\sigma} + \lambda A(1-\alpha)k^\alpha (\xi c + \eta c^\theta)^{-\alpha} (\xi + \eta \theta c^{\theta-1}) = \lambda \quad (19)$$

$$Ak^{\alpha-1} (\xi c + \eta c^\theta)^{1-\alpha} - c/k - \delta = 0 \quad (20)$$

$$A\alpha k^{\alpha-1} (\xi c + \eta c^\theta)^{1-\alpha} - \beta - \delta = 0 \quad (21)$$

Therefore, if $\frac{\alpha}{\eta(\beta+\delta-\alpha\delta)} \left(\frac{\beta+\delta}{A\alpha} \right)^{\frac{1}{1-\alpha}} > \frac{\xi}{\eta}$, then equations (19), (20), and (21) determine the unique steady state:

$$c^* = \left[\frac{\alpha}{\eta(\beta+\delta-\alpha\delta)} \left(\frac{\beta+\delta}{A\alpha} \right)^{\frac{1}{1-\alpha}} - \frac{\xi}{\eta} \right]^{\frac{1}{\theta-1}} \quad (22)$$

$$k^* = c^* \alpha / (\beta + \delta - \delta \alpha) \quad (23)$$

It is easy to verify that the saddle-point stable condition of equation (17) is satisfied when β and/or ξ are small enough to ensure a unique steady state which is saddle-point stable.

If we set the parameters as: $\delta = 0.1$, $\alpha = 0.7$, $\xi = 0.01$, $\eta = 0.5$, $\theta = 0.5$, $\beta = 0.05$, $\sigma = 0.5$, and $A = 1$, then the associated capital stock $k = 2.4316$, the consumption level $c = 0.27709$, the health capital $h = 0.54148$, the output $y = 1.54953$, and $\Lambda = -1.4 \times 10^{-5}$. As a result, the steady state is saddle-point stable. We also present in Table 1 the simulation results of the corresponding equilibrium values of the variables of this economy when we assume different values of A ranging from 1 to 1.5 while assuming other parameters unchanged.²

Based on these simulation results, we have the following findings on the effects of health on economic growth. First, the above results indicate

²Our simulation results indicate that when A is greater than 1.5, we need much less ξ or β to guarantee the existence and stability of the steady state. This is because when A is too large, the condition of the existence and stability of steady state, $[\alpha/(\eta\beta + \eta\delta - \eta\alpha\delta)][(\beta + \delta)/(A\alpha)]^{\frac{1}{1-\alpha}} > \xi/\eta$, can not be satisfied.

TABLE 1.

Simulation results of unique steady state

A	c	k	h	y	$c + \delta k$	Λ	g_y	g_A
1	145.01	1268.85	7.47	271.90	271.90	$-1.4E - 05$		
1.01	157.50	1378.17	7.85	292.40	292.40	$-1.3E - 05$	0.075	0.01
1.03	185.75	1625.28	8.67	348.27	348.27	$-1.1E - 05$	0.28	0.03
1.07	258.16	2258.94	10.62	465.96	465.96	$-7.5E - 06$	0.59	0.07
1.10	330.62	2892.95	12.40	619.92	619.92	$-5.7E - 06$	1.28	0.10
1.50	22468.59	196600.2	299.63	42128.61	42128.61	$-2.9E - 08$	153.94	0.50

Note: g_y denotes the rate of technology progress, g_A the output growth rate and other parameters are the same as defined in the previous part of this paper.

that the economy has a steady state and there is no persistent economic growth in this economy. Hence, even if a rise in consumption and nutrition can improve health capital and hence improve labor productivity, the health capital improvement is not able to induce persistent economic growth. Therefore, the improvement in health capital derived from increment in consumption and nutrition is not the motivation but the by-product of economic growth, which is consistent with what was stated in Boumol (1967) and Zou and Muysken (2001, 2003). Second, from Table 1 we can see that, when there is 1 percent increment of technology level from 1 to 1.01, the output increases by 7.5 percent. In contrast, when technology level increases by 50 percent from 1 to 1.5, the production output increases by about 154 times. Therefore, we find that while improvement in health capital can not introduce persistent economic growth, it is able to enlarge the economic growth driven by exogenous technology, which is consists with Fogel's result. This conclusion is also correct when there are multiple steady states in the economy as what would be discussed in the following section.

3.2. The Existence of Multiple Steady States

If we change the production function to

$$f(k, h) = Ak^\alpha h^{1-\alpha} + Bk^{\omega_3} + Dh^{\omega_4} \quad (24)$$

where $\alpha, \omega_3, \omega_4, A, B$ and D are positive constants. The utility function and the health generation function are still the same as equation (18).

Under the specified functions, we will discuss the existence of steady state and the stability of them. For simplicity, we discuss these using numerical solutions.

Case 1: Set the parameters as: $\theta = 0.5$, $\alpha = 0.5$, $\omega_3 = 0.7$, $\omega_4 = 0.1$, $A = 0.5$, $B = 0.5$, $\delta = 0.15$, $\xi = 0.4$, $\eta = 0.1$, $\beta = 0.1$, and $D = 0.3$. In this

case, we get the steady states values as:

$$\begin{array}{lll} k_1^* = 0.00475, & c_1^* = 0.27499, & h_1^* = 0.16244 \\ k_2^* = 0.18408, & c_2^* = 0.49376, & h_2^* = 0.26777 \\ k_3^* = 8.33639, & c_3^* = 2.59078, & h_3^* = 1.19727 \end{array}$$

and the corresponding Hamiltonian multipliers are $\lambda_1^* = 2.11286$, $\lambda_2^* = 1.66239$, and $\lambda_3^* = 0.88170$ respectively.

We can prove that k_1^* and k_3^* are the saddle-point stable steady states while the second steady state capital stock k_2^* is a critical steady state. If the initial capital stock is less than k_2^* , the capital stock, consumption, and health will converge to the first steady state. In contrast, if the initial capital stock is larger than k_2^* , the capital stock, the consumption, and the health will rise to the third steady state.

Case 2: We consider the situation with low marginal productivity of consumption and we select the parameters as: $\theta = 0.5$, $\alpha = 0.5$, $\omega_3 = 0.7$, $\omega_4 = 0.1$, $A = 0.5$, $B = 0.5$, $\delta = 0.15$, $\xi = 0.1$, $\eta = 0.1$, $\beta = 0.1$, and $D = 0.3$. Now, we can get the steady states as:

$$\begin{array}{lll} k_1^* = 0.00038, & c_1^* = 0.23519, & h_1^* = 0.07202 \\ k_2^* = 0.64704, & c_2^* = 0.65349, & h_2^* = 0.14619 \\ k_3^* = 4.67104, & c_3^* = 1.44939, & h_3^* = 0.26533 \end{array}$$

and the corresponding Hamiltonian multipliers are $\lambda_1^* = 2.21422$, $\lambda_2^* = 1.39389$, and $\lambda_3^* = 0.99177$ respectively.

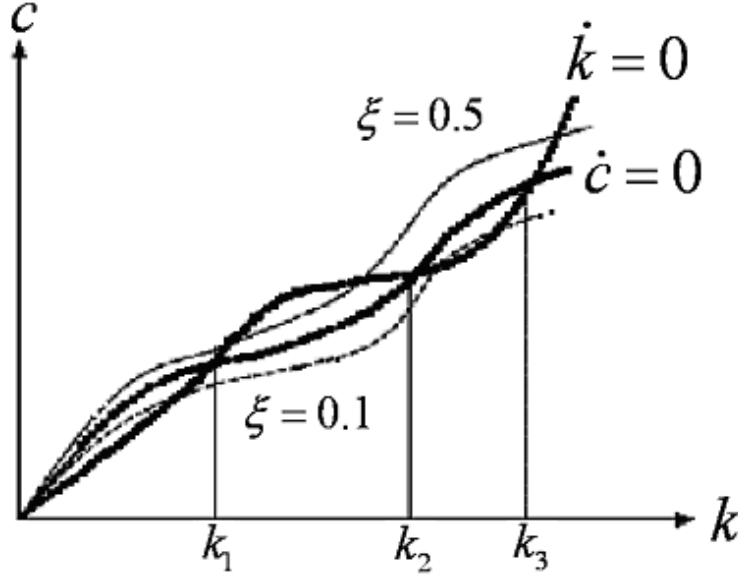
Case 3: We consider the situation with high marginal productivity of consumption and we select the parameters as: $\theta = 0.5$, $\alpha = 0.5$, $\omega_3 = 0.7$, $\omega_4 = 0.1$, $A = 0.5$, $B = 0.5$, $\delta = 0.15$, $\xi = 0.5$, $\eta = 0.1$, $\beta = 0.1$, and $D = 0.3$, we can get the steady states as:

$$\begin{array}{lll} k_1^* = 0.01137, & c_1^* = 0.29984, & h_1^* = 0.20468 \\ k_2^* = 0.09943, & c_2^* = 0.42903, & h_2^* = 0.28002 \\ k_3^* = 9.99119, & c_3^* = 3.08705, & h_3^* = 1.71923 \end{array}$$

and the corresponding Hamiltonian multipliers are $\lambda_1^* = 2.04917$, $\lambda_2^* = 1.77569$, and $\lambda_3^* = 0.84564$ respectively. The results from this situation are very much similar to those of the previous example.

We present the multiple steady states in Figure 1. The solid curves are $\dot{c} = 0$, and $\dot{k} = 0$ when $\xi = 0.4$. These two curves suggest that there are three steady states. In particular, the steady states k_1 and k_3 are saddle-point stable, while k_2 is unstable. With the increment of ξ (say $\xi = 0.5$), the curve $\dot{c} = 0$ will shift up. If ξ decreases, for example to $\xi = 0.1$, the

FIG. 1. The Existence of Multiple Steady States



curve $\dot{c} = 0$ will shift down. For further details on the dynamics, please check Figure 1.

3.3. A Nonmonotonic Health Generation Function

In the previous sections, we analyze the dynamics of the consumption accumulation path and the capital stock accumulation path under the assumption of a monotonic health generation function. In this subsection, we analyze these dynamics under the assumption of a nonmonotonic health generation function. Suppose we define the health generation function as

$$h(c) = \xi c + \eta c^\theta \quad (25)$$

where ξ and $\theta > 1$ are positive constants, $\eta < 0$ is a negative constant. The production function and utility function are specified the same as in equations (24) and (18), respectively:

$$u(c) = \frac{c^{1-\sigma} - 1}{1-\sigma} \quad (26)$$

$$f(k, h) = Ak^\alpha h^{1-\alpha} + Bk^{\omega_3} + Dh^{\omega_4} \quad (27)$$

where $\sigma, \alpha, \omega_3, \omega_4$, and A are positive constants.

For the selected parameters: $\theta = 2, \alpha = 0.5, \omega_3 = 0.6, \omega_4 = 0.1, A = 0.5, B = 0.5, \delta = 0.15, \xi = 0.5, \eta = 0.1, \beta = 0.1$, and $D = 0.3$, we obtain

the critical consumption level $\bar{c} = 2.5$, so we have $h'(c) > 0$ when $c \leq 2.5$; and $h'(c) \leq 0$ otherwise. We get the steady states:

$$\begin{array}{lll} k_1^* = 0.000004, & c_1^* = 0.24946, & h_1^* = 0.11851 \\ k_2^* = 0.30234, & c_2^* = 0.60283, & h_2^* = 0.26508 \\ k_3^* = 4.57422, & c_3^* = 1.63457, & h_3^* = 0.55010 \end{array}$$

and the corresponding Hamiltonian multipliers are $\lambda_1^* = 2.20676$, $\lambda_2^* = 1.49573$, $\lambda_3^* = 0.90285$, respectively.

We can prove that k_1^* and k_3^* are the saddle-point stable steady states capital stocks. The second steady state capital stock k_2^* is a critical steady state. If the initial capital stock is less than k_2^* , the capital stock, the consumption, and the health will converge to the first steady state. In contrast, if the initial capital stock is larger than k_2^* , the capital stock, the consumption, and the health will rise to the third steady state.

4. POLICY ANALYSIS

Introducing government tax to the above model, the budget constraint of the agent can be rewritten as

$$\dot{k} = (1 - \tau_y)f(k, h(c)) - (1 + \tau_c)c - \delta k \quad (2')$$

where τ_y and τ_c are the income tax rate and the consumption tax rate, respectively.

The first-order conditions (8) and (9) can be rewritten as

$$u_c + \lambda(1 - \tau_y)f_h(k, h(c))h'(c) = \lambda(1 + \tau_c) \quad (8')$$

$$\dot{\lambda} = \beta\lambda - \lambda[(1 - \tau_y)f_k(k, h(c)) - \delta] \quad (9')$$

with the same transversality condition as defined in the previous section.

From equation (8'), we can also express the marginal value of physical capital investment as a function of consumption and capital stock

$$\lambda = u_c/D \quad (28)$$

where $D = 1 + \tau_c - (1 - \tau_y)f_h(k, h(c))h'(c)$.

The steady state is characterized by the following two equations:

$$(1 - \tau_y)f(k, h(c)) - (1 + \tau_c)c - \delta k = 0 \quad (29)$$

$$(1 - \tau_y)f_k(k, h(c)) - \delta - \beta = 0 \quad (30)$$

Suppose the steady state exists and is saddle-point stable. We then focus on the effects of the income tax and consumption tax on the steady-state consumption and capital stock. Taking total differentiate with respect to τ_c and τ_y on equations (29) and (30), we get

$$\begin{pmatrix} \beta & (1 - \tau_y)f_h(k, h(c))h'(c) - 1 - \tau_c \\ (1 - \tau_y)f_{kk}(k, h(c)) & (1 - \tau_y)f_{kh}(k, h(c))h'(c) \end{pmatrix} \begin{pmatrix} dk \\ dc \end{pmatrix} \\ = \begin{pmatrix} c \\ 0 \end{pmatrix} d\tau_c + \begin{pmatrix} f \\ -f_k(k, h(c)) \end{pmatrix} d\tau_y \quad (31)$$

The effects of consumption tax on the steady-state capital stock and consumption can be derived as

$$\begin{aligned} \frac{dk}{d\tau_c} &= -\frac{c(1 - \tau_y)f_{kh}(k, h(c))h'(c)}{\Delta} < 0, \\ \frac{dc}{d\tau_c} &= \frac{c(1 - \tau_y)f_{kk}(k, h(c))}{\Delta} < 0 \end{aligned} \quad (32)$$

where

$$\begin{aligned} \Delta &= (1 - \tau_y)f_{kh}(k, h(c))h'(c)\beta \\ &\quad - (1 - \tau_y)f_{kk}(k, h(c))((1 - \tau_y)f_h(k, h(c))h'(c) - 1 - \tau_c) \end{aligned} \quad (33)$$

which is negative from the saddle-point stability condition (14). From equation (32), we find that with the increase of the consumption tax rate, the steady-state capital stock and consumption will decrease.

Similarly, for the effects of the income tax on the steady-state capital stock and consumption, from equation (31), we have

$$\begin{aligned} \frac{dk}{d\tau_y} &= -\frac{1}{\Delta}f(1 - \tau_y)f_{kh}(k, h(c))h'(c) \\ &\quad + f_k(k, h(c))((1 - \tau_y)f_h(k, h(c))h'(c) - 1 - \tau_c) \end{aligned} \quad (34)$$

$$\frac{dc}{d\tau_y} = -\frac{1}{\Delta}\beta f_k(k, h(c)) + f(1 - \tau_y)f_{kk}(k, h(c)) \quad (35)$$

which show that the effects of the income tax rate on the steady-state capital stock and consumption is ambiguous.

In this paper, we identify the negative effects of consumption tax on the long-run consumption level and capital stock. In other words, we find that with the increase in the consumption tax rate, the long-run capital stock

and consumption level will decrease. The reason is that with the increment of the consumption tax rate, the cost of consumption will be increasing, which in turn decreases the long-run consumption level. However, with the decrease of the consumption, the health generation will decrease and the output and investment will also decrease which results in decrease in the capital stock. These causality relationships found in this study are significantly different from what we find in the existing literature. In the traditional literature, such as Rebelo (1991), the consumption tax, which decreases the long-run consumption level, has no effects on the long-run capital stock.

Furthermore, we present the ambiguous effects of the income tax rate on the steady-state capital stock and the consumption, which are different from the negative effects of the income tax rate on the capital stock in the existing literature. The reason behind this discrepancy between our results and those of the existing literature is that as the income tax rate increases, the return on the capital stock will decrease which in turn decreases the steady state capital stock and increases the consumption level. With the increase of the consumption level, the health generation will increase, and the marginal productivity of the capital stock will also increase which leads to the increment of the returns on the capital stock, which will lead to a higher steady state capital stock. The overall effect of the income tax rate on the capital stock will depend on the interaction of the above two effects. Thus, we derive ambiguous effects of the income tax rate on the steady-state capital stock as well as on the long-run consumption level.

5. CONCLUSIONS

In this paper, we first presented a theoretical framework to discuss the consumption path and the capital accumulation path by introducing health as a sector of output production. We assume that health can be generated by private consumption. Based on the simulation results of Case 1 in Section 3.1 we found that, when the improvement in health capital is induced by a rise in consumption, this consumption and nutrition driven health capital is not the motivation but the by-product of economic growth, which is consistent with the conclusion in Boumol (1967) and Zon and Muysken (2001, 2003) concluded. However, we also found that the resulting health capital is able to expand the economic growth driven by exogenous technology, which is consistent with the result of Fogel (1994a, 1994b, and 2002).

Secondly, under the assumption of a nonmonotonic health generation function, we could not derive the uniqueness of steady state, like Kurz (1968). In the given numerical examples, we derived three steady states under some given parameter specifications. The existence of multiple steady

states can be used to explain the economic growth puzzle posed by Lucas (1993): Why would two countries like South Korea and the Philippines, whose wealth and endowment levels were quite close not so long ago, differ so drastically in their recent growth experience.

Lastly, we discussed the effects of consumption tax and income tax on long-run capital stock and consumption. The results obtained from the theoretical framework in our study were different from those found by Rebelo (1991). We found negative effects of the consumption tax rate on the long-run capital stock and the consumption while the effects of the income tax rate on the long-run capital stock and the consumption level were ambiguous.

The model has three important features: (1) treating health as a simple function of consumption, which enable the study of the relationship between health and long-term economic growth in an aggregate macroeconomic model; (2) the existence of multiple equilibria of capital stock, health, and consumption, which is more consistent with the real world situation — rich countries may end up with higher capital stock, better health, and higher consumption level than poor countries; (3) the fundamental proposition of a consumption tax instead of capital taxation based on the traditional growth model does not hold anymore in our model. As long as consumption goods contribute to health formation, the issue of a consumption tax versus an income (or capital) tax should be re-examined.

For future theoretical studies, research should focus on the monetary policy implications based on the framework proposed in our study. We will also extend our model to a two-sector framework to consider simultaneously the physical capital and the human capital. Furthermore, it is interesting to analyze the effects of government expenditures on the long-run capital stock and the consumption with the consideration of health. For future empirical studies, we are interested in studying whether the gap between rich and poor countries is widening, which is suggested by the multiple equilibria framework proposed in our study.

APPENDIX A

The proof of Proposition 1:

1) We proof equation (10) is a necessary condition.

By the Hamilton function, we have

$$\frac{\partial H}{\partial c} = u'(c) + \lambda[f_h(k, h(c))h'(c) - 1] \quad (\text{A.1})$$

$$\frac{\partial H}{\partial k} = \lambda[f_k(k, h(c)) - \delta] \quad (\text{A.2})$$

and

$$\frac{\partial^2 H}{\partial c^2} = u''(c) + \lambda[f_h(k, h(c))h''(c) + f_{hh}(k, h(c))(h'(c))^2] \quad (\text{A.3})$$

$$\frac{\partial^2 H}{\partial c \partial k} = \lambda f_{ck}(k, h(c))h'(c) \quad (\text{A.4})$$

$$\frac{\partial^2 H}{\partial k^2} = \lambda f_{kk}(k, h(c)) \quad (\text{A.5})$$

If the objective function arrives at maximum when (c, k) satisfies the first-order conditions, then the Hamilton function must be nonpositive. Therefore, $\frac{\partial^2 H}{\partial c^2}$ and $\frac{\partial^2 H}{\partial k^2}$ must be nonpositive and the determinant of Hessian second-order matrix must be nonnegative. By assumption (3), $f_{kk} < 0$, in order that $\frac{\partial^2 H}{\partial k^2} \leq 0$, there must be $\lambda \geq 0$, which result in $f_h(k, h(c))h'(c) < 1$.

2) We prove equation (10) is a sufficient condition. First, when $f_h(k, h(c))h'(c) < 1$, then $\lambda > 0$. It is obviously that $\frac{\partial^2 H}{\partial c^2}$ and $\frac{\partial^2 H}{\partial k^2}$ are positive. Second, the determinant of the Haiseir matrix is

$$\begin{aligned} \begin{vmatrix} H_{cc} & H_{ck} \\ H_{kc} & H_{kk} \end{vmatrix} &= \lambda f_{kk} \{u'' + \lambda[h''f_h + h'^2 f_{hh}]\} - (\lambda h' f_{hk})^2 \\ &= \lambda f_{kk} u'' + \lambda^2 [f_{kk} f_h h'' + h'^2 (f_{kk} f_{hh} - f_{hk}^2)] \end{aligned}$$

By the assumption on the utility function, production function and health generation function, the determinant of Hessian matrix must be positive. And hence, (c, k) satisfying equation (6), (8), (9) and transversality condition maximizes the objective function arrives.

APPENDIX B

The condition of saddle-point stability

We linearize system (2') and (11) around the steady state

$$\begin{pmatrix} \frac{dk}{dt} \\ \frac{dc}{dt} \end{pmatrix} = J \begin{pmatrix} k - k^* \\ c - c^* \end{pmatrix} \quad (\text{B.1})$$

where

$$J = \begin{pmatrix} \beta & 10f_h(k, h(c))h'(c) \\ -\frac{\lambda}{\lambda_c} f_{kk} - \frac{\lambda_k}{\lambda_c} \beta & 0 \end{pmatrix} \quad (\text{B.2})$$

is the coefficient matrix associated with the above linear system. The eigenvalues μ_1 and μ_2 of the matrix J satisfy

$$\mu_1 + \mu_2 = \beta \quad (\text{B.3})$$

$$\mu_1\mu_2 = -\frac{\lambda}{\lambda_c}\{\beta h'(c)f_{hk}(k, h(c)) - [h'(c)f_h(k, h(c)) - 1]f_{kk}(k, h(c))\} \quad (\text{B.4})$$

Thus, the saddle-point stability requires that

$$\beta h'(c)f_{hk}(k, h(c)) - [h'(c)f_h(k, h(c)) - 1]f_{kk}(k, h(c)) < 0. \quad (\text{B.5})$$

REFERENCES

- Arora, S., 2001. Health, human productivity, and long-term economic growth. *Journal of Economic History* **61**, 699-749.
- Barro, R. J., 1996. Health and economic growth. Paper presented at the "Senior Policy Seminar on Health, Human Capital and Economic Growth: Theory, Evidence and Policies", Pan American Health Organization and Inter-American Development Bank, Washington, DC.
- Baumol, W. J., 1967. Macroeconomics of unbalanced growth: The anatomy of urban crisis. *American Economic Review* **57**, 415-426.
- Bloom, D. E., D. Canning, and D.T. Jamison, 2004. Health, wealth and welfare. *Finance and Development* **41**, 10-15.
- Bloom, D. E., D. Canning, and J. Sevilla, 2004. The effect of health on economic growth: A production function approach. *World Development* **32**, 1-13.
- Bloom, D. E. and J. G. Williamson, 1998. Demographic transitions and economic miracles in emerging asia. *World Bank Economic Review* **12**, 419-455.
- Bunzel, H. and X. Qiao, 2005. Endogenous lifetime and economic growth revisited. *Economics Bulletin* **5**, 1-8.
- Chakraborty, S., 2004. Endogenous lifetime and economic growth. *Journal of Economic Theory* **116**, 119-137.
- Ehrlich, I. and T. F. Lui, 1991. Intergenerational trade, longevity and economic growth. *Journal of Political Economy* **99**, 1029-1059.
- Fogel, R. W., 1994. Economic growth, population theory, and physiology: The bearing of long-term processes on the making of economic policy. *American Economic Review* **84**, 369-395.
- Fogel, R. W., 1994. The relevance of Malthus for the study of mortality today: Long-run influences on health, mortality, labor force participation, and population growth. NBER Working Paper, h0054.
- Fogel, R. W., 2002. Nutrition, physiological capital, and economic growth. Paper presented at the "Senior Policy Seminar on Health, Human Capital and Economic Growth: Theory, Evidence and Policies," Pan American Health Organization and Inter-American Development Bank, Washington, DC.
- Forster, B.A., 1989. Optimal health investment strategies. *Bulletin of Economic Research* **41**, 45-57.
- Grossman, M., 1972. The demand for health: A theoretical and empirical investigation. NBER, Occasional Paper 119, Columbia University Press.
- Hemmi, N., K. Tabata, and K. Futagami, 2007. The long-term care problem, precautionary saving, and economic growth. *Journal of Macroeconomics* **29**, 60-74.
- Hicks, N. L., 1979. Growth vs. basic needs: Is there a trade-off. *World Development* **7**, 985-994.

- Hosoya, K., 2002. Health, longevity, and the productivity slowdown. Project on Intergenerational Equity, Institute of Economic Research, Hitotsubashi University, Discussion Paper Series, No. 25.
- Hosoya, K., 2003. Tax financed government health expenditure and growth with capital deepening externality. *Economics Bulletin* **5**, 1-10.
- Issa, H., 2003. Human capital, demographic transition, and economic growth. CGBCR Discussion Paper Series. University of Manchester, Manchester.
- Johansson, P.O. and K.G. Lofgren, 1995. Wealth from optimal health. *Journal of Health Economics* **14**, 65-79.
- Kalemli-Ozcan, S., H. E. Ryder, and D. N. Weil, 2000. Mortality decline, human capital investment, and economic growth. *Journal of Development Economics* **62**, 1-23.
- Kurz M., 1968. Optimal economic growth and wealth effects. *International Economic Review* **9**, 348-357.
- Laxminarayan, R., 2004. Does reducing malaria improve household living standards? *Tropical Medicine and International Health* **9**, 267-272.
- Liu, G. G., W. H. Dow, A. Z. Fu, and J. Akin, 2008. Income productivity in China: On the role of health. *Journal of Health Economics* **27**, 27-44.
- Lorentzen, P. L., J. McMillan, and R. Wacziarg, 2008, Death and development. *Journal of Economic Growth* **13**, 81-124.
- Lucas, R.E., 1988. On the mechanics of economic development. *Journal of Monetary Economics* **22**, 3-42.
- Lucas, R.E., 1993. Making a miracle. *Econometrica* **61**, 251-272.
- McDonald, S. and J. Roberts, 2006. AIDS and economic growth: A human capital approach. *Journal of Development Economics* **80**, 228-250.
- Mertzer, D., 1997. Account for future costs in medical cost-effectiveness analysis. *Journal of Health Economics* **16**, 33-64.
- Morand, O. F. 2004. Economic growth, longevity and the epidemiological transition. *European Journal of Health Economics* **5**, 166-174.
- Muurinen, J. M., 1982, Demand for health: A generalised grossman model. *Journal of Health Economics* **1**, 5-28.
- Muysken, J., I. H. Yetkiner, and T. Ziesemer, 1999. Health, labour productivity and growth. MERIT Research Memorandum, No. 099-030, University of Maastricht.
- Ramsey, F., 1928. A mathematical theory of saving. *Economic Journal* **38**, 543-559.
- Rebelo, S., 1991. Long-run policy analysis and long-run growth. *Journal of Political Economy* **99**, 500-521.
- Sachs, J. and A. Warner, 1997. Fundamental sources of long-run growth. *American Economic Review (Papers and Proceedings)* **87**, 184-188.
- Sanso, M. and R. M. Aisa, 2006. Endogenous longevity, biological deterioration and economic growth. *Journal of Health Economics* **25**, 555-578.
- Savedoff, W. D. and T. P. Schultz, 2000. Earnings and the elusive dividends of health. In: *Wealth from Health*, 1-34. Eds. Savedoff, W. D. and T. P. Schultz.
- Shurcke, M., M. McKee, D. Stuckler, R. S. Arce, S. Tsoлова, and J. Mortensen, 2006. The contribution of health to the economy in the european Union. *Public Health* **120**, 994-1001.

- Sohn, B., 2000. Health, nutrition, and economic growth. Ph.D. dissertation, Brown University.
- Strauss, J. and D. Thomas, 1998. Health, nutrition, and economic development. *Journal of Economic Literature* **36**, 766-817.
- Thomas, D. and E. Frankenberg, 2002. Health, nutrition and economic Prosperity: A microeconomic perspective. *Bulletin of the World Health Organization* **80**, 106-113.
- Weil, D.N., 2007. Accounting for the effect of health on economic growth. *Quarterly Journal of Economics* **122**, 1265-1306
- Wheeler, D., 1980. Basic needs fulfillment and economic growth. *Journal of Development Economics* **7**, 435-451.
- Zon, A. H. van and J. Muysken, 2001. Health and endogenous growth. *Journal of Health Economics* **20**, 169-185.
- Zon, A. H. van and J. Muysken, 2003. Health as a principal determinant of economic growth. Working paper, MERIT-Infonomics Research Memorandum.

Public Expenditures, Taxes, Federal Transfers, and Endogenous Growth¹

Liutang Gong²

Guanghua School of Management, Peking University, Beijing, China, 100871

and

Heng-fu Zou

CEMA, Central University of Finance and Economics, Beijing, China 100081

Shenzhen University, Wuhan University, and The World Bank

ABSTRACT

This paper extends the Barro (1990) model with single aggregate government spending and one flat income tax to include public expenditures and taxes by multiple levels of government. It derives the rate of endogenous growth and, with both simulations and special examples, examines how that rate changes with respect to federal income tax, local taxes, and federal transfers. It also discusses the growth and welfare-maximizing choices of taxes and federal transfers.

Keywords: Public expenditures; Taxes; Federal transfers; Endogenous growth.

JEL classification #: E0, H2, H4, H5, H7, O4, R5

1. Introduction

In an endogenous growth model, Barro (1990) has examined the effects on economic growth of aggregate government spending, including both aggregate public consumption and aggregate public investment. Subsequent work has extended Barro's analysis by looking into the composition of government expenditures and economic growth. For example, Easterly and Rebelo (1993) and Devarajan, Swaroop, and Zou (1998) have studied the growth effects of public spending on education, transportation, defense, and social welfare. Glomm and Ravikumar (1994), Hulten (1994), and Devarajan, Xie, and Zou (1998), among many others, have paid particular attention to the

¹We are indebted to the Professors Cuong Le Van and Myrna Wooders for their useful comments and help in revising this paper. This research was sponsored by the National Science Foundation for Distinguished Young Scholars (70725006).

²Mailing address: Liutang Gong, Guanghua School of Management, Peking University, Beijing, 100871, China. Tel: (8610) 6275-7768; Fax: (8610) 6275-3178; E-mail: ltgong@gsm.pku.edu.cn.

association between infrastructure and output growth.³

However, the structure of public expenditures and taxes among different levels of government has a fundamental impact on economic growth in light of the arguments related to fiscal federalism; see Oates (1972, 1973). In fact, the proper assignment of expenditures and taxes among federal and local governments and the proper design of intergovernmental transfers are prerequisites for efficient and equitable public service provision at both the national and local levels. One of the most important goals in establishing a sound intergovernmental fiscal relationship is to promote both local and national economic growth (see also Rivlin 1992, Bird 1993, Gramlich 1993, and Oates 1973).

In view of the important link between the design of intergovernmental fiscal relationships and economic growth, it is natural for us to extend the Barro model and provide an analytical framework for both theoretical and empirical research on the growth effects of public expenditures, taxes, and federal transfers in a federation or in multiple levels of government. This is the main task of our paper.

The remainder of this paper is organized as follows. Section 2 extends the Barro model with one aggregate government spending and one flat income tax to include: (1) public expenditures by both the federal government and local governments; (2) various taxes by both the federal government and local governments; and (3) a federal transfer to a locality⁴. Section 3 derives the rate of endogenous growth. With both simulations and special examples, section 4 examines the change in the rate of endogenous growth with respect to federal income tax, local income tax, and federal transfers. Section 5 derives the optimal federal government income tax rate, local government income tax rate, and the federal matching transfer for the locality. Section 6 presents a more general model with local government consumption tax and property tax. Section 7 concludes the paper.

2. The Model

Following Arrow and Kurz (1970), Barro (1990), and Turnovsky (2000), we

³Of course, the empirical analyses in many of these studies have followed the much-cited work of Aschauer (1989).

⁴The setup of our model is also very different from the dynamic analysis of Zou (1994, 1996) in quite a few respects. First, the focus here is the rate of endogenous growth instead of the traditional long-run analysis of the steady state; second, the representative agent's utility and production function are defined on both federal and local spending instead of only on local spending; third, federal taxation, federal transfer, and federal spending are fully integrated into the

introduce public expenditures by the federal government and local governments into the representative agent's utility function and production function. Federal spending is denoted by f , local public spending by s , and private consumption by c . The instantaneous utility of the representative agent is given by $u(c, f, s)$, which has the following properties:

$$u_c > 0, u_f > 0, u_s > 0, u_{cc} < 0, u_{ff} < 0, u_{ss} < 0. \quad (1)$$

To derive analytical solution for the endogenous growth rate, we extend the utility function of Barro (1990) as follows

$$u(c, f, s) = \frac{c^{1-\sigma} - 1}{1-\sigma} + \frac{f^{1-\sigma} - 1}{1-\sigma} + \frac{s^{1-\sigma} - 1}{1-\sigma}, \quad (2)$$

where $\sigma > 0$ is the inverse of the elasticity of intertemporal substitution.

The representative agent seeks to maximize a discounted utility, as given by

$$U = \int_0^\infty u(c, f, s) e^{-\rho t} dt, \quad (3)$$

where $0 < \rho < 1$ is the constant rate of time preference.

The agent has access to the extended Arrow-Kurz-Barro neoclassical production function

$$y = y(k, f, s), \quad (4)$$

where y is output and k is private capital stock.

The role of government services in both the utility function and the production function was introduced to dynamic analysis of public investment and growth by Arrow and Kurz (1970). The approach to endogenous growth models was popularized by Barro (1990). In recent studies on fiscal decentralization and growth, the Arrow-Kurz-Barro approach to preferences and technology has been extended to different public expenditures by multiple levels of government; see Brueckner (1996), Davoodi and Zou (1998), and Zhang and Zou (1998) for examples. Again, the production function is assumed to have the following standard properties:

$$y_k > 0, y_f > 0, y_s > 0, y_{kk} < 0, y_{ff} < 0, y_{ss} < 0.$$

In this paper, the production function takes the CES form,

$$y = (\alpha k^\theta + \beta f^\theta + \gamma s^\theta)^{\frac{1}{\theta}}, \quad (5)$$

where θ , α , β , and γ are positive constants with $\alpha + \beta + \gamma = 1$.

The federal government levies an income tax at the rate of τ_f , and a typical local government levies a local income tax⁵ (as in the case of state income tax in the United States) at the rate of τ_s . The federal government also makes a transfer to the local government in the form of a matching grant for local public spending at the rate of g . If both levels of government maintain balanced budgets, then their budget constraints can be written as

$$f = \tau_f y - gs \quad (6)$$

and

$$s = \tau_s y + gs, \quad (7)$$

respectively. Hence, federal public spending, f , equals total income tax, $\tau_f y$, minus the transfer to the local government, gs . Local government spending is financed by its income tax, $\tau_s y$ and the grant it receives from the federal government, gs .

Given the tax rates of the two levels of government, the budget constraint of the representative agent can be written as

$$\frac{dk}{dt} = (1 - \tau_f - \tau_s)y(k, f, s) - \delta k - c. \quad (8)$$

The representative agent chooses a consumption path and a capital-accumulation path to maximize his discounted utility in equation (3) subject to constraint (8), and with his initial capital stock given by $k(0) = k_0$.

The Hamiltonian associated with the optimization problem is defined as

$$H = u(c, f, s) + \lambda[(1 - \tau_f - \tau_s)y(k, f, s) - \delta k - c],$$

where λ is the costate variable, and it represents the marginal utility of wealth.

Now, the first-order conditions are

$$\frac{\partial u(c, f, s)}{\partial c} = \lambda, \quad (9)$$

$$\frac{d\lambda}{dt} = \rho\lambda - \lambda[(1 - \tau_f - \tau_s)\frac{\partial y(k, f, s)}{\partial k} - \delta], \quad (10)$$

and the transversality condition (TVC) is

⁵In the following section, we extend the model to a more general framework with consumption tax τ_c and property

$$\lim_{t \rightarrow \infty} \lambda k e^{-\rho t} = 0. \quad (11)$$

Specifically, for the utility function in equation (2) and the production function in equation (5), we rewrite equations (9) and (10) as follows:

$$c^{-\sigma} = \lambda \quad (12)$$

and

$$\frac{\dot{c}}{c} = \frac{1}{\sigma} \{ (1 - \tau_f - \tau_s) (\alpha k^\theta + \beta f^\theta + \gamma s^\theta)^{\frac{1}{\theta}-1} \alpha k^{\theta-1} - \rho - \delta \}. \quad (13)$$

Equation (12) states that the marginal utility of wealth equals the marginal utility of consumption at an optimum. Equation (13) is the familiar Euler equation for consumption with multiple government services and tax rates.

3. The Balanced Growth Rate

Suppose that the economy is on the balanced growth path where private consumption, private capital, federal government expenditure, local government expenditure, and output all grow at the same rate denoted as ϕ , i.e.,

$$\frac{\dot{c}}{c} = \frac{\dot{k}}{k} = \frac{\dot{f}}{f} = \frac{\dot{s}}{s} = \frac{\dot{y}}{y} = \phi. \quad (14)$$

Substituting condition (14) into equations (8) and (13), we obtain

$$\frac{\dot{c}}{c} = \frac{1}{\sigma} \{ (1 - \tau_f - \tau_s) \alpha (\alpha + \beta (\frac{f}{k})^\theta + \gamma (\frac{s}{k})^\theta)^{\frac{1}{\theta}-1} - \rho - \delta \} \quad (15)$$

and

$$\frac{\dot{k}}{k} = (1 - \tau_f - \tau_s) (\alpha + \beta (\frac{f}{k})^\theta + \gamma (\frac{s}{k})^\theta)^{\frac{1}{\theta}} - \rho - \delta - \frac{c}{k}. \quad (16)$$

From equation (15), we have

$$\alpha + \beta (\frac{f}{k})^\theta + \gamma (\frac{s}{k})^\theta = (\frac{\sigma \phi + \rho + \delta}{\alpha (1 - \tau_f - \tau_s)})^{\frac{\theta}{1-\theta}}. \quad (17)$$

Substituting equation (17) into equation (16), we derive the consumption-capital ratio as

$$\frac{c}{k} = (1 - \tau_f - \tau_s) (\frac{\sigma \phi + \rho + \delta}{\alpha (1 - \tau_f - \tau_s)})^{\frac{1}{1-\theta}} - \phi - \rho - \delta. \quad (18)$$

On the other hand, from government budget constraints (6) and (7), and combining

tax τ_k , but we cannot derive the explicit solution for the growth rate.

with equation (16), we obtain

$$\frac{s}{k} = \frac{\tau_s}{1-g} \left(\frac{\sigma\phi + \rho + \delta}{\alpha(1-\tau_f - \tau_s)} \right)^{\frac{1}{1-\theta}} \quad (19)$$

and

$$\frac{f}{k} = \left(\tau_f - \frac{g}{1-g} \tau_s \right) \left(\frac{\sigma\phi + \rho + \delta}{\alpha(1-\tau_f - \tau_s)} \right)^{\frac{1}{1-\theta}}. \quad (20)$$

Using equations (19) and (20), we have

$$\begin{aligned} \beta \left(\frac{f}{k} \right)^\theta + \gamma \left(\frac{s}{k} \right)^\theta &= \beta \left\{ \left(\tau_f - \frac{g}{1-g} \tau_s \right) \left(\frac{\sigma\phi + \rho + \delta}{\alpha(1-\tau_f - \tau_s)} \right)^{\frac{1}{1-\theta}} \right\}^\theta \\ &+ \gamma \left\{ \frac{\tau_s}{1-g} \left(\frac{\sigma\phi + \rho + \delta}{\alpha(1-\tau_f - \tau_s)} \right)^{\frac{1}{1-\theta}} \right\}^\theta. \end{aligned} \quad (21)$$

Substituting equation (21) into equation (17) yields the explicit solution for the growth rate ϕ :

$$\phi = \frac{1}{\sigma} \left[\frac{(1-\tau_f - \tau_s) \alpha^{\frac{1}{\theta}}}{(1-\beta(\tau_f - \frac{g\tau_s}{1-g})^\theta - \gamma(\frac{\tau_s}{1-g})^\theta)^{\frac{1}{\theta}-1}} - \rho - \delta \right]. \quad (22)$$

Equation (22) states that the growth rate is an explicit function of τ_f , τ_s , g , σ , ρ , α , β , γ , and δ .

For the endogenous growth rate ϕ to be positive in (22), it is required that

$$\frac{(1-\tau_f - \tau_s) \alpha^{\frac{1}{\theta}}}{(1-\beta(\tau_f - \frac{g\tau_s}{1-g})^\theta - \gamma(\frac{\tau_s}{1-g})^\theta)^{\frac{1}{\theta}-1}} - \delta > \rho. \quad (23)$$

At the same time, the TVC (11) gives

$$\rho > (1-\sigma) \left[\frac{(1-\tau_f - \tau_s) \alpha^{\frac{1}{\theta}}}{(1-\beta(\tau_f - \frac{g\tau_s}{1-g})^\theta - \gamma(\frac{\tau_s}{1-g})^\theta)^{\frac{1}{\theta}-1}} - \delta \right]. \quad (24)$$

Equations (23) and (24) present the condition for endogenous growth.

Now, the optimal growth paths for capital accumulation, $k(t)$, consumption, $c(t)$, federal government spending, $f(t)$, local government spending, $s(t)$, and output, $y(t)$ are derived as follows

$$k(t) = k(0)e^{\phi t}, \quad c(t) = c(0)e^{\phi t}, \quad f(t) = f(0)e^{\phi t}, \quad s(t) = s(0)e^{\phi t}, \quad y(t) = y(0)e^{\phi t}, \quad (25)$$

where the initial capital stock $k(0)$ is given, but the initial federal spending $f(0)$, local government spending $s(0)$, initial consumption $c(0)$, and initial output $y(0)$, will de

determined by the model.

First, from equations (6) and (7), we obtain

$$s(0) = \frac{\tau_s y(0)}{1-g} \quad (26)$$

and

$$f(0) = (\tau_f - \frac{g\tau_s}{1-g})y(0). \quad (27)$$

Now, from equation (5), we have

$$y(0)^\theta = \alpha k(0)^\theta + \beta(\tau_f - \frac{g\tau_s}{1-g})^\theta y(0)^\theta + \gamma(\frac{\tau_s}{1-g})^\theta y(0)^\theta.$$

Thus, we obtain the initial output as a function of the initial capital stock, $k(0)$ as follows:

$$y(0)^\theta = \frac{\alpha k(0)^\theta}{1 - \beta(\tau_f - \frac{g\tau_s}{1-g})^\theta - \gamma(\frac{\tau_s}{1-g})^\theta}. \quad (28)$$

With $y(0)$ given in equation (28), $f(0)$ and $s(0)$ can be determined by equations (26) and (27), respectively; $c(0)$ can be determined by the budget constraint of the agent:

$$c(0) = (1 - \tau_f - \tau_s)y(0) - (\delta + \phi)k(0). \quad (29)$$

With the aid of explicit solutions for the growth rate, we can analyze the effects on growth of the federal government's income tax, the local government's income tax, and the federal government's matching transfer. Using the explicit paths of the capital accumulation, consumption, and government's spending, we can derive the social welfare function, and then we can derive the optimal tax rate and government transfer to maximize the social welfare. We will process these in the next section.

4. Effects of Taxes and Federal Transfers

Differentiating equation (22) with respect to the federal government income tax rate, τ_f , local government's income tax rate, τ_s , and the federal matching grant for locality, g , we have

$$\frac{\partial \phi}{\partial \tau_f} = \frac{1}{\sigma} \left[\frac{(1 - \tau_f - \tau_s)(1 - \theta) \alpha^{\frac{1}{\theta}} \beta (\tau_f - \frac{g\tau_s}{1-g})^{\theta-1}}{(1 - \beta(\tau_f - \frac{g\tau_s}{1-g})^\theta - \gamma(\frac{\tau_s}{1-g})^\theta)^{\frac{1}{\theta}}} - \frac{\alpha^{\frac{1}{\theta}}}{(1 - \beta(\tau_f - \frac{g\tau_s}{1-g})^\theta - \gamma(\frac{\tau_s}{1-g})^\theta)^{\frac{1}{\theta}-1}} \right], \quad (30)$$

$$\frac{\partial \phi}{\partial \tau_s} = -\frac{1}{\sigma} \frac{\alpha^{\frac{1}{\theta}}}{(1 - \beta(\tau_f - \frac{g\tau_s}{1-g})^\theta - \gamma(\frac{\tau_s}{1-g})^\theta)^{\frac{1}{\theta}-1}} - \frac{1 - \tau_f - \tau_s}{\sigma} \frac{\alpha^{\frac{1}{\theta}} [\beta(\tau_f - \frac{g\tau_s}{1-g})^{\theta-1} g - \gamma(\frac{\tau_s}{1-g})^{\theta-1}] \frac{1-\theta}{1-g}}{(1 - \beta(\tau_f - \frac{g\tau_s}{1-g})^\theta - \gamma(\frac{\tau_s}{1-g})^\theta)^{-\frac{1}{\theta}}}, \quad (31)$$

and

$$\frac{\partial \phi}{\partial g} = \frac{1 - \tau_f - \tau_s}{\sigma} \frac{\alpha^{\frac{1}{\theta}} [\beta(\tau_f - \frac{g\tau_s}{1-g})^{\theta-1} - \gamma(\frac{\tau_s}{1-g})^{\theta-1}] \frac{1-\theta}{(1-g)^2}}{(1 - \beta(\tau_f - \frac{g\tau_s}{1-g})^\theta - \gamma(\frac{\tau_s}{1-g})^\theta)^{-\frac{1}{\theta}}}. \quad (32)$$

Equations (30), (31), and (32) state the ambiguous effects on growth of the federal government income tax rate, τ_f , local government's income tax rate, τ_s , and the federal matching grant for locality, g . For the intuition, we present some numerical solutions.

(Insert Figure 1 about here)

Figure 1 shows the relationship between the rate of endogenous growth, ϕ , and the federal government's income tax rate, τ_f , when the following base values are used for the structure of local government taxation and federal transfer: a local income tax at 10 percent: $\tau_s = 0.10$, and a federal matching grant at 50 percent: $g = 0.5$. We assume the following values for preference and technology parameters: $\alpha = 0.5$, $\rho = 0.05$, $\delta = 0.05$, $\sigma = 2$, and $\theta = 0.4$. For the parameters of β and γ , which represent the marginal productivity of federal government spending and local government spending, respectively, we consider three cases: $\beta = \gamma = 0.25$; $\beta = 0.35$, $\gamma = 0.15$; and $\beta = 0.15$, $\gamma = 0.35$. In all three cases, Figure 1 presents typical Laffer curves relating the growth rate to federal income tax. In the case of equal marginal productivity of federal government expenditure and local government expenditure, given local tax, federal transfer, and all other parameters in our model, a rise in federal income tax will increase the growth rate before the tax rate hits around 13 percent. In fact, when the federal income tax rate rises from zero to 10 percent, the growth rate rises from zero percent to almost 4.4 percent. Further increases in the federal income tax rate above 13 percent will reduce the growth rate. Just before the federal income tax rate reaches a high of 60 percent (note that the local income tax rate is assumed to be 10 percent), the growth rate is around zero.

The explanation for this Laffer curve is as follows. A change in federal income tax has three effects. First, a higher federal income tax directly reduces the return on private capital and the growth rate directly. Second, a larger tax revenue implies higher federal

expenditure, which is assumed to increase both private utility and private productivity. The rising productivity of private capital raises the growth rate. Third, at the same time, a larger tax revenue can lead to a larger federal transfer to local government, whose public services are also utility- and productivity-enhancing. When the federal income tax rate is initially very small, the second and third forces dominate. When the federal income tax is already high, the first force dominates.

For the effects of federal government expenditure and local government expenditure, we find that as the marginal productivity of local government expenditure increases, the growth rate decreases before the critical point of federal government income tax rate $\tau_f = 0.30$. The critical point of that rate, which reaches the maximum growth rate, decreases. In fact, from equation (22), we have

$$\frac{\partial \phi}{\partial \gamma} = \frac{1}{\sigma} \frac{1-\theta}{\theta} \frac{(1-\tau_f-\tau_s)\alpha^{\frac{1}{\theta}}((\tau_f - \frac{g\tau_s}{1-g})^{\theta} - (\frac{\tau_s}{1-g})^{\theta})}{(1-\beta(\tau_f - \frac{g\tau_s}{1-g})^{\theta} - \gamma(\frac{\tau_s}{1-g})^{\theta})^{\frac{1}{\theta}}}. \quad (33)$$

Thus, when $0 < \theta < 1$, we have

$$\text{sgn}(\frac{\partial \phi}{\partial \gamma}) = \text{sgn}((\tau_f - \frac{g\tau_s}{1-g})^{\theta} - (\frac{\tau_s}{1-g})^{\theta}).$$

In this special case, $\frac{(1+g)\tau_s}{1-g} = 0.3$. Hence, when $\tau_f > 0.3$, we have $\frac{\partial \phi}{\partial \gamma} > 0$; and when $\tau_f < 0.3$, we have $\frac{\partial \phi}{\partial \gamma} < 0$. This is shown in Figure 1.

(Insert Figure 2 about here.)

Figure 2 shows a similar picture of the relationship between the growth rate, ϕ , and local income tax rate, τ_s using the following base values for the structure of federal income tax, local taxes other than local income tax, and federal transfer: a federal income tax at 20 percent: $\tau_f = 0.20$, and a federal matching grant at 50 percent: $g = 0.5$. We also assume the following values for preference and technology parameters: $\alpha = 0.5$, $\rho = 0.05$, $\delta = 0.05$, $\sigma = 2$, and $\theta = 0.4$; and consider three cases: $\beta = \gamma = 0.25$; $\beta = 0.35$, $\gamma = 0.15$; and $\beta = 0.15$, $\gamma = 0.35$. We find Laffer curves similar to those in Figure 1.

For the equal marginal productivity of federal and local government expenditure, as the base federal income tax is already at a relatively high rate of 20 percent, the growth

rate is rising with local income tax until τ_s reaches about 5 percent. When the local income tax rate is set at 20 percent, the growth rate is zero. Because the local government receives a matching grant from the federal government at a rate of 30 percent, and because it also raises tax revenues from consumption tax and property tax, the local government can still finance its productive public expenditures without resorting to income tax. This is why the growth rate in Figure 2 is still above 3 percent even though local income tax is zero.

We present similar effects of federal government expenditure and local government expenditure, in this case $\frac{(1-g)\tau_f}{1+g} = 0.0667$. We find that as the marginal productivity of local government expenditure increases, the growth rate decreases before the critical point of the government income tax rate $\tau_s = 0.0667$. The critical point of local government income tax rate, which reaches the maximal growth rate, decreases.

(Insert Figure 3 about here.)

Figure 3 relates the growth rate, ϕ , to the federal matching grant for locality, g , based on a federal income tax of 20 percent and a local income tax at 10 percent. Again, we assume the following values for preference and technology parameters: $\alpha = 0.5$, $\rho = 0.05$, $\delta = 0.05$, $\sigma = 2$, and $\theta = 0.4$; and consider three cases: $\beta = \gamma = 0.25$; $\beta = 0.35$, $\gamma = 0.15$; and $\beta = 0.15$, $\gamma = 0.35$.

We obtain three different effects on growth of a federal matching grant for the local government: when the marginal productivity of federal government spending is larger than the marginal productivity local government spending, i.e. $\beta < \gamma$, the federal government matching transfer will decrease the growth rate. When the two government have the same marginal productivity, there is a non-evidence effect of the matching transfer on growth before $g = 0.5$. When local government expenditure has relatively larger marginal productivity, we find the contrasting solution whereby as the federal matching transfer increases, the growth rate increases before $g = 0.6$.

We find that the effects of a federal matching transfer on growth can be negative when it is too large (say, $g = 0.6$). We have selected the tax base for the local government income tax rate as $\tau_s = 0.1$, and thus the local government has already

obtained an amount of revenue from its income tax. If the matching transfer is too high, then the federal government should pay for the major part of local government expenditure, and the local government's income tax will be in surplus. This will harm economic growth.

Similarly, we can show the effects of the marginal productivity of local government expenditure under the selected parameters, with the critical point $\frac{\tau_f - \tau_s}{\tau_f + \tau_s} = g^* = \frac{1}{3}$. Thus, when $g < g^*$, the effect of local government expenditure on growth is negative; when, $g > g^*$, the effect is positive.

5. Optimal Taxes and Transfers

5.1 Growth Maximization and Welfare Maximization

In the last section, we numerically presented the relationships among τ_f , τ_s , g , and the growth. Recall that Barro (1990) shows that maximizing social welfare is equivalent to maximizing the rate of growth, and the optimal tax rate equals the marginal contribution of government expenditure. To compare our solutions with that of Barro (1990), we reexamine the conclusions we have drawn by using the special production function. In this section, we specify the production function as the Cobb-Douglas production function, which amounts to set $\theta = 0$ in the CES production function, namely,

$$y = k^\alpha f^\beta s^\gamma, \quad (34)$$

where α , β , and γ are positive constants with $\alpha + \beta + \gamma = 1$.

Hence, the explicit balanced growth rate expressed in equation (22) has the following form,

$$\phi = \frac{1}{\sigma} \left[(1 - \tau_f - \tau_s) \alpha \left(\tau_f - \frac{g\tau_s}{1-g} \right)^{\frac{\beta}{\alpha}} \left(\frac{\tau_s}{1-g} \right)^{\frac{\gamma}{\alpha}} - \rho - \delta \right]. \quad (35)$$

On the other hand, if we substitute the growth paths (25) for consumption, federal spending, and local government spending into the utility function in (2), the agent's welfare is given as

$$\begin{aligned}
U &= \int_0^\infty \left[\frac{c(0)^{1-\sigma} e^{\phi(1-\sigma)t} - 1}{1-\sigma} + \frac{f(0)^{1-\sigma} e^{\phi(1-\sigma)t} - 1}{1-\sigma} + \frac{s(0)^{1-\sigma} e^{\phi(1-\sigma)t} - 1}{1-\sigma} \right] e^{-\rho t} dt \\
&= \frac{c(0)^{1-\sigma} + f(0)^{1-\sigma} + s(0)^{1-\sigma}}{(\rho - (1-\sigma)\phi)(1-\sigma)} + \frac{3}{\rho(1-\sigma)},
\end{aligned} \tag{36}$$

where we have used the TVC (11) to obtain $\rho - (1-\sigma)\phi > 0$.

Here, $f(0)$ and $s(0)$ are still determined by equations (26) and (27). $y(0)$ is determined by substituting equations (26) and (27) into equation (34), i.e,

$$y(0) = \left(\tau_f - \frac{g\tau_s}{1-g} \right)^{\frac{\beta}{\alpha}} \left(\frac{\tau_s}{1-g} \right)^{\frac{\gamma}{\alpha}} k(0). \tag{37}$$

Then, $c(0)$ can be determined by equation (29).

(Insert Figure 4 about here!)

It is simple to see that in equation (36) the agent's welfare is an increasing function of the economic growth rate⁶, ϕ . However⁷, because $c(0)$, $f(0)$, and $s(0)$ also depend on τ_f , τ_s , and g , welfare maximization may not be equivalent to growth maximization. This is shown in Figure 4, where we select the parameters as: $\alpha = 0.5$, $\beta = 0.35$, $\gamma = 0.15$, $\rho = 0.05$, $\delta = 0.05$, $\sigma = 1.1$, $g = 0.5$, and $\tau_s = 0.05$. We find that the optimal federal income tax rate, which maximizes growth is around 0.4. However, the optimal federal income tax rate, which reaches the maximum welfare, is around 0.5. Therefore, growth- and the welfare-maximizing taxes are not equivalent in this case.

5.2 The Growth-maximizing Taxes

We now focus on finding the optimal taxes for growth maximization. In fact, differentiating equation (35) with respect to τ_f , τ_s , and g yields

$$\frac{\partial \phi}{\partial \tau_f} = \frac{1}{\sigma} \left(\tau_f - \frac{g\tau_s}{1-g} \right)^{\frac{\beta}{\alpha}-1} \left(\frac{\tau_s}{1-g} \right)^{\frac{\gamma}{\alpha}} \left[-\alpha \left(\tau_f - \frac{g\tau_s}{1-g} \right) + (1-\tau_f-\tau_s)\beta \right] = 0, \tag{38}$$

⁶In fact, differentiating on equation (36) with respect to ϕ yields

$$\frac{\partial U}{\partial \phi} = \frac{\left(\frac{\sigma-\alpha}{\alpha} \phi + \frac{1-\alpha}{\alpha} (\rho + \delta) \right)^{-\sigma} \left[\frac{\sigma-\alpha}{\alpha} (\rho + \sigma\phi) + \frac{1-\alpha}{\alpha} (\rho + \delta) \right]}{(\rho - (1-\sigma)\phi)^2}.$$

Therefore, we have $\frac{\partial U}{\partial \phi} > 0$ when $\sigma > \alpha$.

⁷We thank the associate editor very much to point out this.

$$\begin{aligned} \frac{\partial \phi}{\partial \tau_s} &= \frac{1}{\sigma} (\tau_f - \frac{g\tau_s}{1-g})^{\frac{\beta}{\alpha}-1} (\frac{\tau_s}{1-g})^{\frac{\gamma}{\alpha}-1} \{-\alpha(\tau_f - \frac{g\tau_s}{1-g})(\frac{\tau_s}{1-g}) \\ &\quad + (1-\tau_f - \tau_s)[\beta(\frac{\tau_s}{1-g})(-\frac{g}{1-g}) + \gamma(\tau_f - \frac{g\tau_s}{1-g})\frac{1}{1-g}]\} \\ &= 0, \end{aligned} \quad (39)$$

and

$$\frac{\partial \phi}{\partial g} = \frac{1}{\sigma} (1-\tau_f - \tau_s)\beta(\tau_f - \frac{g\tau_s}{1-g})^{\frac{\beta}{\alpha}-1} (\frac{\tau_s}{1-g})^{\frac{\gamma}{\alpha}} \frac{1}{1-g} [\beta(-\frac{\tau_s}{1-g}) + (\tau_f - \frac{g\tau_s}{1-g})\gamma] = 0. \quad (40)$$

Thus, we have

$$\alpha(\tau_f - \frac{g\tau_s}{1-g}) = (1-\tau_f - \tau_s)\beta, \quad (38')$$

$$\alpha\tau_s(\tau_f - \frac{g\tau_s}{1-g}) = (1-\tau_f - \tau_s)[\beta\tau_s(-\frac{g}{1-g}) + \gamma(\tau_f - \frac{g\tau_s}{1-g})], \quad (39')$$

and

$$\beta\frac{\tau_s}{1-g} = (\tau_f - \frac{g\tau_s}{1-g})\gamma. \quad (40')$$

Equation (40') yields the same expression as equations (38') and (39'). Hence, we know that optimal choices of τ_f , τ_s , and g are interdependent. The choice of the federal matching grant is endogenous in the following sense: once g is chosen from the interval $(0,1)$, federal income tax and local income tax are determined by (38') and (39'). From equations (38') and (39'), we obtain the optimal tax rates as⁸

$$\tau_f = \beta + g\gamma \quad (41)$$

and

$$\tau_s = \gamma - g\gamma, \quad (42)$$

for the federal government and local government, respectively.

Once g is given in the interval $(0,1)$, the federal and local income taxes are determined by their productiveness and the matching rate multiplied by the productivity of local public spending. The aggregate optimal tax rate is just the sum of the

⁸The Jacobian matrix at the optimal taxes can be derived as

$$\begin{pmatrix} \frac{1}{\sigma} \beta^{\frac{\beta}{\alpha}-1} \gamma^{\frac{\gamma}{\alpha}} (-\alpha - \beta) & \frac{1}{\sigma} \beta^{\frac{\beta}{\alpha}-1} \gamma^{\frac{\gamma}{\alpha}} (\frac{\alpha g}{1-g} - \beta) \\ \frac{1}{\sigma} \beta^{\frac{\beta}{\alpha}-1} \gamma^{\frac{\gamma}{\alpha}} (\frac{\alpha g}{1-g} - \beta) & \frac{1}{\sigma} \beta^{\frac{\beta}{\alpha}-1} \gamma^{\frac{\gamma}{\alpha}} (\frac{-\alpha\beta - \alpha\gamma g^2}{(1-g)^2} - \beta\gamma) \end{pmatrix}.$$

It is easy to prove that this matrix is negative definite. Therefore, the second-order conditions are satisfied.

productiveness of federal and local expenditures:

$$\tau_s + \tau_f = \beta + \gamma . \quad (43)$$

With the choices of tax rates and transfer specified in equations (41) and (42), the growth-maximizing growth rate is

$$\phi = \frac{1}{\sigma} [\alpha^2 \beta^{\frac{\beta}{\alpha}} \gamma^{\frac{\gamma}{\alpha}} - \rho - \delta] .$$

We should say something about the effects of optimal government matching transfer and effects of the matching transfer on the growth. In the last section, we showed the government matching transfer can affect growth but the optimal choices of government matching transfer and the tax rates are interdependent. This occurs because, from equations (6) and (7), we have

$$f + s = (\tau_s + \tau_f)y .$$

The government transfer becomes an independent variable. We can also derive the equation with only the federal government income tax rate and government matching transfer; thus, the local government income tax rate becomes an independent variable.

The effects of government matching transfer in the last section are based on the selected federal and local income tax rates. Thus, we obtain the effects shown in Figure 3. In addition, given the local income tax rate, from equations (38') and (40') we can determine the optimal choices for the government matching transfer and federal income tax rate.

6. A More General Framework

We can extend our analytical framework to a more general one and consider more tax rates. The same set up is used for the federal government but we introduce two more taxes for the typical local government. It now levies three taxes: a local income tax (such as the case of state income tax in the United States) at the rate of τ_s , a consumption tax τ_c and a property tax (capital tax in our model) τ_k . The federal government also makes a transfer to the local government in the form of a matching grant for local public spending at the rate of g . If both levels of government maintain balanced budgets, then budget constraints (6) and (7) can be written as

$$f = \tau_f y - gs \quad (44)$$

and

$$s = \tau_s y + \tau_k k + \tau_c c + g s, \quad (45)$$

respectively.

Hence, the federal government public spending, f , is equal to its total income tax, $\tau_f y$, minus its transfer to the local government, $g s$. The local government's spending is financed by its income tax, $\tau_s y$, its property tax, $\tau_k k$, its consumption tax, $\tau_c c$, and the grant it receives from the federal government, $g s$.

In a similar way, we derive a highly nonlinear equation for the growth rate ϕ :

$$\frac{\sigma\phi + \rho + \delta + \tau_k}{\alpha(1 - \tau_f - \tau_s)} = (\alpha + \psi(\phi, \tau_f, \tau_s, \tau_k, \tau_c, g, \sigma, \rho, \alpha, \beta, \gamma, \delta))^{\frac{1}{\theta}-1}, \quad (46)$$

where $\psi(\phi, \tau_f, \tau_s, \tau_k, \tau_c, g, \sigma, \rho, \alpha, \beta, \gamma, \delta)$ is given by

$$\begin{aligned} & \psi(\phi, \tau_f, \tau_s, \tau_k, \tau_c, g, \sigma, \rho, \alpha, \beta, \gamma, \delta, \theta) \\ &= \beta \left\{ \left[\left(\tau_f - \frac{g}{1-g} \tau_s \right) - \frac{g\tau_c}{1-g} \frac{1 - \tau_f - \tau_s}{1 + \tau_c} \right] \left(\frac{\sigma\phi + \rho + \delta + \tau_k}{\alpha(1 - \tau_f - \tau_s)} \right)^{\frac{1}{1-\theta}} \right. \\ & \quad \left. - \frac{g\tau_k}{1-g} + \frac{g\tau_c}{1-g} \frac{\rho + \delta + \tau_k + \phi}{1 + \tau_c} \right\}^\theta + \gamma \left\{ \frac{\tau_s}{1-g} \left(\frac{\sigma\phi + \rho + \delta + \tau_k}{\alpha(1 - \tau_f - \tau_s)} \right)^{\frac{1}{1-\theta}} \right. \\ & \quad \left. + \frac{\tau_k}{1-g} + \frac{\tau_c}{1-g} \frac{1 - \tau_f - \tau_s}{1 + \tau_c} \left(\frac{\sigma\phi + \rho + \delta + \tau_k}{\alpha(1 - \tau_f - \tau_s)} \right)^{\frac{1}{1-\theta}} + \frac{\tau_c}{1-g} \frac{\rho + \delta + \tau_k + \phi}{1 + \tau_c} \right\}^\theta. \end{aligned}$$

Note that ϕ appears on both sides of equation (46). Therefore, the growth rate is implicitly defined as a function of τ_f , τ_s , τ_k , τ_c , g , σ , ρ , α , β , γ , and δ .

For the endogenous growth rate to be positive, we must impose $\phi > 0$, and from the TVC

$$\rho - (1 - \sigma)\phi > 0,$$

which is also the condition for a bounded discounted utility over the infinite horizon.

Given such an extended framework with government expenditures and taxes by two levels of government and intergovernmental transfer, we cannot hope that growth-maximizing choices of tax rates and the transfer rate will be consistent with the welfare-maximizing ones. The simple case in Barro's (1990) analysis whereby growth maximization coincides with welfare maximization disappears here. In fact, when both a local consumption tax and a local property tax are present, welfare is a complicated function of the growth rate, which in turn is a complicated function of various taxes and

the federal transfer, as shown in equation (46).

7. Conclusion

This paper has extended the Barro (1990) model with single aggregate government spending and one flat income tax to include public expenditures and taxes by multiple levels of government. We have derived the rate of endogenous growth under quite general specifications of preferences and production technology. With simulations, we have examined how the rate of endogenous growth changes with respect to federal income tax, local income tax, and federal transfer. We have also discussed growth-maximizing choices of income taxes and federal transfer. In addition, we extend our model to a more general framework including a local consumption tax and local property tax. A preliminary simulation analysis has shown that the local property tax has the largest negative impact on the rate of economic growth, whereas a local consumption tax is always growth enhancing. This finding contrasts with that of Rebelo (1991), who shows that a consumption tax has no effect on the growth rate.

The model in this paper sets up a positive framework for evaluating how the assignment of taxes and expenditures among different levels of government and intergovernmental transfers affect economic growth. Our analysis also sheds light on the role of intergovernmental transfers in regional economic growth. If a local government has sufficient revenue base, federal transfers seem to reduce the growth rate. Even if local revenue is not sufficient, the rise in the rate of federal transfer increases the growth rate to a very modest degree. Of course, the model is also useful for normative discussions of the welfare- and growth-maximizing choices of taxes, transfers, and expenditures in the context of fiscal federalism.

In future, we will add two more dimensions: one will be to follow Arrow and Kurz (1970) and introduce public consumption and public capital accumulation at both the federal and local levels into the endogenous growth model; the other will be to formulate a game-theoretical growth model and allow strategic interactions between the federal government and multiple local governments in the choices of taxes, public expenditures, and intergovernmental transfers.

References

Arrow, K., and M. Kurz (1970) *Public Investment, the Rate of Return and Optimal Fiscal Policy*, Johns Hopkins University Press.

Aschauer, D. (1989) "Is government spending productive?" *Journal of Monetary Economics* **23**, 177-200.

Barro, R.J. (1990) "Government spending in a simple model of endogenous growth" *Journal of Political Economy* **98**, S103-S125.

Bird, R. (1993) "Threading the fiscal labyrinth: Some issues in fiscal decentralization" *National Tax Journal* **46**, 207-227.

Brueckner, J. (1996) "Fiscal federalism and capital accumulation" *Mimeo*. University of Illinois at Urbana-Champaign.

Davoodi, H. and H. Zou (1998) "Fiscal decentralization and economic growth: A cross-country study" *Journal of Urban Economics* **43**, 244-257.

Devarajan, S, V. Swaroop, and H. Zou (1996) "The composition of government expenditure and economic growth" *Journal of Monetary Economics* **37**, 313-344.

Devarajan, S, D. Xie, and H. Zou (1998) "Should public capital be subsidized or provided?" *Journal of Monetary Economics* **41**, 319-331.

Easterly, W. and S. Rebelo (1993) "Fiscal policy and economic growth: An empirical investigation" *Journal of Monetary Economics* **32**, 417-458.

Glomm, G. and B. Ravikumar (1994) "Public investment in infrastructure in a simple growth model" *Journal of Economic Dynamics and Control* **18**, 1173-1187.

Gramlich, E. (1993) "A policy maker's guide to fiscal decentralization" *National Tax Journal* **46** 229-235.

Hulten, C. (1994) "Optimal growth with infrastructure capital: Theory and implications for empirical modeling" *Mimeo*. University of Maryland.

King, R.G. and S. Rebelo (1990) "Public policy and economic growth: Developing neoclassical implications" *Journal of Political Economy* **98**, S126-S150.

Oates, W. (1972) *Fiscal Federalism*, New York, Harcourt Brace Jovanovich.

Oates, W. (1993) "Fiscal decentralization and economic development" *National Tax Journal* **46**, 237-243.

Rebelo, S. (1991) "Long-run policy analysis and long-run growth" *Journal of Political Economy* **99**, 500-521.

Rivlin, A. (1992) *Reviving the American Dream: The Economy, the States, and the Federal Government*, Brookings Institution, Washington, D.C.

Turnovsky, S. (2000) *Methods of Macroeconomic Dynamics*, MIT Press.

Zhang, T. and H. Zou (1998) "Fiscal decentralization, public spending, and economic growth in China" *Journal of Public Economics* **67**, 221-240.

Zou, H. (1995) "Dynamic effects of federal grants on local spending" *Journal of Urban Economics* **36**, 98-115.

Zou, H. (1996) "Taxes, federal grants, local public spending, and growth" *Journal of Urban Economics* **39**, 303-317.

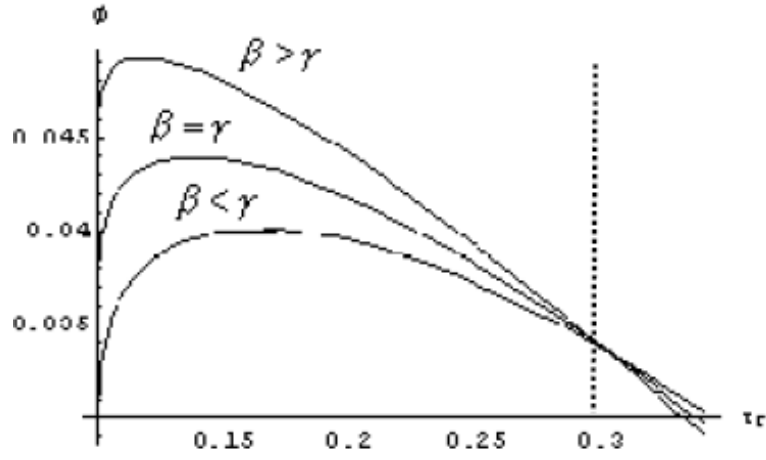


Figure 1: Growth rate versus federal government income tax rate. The parameters are: $\alpha=0.5$, $\theta=0.4$, $\rho=0.05$, $\delta=0.05$, $\sigma=2$, $g=0.5$, and $\tau_s=0.1$; in the case of $\beta > \gamma$: $\beta=0.35$, $\gamma=0.15$; $\beta=\gamma$: $\beta=0.25$, $\gamma=0.25$; $\beta < \gamma$: $\beta=0.15$, $\gamma=0.35$.

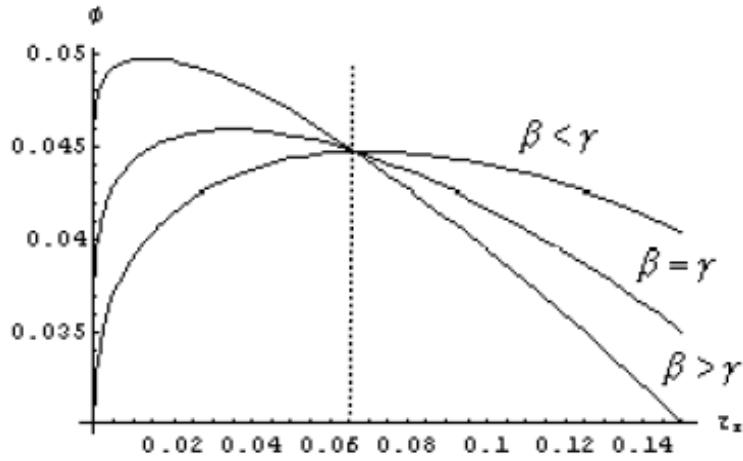


Figure 2: Growth rate versus local government income tax rate. The parameters are: $\alpha=0.5$, $\theta=0.4$, $\rho=0.05$, $\delta=0.05$, $\sigma=2$, $g=0.5$, and $\tau_f=0.2$; in the case of $\beta > \gamma$: $\beta=0.35$, $\gamma=0.15$; $\beta=\gamma$: $\beta=0.25$, $\gamma=0.25$; $\beta < \gamma$: $\beta=0.15$, $\gamma=0.35$.

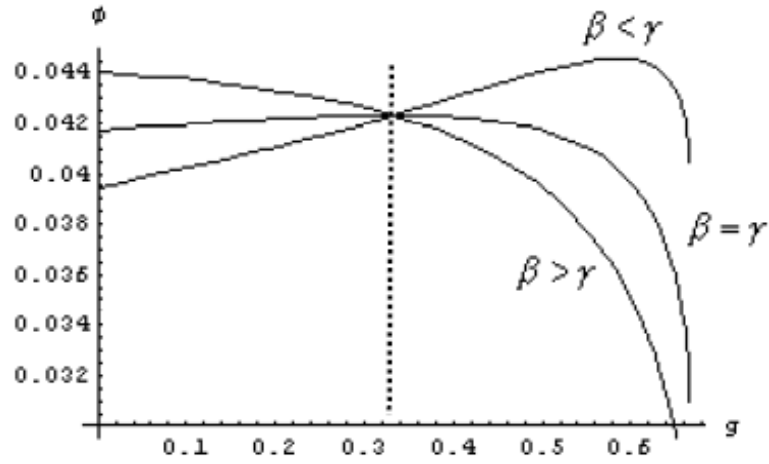


Figure 3: Growth rate versus federal government matching grant for locality g . The parameters are: $\alpha=0.5$, $\theta=0.4$, $\rho=0.05$, $\delta=0.05$, $\sigma=2$, $\tau_f=0.2$, and $\tau_s=0.1$; in the case of $\beta > \gamma$: $\beta=0.35$, $\gamma=0.15$; $\beta=\gamma$: $\beta=0.25$, $\gamma=0.25$; $\beta < \gamma$: $\beta=0.15$, $\gamma=0.35$

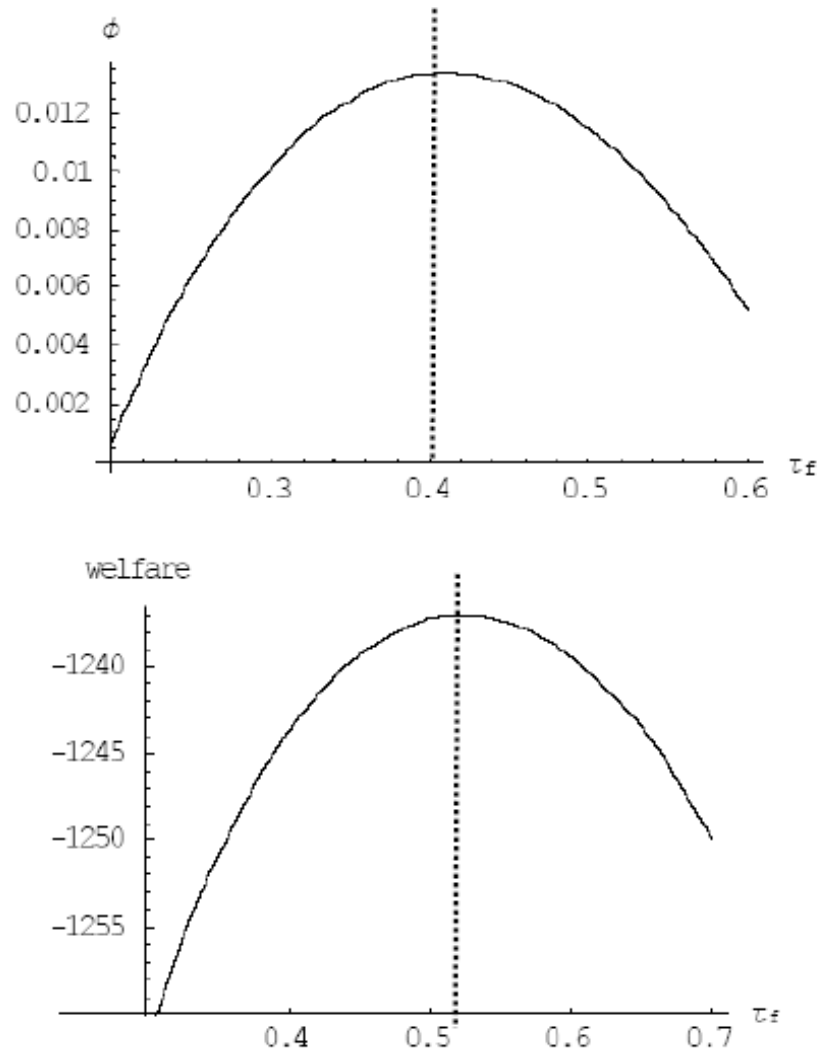


Figure 4: Growth maximization and welfare maximization. The parameters are: $\alpha = 0.5$, $\beta = 0.35$, $\gamma = 0.15$, $\rho = 0.05$, $\delta = 0.05$, $\sigma = 1.1$, $g = 0.5$, and $\tau_s = 0.05$.

Social Status, the Spirit of Capitalism, and the Term Structure of Interest Rates in Stochastic Production Economies¹

Liutang Gong

Guanghua School of Management, Peking University, Beijing, 100871, China

Institute for Advanced Study, Wuhan University, Wuhan, 430072, China

Ltgong@gsm.pku.edu.cn

Yulei Luo

School of Economics and Finance, University of Hong Kong, Hong Kong

Yluo@econ.hku.hk

Heng-fu Zou²

Institute for Advanced Study, Wuhan University, Wuhan, 430072, China

Development Research Group, The World Bank, Washington, DC 20433, USA

Hzhou@worldbank.org

Abstract

This paper studies capital accumulation and equilibrium interest rates in stochastic production economies with the concern of social status. Given a specific utility function and production function, explicit solutions for capital accumulation and equilibrium interest rates have been derived. With the aid of steady-state distributions for capital stock, the effects of fiscal policies, social-status concern, and stochastic shocks on capital accumulation and equilibrium interest rates have been investigated. A significant finding of this paper is the demonstration of multiple stationary distributions for capital stocks and interest rates with the concern of social status.

Key Words: Stochastic growth; Social status; Fiscal policies; Interest rates.

JEL Classification: E0, G1, H0, O0.

¹Project 70271063 supported by the National Natural Science Foundation of China.

²Mailing address: Heng-fu Zou, Development Research Group, the World Bank, 1818 H St. NW, Washington, DC 20433, USA. Email: Hzhou@worldbank.org

Social Status, the Spirit of Capitalism, and the Term Structure of Interest Rates in Stochastic Production Economies

Abstract

This paper studies capital accumulation and equilibrium interest rates in stochastic production economies with the concern of social status. Given a specific utility function and production function, explicit solutions for capital accumulation and equilibrium interest rates have been derived. With the aid of steady-state distributions for capital stock, the effects of fiscal policies, social-status concern, and stochastic shocks on capital accumulation and equilibrium interest rates have been investigated. A significant finding of this paper is the demonstration of multiple stationary distributions for capital stocks and interest rates with the concern of social status.

Key Words: Stochastic growth; Social status; Fiscal policies; Interest rates.

JEL Classification: E0, G1, H0, O0.

1. Introduction

Capital accumulation, interest rates, fiscal policies, and asset pricing under uncertainty have been studied extensively since the 1960s, e.g., Phelps (1962), Levhari and Srinivisan (1969), Brock and Mirman (1972), Mirrlees (1965). Merton (1975) studied the asymptotic theory of growth under uncertainty, and Foldes (1978) explored optimal saving with risk in continuous time. As for the term structure of interest rates, Cox, Ingersoll, and Ross (1981, 1985) considered the equilibrium theory of the term structure of interest rates, and presented the general theory for interest rates in a production economy. Sunderason (1983) provided a plausible equilibrium model, in which the assumption of a constant interest rate is valid. Bhattacharya (1981), Constantinides (1980), and Stapleton and Subrahmanyam (1978) also studied these topics and presented the conditions for a constant interest rate. Constantinides (1980) showed that the term structure of interest rate evolves deterministically over time under the assumptions of perfect capital markets, homogeneous expectations, and the state independent utility. Sunderason (1984) also derived the conclusion of a constant interest rate under Constantinides (1980)'s assumptions on capital markets, expectations, and utility. For the effects of fiscal policies on capital accumulation, interest rates, and asset pricing in stochastic economies, Eaton (1981), Turnovsky (1993, 1995), Grinols and Turnovsky (1993, 1994), and Obstfeld (1994) introduced taxations and government expenditure into the stochastic continuous-time growth and asset-pricing models. Under a linear production technology and other specified assumptions on preferences and stochastic shocks, they have derived explicit solutions of growth rates of consumption, savings, and equilibrium returns on assets.

In all these neoclassical models of capital accumulation, interest rates and asset pricing models, wealth accumulation is often taken to be solely driven by one's desire to increase consumption rewards. The representative agent chooses his consumption path to maximize his discounted utility, which is defined only on consumption. This motive is important for wealth accumulation. It is, however, not the only motive. Because man is a social animal, he also accumulates wealth to gain prestige, social status, and power in the society; see Frank (1985), Cole, Mailath and Postlewaite (1992, 1995), Fershtman and

Weiss (1993), Zou (1994, 1995), Bakshi and Chen (1996), and Fershtman, Murphy and Weiss (1996). In these wealth-is-status models, the representative agent accumulates wealth not only for consumption but also for wealth-induced status. Mathematically, in light of the new perspective, the utility function can be defined on both consumption, c , and wealth, w : $u(c_t, w_t)$. Another interpretation of these models is in line of the spirit of capitalism in the sense of Weber (1958): capitalists accumulate wealth for the sake of wealth³.

With the wealth-is-status and the-spirit-of-capitalism models, many authors mentioned above have tried to explore diverse implications for growth, savings, interest rates, and asset pricing. Cole, Mailath, and Postlewaite (1992) have demonstrated how the presence of social-status concern leads to multiple equilibria in long-run growth. Zou (1994, 1995) has studied the spirit of capitalism and long-run growth and showed that a strong capitalistic spirit can lead to unbounded growth of consumption and capital even under the neoclassical assumption of production technology. Gong and Zou (2002) have studied fiscal policies, asset pricing, and capital accumulation in a stochastic model with the spirit of capitalism. Bakshi and Chen (1996) have explored empirically the relationship between the spirit of capitalism and stock market pricing and offered an attempt towards the resolution of the equity premium puzzle in Mehra and Prescott (1985). Smith (2001) has studied the effects of the spirit of capitalism on asset pricing and has shown that when investors care about status they will be more conservative in risk taking and more frugal in consumption spending. Furthermore, stock prices tend to be more volatile with the presence of the spirit of capitalism.

This paper explores capital accumulation and equilibrium interest rates in a stochastic model with the spirit of capitalism and with diminishing return to scale technology. Under a CES utility function defined on both consumption and wealth accumulation and a Cobb-Douglas production function, explicit solutions for capital accumulation and equilibrium interest rates have been derived. These multiple optimal paths and stationary distributions of capital stock and interest rates are quite significantly different from many existing neoclassical models. With the aid of the steady-state

³See Cole, Mailath, and Postlewaite (1992), Zou (1994); and Bakshi and Chen (1996) for details.

distributions for capital stock, the effects of fiscal policies on the long-run economy and the equilibrium interest rates have been investigated. In particular, the equilibrium interest rates are constant when the technology is linear and when the utility function is extended to include the wealth-is-status concern. Moreover, the equilibrium interest rates are a mean reserve process with these special assumptions.

This paper is organized as follows: in section 2, we set up a stochastic growth model in a production economy with the social-status concern. Allowing some special utility function and production function with selected parameters, explicit solutions for the optimal paths and stationary distributions of consumption, capital accumulation and interest rates have been derived in section 3. With the aid of the steady-state distribution of the endogenous variables, the effects of fiscal policies, production shocks, and the spirit of capitalism on the long-run economy have been examined in section 4. In section 5, we present the equilibrium interest rates under both a nonlinear technology and a linear technology and analyze the dynamic behavior of equilibrium interest rates and discuss the effects of fiscal policies and stochastic shocks on the interest rates. In section 6, we present some examples to show the existence of multiple stationary distributions of optimal capital accumulation and equilibrium interest rates. We conclude our paper in section 7.

2. The model

Following Eaton (1981) and Smith (2001), we assume that output y is given by

$$dy = f(k)dt + \varepsilon kdz, \quad (1)$$

where z is the standard Brown motion, ε is the stochastic shocks of production.

Equation (1) asserts that the accumulated flow of output over the period $(t, t + dt)$, given by the right-hand side of this equation, consists of two components. The deterministic component is described as the first term on the right-hand side, which is the firm's production technology and has been specified as a neoclassical production function, $f(k)$. The second part is the stochastic component, εkdz , which can be viewed as the shocks to the production and assumed to be temporally independent, normally distributed.

Suppose the government levy an income tax and a consumption tax. Then, the

agent's budget constraint can be written as⁴

$$dk = ((1 - \tau)f(k) - (1 + \tau_c)c)dt + (1 - \tau')\varepsilon k dz \quad (2)$$

where τ and τ' are the tax rates on the deterministic component of capital income and stochastic capital income, respectively, and τ_c is the consumption tax rate.

With the social status concern, the utility function can be written as $u(c, k)$. Suppose the marginal utilities of consumption and capital stock are positive, but diminishing, i.e.

$$u_1(c, k) > 0, u_2(c, k) > 0, u_{11}(c, k) < 0, u_{22}(c, k) < 0 \quad (3)$$

The representative agent is to choose his consumption level and capital accumulation path to maximize his expected discounted utility, namely,

$$\max E_0 \int_0^\infty u(c, k) e^{-\rho t} dt$$

subject to a given initial capital stock $k(0)$ and the budget constraint (2). Where $0 < \rho < 1$ is the discount rate.

Associated with the above optimization problem, the value function $J(k, t)$ is defined as

$$J(k, t) = \max E_t \int_t^\infty u(c, k) e^{-\rho t} dt$$

subject to the given initial capital stock $k(t)$ and the budget constraint (2). Define the

⁴Merton (1975) assumed that output is produced by a strictly concave production function, $Y = AF(K, L)$, where $K(t)$ denotes capital stock, $L(t)$ the labor force, and $A(t)$ is technology progress. Production is

$$Y = AF(K, L)$$

and the labor force follows

$$dL = aLdt + \varepsilon Ldz,$$

where z is the standard Brown motion.

Defining the capital-labor ratio $k = K / L$, from the Itô's Lemma, we can derive the capital accumulation equation similar to equation (2). Or we assume the technology progress follows

$$dA / A = adt + \varepsilon dz$$

Defining the efficiency capital $\tilde{k} = K / (AL)$, we can also derive the capital accumulation similar to equation (2).

current value function $X(k)$ as

$$X(k) = J(k, t)e^{\rho t} \quad (4)$$

The recursive equation associated with the above optimization problem is

$$\max_c \{u(c, k) - \rho X(k) + X'(k)((1-\tau)f(k) - (1+\tau_c)c) + \frac{1}{2} X''(k)(1-\tau')^2 \varepsilon^2 k^2\} = 0$$

Therefore, we get the first-order condition

$$u_c(c, k) = (1+\tau_c)X'(k) \quad (5)$$

and the Bellman equation

$$u(c, k) - \rho X(k) + X'(k)((1-\tau)f(k) - (1+\tau_c)c) + (1-\tau')^2 \frac{1}{2} X''(k) \varepsilon^2 k^2 = 0 \quad (6)$$

Equation (5) states that the marginal utility of consumption equals the after-tax marginal utility of capital stock. Equation (6) determines the value function $X(k)$. In the next section, we will specify the utility function and the production function to present an explicit solution for the value function.

3. An explicit solution

3.1 The explicit solution under the separable utility function

In order to derive an explicit solution, we specify the utility function as⁵

$$u(c, k) = \frac{c^{1-\sigma}}{1-\sigma} + \xi \frac{k^{1-\sigma}}{1-\sigma}, \quad (7)$$

where $\sigma > 0$ is the constant relative risk aversion, and it also represents the elasticity of intertemporal substitution. $\xi \geq 0$ measures the investor's concern with his social status or measures his spirit of capitalism. The larger the parameter ξ , the stronger the agent's spirit of capitalism or concern for social status.

The production function is specified as

$$f(k) = Ak^\alpha \quad (8)$$

where $A > 0$ and $0 < \alpha < 1$ are positive constants.

For the special utility function and production function in equations (7) and (8), we

⁵In Appendix B, we present a similar analysis when the utility function is non-separable.

conjecture that the value function takes the following form

$$X(k) = a + \frac{b^{-\sigma} k^{1-\sigma}}{1-\sigma} \quad (9)$$

where a and b are constants, and they are to be determined.

Under the specified value function in equation (9), we rewrite the first-order condition (5) as

$$c^{-\sigma} = (1 + \tau_c) X_k = (1 + \tau_c) b^{-\sigma} k^{-\sigma}$$

namely,

$$c = (1 + \tau_c)^{-\frac{1}{\sigma}} b k \quad (10)$$

Upon the relationship (10), the Bellman equation (6) is reduced to

$$\begin{aligned} & \frac{(1 + \tau_c)^{1-\frac{1}{\sigma}} b^{1-\sigma} k^{1-\sigma}}{1-\sigma} + \xi \frac{k^{1-\sigma}}{1-\sigma} - \rho \left(a + \frac{b^{-\sigma} k^{1-\sigma}}{1-\sigma} \right) + b^{-\sigma} k^{-\sigma} ((1-\tau) A k^\alpha - (1 + \tau_c)^{1-\frac{1}{\sigma}} b k) \\ & + \frac{1}{2} X_{kk} \varepsilon^2 (1 - \tau')^2 k^2 = 0 \end{aligned} \quad (11)$$

If $\alpha = 1$, from equation (11), we have $a = 0$ and b is determined by the following equation

$$0 = \sigma (1 + \tau_c)^{1-\frac{1}{\sigma}} b + \xi b^\sigma - (\rho - (1-\tau) A (1-\sigma) + \frac{1}{2} (1-\sigma) \sigma (1-\tau')^2 \varepsilon^2)$$

In general, for the case of $\alpha \neq 1$, we cannot determine the constants a and b . Following Xie (1994), we specified the parameters as $a = \sigma$, then from equation (11), we have

$$a = \frac{(1-\tau) A}{\rho b^\sigma}$$

and b is determined by

$$0 = (1 + \tau_c)^{1-\frac{1}{\sigma}} \sigma b + \xi b^\sigma - \left(\rho + \frac{1}{2} (1-\sigma) \sigma (1-\tau')^2 \varepsilon^2 \right).$$

Summarizing the discussions above, we have

Proposition 1. *Under the special utility function and production function in equations (7) and (8), if $\alpha = 1$, then the explicit solutions for the economy system are*

$$\frac{c}{k} = (1 + \tau_c)^{-\frac{1}{\sigma}} b \quad (12)$$

$$dk = ((1-\tau)Ak - (1+\tau_c)^{1-\frac{1}{\sigma}}bk)dt + (1-\tau')\varepsilon kdz \quad (13)$$

and the TVC

$$\lim_{t \rightarrow \infty} E(X'(k)e^{-\rho t}) = 0 \quad (14)$$

where b is determined by

$$0 = (1+\tau_c)^{1-\frac{1}{\sigma}}\sigma b + \xi b^\sigma - (\rho - (1-\tau)A(1-\sigma) + \frac{1}{2}(1-\sigma)\sigma(1-\tau')^2\varepsilon^2). \quad (15)$$

If $\alpha \neq 1$ and $\alpha = \sigma$, then the explicit solutions for the economy are

$$\frac{c}{k} = (1+\tau_c)^{-\frac{1}{\sigma}}b, \quad (12')$$

$$dk = ((1-\tau)Ak^\alpha - (1+\tau_c)^{1-\frac{1}{\sigma}}bk)dt + (1-\tau')\varepsilon kdz \quad (13')$$

and the TVC (14) holds, whereas b is determined by

$$0 = (1+\tau_c)^{1-\frac{1}{\sigma}}\sigma b + \xi b^\sigma - (\rho + \frac{1}{2}(1-\sigma)\sigma(1-\tau')^2\varepsilon^2). \quad (15')$$

When $\alpha = 1$, the capital stock follows the stochastic growth path (13), and we get the mean growth rate for the capital stock

$$E\left(\frac{dk}{k}\right) = ((1-\tau)A - (1+\tau_c)^{1-\frac{1}{\sigma}}b)dt$$

It is easy to show from equation (12') and the production function that the mean growth rates for consumption level, output, and capital stock are equal. Let us denote the common mean growth rate as ϕ , which is given by

$$\phi = (1-\tau)A - (1+\tau_c)^{1-\frac{1}{\sigma}}b.$$

From the expression above, it is clear that capital income taxation, consumption taxation, stochastic shocks, and various preference and production parameters jointly determine the growth rate of the economy. Please also note that when the parameters satisfy the condition $\alpha = \sigma$, the deterministic income tax rate has no effects on the equilibrium consumption-capital stock ratio.

3.2 Steady-state distributions for endogenous variables

Similar to the certainty model, we will examine the existence and the properties of the steady state economy. As in Merton (1975), we are seeking the conditions under which there is a unique stationary distribution for the capital stock k , which is time and

initial condition independent.

From equation (13'), the capital stock follows the following stochastic process,

$$dk = ((1-\tau)Ak^\alpha - (1+\tau_c)^{1-\frac{1}{\sigma}}bk)dt + (1-\tau')\varepsilon k dz + b(k)dt + (a(k))^{1/2}dz, \quad (16)$$

where we denote $a(k) = (1-\tau')^2 \varepsilon^2 k^2$ and $b(k) = (1-\tau)Ak^\alpha - (1+\tau_c)^{1-\frac{1}{\sigma}}bk$.

Let $\pi_k(k)$ be the steady-state density function for the capital stock. As in Merton (1975), $\pi_k(k)$ exists and it can be shown to be

$$\pi_k(k) = \frac{m}{a(k)} \exp \int_0^k \frac{2b(x)}{a(x)} dx,$$

where m is a constant chosen so that $\int_0^\infty \pi_k(x) dx = 1$.

Substituting the expressions for $a(k)$ and $b(k)$, we have

$$\begin{aligned} \pi_k(k) &= \frac{m}{(1-\tau')^2 \varepsilon^2 k^2} \exp \int_0^k \frac{2((1-\tau)Ax^\alpha - (1+\tau_c)^{1-\frac{1}{\sigma}}bx)}{(1-\tau')^2 \varepsilon^2 x^2} dx \\ &= mk^{-\frac{2(1+\tau_c)^{1-\frac{1}{\sigma}}b + (1-\tau')^2 \varepsilon^2}{(1-\tau')^2 \varepsilon^2}} \exp\left(-\frac{2A(1-\tau)}{(1-\alpha)(1-\tau')^2 \varepsilon^2} k^{\alpha-1}\right) \end{aligned}$$

Defining variable $R = k^{\alpha-1}$, we have

$$\pi_R(R) = \pi_k(k) / \left| \frac{dR}{dk} \right| = \frac{m}{1-\alpha} R^{\gamma-1} \exp(-\beta R),$$

where $\gamma = 2 \frac{(1+\tau_c)^{1-\frac{1}{\sigma}}b + (1-\tau')^2 \varepsilon^2}{(1-\alpha)(1-\tau')^2 \varepsilon^2} > 0$, $\beta = \frac{2A(1-\tau)}{(1-\alpha)(1-\tau')^2 \varepsilon^2} > 0$, and b is determined by equation (12').

Therefore, we have

$$m = \frac{(1-\alpha)\beta^\gamma}{\Gamma(\gamma)},$$

where $\Gamma(\cdot)$ is the gamma function⁶.

Thus, the steady-state distribution for the capital stock is

⁶The Gamma function $\Gamma(\alpha)$ is defined as

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx.$$

$$\pi(k) = \begin{cases} 0, & k \leq 0 \\ \frac{(1-\alpha)\beta^\gamma}{\Gamma(\gamma)} k^{-\frac{2((1+\tau_c)^{-\frac{1}{\sigma}}b+(1-\tau')^2e^2)}{(1-\tau')^2e^2}} \exp(-\beta k^{-(1-\alpha)}), & k > 0 \end{cases} \quad (17)$$

and the moment-generating function

$$\Phi_k(\theta) = E\{k^\theta\} = \frac{\Gamma(\gamma - \theta/(1-\alpha))}{\Gamma(\gamma)} \beta^{\theta/(1-\alpha)} \quad (18)$$

The steady-state distribution for $R = k^{\alpha-1}$ is

$$\pi(R) = \begin{cases} 0, & R \leq 0 \\ \frac{\beta^\gamma}{\Gamma(\gamma)} R^{\gamma-1} \exp(-\beta R), & R > 0 \end{cases} \quad (19)$$

and the moment-generating function

$$\Phi_R(\theta) = E\{R^\theta\} = \frac{\Gamma(\gamma + \theta)}{\Gamma(\gamma)} \beta^{-\theta} \quad (20)$$

The steady-state distribution for $y (= Ak^\alpha)$ is

$$\pi(y) = \begin{cases} 0, & y \leq 0 \\ \frac{\eta\beta^\gamma}{A\Gamma(\gamma)} \left(\frac{y}{A}\right)^{-(\eta\gamma+1)} \exp(-\beta(\frac{y}{A})^{-\eta}), & y > 0 \end{cases} \quad (21)$$

and the moment-generating function

$$\Phi_y(\theta) = E\{y^\theta\} = \frac{A^\theta \Gamma(\gamma - \theta/\eta)}{\Gamma(\gamma)} \beta^{\theta/\eta} \quad (22)$$

With the aid of these steady-state distributions and moment-generating functions, we can derive quite a few long-run properties of our endogenous variables in the following section of comparative static analysis.

4. Comparative static analysis

With the given steady-state distributions of the capital stock, output, and the interest rate, we can derive the long-run expected values of capital stock, consumption level, and output

$$E(k) = \int_0^\infty k \pi(k) dk = \frac{\Gamma[\gamma - 1/(1-\alpha)]}{\Gamma(\gamma)} \beta^{1/(1-\alpha)}, \quad (23a)$$

$$E(c) = (1 + \tau_c)^{-\frac{1}{\sigma}} b E(k), \quad (23b)$$

$$E(y) = \frac{A\Gamma(\gamma-1/\eta)}{\Gamma(\gamma)} \beta^{1/\eta}, \quad (23c)$$

where b, β, η , and γ are presented in section 3 above, and $\Gamma(\cdot)$ is the Gamma function.

4.1 Effects of uncertainty on expected capital, output, and consumption

As in Zou (1994), the modified golden rule for the long-run capital stock in a deterministic model with the spirit of capitalism can be derived as

$$(1-\tau)f'(k) = \rho - (1+\tau_c) \frac{u_k(c, k)}{u_c(c, k)}$$

With our special utility function and production function and with the parameter condition of $\alpha = \sigma$, we have

$$(1-\tau)A\alpha k^{\alpha-1} + (1+\tau_c)\xi\left(\frac{1-\tau}{1+\tau_c} Ak^{\alpha-1}\right)^\sigma = \rho$$

and the associated consumption level and output are

$$c^* = \frac{1-\tau}{1+\tau_c} A(k^*)^\alpha, \quad y^* = A(k^*)^\alpha$$

Comparing with the uncertainty case, we have

$$E(k) < k^*, \quad E(c) < c^*, \quad E(y) < y^*$$

Thus, the long-run expected capital stock, expected consumption level, and expected output are smaller than the deterministic steady-state ones, respectively. This is because that the output is a strictly concave function of the capital stock, and Jensen's inequality implies that an increase in capital risk must reduce the expected capital stock and expected output. The fall in the expected output results in a fall in the expected consumption.

4.2 Effects of the spirit of capitalism

In our special utility function, we know that the parameter ξ measures the representative agent's concern with his social status or his spirit of capitalism. Because we have specified the parameters as $\alpha = \sigma$, we have $\sigma \in [0, 1]$.

(Please insert figure 1 about here)

Figure 1 presents the effects of the spirit of capitalism on the economy under the cases of $\alpha = 1$ (the solid line) and $\alpha \neq 1$ and $\alpha = \sigma$ (the star line). It is easy to see that with a

stronger spirit of capitalism, the long-run expected capital stock, consumption level, and output will be higher.

4.3 Effects of production shocks

(Please insert figure 2 about here)

Figure 2 shows that with increasing production shocks, the long-run expected capital stock, output, and consumption will be decreasing. Therefore, uncertainty in production reduces investment, output and consumption. This result is rather clear-cut because other related studies have indicated an ambiguous result of production shocks on investment and output, see Turnovsky (1993, 2000), Obstfeld (1994), and Gong and Zou (2002).

4.4 Effects of fiscal policies

(Please insert figure 3 about here)

The solid line in figure 3 shows the effects of income tax rate on the long-run economy. From which, we find that with a rise in the deterministic income tax rate, the long-run capital stock, output, and consumption will be decreasing (solid lines in figure 3). The effects of stochastic income tax rate (starred lines in figure 3) on the economy are just opposite to the effects of deterministic income tax rate: A rising stochastic income tax rate raises expected capital stock, output and consumption.

As for the effects of consumption tax rate on the economy, from the circled line in figure 3, we find that with an increasing consumption tax rate, the long-run expected capital stock and output will be rising, whereas the long-run expected consumption will be decreasing. This is true because a rising consumption tax raises the cost of consumption, which leads to a reduction in consumption and an increase in investment, capital stock and output. Please note that this positive effect of a consumption tax rate on capital accumulation and output is a significant feature of stochastic growth model. In the traditional, deterministic literature such as Rebelo (1990), a consumption tax has no effect on the long-run capital accumulation.

5. Equilibrium interest rates

From Cox, Ingersoll, and Ross (1985), we know that the equilibrium interest rates can be written as

$$r = \rho - \frac{L(X_k)}{X_k},$$

where $L(\cdot)$ is the differential operator.

Thus we have⁷

Proposition 2. *With the utility function and technology in equations (7) and (8), the equilibrium interest rate is given by*

$$r = \rho + \sigma((1-\tau)Ak^{\alpha-1} - (1+\tau_c)^{1-\frac{1}{\sigma}}b) - \frac{1}{2}\varepsilon^2(1-\tau')^2\sigma(\sigma+1) \quad (24)$$

when $\alpha \neq 1$ and $\alpha = \sigma$, where b is determined by equation (15'). Furthermore, the equilibrium interest rate is given by

$$r = \rho + \sigma((1-\tau)A - (1+\tau_c)^{1-\frac{1}{\sigma}}b) - \frac{1}{2}\varepsilon^2(1-\tau')^2\sigma(\sigma+1) \quad (25)$$

when $\alpha = 1$, where b is determined by equation (15).

From proposition 2, the equilibrium interest rate is a constant when the technology is linear. Among many existing studies, Sundareson (1983, 1984) has also presented a constant interest rate for a constant absolute risk aversion utility function in an infinite horizon dynamic portfolio and consumption choice problem. Our model obtains the same result while allowing the utility function to be dependent on both consumption and

⁷With the utility function and technology in equations (B1) and (8), the equilibrium interest rate is given by

$$r = \rho + (\sigma + \lambda)((1-\tau)Ak^{\alpha-1} - (1+\tau_c)^{1-\frac{1}{\sigma}}b) + \frac{1}{2}(1-\tau')^2\varepsilon^2(\sigma + \lambda)(-\sigma - \lambda - 1)$$

when $\alpha \neq 1$ and $\alpha = \sigma + \lambda$, where b is determined by

$$b = \frac{\rho^{\frac{1-\sigma}{1-\sigma-\lambda}} + \frac{1}{2}(1-\sigma)(\sigma + \lambda)(1-\tau')^2\varepsilon^2}{(1+\tau_c)^{1-\frac{1}{\sigma}}\sigma}.$$

On the other hand, the equilibrium interest rate is given by

$$r = \rho + (\sigma + \lambda)((1-\tau)A - (1+\tau_c)^{1-\frac{1}{\sigma}}b) + \frac{1}{2}(1-\tau')^2\varepsilon^2(\sigma + \lambda)(-\sigma - \lambda - 1)$$

when $\alpha = 1$, where b is determined by

$$b = \frac{\rho^{\frac{1-\sigma}{1-\sigma-\lambda}} - (1-\tau)A + \frac{1}{2}(1-\sigma)(\sigma + \lambda)(1-\tau')^2\varepsilon^2}{(1+\tau_c)^{1-\frac{1}{\sigma}}\sigma}.$$

wealth.

Comparative static analysis shows that, with a rise of technology shocks, the equilibrium interest rate will be decreasing; with a rise in the deterministic income tax rate, the equilibrium interest rate will be increasing; but the equilibrium interest rate will be decreasing with a rise of the stochastic income tax rate. Also, we find that with the increase of the consumption tax rate, the equilibrium interest rate will be increasing; please see figure 4.

When $\alpha \neq 1$, the equilibrium interest rate is stochastic, not a constant anymore. Using the expression for the equilibrium interest rate, the dynamics for the capital stock can be rewritten as

$$\frac{dk}{k} = \frac{1}{\sigma}(r - \rho + \frac{1}{2}\varepsilon^2(1 - \tau')^2\sigma(\sigma + 1))dt + \varepsilon dz \quad (26)$$

Thus, the dynamics of the interest rate is

$$\begin{aligned} dr = & \left[\frac{1}{\sigma}(r - \rho + \frac{1}{2}\varepsilon^2(1 - \tau')^2\sigma(\sigma + 1)) + (1 + \tau_c)^{\frac{1-\alpha}{\sigma}}b \right] \\ & \times \left\{ (r - \rho + \frac{1}{2}\varepsilon^2(1 - \tau')^2\sigma(\sigma + \alpha - 1))dt + (1 - \tau')\varepsilon dz \right\} \end{aligned} \quad (27)$$

If $\varepsilon = 0$ and $\xi = 0$, we have

$$\frac{dk}{k} = \frac{1}{\sigma}(r - \rho)dt, \quad \frac{dr}{r} = \frac{1}{\sigma}(r - \rho)dt$$

These are dynamic accumulation paths for the capital stock and the interest rate without production shocks and the spirit of capitalism. It is obviously that the equilibrium interest rate will convergent to ρ .

Equations (26) and (27) can be used to study the behavior of the interest rate in this economy. For example, when the initial interest rate is very high, say it is larger than $\rho - \frac{1}{2}\varepsilon^2(1 - \tau')^2\sigma(\sigma + 1)$, then the capital stock will be growing. And from the expression for the interest rate, it will go down. If the initial interest rate is lower enough, say lower than $\rho - \frac{1}{2}\varepsilon^2(1 - \tau')^2\sigma(\sigma + 1)$, then the capital stock will be increasing, thus the interest rate will be go up. Thus, the equilibrium interest rate will fluctuate around a value depending on ρ , ε^2 , and σ .

Similarly, we can find the stationary distribution for the interest rate. For simplicity,

we define $\bar{r} = \alpha(1-\tau)Ak^{\alpha-1}$, the steady-state distribution and the moment-generating function for variable $\hat{r}$ can be found as

$$\pi(\bar{r}) = \begin{cases} 0, & \bar{r} \leq 0 \\ \frac{(\frac{\beta}{\alpha(1-\tau)A})^\gamma}{\Gamma(\gamma)} \bar{r}^{\gamma-1} \exp(-\frac{\beta}{\alpha(1-\tau)A} \bar{r}), & \bar{r} > 0 \end{cases} \quad (28)$$

$$\Phi_{\bar{r}}(\theta) = E\{\bar{r}^\theta\} = \frac{\Gamma(\gamma+\theta)}{\Gamma(\gamma)} (\frac{\beta}{\alpha(1-\tau)A})^{-\theta} \quad (29)$$

Thus, we get the long-run behavior of the equilibrium interest rate

$$E(r) = \frac{\Gamma(\gamma+1)}{\Gamma(\gamma)} (\frac{\beta}{\alpha(1-\tau)A})^{-1} + \rho - \sigma(1+\tau_c)^{1-\frac{1}{\sigma}} b - \frac{1}{2} \varepsilon^2 (1-\tau')^2 \sigma(\sigma+1), \quad (30)$$

where b is determined by equation (15).

(Please insert figure 4 about here)

From figure 4, we know that the long-run expected interest rate will be decreasing with a rise in technology shocks and the deterministic income tax rate (the star line in figure 4c). At the same time, the stochastic income tax rate, the consumption tax rate, and the spirit of capitalism all have positive effects on the long-run expected interest rate.

6. Multiple optimal paths and stationary distributions

From equation (15'), we cannot determine the unique solution for variable b . In this section, we examine the existence of multiple solutions for the consumption-capital ratio for a few selected parameters. Because there exists a unique path for the capital accumulation associated with the consumption-capital ratio, there will exist a unique steady-state distribution associated with each path⁸. Below, we will present examples to show the existence of multiple optimal paths or stationary distributions and their associated long-run expected capital stocks, consumption levels, equilibrium interest rates, and output.

If we select the parameters as $A = 0.5$, $\alpha = \sigma = 0.6$, $\tau = 0.3$, $\tau' = 0.3$, $\xi = 0.1$, $\rho = 0.1$, $\tau_c = 0$, and let ε^2 vary from 0.5, 1, and 1.1, and we get the following results

⁸For the non-separable utility function in (B1) in Appendix B, we can determine the unique steady state.

Table 1: Multiple optimal paths when $\alpha = \sigma = 0.6$

	$\varepsilon^2 = 0.5$			$\varepsilon^2 = 1$			$\varepsilon^2 = 1.1$		
	path 1	path 2	path 3	path 1	path 2	path 3	path 1	path 2	path 3
c/k	0.0210	0.3519	0.1601	0.4217	0.2010	0.0430	0.2093	0.4355	0.0476
$E(k)$	150.1	0.8	4.4	0.4	2.0	25.8	1.7	0.4	20.1
$E(r)$	0.0559	0.0559	0.0559	0.0118	0.0118	0.0118	0.003	0.003	0.003
$E(c)$	3.1568	0.2732	0.6984	0.1810	0.3967	1.1067	0.3619	0.1688	0.9566
$E(y)$	4.5098	0.3902	0.9977	0.2585	0.5667	1.5811	0.5171	0.2411	1.3665

For the case of a linear technology, i.e., $\alpha = 1$, we have derived the mean growth rate of the economy and the equilibrium interest rate as follows

$$\phi = (1 - \tau)A - (1 + \tau_c)^{\frac{1}{1-\sigma}} b$$

$$r = \rho + \sigma((1 - \tau)A - (1 + \tau_c)^{\frac{1}{1-\sigma}} b) - \frac{1}{2} \varepsilon^2 (1 - \tau')^2 \sigma(\sigma + 1)$$

where b is determined by equation (15).

In this case, we select the parameters as: $\alpha = 1$, $A = 0.43$, $\sigma = 0.6$, $\tau = 0.3$, $\tau' = 0.3$, $\rho = 0.21$, and $\tau_c = 0$. When $\xi = 0$, we have a unique path or stationary distribution for consumption-capital ratio, the growth rate, and the equilibrium interest rate. When $\xi = 0.025$, we have three stationary distributions for these variables. See Table 2 for details.

Table 2: Multiple optimal paths when $\alpha = 1$

	$\xi = 0$	$\xi = 0.025$		
	path 1	path 1	path 2	path3
c/k	0.2473	0.2245	0.2578	0.2301
mean growth rate	0.0537	0.0765	0.0432	0.0709
interest rate	0.007	0.0207	0.0007	0.0174

Finally, we select the parameters as: $\alpha = 1$, $A = 0.46$, $\sigma = 0.6$, $\tau = 0.3$, $\tau' = 0.3$, $\rho = 0.25$, and $\tau_c = 0$. That is to say, we only change the values of A and the discount rate

of ρ slightly. Again, we have multiple expected values or multiple stationary distributions in the economy. See details in Table 3:

Table 3: Multiple optimal paths when $\alpha = 1$

	$\xi = 0$	$\xi = 0.025$		
	path 1	path 1	path 2	path 3
c/k	0.3000	0.2744	0.3119	0.2806
mean growth rate	0.0220	0.0476	0.0101	0.0414
interest rate	0.0280	0.0434	0.0209	0.0397

This existence of multiple stationary distributions for asset accumulation and interest rates is significant different from the unique stationary distribution in Brock and Mirman (1972), Merton (1973), Lucas (1978), Brock (1982), Cox, Ingersoll and Ross (1985), and many other classical models on stochastic capital theory and the term structure of interest rates. In fact, multiple stationary distributions in asset markets and returns may provide a more realistic picture of the real world because it admits the rationality and plausibility of different expectations and heterogeneity, though our model is still in line with the representative agent framework.

7. Conclusion

This paper has studied capital accumulation and the equilibrium interest rates in stochastic production economies with the spirit of capitalism. Under the specified utility function, production function, and selected parameters, we have presented the explicit solutions for consumption and capital accumulation. With the aid of steady-state distributions for the capital stock, we presented the effects of fiscal policies, the spirit of capitalism, and stochastic shocks on the long-run economy.

We find that the long-run capital stock, output, and consumption level in the uncertainty case are less than those ones in the deterministic case; the long-run interest rate in the uncertainty case is larger than the deterministic case. These conclusions are different from the one presented by Merton (1975), similar to the one in Smith (1998).

As for the effects of the spirit of capitalism on the long-run economy, we find that

with the increase of the spirit of capitalism, the long-run interest rate, consumption level, and output will be decreasing. The effect of the spirit of capitalism on the long-run capital stock will be negative when the production shocks are small; its effect will be positive when the production shocks are larger. These findings are different from the ones in Gong and Zou (2001, 2002), and Zou (1994, 1995).

The effects of the income tax rate on the long-run economy have also been investigated in this paper, and we have found the similar effects of a deterministic income tax and a stochastic income tax rate presented by Gong and Zou (2002), Turnovsky (1993, 2000). With the rise in the deterministic income tax rate, the long-run capital stock, output, and consumption level will be decreasing, but the interest rate will be increasing. The effects of the stochastic income tax rate on the long-run economy are just opposite to the effects of the deterministic income tax rate on the long-run economy. We have also shown that the consumption tax rate will affect the long-run expected capital stock, output, and consumption, which are different from the ones in the traditional, deterministic models.

The equilibrium interest rate has been shown under a linear technology and a nonlinear technology, respectively. When the production technology is linear, we can still obtain a constant interest rate for this stochastic model with the spirit of capitalism and social status. This result is similar to the one in Sunderason (1983), who presented the conclusion of constant interest rate under the assumption of CES utility function and a linear technology. Of course, his utility function is independent of the state variable of capital stock. But, with a nonlinear technology, we find that the interest rate follows the mean reserve process and fluctuates around a value depending on the parameters of ρ and σ .

Finally, the existence of multiple stochastic optimal paths or multiple stationary distributions for capital accumulation is presented in this paper. This is a main feature of a model with the spirit of capitalism or social-status concern. Associated with the multiple stationary distributions for capital accumulation, there exist multiple expected interest rates. This line of investigation enriches our understanding of the complexity of asset markets and the term structure of interest rates.

This paper considers an economy with one consumption good and production

technology. A first extension of this paper is to follow Sunderason (1983) to study the equilibrium interest rate in an economy with many consumption goods and production technologies. Secondly, this paper has not considered monetary policy, and we should follow Grinols and Turnovsky (1998) and extend this model to a monetary one with the spirit of capitalism. Thirdly, we can extend this model to consider habit formation, catching up with the Joneses, and the non-expected utility.

Appendix A: The steady-state distribution for a diffusion process

We follow Merton (1975) and consider the steady-state distribution for a diffusion process. Let $X(t)$ be the solution to the Itô equation

$$dx = b(x)dt + (a(x))^{1/2} dz ,$$

where $a(\cdot)$ and $b(\cdot)$ are twice-differentiable function on $[0, \infty)$ and independent of t with $a(x) > 0$ and $a(0) = b(0) = 0$.

The steady-state distribution will always exist, and it can be expressed as

$$\pi(x) = \frac{m}{a(x)} \exp \int_0^x \frac{2b(y)}{a(y)} dy ,$$

where m is chosen such that $\int_0^\infty \pi(x) dx = 1$.

Appendix B: The case of non-separable utility function

If we specified the utility function as in Bakshi and Chen (1996), Gong and Zou (2002)

$$u(c, k) = \frac{c^{1-\sigma} k^{-\lambda}}{1-\sigma}, \quad (B1)$$

where σ is the constant absolute risk aversion, and it is assumed $\sigma > 0$, and $\lambda \geq 0$ when $\sigma \geq 1$, and $\lambda < 0$ otherwise; $|\lambda|$ measures the investor's concern with his social status or measures his spirit of capitalism. The larger the parameter $|\lambda|$, the stronger the agent's concern for social status.

For the specified utility function and production function, we conjecture that the value function takes the following form

$$X(k) = a + \frac{b^{-\sigma} k^{1-\sigma-\lambda}}{1-\sigma-\lambda},$$

where a and b are constant, and they are to be determined as follows.

From the first-order condition, we have

$$c = (1 + \tau_c)^{-\frac{1}{\sigma}} b k$$

and the Bellman equation (6) is reduced to

$$\begin{aligned} & \frac{(1 + \tau_c)^{1-\frac{1}{\sigma}} b^{1-\sigma} k^{1-\sigma-\lambda}}{1-\sigma} - \rho \left(a + \frac{b^{-\sigma} k^{1-\sigma-\lambda}}{1-\sigma-\lambda} \right) + b^{-\sigma} k^{-\sigma-\lambda} ((1-\tau) A k^\alpha - (1 + \tau_c)^{1-\frac{1}{\sigma}} b k) \\ & + (1 - \tau')^2 \frac{1}{2} X_{kk} \varepsilon^2 k^2 = 0. \end{aligned}$$

If $\alpha = 1$, from the above equation, we have

$$\begin{aligned} a &= 0, \\ b &= \frac{\rho^{\frac{1-\sigma}{1-\sigma-\lambda}} - (1-\tau)A + \frac{1}{2}(1-\sigma)(\sigma+\lambda)(1-\tau')^2 \varepsilon^2}{(1 + \tau_c)^{1-\frac{1}{\sigma}} \sigma} \end{aligned}$$

In generally, for the case of $\alpha \neq 1$, we cannot determine the constants a and b . Following Xie (1994), we specified the parameters as $\alpha = \sigma + \lambda$, then, we have

$$\begin{aligned} a &= \frac{(1-\tau)A}{\rho b^\sigma}, \\ b &= \frac{\rho^{\frac{1-\sigma}{1-\sigma-\lambda}} + \frac{1}{2}(1-\sigma)(\sigma+\lambda)(1-\tau')^2 \varepsilon^2}{(1 + \tau_c)^{1-\frac{1}{\sigma}} \sigma}. \end{aligned}$$

The remaining discussions are similar to ones in the main text.

References:

Bakshi, G. S. and Z. Chen (1996). The spirit of capitalism and stock-market prices. *The American Economic Review* 86: 133-157.

Black, F. (1990). Mean reversion and consumption smoothing. *The Review of Finance Studies* 3:107-114.

Bhattacharya, S. (1981). Notes on multi-period valuation and the pricing of options. *The Journal of Finance* 36(March):163-180.

Brock, W. (1982). Asset prices in a production economy. In *The Economics of Information and Uncertainty*, edited by J. J. McCall, pp. 1-43. University of Chicago Press.

Brock, W. and L. Mirman (1972). Optimal economic growth and uncertainty: The discounted case. *Journal of Economic Theory* 4:479-513.

Cole, H., G. Mailath, and A. Postlewaite (1992). Social norms, saving behavior, and growth. *Journal of Political Economy* 100: 1092-1125.

Cole, H., G. Mailath, and A. Postlewaite (1995). Incorporating concern for relative wealth into economic models. *Quarterly Review*. Federal Reserve Bank of Minneapolis, vol.19: 12-21.

Constantinides, G. M. (1980). Admissible uncertainty in the intertemporal asset pricing model. *Journal of Finance Economics* 8(March):71-78.

Cox, J., J. Ingersoll, and S. Ross (1985) A theory of the term structure of interest rates. *Econometrica* 53: 385-408.

Cox, J., J. Ingersoll, and S. Ross (1981) A reexamination of traditional hypotheses about the term structure of interest rates. *The Journal of Finance* 36(September):769-799.

Eaton, J. (1981). Fiscal policy, inflation, and the accumulation of risk capital. *Review of Economic Studies* 48: 435-445.

Fershtman, C. and Y. Weiss (1993). Social status, culture and economic performance. *Economic Journal* 103: 946-59.

Fershtman, C., Murphy, K., and Y. Weiss (1996). Social status, education, and growth. *Journal of Political Economy* 106: 108-132.

Foldes, L. (1978). Optimal saving and risk in continuous time. *Review of Economic Studies* 45:39-65.

Gong, Liutang and Heng-fu Zou (2001). Money, social status, and capital accumulation in a cash-in-advance model. *Journal of Money, Credit, and Banking* 33(2): 284-293.

Gong, Liutang and Heng-fu Zou (2002). Direct preferences for wealth, the risk premium puzzle, growth, and policy effectiveness. *Journal of Economic Dynamics and Control* 26: 243-270.

Grinols, E. and S. Turnovsky (1993). Risk, the financial market, and macroeconomic equilibrium. *Journal of Economic Dynamics and Control* 17:1-36

Kurz, M. (1968). Optimal economic growth and wealth effects. *International Economic Review* 9:348-357.

Levhari, D. and T. Srinivisan (1969). Optimal saving under uncertainty. *Review of Economic Studies* 36:153-163.

Lucas, R. E. (1978). Asset prices in an exchange economy. *Econometrica* 46: 1426-1445.

Mehra, R. and E. C. Prescott (1985). The Equity premium: A puzzle. *Journal of Monetary Economy* 15: 145-161.

Merton, R.C.(1971). Optimal consumption and portfolio rules in a continuous-time model. *Journal of Economic Theory* 3: 373-413.

Merton, R.C. (1973). An intertemporal capital asset pricing model. *Econometrica* 41: 867-887.

Merton, R. C. (1975). An asymptotic theory of growth under uncertainty. *Review of Economic Studies* 42: 375-393.

Mirlees, J. A. (1965). Optimum accumulation under uncertainty. *Working Paper*.

Obstfeld, M. (1994). Risk-taking, global diversification, and growth. *The American Economic Review* 84: 1310-29.

Phelps, E. S. (1962). The accumulation of risk capital: A sequential utility analysis. *Econometrica* 30:729-743.

Ramsey, F. (1928). A mathematical theory of saving. *Economic Journal* vol 38:543-559.

Rebelo, S. (1991). Long-run policy analysis and long-run growth. *Journal of Political Economy* 99: 500-521.

Smith, W.T. (2001). How does the spirit of capitalism affect stock market prices? *Review of Financial Studies* 14(4):1215-1232.

Stapleton, R. and M. G. Subrahmanyam (1978). A multiperiod equilibrium asset pricing model. *Econometrica* 46(September):1077-1096.

Sunderason, M. (1983). Constant absolute risk aversion preferences and constant equilibrium interest rates. *Journal of Finance* 38(1):205-212.

Sunderason, M. (1984). Consumption and equilibrium interest rates in stochastic production economies. *Journal of Finance* 39(1):77-92.

Turnovsky, S. (1993). Macroeconomic policies, growth, and welfare in a stochastic economy. *International Economic Review* 35: 953-981.

Turnovsky, S. (2000). *Methods of Macroeconomic Dynamics*. 2nd ed. MIT Press.

Xie, D. (1994). Divergence in economic performance: Transitional dynamics with multiple equilibria. *Journal of Economic Theory* 63: 97-112.

Zou, H. (1994). The spirit of capitalism and long-run growth. *European Journal of Political Economy* 10: 279-93.

Zou, H. (1995). The spirit of capitalism and savings behavior. *Journal of Economic Behavior and Organization* 28: 131-143.

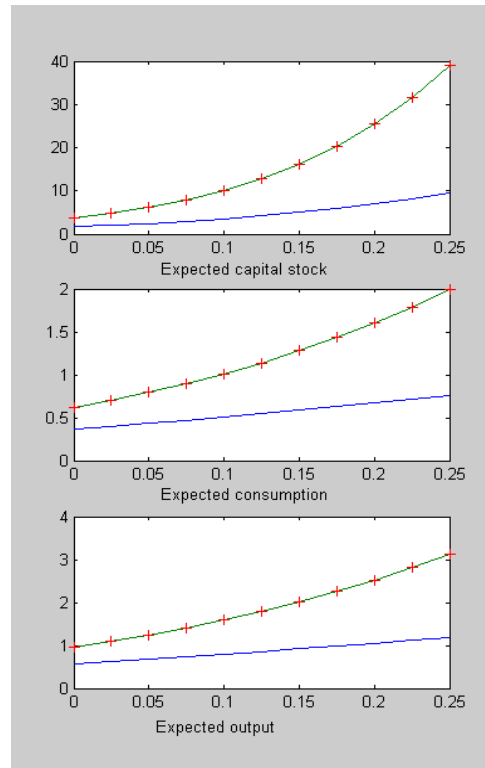


Figure 1: The effects of the spirit of capitalism on the long-run economy.

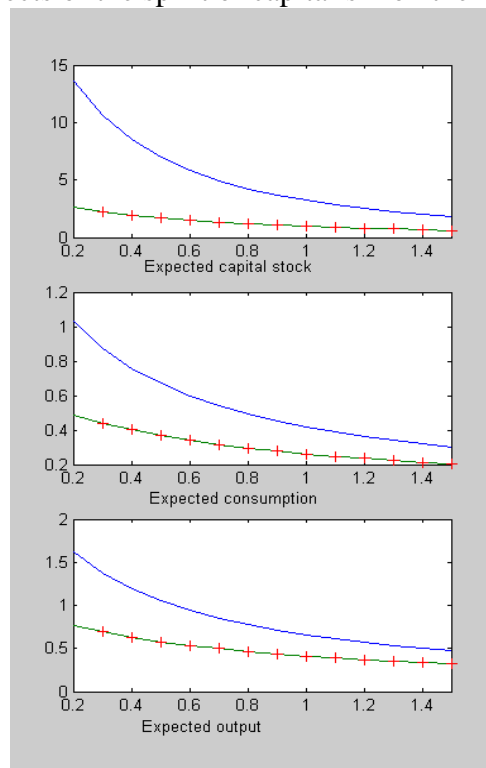


Figure 2: The effects of production shocks on the long-run economy.

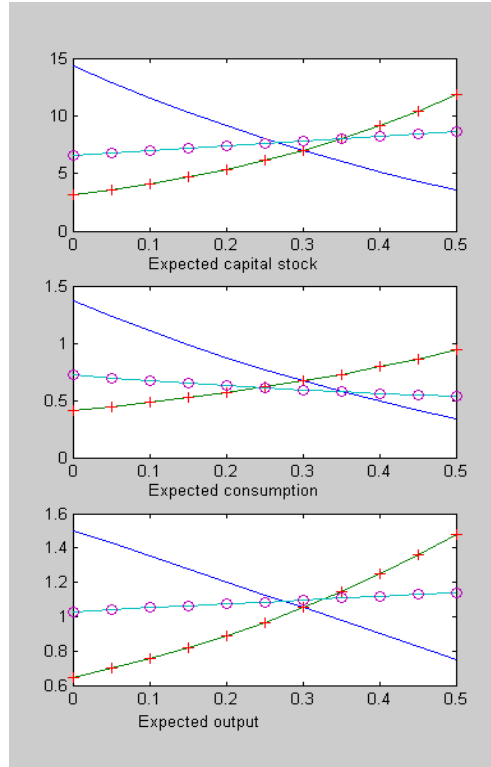


Figure 3: The effects of the deterministic income tax rate, the stochastic income tax rate, and the consumption tax rate on the long-run economy.

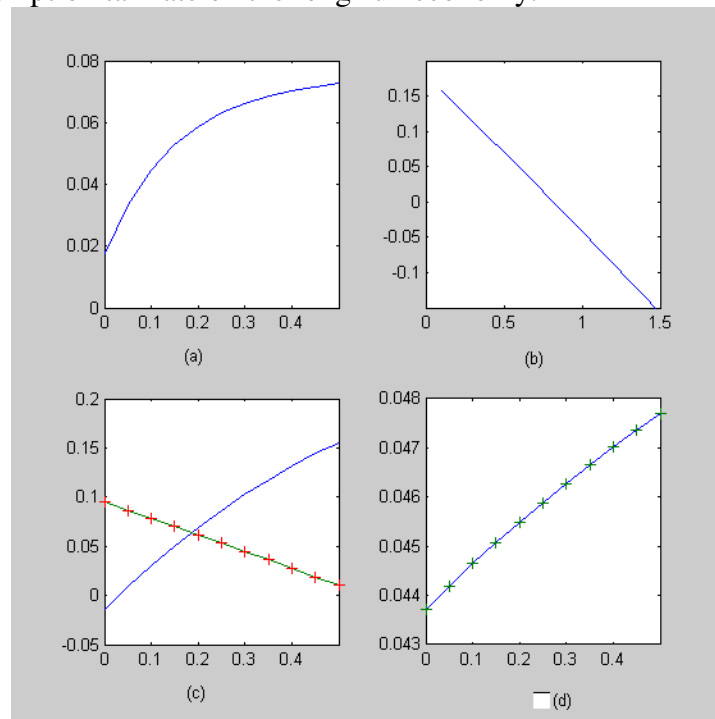


Figure 4: (a) Effects of the spirit of capitalism on the interest rate; (b) Effects of production shocks on the interest rate; (c) Effects of the deterministic income tax rate and the stochastic income tax rate on the interest rate; (d) Effects of the consumption tax rate on the interest rate.



Inflation aversion and macroeconomic policy in a perfect foresight monetary model[☆]

Gaowang Wang^{a,*}, Heng-fu Zou^{a,b,c}

^a School of Economics and Management, Institute for Advanced Study, Wuhan University, Wuhan, 430072, China

^b China Economics and Management Academy, Central University of Finance and Economics, Beijing, 100081, China

^c Institute for Advanced Study, Shenzhen University, Shenzhen, 518060, China

ARTICLE INFO

Article history:

Accepted 3 March 2011

JEL classification:

E31

E5

O41

Keywords:

Inflation aversion

Endogenous time preference

Monetary superneutrality

Optimum quantity of money

ABSTRACT

This paper reexamines monetary non-superneutrality and the optimality of the optimum quantity of money in the money-in-utility Sidrauski model with endogenous fluctuations of the time preference by introducing inflation aversion. It is shown that the long-run superneutrality of the standard Sidrauski model does not hold, and Friedman's optimum quantity of money is not optimal.

© 2011 Elsevier B.V. All rights reserved.

1. Introduction

Inflation, a term familiar to economists, policy makers and common citizens, makes people impatient, anxious, nervous and less confident. Many economists have studied the economic and psychological costs of inflation. Keynes (1936) points out that inflation leads to economic, social and institutional uncertainty and strikes at confidence. Much earlier than Keynes, Bohm-Bawerk (1891) says that inflation increases the time discount rate. Facing high inflation in the late 1960s and 1970s in the United States, Katona (1975) tells us that, with high inflation, even if real income has remained constant or increased substantially, people still feel cheated, and psychologically they regard inflation as a “bad thing”. At the same time, Fabricant (1976) states that the uncertainty and anxiety from inflation makes more impatience and a large time discount rate and that high inflation makes rational calculation more difficult or impossible and makes people possess even less “adequate power to imagine and to abstract” the future. Burns (1978) also writes that “by causing disillusionment and breeding discount, inflation excites doubts among people about themselves, about the competence of their government, and about the free enterprise system itself.” More recently, Shiller (1996) has written that “it was very easy to see why

people dislike inflation: people think inflation erodes their standard of living”; and that “this standard of living effect is not the only perceived cost of inflation among non-economists: other perceived costs are tied up with issues of exploitation, political instability, loss of morale, and damage to national prestige.” All these statements and assessments lead to the same conclusion: inflation tends to impair the patience and confidence of the people.

In order to model this negative effect of anticipated inflation on patience, we take the time preference rate as an increasing function of the inflation rate endogenously, and name it “inflation aversion”. The objective of this paper is to investigate the macroeconomic implication of this “stylized” psychological fact. Actually, Stockman (1981) has given some hints on this modeling strategy in the first footnote of his paper. He says that “if inflation affects β (the time preference rate) in the steady state, then any effect of inflation on the capital stock is possible, depending upon how inflation affects this particular aspect of ‘tastes’.” Stockman's analysis had been anticipated by Keynes (1936) who had attached great importance to this psychological characteristics of human nature and states the endogenous fluctuation of the rate of time-discounting (page 93), “The state of confidence, as they term it, is a matter to which practical men always pay the closest and most anxious attention. But economists have not analysed it carefully and have been content, as a rule, to discuss it in general terms. In particular it has not been made clear that its relevance to economic problems comes in through its important influence on the schedule of the marginal efficiency of capital. There are not two separate factors affecting the rate of investment, namely, the schedule of the marginal efficiency of capital and the state of

[☆] The authors are greatly indebted to the two anonymous referees and editor for their helpful comments and suggestions. All remaining errors are our own.

* Corresponding author. Tel.: +86 1062288992; fax: +86 1062288376.

E-mail addresses: wanggaowang@gmail.com (G. Wang), zouhengfu@gmail.com (H. Zou).

confidence. The state of confidence is relevant because it is one of the major factors determining the former, which is the same thing as the investment demand-schedule.”

A large literature has examined the relationship between endogenous time preferences and monetary superneutrality. Uzawa (1968) sets up an infinitely-lived-representative-agent model with an endogenous time preference to replicate the Mundell–Tobin effect. By assuming that the rate of time preference is an increasing and convex function of the current level of utility, he shows that monetary growth raises savings and the capital stock. Using Uzawa’s time preference, Obstfeld (1981) further examines the long-run monetary non-superneutrality in a small open economy. Epstein and Hynes (1983) have also examined monetary superneutrality in Sidrauski (1967) model and concluded that a higher rate of monetary expansion increases the steady-state levels of consumption and capital stock, and reduces the steady-state level of real balances. Recently, in a growth model with the Marshallian time preference, Gootzeit et al. (2002) show that a permanent increase in government expenditure causes “super-crowding-out” of consumption and lowers the steady-state capital stock. By modeling time preference as an increasing function of real wealth, Kam (2005) has also reexamined the existence of the Tobin effect.¹

And ever since Friedman puts forward his famous rule for the optimum quantity of money,² many economists have examined its optimality. It has been shown to be optimal in monetary economies with monopolistic competition (Ireland, 1996) and, under certain circumstances, in a variety of monetary economies where government levies other distorting taxes (Chari et al., 1996; Gahvari, 2007; Da Costa and Werning, 2008). However, there exist several cases where the Friedman rule is not optimal. These include economies with cash-in-advance constraints (Stockman, 1981; Abel, 1985; Ellison and Rankin, 2007); economies with time inconsistency of monetary and fiscal policy (Alvarez et al., 2004), economies with intergenerational wealth effects of monetary growth (Gahvari, 1988, 2007); economies with redistributive effects of monetary growth (Bhattacharya et al., 2005), and economies with strong Tobin effects (Bhattacharya et al., 2009).

The paper incorporates “inflation aversion” into the standard Sidrauski (1967) model and reexamines monetary superneutrality and the optimality of Friedman’s rule for optimum quantity of money. Again, “inflation aversion” means that inflation causes people to become more impatient and they increase their subjective discount rate. The formal model of inflation aversion is presented in Section 2. In Section 3, we show the dynamics of the system and study the properties of the steady state. Comparative dynamics are analyzed in Section 4, and a summary of our main findings concludes the paper.

2. The model

2.1. The endogenous time preference with inflation aversion

As is well known, the time preference rate is a measure of the agent’s patience in common sense. And in the continuous-time model, the larger the time discount rate, the less patience the agent. Usually the time discount rate is assumed to be an exogenously given, positive constant. In order to investigate the possible economic effects of the psychological aversion of inflation, we assume that the time

preference rate of the representative individual is a strictly increasing, strictly concave function of the expected inflation rate. That is,

$$\rho_t = \rho(\pi_t), \quad (1)$$

which satisfies

$$\rho'(\pi_t) > 0, \rho''(\pi_t) < 0, \rho(0) = \rho_f. \quad (2)$$

Assumptions (1) and (2) make the time preference rate endogenous, and they imply the higher the inflation rate, the less patience the individual. But notice that the decrease in the patience is at a decreasing rate. Moreover, the discount rate is a positive constant if the inflation rate is zero, just like a “Fisherian” consumer with a constant rate of time preference, i.e., $\rho(0) = \rho_f$. Furthermore, it is also assumed that the time discount factor of the individual at time t depends not only on the current level of inflation, but also on the entire path of past inflation $\{\pi_v\}_{v=0}^t$, namely,

$$\Delta_t = \int_v^t \rho(\pi_v) dv. \quad (3)$$

Then the modeling strategy has generated a new state variable, the real time discount factor Δ_t . Differentiating Δ_t with respect to t in Eq. (3), we obtain the dynamic accumulation equation of the time discount factor, namely,

$$\dot{\Delta}_t = \rho(\pi_t). \quad (4)$$

With these new elements introduced, this paper will reexamine the Sidrauski model and the long-run effects of the monetary policy.³

2.2. The Sidrauski model with inflation aversion

2.2.1. Consumer’s behavior

The representative individual’s optimization problem is to maximize

$$W = \int_{t=0}^{\infty} [u(c_t, m_t)] e^{-\Delta_t} dt \quad (5)$$

subject to the budget constraint

$$\dot{a}_t = r_t k_t + w_t - c_t - \pi_t m_t + \tau_t, \quad (6)$$

and wealth constraint

$$a_t = k_t + m_t, \quad (7)$$

plus the no-Ponzi-game condition

$$\lim_{t \rightarrow \infty} a_t \exp\left(-\int_v^t r_v dv\right) = 0, \quad (8)$$

where c_t , m_t , k_t , and a_t are consumption, real money balances, physical capital stock, and total wealth, respectively; r_t and w_t are the real interest rate and real wages; Δ_t and π_t are the time discount factor and the expected rate of inflation; and τ_t denotes lump-sum real money transfer payments. The stock constraint requires that the total wealth a_t be allocated between capital k_t and real balances m_t . And the no-Ponzi-game condition rules out unlimited borrowings. The instantaneous utility function $U_t = u(c_t, m_t)$ is assumed to be well-behaved, satisfying $u_c > 0$, $u_m > 0$, $u_{cc} < 0$, $u_{mm} < 0$, $u_{cc}u_{mm} - u_{cm}^2 > 0$ and the Inada conditions. Following Sidrauski (1967) and Fisher (1979), we assume

¹ Actually, many papers have examined the recursive structure of the endogenous time preferences, such as Obstfeld (1990) and Epstein (1983, 1987).

² Friedman (1969) argues that a positive nominal interest rate represents a distortionary tax on real money balances. To reach the first-best, the distortion should be removed and the nominal interest rate should be set to zero. This prescription is known as the Friedman rule for the optimum quantity of money.

³ For simplicity, we just consider the case without population growth.

that both commodities are not inferior.⁴ Furthermore, to reach a definitive conclusion, following Calvo (1979), we assume that consumption and real money balances are Edgeworth-complementary, i.e., $u_{cm} > 0$.⁵ Intuitively, an increase in real balances raises the marginal valuation of consumption and increases consumption; and a lower level of money holdings decreases the marginal valuation of consumption and lowers consumption. Hence, in the steady state, consumption and real money balances move in the same direction.

To proceed, the optimization problem of the representative consumer is to maximize Eq. (5), subject to Eqs. (6), (4), (7) and (8). The Hamiltonian associated with this problem is

$$H = u(c_t, m_t)e^{-\Delta t} + \tilde{\lambda}_t[r_t k_t + w_t - c_t - \pi_t m_t + \tau_t] + \tilde{\mu}_t \rho(\pi_t) + \tilde{q}_t(k_t + m_t - a_t), \quad (9)$$

where $\tilde{\lambda}_t$ and $\tilde{\mu}_t$ are the multiplier associated with the constraints (6) and (4), representing the shadow values of wealth and time discount factor, respectively; $\tilde{q}_t$ is the Lagrangian multiplier attached to the stock constraint (7), representing the marginal value of total wealth.⁶

The first-order conditions for a maximum are given by Eqs. (10)–(13) together with the transversality conditions:

$$u_c(c, m)e^{-\Delta} = \tilde{\lambda}, \quad (10)$$

$$u_m(c, m)e^{-\Delta} = (r + \pi)\tilde{\lambda}, \quad (11)$$

$$\dot{\tilde{\lambda}} + r\tilde{\lambda} = 0, \quad (12)$$

$$u(c, m)e^{-\Delta} = \dot{\tilde{\mu}}. \quad (13)$$

$$\lim_{t \rightarrow \infty} e^{-\Delta t} \tilde{\lambda} k = 0, \lim_{t \rightarrow \infty} e^{-\Delta t} \tilde{\mu} \Delta = 0. \quad (14)$$

Eqs. (10) and (11) are two intratemporal optimality conditions, implying that the marginal utility of consumption and (or) real balances equals the real marginal valuation of wealth; Eqs. (12) and (13) are two Euler equations, which determine the intertemporal choices of consumption and real money balances; and Eq. (12) is the Keynes–Ramsey condition, which implicitly shows that the marginal rate of substitution between consumption at two points of time must equal the marginal rate of transformation.

Now let us define the current-value Hamiltonian multipliers λ and μ as a product of their corresponding present-value Hamiltonian multipliers and $e^{\Delta t}$:

$$\lambda = e^{\Delta t} \tilde{\lambda}, \mu = e^{\Delta t} \tilde{\mu}. \quad (15)$$

Taking the derivative of Eq. (15) with respect to t , we have:

$$\dot{\lambda} = [\dot{\tilde{\lambda}} - \rho(\pi)\tilde{\lambda}]e^{-\Delta}, \dot{\mu} = [\dot{\tilde{\mu}} - \rho(\pi)\tilde{\mu}]e^{-\Delta}. \quad (16)$$

Substituting Eq. (10) into Eq. (15) leads to

$$u_c(c, m) = \lambda. \quad (17)$$

Putting Eqs. (16), (15) and (17) into Eq. (10) gives rise to

$$\dot{\lambda} = -[r - \rho(\pi)]\lambda. \quad (18)$$

⁴ It is not hard to prove that the normality of the two goods is equivalent to the following two conditions, respectively, $u_{mm} - \frac{u_{cm}u_m}{u_c} < 0$, $\frac{u_{cc}u_m}{u_c} - u_{cm} < 0$.

⁵ Wang and Yip (1992) called the assumption Pareto complementarity between consumption and money.

⁶ For notational simplicity, we will omit the time subscript in the following mathematical presentations.

Taking the derivative of Eq. (17) with respect to t , and using Eqs. (17) and (18) lead to

$$\dot{c} = -[r - \rho(\pi)] \frac{u_c(c, m)}{u_{cc}(c, m)} - \frac{u_{cm}(c, m)}{u_{cc}(c, m)} \dot{m}. \quad (19)$$

Eqs. (10) and (11) imply that:

$$\frac{u_m(c, m)}{u_c(c, m)} = (r + \pi). \quad (20)$$

Hence, at optimum the marginal rate of substitution between consumption and real money balances is equal to the nominal interest rate, which is the price of monetary services or the opportunity cost of holding money.

Finally, Eqs. (13) and (16) together imply

$$\dot{\mu} = u(c, m) + \rho(\pi)\mu. \quad (21)$$

2.2.2. Behavior of the firm

It is assumed that the production function of the firm is well behaved, namely, $f(0) = 0, f(\infty) = \infty, f(0 < k < \infty) > 0, f'(0) = \infty, f'(\infty) = 0, f'(0 < k < \infty) > 0, f''(k) < 0$, and that factor markets are competitive.⁷ Accordingly,

$$r = f'(k), w = f(k) - kf'(k). \quad (22)$$

That is to say, the market interest rate equals the marginal productivity of capital and the market wage rate equals the marginal productivity of labor.

2.2.3. Macroeconomic equilibrium

In order to complete the system, we introduce the government's behavior. It is assumed that the government maintains a constant rate of monetary expansion

$$\frac{\dot{M}}{M} = \theta \quad (23)$$

and keeps its budget balanced:

$$\tau + g = \frac{\dot{M}}{P}, \quad (24)$$

where θ and g are two constants denoting the monetary growth rate and government expenditure, respectively. By the definition of real money balances, $m = \frac{M}{P}$. Substituting Eq. (23) into Eq. (24) results in

$$\tau + g = \theta m. \quad (25)$$

We impose the assumption of perfect foresight which says that the expected rate of inflation is equal to the real rate of inflation, namely,

$$\frac{\dot{P}}{P} = \pi. \quad (26)$$

Taking the derivative of $m = \frac{M}{P}$ with respect to t , rearranging, and substituting Eqs. (23) and (26) into it, we have

$$\dot{m} = (\theta - \pi)m. \quad (27)$$

⁷ For simplicity, we assume that the rate of depreciation for capital is zero.

Putting Eq. (22) into Eq. (20) and rearranging them,

$$\pi = \frac{u_m(c, m)}{u_c(c, m)} - f'(k). \quad (28)$$

From Eq. (28), we solve π as a function of c, m , and k , i.e., $\pi_t = \pi(c, k, m)$. And it is easy to show that

$$\pi_c = \frac{u_{mc}u_c - u_{cc}u_m}{u_c^2} > 0, \pi_m = \frac{u_{mm}u_c - u_{cm}u_m}{u_c^2} < 0, \pi_k = -f''(k) > 0. \quad (29)$$

Putting $\pi_t = \pi(c, k, m)$ into Eq. (27) gives the dynamics of real money balances

$$\dot{m} = (\theta - \pi(c, k, m))m. \quad (30)$$

Substituting Eqs. (7), (22), (25), and (27) into Eq. (6) results in the dynamic equation of physical capital accumulation

$$\dot{k} = f(k) - c - g. \quad (31)$$

Putting Eqs. (22), (30), and $\pi_t = \pi(c, k, m)$ into Eq. (19) gives the dynamic equation of consumption

$$\dot{c} = -[f'(k) - \rho(\pi(c, k, m))] \frac{u_c(c, m)}{u_{cc}(c, m)} - \frac{u_{cm}(c, m)}{u_{cc}(c, m)} (\theta - \pi(c, k, m))m. \quad (32)$$

Therefore, Eqs. (30)–(32) describe the whole dynamics of the model.

2.3. Dynamics and the steady state

2.3.1. The steady state

In the steady state (c^*, k^*, m^*) , $\dot{c} = \dot{k} = \dot{m} = 0$, namely,

$$f'(k^*) = \rho(\pi(c^*, k^*, m^*)), \quad (33)$$

$$f(k^*) = c^* + g, \quad (34)$$

$$\theta = \pi(c^*, k^*, m^*). \quad (35)$$

Eq. (33) gives the familiar modified golden-rule level of capital accumulation, which shows that, in the steady state, the marginal product of physical capital equals the subjective time preference rate; Eq. (34) tells that the steady-state production can be divided into two parts: one is the steady-state level of consumption, and the other is the exogenous level of government expenditure; and Eq. (35) shows that the steady-state level of inflation is equal to the exogenous level of monetary growth.

Furthermore, it is easy to see the existence and uniqueness of the steady state from the steady-state Eqs. (33)–(35) and the basic assumptions of the model.

2.3.2. Stability of the steady state

To examine the local stability of the steady state, we linearize Eqs. (30)–(32) around the steady state (c^*, k^*, m^*)

$$\begin{bmatrix} \dot{c} \\ \dot{k} \\ \dot{m} \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ -1 & f'(k^*) & 0 \\ -\pi_c^* m^* & -\pi_k^* m^* & -\pi_m^* m^* \end{bmatrix} \begin{bmatrix} c - c^* \\ k - k^* \\ m - m^* \end{bmatrix}, \quad (36)$$

where

$$\begin{aligned} a_{11} &= \frac{\pi_c^* [u_c^* \rho'(\theta) + u_{cm}^* m^*]}{u_{cc}^*} < 0, \\ a_{12} &= \frac{-u_c^* [f''(k^*) - \pi_k^* \rho'(\theta)]}{u_{cc}^*} + \frac{u_{cm}^* \pi_k^* m^*}{u_{cc}^*} < 0, \\ a_{13} &= \frac{\pi_m^* [u_c^* \rho'(\theta) + u_{cm}^* m^*]}{u_{cc}^*} > 0. \end{aligned}$$

Let us define the Jacobian matrix of the linearized system as J . It is not hard to find that

$$\prod_{i=1}^3 \lambda_i = \det(J) = \frac{u_c^* f''(k^*) \pi_m^* m^*}{u_{cc}^*} < 0. \quad (37)$$

Eq. (37) implies that there exists one negative real eigenvalue or three eigenvalues with negative real parts. The trace of the Jacobian matrix is

$$\sum_{i=1}^3 \lambda_i = \text{tr}(J) = f'(k^*) + \frac{(u_{mc}^* u_c^* - u_{cc}^* u_m^*) \rho'(\theta) + (u_{mc}^{*2} - u_{cc}^* u_{mm}^*) m^*}{u_c^* u_{cc}^*}, \quad (38)$$

and we cannot decide its sign on the basis of the assumptions of the model. In order to guarantee the saddle-point stability of the steady state, we impose the following assumption:

$$\text{tr}(J) > 0. \quad (39)$$

If condition (39) holds, then there exists a unique negative eigenvalue corresponding to the unique predetermined variable k . Hence, the steady state is a saddle point.

Notice that, the condition (39) is not stringent at all. In addition, the curvature of the time preference function plays no role in the determination of stability, since the second derivative of the time preference function does not enter the Jacobian matrix J . For sure, let us see three numerical examples.

Example 1. Assume the utility function is separable in consumption and real balances for simplicity: $u(c, m) = \log c + \log m$. Let the production be a Cobb–Douglas technology: $f(k) = k^{0.35}$. And define the time preference as a concave function of the inflation rate: $\rho(\pi) = \log(\pi + 1.2)$. With $\theta = 0.001$, the unique steady state is: $k = 2.7083$, $c = 1.4172$, $m = 7.6959$, $\pi = 0.001$, $\rho = 0.1832$ and the corresponding eigenvalues are: $-0.2921, 0.4103, 0.0957$. Then $\text{tr}(J) = 0.2139 > 0$, and condition (39) is satisfied.

Example 2. Let the utility function, the production function and θ be the same as in Example 1. Let the time discount rate be: $\rho(\pi) = \log(\pi + 0.01)$. Then, the unique steady state is given by: $k = 205.0257$, $c = 6.4437$, $m = 536.9720$, $\pi = 0.001$, $\rho = 0.011$ and the corresponding eigenvalues are: $-0.0194, 0.0248, 0.0056$. Now $\text{tr}(J) = 0.0100 > 0$, and condition (39) holds again.

Example 3. Keep everything the same as in Example 1 expect for the time discount rate: $\rho(\pi) = \exp(\pi) - 0.998$. Then, the unique steady state is given by: $k = 1312.8879$, $c = 12.9698$, $m = 3242$, $\pi = 0.001$, $\rho = 0.0056$ and the three eigenvalues are: $-0.0054, 0.0065, 0.00519$. It is obvious the sum of the three eigenvalues is positive $\text{tr}(J) = 0.00629 > 0$ as required by condition (39).

Therefore, we have the following proposition.

Proposition 1. *In the Sidrauski model with inflation aversion, if $\text{tr}(J) > 0$, the steady state is locally saddle-point stable.*

3. Macroeconomic policy analysis

3.1. Long-run effects of monetary policy

3.1.1. Monetary non-superneutrality

Totally differentiating Eqs. (33)–(35) give us a three-dimensional linear system as follows:

$$\begin{bmatrix} \rho'(\theta)\pi_c^* & \rho'(\theta)\pi_k^* - f''(k^*) & \rho'(\theta)\pi_m^* \\ 1 & -f'(k^*) & 0 \\ \pi_c^* & \pi_k^* & \pi_m^* \end{bmatrix} \begin{bmatrix} dc^* \\ dk^* \\ dm^* \end{bmatrix} = \begin{bmatrix} 0 \\ -dg \\ d\theta \end{bmatrix}. \quad (40)$$

Setting $dg = 0$ and applying Cramer's rule, we obtain

$$\frac{dc^*}{d\theta} = \frac{\rho'(\theta)f'(k^*)}{f''(k^*)} < 0, \quad (41)$$

$$\frac{dk^*}{d\theta} = \frac{\rho'(\theta)}{f''(k^*)} < 0, \quad (42)$$

$$\frac{dm^*}{d\theta} = \frac{f''(k^*) - \rho'(\theta)\pi_k^* - \rho'(\theta)\pi_c^*f'(k^*)}{\pi_m^*f''(k^*)} < 0. \quad (43)$$

Therefore, we have the following proposition.

Proposition 2. *A permanent increase in the monetary growth rate decreases the steady-state consumption, capital accumulation and real balance holdings. That is to say, money is not superneutral in the Sidrauski model with inflation aversion.*

In the standard Sidrauski model with a constant time preference rate, the steady-state levels of capital stock and consumption are given by the same conditions as those in the nonmonetary Ramsey model, and, they are independent of the monetary growth rate. That is to say, money is superneutral in the long run. However, in the Sidrauski model with inflation aversion, the time preference rate depends on the inflation rate endogenously, and hence, the steady-state level of capital depends on the inflation rate (or money growth rate). The logic for the failure of superneutrality is as follows: an increase in the rate of money growth raises the rate of time preference in the steady state ($\rho'(\theta) > 0$), the steady-state interest rate increases ($\frac{dr}{d\theta} = \rho'(\theta) > 0$), and then the steady-state stock of capital falls because of diminishing returns and higher rental costs. In turn, the fall in the capital stock reduces net output and consumption. These conclusions are just what Burns (1978) had said: inflation “weakens the willingness to save. It drives up the level of interest rates. It affects adversely both stock prices and bond prices.”

As inflation rises, the opportunity cost of holding money is higher. Hence, the steady-state level of real balances decreases. In the Sidrauski model, the money demand function is $m = m(\rho, \theta)$ and the effect of a positive monetary disturbance is negative, i.e., $\frac{dm^*}{d\theta} = m_\theta^*(\rho, \theta) < 0$. But, in the model with inflation aversion, the money demand function is $m = m(\rho(\theta), \theta)$ and a negative item ($m_\rho^*(\rho, \theta)\rho'(\theta) = -\frac{\rho'(\theta)[\pi_c^*f'(k^*) + \pi_k^*]}{\pi_m^*f''(k^*)}$) is added to $\frac{dm^*}{d\theta}$, namely,

$$\frac{dm^*}{d\theta} = m_\rho^*(\rho, \theta)\rho'(\theta) + m_\theta^*(\rho, \theta) = -\frac{\rho'(\theta)[\pi_c^*f'(k^*) + \pi_k^*]}{\pi_m^*f''(k^*)} + m_\theta^*(\rho, \theta) < 0.$$

Hence, the total effect includes both the original Keynesian part, $m_\theta(\rho, \theta)$ and the new part coming from “inflation aversion”, $m_\rho(\rho, \theta)\rho'(\theta)$

($= -\frac{1}{\pi_m^*}[\pi_c^*f'(k^*) + \pi_k^*f''(k^*)]\rho'(\theta)$). Therefore, the negative effect on real money balances of a positive monetary disturbance is stronger.

The negative effect of inflation is strong enough so that both consumption and physical capital decrease, which is similar to Stockman (1981) and Abel (1985), but is different from the positive effect of inflation in Tobin (1965), Uzawa (1968), Epstein and Hynes (1983), Obstfeld (1981), and Kam (2005). And these conclusions affirm the theoretical conjecture of Stockman (1981) and the empirical evidences provided by Fischer (1993).⁸

3.1.2. The optimum quantity of money

To examine the optimality of Friedman's rule for optimum quantity of money, let us write down the steady-state utility:

$$W^* = \int_0^\infty e^{-\rho(\theta)t} u(c^*, m^*) dt = \frac{u(c^*, m^*)}{\rho(\theta)}. \quad (44)$$

Taking the derivative of W with respect to θ in Eq. (44) yields

$$\frac{dW^*}{d\theta} = \frac{[u_c^* \frac{dc^*}{d\theta} + u_m^* \frac{dm^*}{d\theta}]\rho(\theta) - \rho'(\theta)u(c^*, m^*)}{\rho^2(\theta)} < 0. \quad (45)$$

It is easy to find that the total effect of a permanent increase in monetary growth on the equilibrium welfare can be divided into three negative parts: a decrease in utility owing to a lower consumption, $u_c^* \frac{dc^*}{d\theta} \rho(\theta)$; a decrease in utility due to lower real balances, $u_m^* \frac{dm^*}{d\theta} \rho(\theta)$; and a decrease in utility due to increased impatience, $-\rho'(\theta)u(c, m)$. Altogether, Eq. (45) tells us that an increase in the monetary growth rate cuts the steady-state welfare. Therefore, the equilibrium welfare can be improved by reducing the rate of monetary growth. That is to say, Friedman's rule for optimum quantity of money is not optimal in the economy. In fact, we can explain this in another way. Suppose that the Friedman rule still holds, that is, the nominal interest rate is equal to zero. From Eq. (20), we have $u_m = 0$. Putting it into Eq. (45) leads to

$$\frac{dW^*}{d\theta} \left(= \frac{u_c^* \frac{dc^*}{d\theta} \rho(\theta) - \rho'(\theta)u(c^*, m^*)}{\rho^2(\theta)} \right) < 0. \text{ This inequality implies that}$$

the steady-state level of welfare can be improved all along by reducing the monetary growth rate. The optimum quantity of money may be setting $\theta = -\infty$, which is unreasonable and impossible. This implies that the optimum quantity of money in Friedman's rule does not hold in our model.

3.2. Long-run effects of fiscal policy

3.2.1. Purely crowding out of consumption

Similar to Section 3.1.1, setting $d\theta = 0$ in Eq. (40) and applying Cramer's rule lead to

$$\frac{dc^*}{dg} = -1, \quad (46)$$

$$\frac{dk^*}{dg} = 0, \quad (47)$$

$$\frac{dm^*}{dg} = \frac{\pi_c^*}{\pi_m^*} < 0. \quad (48)$$

⁸ Fischer (1993) demonstrates a 1% rise in inflation can cost an economy on the order of 0.1% in its rate of growth.

Proposition 3. *An increase in government expenditure reduces the steady-state consumption and real balance holdings, whereas it has no effect on the steady-state capital stock.*

It is easy to see from Eqs. (46) and (47) that the long-run effects of positive government disturbances are the same as the nonmonetary Ramsey model: an increase of government expenditure crowds out private consumption one-to-one, and it has no effect on the long-run capital accumulation. And the negative effect on real money balances of an increase in government expenditures can be explained intuitively. The budget constraint of the government says that the income of inflation tax $\frac{M}{P} (= \theta m)$ can be divided into two parts: the expenditure on the government consumption, g , and lump-sum transfers to the private sector, τ . If keeping τ and m constant and increasing g , the monetary authorities must increase θ and, hence, the inflation rate π . Then, a private individual with inflation aversion becomes more impatient, and he increases current consumption and cuts real money balances.

3.2.2. The effect on the steady-state welfare

Taking the derivative of W with respect to θ in Eq. (44) gives rise to

$$\frac{dW^*}{d\theta} = \frac{u_c^* \frac{dc^*}{d\theta} + u_m^* \frac{dm^*}{d\theta}}{\rho(\theta)} = \frac{-[u_c^* - u_m^* \pi_c^* f'(k^*)]}{\rho(\theta)} < 0. \quad (49)$$

The equation above implies that an increase of government expenditure reduces the steady-state welfare. The negative effects can be divided into two parts: one is from the decrease in consumption ($\frac{-u_c^*}{\rho(\theta)}$), and the other is from lower real balance holdings ($\frac{u_m^* \pi_c^* f'(k^*)}{\rho(\theta)}$). Then, with less patience and higher opportunity costs of holding money, consumers increase consumption, reduce savings and lower the holdings of money. In the long run, the steady-state levels of consumption and real money balances are reduced, and so is the welfare.

Furthermore, it may be interesting to examine the long-run effects of a fiscal expansion on the equilibrium inflation rate. Totally differentiating $\pi_t = \pi(c, k, m)$ with respect to g results in

$$\frac{d\pi^*}{dg} = \pi_c^* \frac{dc^*}{dg} + \pi_k^* \frac{dk^*}{dg} + \pi_m^* \frac{dm^*}{dg} = -\pi_c^* + \pi_m^* \frac{dm^*}{dg}. \quad (50)$$

First, we notice that the term $-\pi_c^*$ is negative from Eq. (29). Next the second term $\pi_m^* \frac{dm^*}{dg}$ is positive from Eqs. (29) and (46). Hence, the total effect is ambiguous.

4. Summary

In this short paper, by introducing the inflation rate into the representative agent's time preference rate, we have reexamined the effects of monetary and fiscal policies in the money-in-utility model. The comparative static analysis has demonstrated: neither monetary superneutrality nor Friedman's rule for optimum quantity of money holds. Specifically, with an increase of the money growth rate, the steady-state consumption, physical capital stock, real money balance

holdings, and welfare all decrease. In addition, with a rise in government expenditure, the steady-state consumption, real money balances, and welfare will be reduced, whereas the steady-state capital stock remains unchanged.

References

- Abel, A., 1985. Dynamic behavior of capital accumulation in a cash-in-advance model. *Journal of Monetary Economics* 10, 55–71.
- Alvarez, F., Kehoe, P., Neumeyer, P., 2004. The time consistency of optimal monetary and fiscal policies. *Econometrica* 72, 541–567.
- Bhattacharya, J., Haslag, J., Martin, A., 2005. Heterogeneity, redistribution, and the Friedman rule. *International Economic Review* 46, 437–454.
- Bhattacharya, J., Haslag, J., Martin, A., 2009. Optimal monetary policy and economic growth. *European Economic Review* 53, 210–221.
- Bohm-Bawerk, E., 1891. *The Positive Theory of Capital*. (W. Smart, Trans.). Macmillan, London.
- Burns, A., 1978. *The Condition of the American Economy*. American Enterprise Institute for Public Policy Research, Washington, D.C.
- Calvo, G., 1979. On models of money and perfect foresight. *International Economic Review* 20, 83–103.
- Chari, V., Christiano, L., Kehoe, P., 1996. Optimality of the Friedman rule in economies with distorting taxes. *Journal of Monetary Economics* 37, 203–223.
- Da Costa, C., Werning, I., 2008. On the optimality of the Friedman rule with heterogeneous agents and nonlinear income taxation. *Journal of Political Economy* 116 (1).
- Ellison, M., Rankin, N., 2007. Optimal monetary policy when lump-sum taxes are unavailable: a reconsideration of the outcomes under commitment and discretion. *Journal of Economic Dynamics and Control* 31, 219–243.
- Epstein, L., 1983. Stationary cardinal utility and optimal growth under uncertainty. *Journal of Economic Theory* 31, 133–152.
- Epstein, L., 1987. The global stability of efficient intertemporal allocations. *Econometrica* 55, 329–355.
- Epstein, L., Hynes, J., 1983. The rate of time preference and dynamic economic analysis. *Journal of Political Economy* 91 (4).
- Fabricant, S., 1976. *Economic Calculation under Inflation: the Problem in Perspective*. Economic Calculation under Inflation. Liberty Press.
- Fischer, S., 1993. The role of macroeconomic factors in growth. *Journal of Monetary Economics* 32, 485–512.
- Fisher, S., 1979. Capital accumulation on the transition path in a monetary optimizing model. *Econometrica* 47, 1433–1439.
- Friedman, M., 1969. *The Optimum Quantity of Money and Other Essays*. Aldine, Chicago.
- Gahvari, Firouz, 1988. Lump-sum taxation and the superneutrality and optimum quantity of money in life cycle growth models. *Journal of Public Economics* 36, 339–367.
- Gahvari, F., 2007. The Friedman rule: old and new. *Journal of Monetary Economics* 54, 581–589.
- Gootzeit, M., Schneider, J., Smith, W., 2002. Marshallian recursive preferences and growth. *Journal of Economic Behavior and Organization* 49, 381–404.
- Ireland, P., 1996. The role of countercyclical monetary policy. *Journal of Political Economy* 104, 704–723.
- Kam, E., 2005. A note on time preference and the Tobin effect. *Economics Letters* 89, 127–132.
- Katona, G., 1975. *Psychological Economics*. Elsevier, New York.
- Keynes, M.J., 1936. *The General Theory of Employment, Interest, and Money*. A Harvest/HBJ Book. Harcourt Brace Jovanovich, Publishers.
- Obstfeld, M., 1981. Macroeconomic policy, exchange-rate dynamics, and optimal asset accumulation. *Journal of Political Economy* 89, 1142–1161.
- Obstfeld, M., 1990. Intertemporal dependence, impatience, and dynamics. *Journal of Monetary Economics* 26, 45–75.
- Shiller, R., 1996. *Why Do People Dislike Inflation?* NBER Working Paper.
- Sidrauski, M., 1967. Rational choice and patterns of growth in a monetary economy. *The American Economic Review* 57, 534–544.
- Stockman, A., 1981. Anticipated inflation and the capital stock in a cash-in-advance economy. *Journal of Monetary Economics* 8, 387–393.
- Tobin, J., 1965. Money and economic growth. *Econometrica* 33, 671–684.
- Uzawa, H., 1968. Time preference, the consumption function, and optimal asset holdings. In: Wolfe, J.N. (Ed.), *Value, Capital and Growth: Papers In Honor of Sir John Hicks*. University of Edinburgh Press.
- Wang, P., Yip, C., 1992. Alternative approaches to money and growth. *Journal of Money, Credit and Banking* 24, 553–562.

Inflation Aversion

Heng-fu Zou

*Central University of Finance and Economics
Shenzhen University
Wuhan University
Peking University*

Liutang Gong

Guanghua School of Management, Peking University

and

Xinsheng Zeng^{*}

Wuhan University

With inflation aversion, an increase in the monetary growth rate decreases the steady-state value of capital stock, consumption, and real balance holding.

Key Words: Inflation aversion; Capital accumulation ; Money.

JEL Classification Numbers: D91, E63, O23.

1. INTRODUCTION

In optimal growth models, positive time preferences, i.e. people systematically discount utility derived from future consumption, is taken for granted [Olson and Bailey (1981)]. In addition, a further modeling of time preferences by Uzawa (1968), where the time preference is an increasing function of current utility, has been increasingly used in growth and asset accumulation models. See Obstfeld (1981), (1982) and (1989) among others. While most economists do not question the existence of positive

^{*} A very preliminary version of inflation aversion in this paper was developed by Xinsheng Zeng and Heng-fu Zou in 1993. Please see Xinsheng Zeng's Ph.D. thesis at Boston University, 1993.

time preference, they have raised some doubt about Uzawa's assumption. For example, Blanchard and Fischer(1989) state that, for Uzawa's specification, "in steady state, a higher level of consumption implies a higher rate of time preference. The assumption is difficult to defend a priori; indeed, we usually think it is the rich who are more likely to be patient. ... The Uzawa function ... is not particularly attractive as a description of preferences and is not recommended for general use" (pp.72-75).

Here in a monetary growth model, we attempt to establish a link between the time preference and inflation. We will demonstrate that, if there exists positive time preference in the real world, inflation is an important element in determining its magnitude. The argument for defining the time preference as an increasing function of inflation is much stronger than for Uzawa's specification. With this new definition of time discount rate, we will show that inflation reduces long-run capital accumulation, consumption and real balance holdings.

2. ENDOGENOUS TIME PREFERENCE AND INFLATION AVERSION

While the time preference depends on various factors in a society, inflation is an important factor leading to social and economic instability and disorder. Thus, it is convincing to define the time preference, denoted as δ , to be an increasing function of expected inflation rate, denoted as π . Namely, $\delta = \delta(\pi)$, $\delta'(\pi) \geq 0$, and $\delta''(\pi) < 0$. Obviously enough we call this definition as inflation aversion. We first present the theory of endogenous time preference in Rae (1834), Bohm-Bawerk (1959), Fisher (1930), and then argue why this definition is reasonable.

The time preference theory has its direct origin in Rae (1834) and Bohm-Bawerk (1959); for this, Irving Fisher dedicated The Theory of Interest (1930) in their memory. Rae calls the time preference as the effective desire of accumulation, which has the following definition:

"The determination to sacrifice a certain amount of present good, to obtain another greater amount of good, at some future period, may be termed the effective desire of accumulation. All men may be said to have a desire of this sort, for all men prefer a greater to a less; but to be effective it must prompt to action." (Rae, 1834, p.119)

What determines this effective desire of accumulation? Rae mainly lists the following three elements: "1. The prevalence throughout the society, of the social and benevolent affections... 2. The extent of the intellectual powers, and the consequent prevalence of habits of reflection, and prudence, in the minds of the members of the society. 3. The stability of the condition of the affairs of the society, and the reign of law and order throughout it." (Rae, 1834, *New Principles of Political Economy*, pp.124.) In this list,

Rae does not say anything about how does inflation influence the time preference. But it is quite clear from Rae's long discussion on "the social and benevolent affection" and the desire of the social stability in the moral sense that he admits the role of "inflation aversion" in strengthening the effective desire of accumulation in modern time. When he takes "the money-making spirit" as the main element of the social affection and the instability of the society, he says "(the love of) money is the root of all evil, and infallibly leads to wickedness", "these feelings, therefore, investing the concerns of futurity with a lively interest to the individual, and giving a continuity to the existence and projects of the race, must tend to strengthen very greatly the effective desire of accumulation."

The time preference theory was fully developed by Eugen von Bohm-Bawerk. Indeed, Olson and Baily (1981) are right, "the clearest conception of positive time preference that we have been able to find was in Bohm-Bawerk's original account." According to Bohm-Bawerk,

"we feel less concerned about future sensation of joy and sorrow simply because they do lie in the future, and the lessening of our concern is in proportion to the remoteness of that future. Consequently we accord to goods which are intended to serve future ends a value which falls short of the true intensity of their future marginal utility. We systematically undervalue our future wants and also the means which serve to satisfy them. That is a fact of that there can be no doubt." (Bohm-Bawerk, *Capital and Interest*, Vol. II, p.268).

Bohm-Bawerk provides three causes for this positive time preference: (1) "the fragmentary nature of the imaginary picture that we construct of the future state of our wants" (p.269); (2) "a failure of will power and losing control over ourselves in facing immediate enjoyment" (p.269); and (3) "consideration of the brevity and uncertainty of human life." (p.270)

The causes listed by Bohm-Bawerk in determining positive time preference manifest themselves fully in an inflationary world. First, inflation leads to economic, social and institutional uncertainty, and causes "disillusionment and discontent" (Arthur Burns) in the society and "strikes at confidence" (John Maynard Keynes) of the people. All those uncertainty and anxiety, of course, result in more impatience and larger time discount rate. We cite two excellent quotations of Keynes and Burns from Fabricant (1976).

For John Maynard Keynes,

"There is no subtler, no sure means of overturning the existing basis of society than to debauch the currency. The process engages all the hidden forces of economic law on the side of destruction, and does it in a manner which not one man in a million is able to diagnose ... [The] arbitrary arrangement of riches [caused by inflation] strikes not only at security but at confidence in the equity of the existing distribution of wealth ... All

permanent relations between debtors and creditors, which form the ultimate foundation of capitalism, become so utterly disordered as to be almost meaningless; and the process of wealth-getting degenerates into a gamble and a lottery." (Keynes, 1919, *The Economic Consequences of the Peace*, pp. 235-248.)

Chairman Arthur Burns of the Federal Reserve Board warns "the menace of inflation":

"Concerned as we all are about the economic consequences of inflation, there is even greater reason for concern about the impact on our social and political institution. We must not risk the social stress that persistent inflation breeds. Because of its capricious effects on the income and wealth of a nation's families and businesses, inflation inevitably causes disillusionment and discontent. ... Discontent bred by inflation can provoke profoundly disturbing social and political change, as the history of other nations teaches. I do not believe I exaggerate in saying that the ultimate consequence of inflation could well be a significant decline of economic and political freedom for the American people."

Secondly, while high inflation makes rational economic calculation more difficult or impossible, it makes people possess even less "adequate power to imagine and to abstract" the future world. This corresponds cause (1) of positive time preference pointed out by Bohm-Bawerk. Indeed expected high inflation often leads people to perceive the future in dark color, and people may enjoy more today at the sacrifice of future consumption. This weakness of human will in facing high inflation results in large time discount rate.

Thirdly, the obvious thing about inflation is that with high inflation, even if their real income has kept constant or increased substantially, people still feel being cheated and psychologically they regard inflation as a "bad thing" [Katona (1975)]. This psychological "irrationality" has been proved again and again in experience. According to Katona (1980), in America "there is no doubt that most people consider inflation an evil. In the late 1970s many more Americans said that inflation was the most serious problem confronting them. When asked which causes more serious hardship, inflation or unemployment, about two-thirds of the respondents in 1979 named inflation and one-fourth mentioned unemployment. This despite the fact that ... very many Americans did not feel hurt by inflation" (p.81). Recent experience in China provides alarming signal about how dangerous the high inflation would be, even accompanied by rapid income growth in the decade of economic reforms. People suffer most psychologically from inflation, and if the future is an inflation world, there is no way to stop people from discounting the future heavily.

To sum up, the assumption of time preference as a positive function of expected inflation is quite convincing to us. The only thing we feel

strange is why this inflation aversion approach has not been widely used in monetary growth and asset accumulation models.

3. THE INFLATION AVERSION MODEL AND ITS RESULTS

A representative family, whose size grows at natural rate n , maximizes the discounted utility over an infinite horizon ,

$$W = \int_0^\infty u(c, m) e^{-\int_0^t \delta(\pi(s)) ds} dt \quad (1)$$

where c is per-capita consumption, m is per-capita real balances holding, π is the expected inflation rate, and δ is the time preference generation function of inflation rate with following properties: $0 < \delta_{\min} \leq \delta(\pi) < 1$, $\delta'(\pi) \geq 0$, and $\delta''(\pi) < 0$, and there exists an inflation rate such that δ retains its minimum $\delta_{\min}$. The instantaneous utility function u is increasing, concave, and continuously differentiable in c and m , namely

$$u_c > 0, u_m > 0, u_{cc} < 0, u_{mm} < 0.$$

Output is produced by a standard neoclassical production function, $f(k)$, and k is per capita capital stock, and $f'(k) > 0$, $f''(k) < 0$.

There are two assets in the representative family's portfolios: money and capital. The dynamic budget constraint for the family is

$$\dot{a} = f(k) + x - c - nk - (\pi + n)m \quad (2)$$

$$a = k + m \quad (3)$$

where a is the total asset, and x is real transfer from the government; and a dot over a variable denotes time derivative.

Under the budget constraints (2), (3), and the given initial capital stock $k(0)$ and $m(0)$, the representative agent is to select the money holding path, capital stock accumulation path, and the consumption level to maximize the discounted utility expressed in equation (1).

Let

$$\Delta_t = \int_0^t \delta(\pi(s)) ds \quad (4)$$

and so,

$$d\Delta_t = \delta(\pi(t))dt \quad (4')$$

Upon substitution, equations (1) and (2) can be written as

$$W = \int_0^\infty \frac{u(c, m)}{\delta(\pi)} e^{-\Delta_t} dt \quad (5)$$

$$\frac{da}{d\Delta} = \frac{f(k) + x - c - nk - (\pi + n)m}{\delta(\pi)} \quad (6)$$

Thus the optimization problem can be reduced to maximize the function in equation (5) subject to the constraints (6) and (3), with the initial capital stock $k(0)$ and real balances holding $m(0)$ are given.

Associated with the optimization problem, the Hamiltonian is defined as

$$H = \frac{u(c, m)}{\delta(\pi)} + \nu_1 \frac{f(k) + x - c - nk - (\pi + n)m}{\delta(\pi)} + \nu_2(a - k - m) \quad (7)$$

where ν_1 is the Hamilton multiplier associated with the equation (6), and ν_2 is the Lagrange multiplier associated with the wealth constraint (3).

The necessary conditions for optimization are

$$u_c = \nu_1 \quad (8)$$

$$\nu_1 \frac{f'(k) - n}{\delta(\pi)} = \nu_2 \quad (9)$$

$$\frac{u_m(c, m)}{\delta(\pi)} - \nu_1 \frac{\pi + n}{\delta(\pi)} = \nu_2 \quad (10)$$

$$\frac{d\nu_1}{d\Delta} = \nu_1 - \nu_2 \quad (11)$$

and the transversality condition $\lim_{\Delta \rightarrow \infty} e^{-\Delta_t} a \nu_1 = 0$.

Substituting equations (8) and (9) into equations (10) and (11), and using equation (4'), we get

$$u_m = u_c(f'(k) + \pi) \quad (12)$$

$$\frac{du_c}{dt} = u_c(\delta(\pi) + n - f'(k)) \quad (13)$$

By the definition of the per-capita real balance, we get the accumulation equations for the real balance

$$\dot{m} = m(\theta - \frac{\dot{p}}{p} - n) \quad (14)$$

where θ is the growth rate of nominal money supply, p is the price level.

On the perfect foresight path, the expected inflation rate is equal to the actual inflation rate:

$$\frac{\dot{p}}{p} = \pi \quad (15)$$

Government revenue comes from money creation and makes transfer, x , to the representative agent, so, we have

$$x = \theta m \quad (16)$$

Summarizing the discussion above, the full dynamics of the economy can be described by

$$\dot{c} = -\frac{u_{cm}}{u_{cc}}[f'(k) + \theta - n - \frac{u_m}{u_c}]m + \frac{u_c}{u_{cc}}[\delta(\frac{u_m}{u_c} - f'(k)) + n - f'(k)] \quad (17)$$

$$\dot{m} = m(\theta - \frac{u_m}{u_c} + f'(k) - n) \quad (18)$$

$$\dot{k} = f(k) - nk - c \quad (19)$$

Using equations (17), (18), and (19), we can analyze the dynamic characters of consumption, capital stock, and the real balance, and from equation (12), we can determine inflation rate.

4. LONG-RUN EFFECTS

In this section, we analyze the long-run effects of money growth rate on the economy¹. The steady-state value (c^*, m^*, k^*) of consumption, real balances, and the capital stock, reaches when $\dot{c} = \dot{m} = \dot{k} = 0$, hence,

$$f'(k^*) = \delta(\theta - n) + n \quad (20)$$

¹Appendix A shows that the dynamic system of economy (17)-(19) is saddle-point stable.

$$\frac{u_m(c^*, m^*)}{u_c(c^*, m^*)} = \theta + f'(k^*) - n \quad (21)$$

$$f(k^*) - nk^* - c^* = 0 \quad (22)$$

Equation (20) says that optimal long-run capital is determined by equating the marginal productivity of capital to the sum of discount rate and population growth; equation (21) is the optimal condition for money holding: marginal rate of substitution between real balances and consumption is equal to the ratio of the cost of money holding, $\theta + f'(k^*) - n = f'(k^*) + \pi$, over the cost of consumption, which is one. Except for the dependence of time preference on inflation rate, all these steady state conditions are identical to Sidrauski's (1967).

From (20), it is very easy to see that increase in inflation reduces long-run capital stock:

$$\frac{dk}{d\theta} = \frac{\delta'(\theta - n)}{f''(k^*)} < 0. \quad (23)$$

The reason is quite simple, high inflation leads to more impatience, and the representative family discounts further the future consumption and increase its current consumption; in the end, saving and capital stock will be reduced in the new equilibrium.

The effect of inflation on long-run consumption is negative:

$$\frac{dc}{d\theta} = (f'(k^*) - n) \frac{dk}{d\theta} = \delta \delta'(\theta - n) \frac{dk}{d\theta} < 0. \quad (24)$$

Here we have used steady state condition (20) to get the second equality in (24). The reason for this result is following: in the long run, optimal capital stock is determined by the modified golden rule, and is less than the golden rule level; and there does not exist dynamic inefficiency (overaccumulation of capital) in this economy. Hence any reduction in the capital stock caused by inflation leads to reduction in consumption.

The effect of inflation on real balance holdings is also negative:

$$\frac{dm}{d\theta} = \frac{u_{cc} - u_{cm}}{u_{mm} - u_{cm}} \frac{dc}{d\theta} + \frac{u_c(\delta'(\theta - n) + 1)}{u_{mm} - u_{cm}} \quad (25)$$

Both terms on the right hand side of (25) are negative, for u_{cm} is positive and u_{cc} , u_{mm} are negative. Real balances are reduced because money is more costly to hold (substitution effect), and income is lower as a result of reduced capital (income effect).

Therefore inflation is an "evil" which brings about a high time discount rate and low instantaneous utility by reducing both consumption and real

balances in the long run. if the government intends to maximize the steady state welfare of the representative family, the simple rule is to choose an inflation rate which minimizes the time discount rate. In this case, Milton Friedman's (1969) rule may not be right as consumption and capital accumulation, unlike the original Sidrauski's model, are decreasing function of inflation, and

$$\frac{du}{d\theta} = u_c \frac{dc}{d\theta} + u_m \frac{dm}{d\theta} \quad (26)$$

If the minimum of time discount rate, $\delta_{\min}$, is obtained before the deflation rate reaches " $-f'(k)$ ", it is still desirable to deflate further following Friedman's prescription, $\theta = -\delta(\theta - n)$, the optimal growth rate of money supply equals the inverse of time preference.

5. CONCLUDING REMARKS

We believe that inflation aversion will be widely used in monetary economies and international finance.

APPENDIX A

The Stability of the Model

Linearizing (17), (18) and (19) around the steady state values:

$$\begin{pmatrix} \dot{c} \\ \dot{m} \\ \dot{k} \end{pmatrix} = A \begin{pmatrix} c - c^* \\ m - m^* \\ k - k^* \end{pmatrix} \quad (A.1)$$

where

$$A = \begin{pmatrix} \kappa J_1 & \kappa J_2 & -f''(\frac{u_c}{u_{cc}}(\delta' + 1) + \frac{u_{cm}}{u_{cc}}m) \\ -mJ_1 & -mJ_2 & mf''(k) \\ -1 & 0 & \delta \end{pmatrix}$$

and $\kappa = (\frac{u_c}{u_{cc}}\delta' + \frac{u_{cm}}{u_{cc}}m)$, $J_1 = \frac{u_{mc}u_c - u_m u_{cc}}{u_c^2}$, and $J_2 = \frac{u_{mm}u_c - u_m u_{cm}}{u_c^2}$. It is assumed that $J_1 > 0$ and $J_2 < 0$.

Denote the 3×3 matrix as A and denote the three characteristic roots as λ_1 , λ_2 and λ_3 . It is known that the product of the three characteristic roots of the system is given by the determinant of matrix A , and the sum of the three roots is given by the trace of A . We first calculate the determinant of A , $\det(A)$:

$$\lambda_1 \lambda_2 \lambda_3 = \det(A) = m \frac{u_c}{u_{cc}} f''(k) J_2 < 0 \quad (A.2)$$

So the system has either one negative root or three negative roots. The trace of A does not give us clear sign:

$$\lambda_1 + \lambda_2 + \lambda_3 = \text{tr}(A) = \kappa J_1 - m J_2 + \delta \quad (\text{A.3})$$

where the second and the third terms on the right hand side are positive, but the first term is negative, given $\kappa < 0$, $J_1 > 0$ and $J_2 < 0$.

We also know that the sum of the three second order principal minors

$$\begin{aligned} & \lambda_1 \lambda_2 + \lambda_1 \lambda_3 + \lambda_2 \lambda_3 \\ = & \det \begin{pmatrix} \kappa J_1 & \kappa J_2 \\ -m J_1 & -m J_2 \end{pmatrix} + \det \begin{pmatrix} -m J_2 & m f''(k) \\ 0 & \delta \end{pmatrix} \\ & + \det \begin{pmatrix} \kappa J_1 & -f''(\frac{u_c}{u_{cc}}(\delta' + 1) + \frac{u_{cm}}{u_{cc}}m) \\ -1 & \delta \end{pmatrix} \\ = & -\delta m J_2 + \delta \kappa J_1 - f''(\kappa + \frac{u_c}{u_{cc}}) < 0 \end{aligned} \quad (\text{A.4})$$

This is because that the third term on the right hand side of the second equality of (A.4) are negative, and the sum of the first term and the second term is also negative. To see the latter, the sum of these two terms is given by

$$-\delta m J_2 + \delta \kappa J_1 = \frac{u_c}{u_{cc}} \delta' \delta J_1 + (J_1 - u_{cc} J_2 / u_{cm}) m u_{cm} \delta / u_{cc} \quad (\text{A.5})$$

In (A5), we only need to show that the term in the bracket is positive:

$$(J_1 - u_{cc} J_2 / u_{cm}) = \frac{1}{u_c^2 u_{cm}} (u_{mc} u_{cm} - u_{cc} u_{mm}) u_c > 0 \quad (\text{A.6})$$

From (A.2) and (A.4), it is very easy to see that there exist only one negative root because the existence of three negative roots will contradict (A.4). Our system has one state variable k and two jumping variables (c and m), so there exists a unique perfect foresight path converging to the steady state.

REFERENCES

- Orphanides, A. and Solow, R., 1990. Money, inflation and growth. In *Handbook of Monetary Economics*, ed. by B.M. Friedman and F.H. Hahn, North-Holland.
- Blanchard, O. J. and Fischer S., 1989. *Lecture on Macroeconomics*. MIT Press.
- Fabricant, S., 1976. Economic calculation under inflation: the problem in perspective. In *Economic Calculation under Inflation*, Liberty Press.

- Fisher, I., 1930. *The Theory of Interest*. The Macmillan Company, New York.
- Fischer, S., 1979. Capital accumulation on the transition path in a monetary optimization model. *Econometrica* **47**, 1433-1439.
- Friedman, M., 1969. The optimum quantity of money. In M. Friedman, *The Optimum Quantity of Money and Other Essays*. Chicago, Aldine.
- Katona, G., 1975. *Psychological Economics*. New York: Elsevier.
- Katona, G., 1980. *Essays on Behavioral Economics*. The University of Michigan.
- Gong, L. and Zou, H., 2001. Money, social status, and capital accumulation in a cash-in-advance model. *Journal of Money, Credit, and Banking* **33**, 284-293.
- Obstfeld, M., 1981. Macroeconomic policy, exchange-rate dynamics, and optimal asset accumulation. *Journal of Political Economy* **89**, 1142-61.
- Obstfeld, M., 1982. Aggregate spending and the terms of trade: Is there a Laursen-Metzler effect? *Quarterly Journal of Economics* **97**, 251-270.
- Obstfeld, M., 1989. Intertemporal dependence, impatience, and dynamics. NBER Working Paper No. 3028.
- Olson, M. and Baily, M. 1981. Positive time preference. *Journal of Political Economy* **89**, 1-25.
- Rae J., 1964. *New Principles of Political Economy*. (1st edition, 1834), reprinted by Augustus M. Kelly, Publisher.
- Sidrauski, M., 1967. Rational choice and patterns of growth in a monetary economy. *American Economic Review, Papers and Proceedings* **57**, 534-544.
- Simonsen, M. H. and R. P. Cysne, 2001. Welfare costs of inflation and interest-bearing money. *Journal of Money, Credit, and Banking* **33**, 90-100.
- Uzawa, H., 1968. Time preference, the consumption function, and optimum asset holdings. In J. N. Wolfe, ed., *Capital and Growth: Paper in Honor of Sir John Hicks*. Chicago: Aldine.
- von Bohm-Bawerk, E. 1959. *Capital and Interest*, Vol. II. Libertarian Press.

Inflation Aversion and the Optimal Inflation Tax

Gaowang Wang

School of Economics and Management, Wuhan University, Wuhan, 430072, China
E-mail: wanggaowang@gmail.com

and

Heng-fu Zou^{*}

CEMA, Central University of Finance and Economics, Beijing
China School of Advanced Study (SAS), Shenzhen University, Shenzhen
IAS, Wuhan University
GSM, Peking University
E-mail: zouhengfu@gmail.com

The optimal inflation tax is reexamined in the framework of dynamic second best economy populated by individuals with inflation aversion. A simple formula for the optimal inflation rate is derived. Different from the literature, it is shown that if the marginal excess burden of other distorting taxes approaches zero, Friedman's rule for optimum quantity of money is not optimal, and the optimal inflation tax is negative; if the marginal excess burden of other taxes is nonzero, the optimal inflation rate is indeterminate and relies on the tradeoffs between the impatience effect of inflation and the effects of other economic forces in the monetary economy.

Key Words: Inflation aversion; Optimal inflation tax; Second best taxation; The friedman rule.

JEL Classification Numbers: E31, E41, E52, H21.

^{*} Corresponding author.

1. INTRODUCTION

The literature on the optimal inflation tax¹ has drawn two different conclusions roughly: the inflation tax (i.e., the nominal interest rate in the paper) should be positive or zero. In the beginning, the result of the zero inflation tax is in the first best environment and the result of the positive inflation tax is in the second best one. In a first best environment where lump-sum taxes are available, Friedman (1969) proposes a monetary policy rule that generates a zero nominal interest rate, corresponding to a zero inflation tax and to a negative rate of inflation. Sidrauski (1967) and Turnovsky & Brock (1980) have also produced the result of the zero nominal interest rate in the first best framework. And in a framework of second best taxation, Chamley (1985a) proves that when the marginal excess burden of other distorting taxes approaches zero, the model degenerates as a first best one, and the optimal inflation tax is zero. By optimizing the inflation rate together with other distortionary taxes and exogenous factor prices, Phelps (1973) argues that “the optimal inflation tax is positive” and Friedman’s rule is unlikely to be optimal in an economy without lump-sum taxes. Chamley (1985a) extends Phelps (1973) to a general equilibrium model with capital and draws the same conclusion under the condition that the marginal excess burden of other distorting taxes is nonzero. The intuition for these studies is based on the assumption that money is a consumption good. In the framework of first best, based on the rule for the equality of marginal benefit and marginal cost, the nominal interest rate should be zero, because the cost of supplying money is negligible. And in the second best framework with distorting taxes, money is a consumption good that should be taxed, just as other consumption goods, based on the theorem of uniform taxation derived by Atkinson and Stiglitz (1972). That is, the optimal inflation tax is positive, since inflation is the method of taxing cash balances by printing money.

However, many studies have proved the validity of the Friedman rule in the framework of second best. Chari, Christiano & Kehoe (1996) and Chari and Kehoe (1999) establish that if the utility function satisfies a few simple homotheticity and separability conditions, the Friedman rule is optimal in three standard monetary economies (a cash-credit model, a

¹There are several different measures of the inflation tax. Friedman (1948) and Bailey (1956) identified the inflation tax revenue as the rate of inflation multiplied by the real value of the (outside) quantity of money, $\pi M/p$; Marty (1967, 1973) proposed to measure the inflation tax by the rate of growth of the money supply times real balances, $(\pi + g)M/p$, where g is the real growth rate; Friedman (1971) endorsed the total inflation tax as the money-supply growth rate times real balances, $(\dot{M}/M)M/p$; Phelps (1972, 1973) and Correia and Teles (1999) used the nominal interest rate multiplied by real balances, $(\pi + r)M/p$, where r is the real interest rate. And the paper follows the last one.

money-in-utility-function model, and a shopping time model) with distorting taxes. Correia and Teles (1996) show that the Friedman rule is the optimal solution in those monetary models with homogeneous transactions cost functions; furthermore, Correia and Teles (1999) argue that the Friedman rule is a general result in the set-up where liquidity is modeled as a final good. In an economy with heterogeneous agents subjecting to non-linear taxation of labor income, Da Costa and Werning (2008) find that the Friedman rule is optimal when combined with a nondecreasing labor income tax. These studies present different sufficient conditions for the optimality of the Friedman rule in the monetary economy with distorting taxation, in contrast to the results of a positive inflation tax derived by Phelps (1973) and Chamley (1985a) in a second best framework. In the paper, we draw the conclusion that the optimal inflation tax is indeterminate, and it relies on the particular environment, just as what Siegel (1978) had stressed the indeterminacy of the optimal tax structure in the general equilibrium framework and what Drazen (1979) had stated that it appeared difficult to say even whether the optimal inflation rate would be positive or negative.

In our opinion, the consistency of the literature comes from the simplified assumption that money is just an ordinary consumption good. Actually, money is a kind of commodity whose production is executed by government monopolistically in most of the nations. The revenue from the creation of money belongs to government, and the excess levy of the inflation tax would activate the printing of money rather than discourage it. Moreover, inflation is a common phenomenon closely relating to our everyday lives and tends to impair the patience and confidence of the people². Following Zou, Gong and Zeng (2011) and Wang and Zou (2011), the paper conceptualizes the important psychological effect of inflation as “inflation aversion” and examines its effect on the optimal inflation tax. With inflation aversion in our model, we need to consider the following tradeoffs: (1) the cost-benefit analysis of money being a production good, (2) the efficiency cost of other distorting taxes and the impatience cost of inflation, (3) the revenues of money creation and the psychological cost of inflation, (4) the utility effect of money and the impatience effect of inflation, and (5) the holdings of money and other financial assets. Fortunately enough, a simple formula for the optimal inflation rate is derived, even with so many tradeoffs. Different from the literature, it is shown that if the marginal excess burden of other distorting taxes approaches zero, then Friedman’s rule for optimum quantity of money is optimal and the optimal inflation tax is negative; if the marginal excess burden of other taxes is nonzero, the sign of the nom-

²Many economists have studied the economic and psychological costs of inflation, such as Bohm-Bawerk (1891), Keynes (1936), Katona (1975), Fabricant (1976), Burns (1978), and Shiller (1996).

inal interest rate is indeterminate and relies on the particular economic tradeoffs of the monetary economy.

The paper is organized as follows. Section 2 lays down a second best monetary model with inflation aversion and with separability between consumption and money and it derives the main results of the paper. In Section 3, the simple model is generalized to the case with a nonseparable utility function. The concluding remarks are presented in section 4.

2. THE DYNAMIC MODEL WITH INFLATION AVERSION

2.1. The Model with Separable Utility Functions

Following the inflation aversion concept in Zou, Zeng and Gong (2011) and Wang and Zou (2011), it is assumed that the time preference rate of the representative individual is a strictly increasing and concave function of the current expected inflation rate, namely,

$$\rho_t = \rho(\pi_t), \rho'(\pi_t) > 0, \rho''(\pi_t) < 0, \quad (1)$$

which imply that the patience of an individual changes with inflation; and the higher the inflation, the less patient the individual is. Correspondingly, the time discount factor of time t , Δ_t , is an implicit function of the entire orbit of the past expected inflation rate, i.e., $\Delta_t = \int_{s=0}^t \rho(\pi_s) ds$, whose derivative is

$$\dot{\Delta}_t = \rho(\pi_t). \quad (2)$$

Let us first consider the case of the separability between consumption c , labor l , and money m : $\tilde{u}(c, l, m) = U(c, l) + v(m)$, where both $U(c, l)$ and $v(m)$ are concave. The objective function of the representative individual is

$$\int_{t=0}^{\infty} e^{-\Delta_t} [U(c_t, l_t) + v(m_t)] dt. \quad (3)$$

All quantities are measured per capita. The total financial assets of the individual a_t are allocated among capital k_t , bonds b_t , and real money balances m_t :

$$a_t = k_t + b_t + m_t. \quad (4)$$

Output is produced with the standard neoclassical production technology utilizing two inputs, capital k_t and labor l_t : $y_t = f(k_t, l_t)$. The gross factor prices are determined by the marginal productivities:

$$r_t = f_k(k_t, l_t), w_t = f_l(k_t, l_t). \quad (5)$$

Endowed with perfect foresight, the representative individual takes these competitive factor prices as given.

The government finances an exogenous stream of public consumption by a labor tax and the creation of fiat money. If the flow of receipts and expenditures does not coincide in the efficient solution, the government issues or trades bonds between different instants at the interest rate r_t . Since there is no uncertainty, bonds are perfectly substitutable with capital and have the same rate of return r_t . In the second-best framework, the initial level of the debt, b_0 , must be taken as exogenously given. Therefore, the variations of the debt or the budget constraint of government is

$$\dot{b}_t = r_t b_t + g_t - (w_t - \bar{w}_t)l_t - (\dot{m}_t + \pi_t m_t), \quad (6)$$

where $\bar{w}_t$ represents the net wage rate ($(w_t - \bar{w}_t)$ can be seen as the labor tax rate), g_t is the level of public consumption, and $\dot{m}_t + \pi_t m_t$ is the level of revenues generated by the creation of money. Setting the growth rate of money as a constant, θ , we have $m_t = \frac{M_t}{P_t}$ and $\dot{m}_t = (\theta - \pi_t)m_t$. The problem of the government is to determine the policies of taxation and inflation which optimize the individual's utility subject to the government's budget constraint and the feasibility constraint of the economy.

2.2. The Problem of Second Best

In the standard second-best problem, the policy maker has to take into account the constraints implied by the optimizing behavior of the private sector. The representative individual's problem is to maximize (3), subject to (2), (4), and his budget constraint

$$\dot{a}_t = r_t a_t + \bar{w}_t l_t - (r_t + \pi_t)m_t - c_t, \quad (7)$$

taking $\{r_t, \bar{w}_t, g_t\}_{t=0}^{\infty}$ and a_0 as given.

To proceed, the Hamiltonian is

$$\begin{aligned} H = & e^{-\Delta t} \{U(c_t, l_t) + v(m_t) + q_t[r_t a_t + \bar{w}_t l_t - (r_t + \pi_t)m_t - c_t] \\ & - \kappa_t \rho(\pi_t) + \eta_t(a_t - k_t - b_t - m_t)\}, \end{aligned}$$

where q_t and $-\kappa_t$ are two Hamiltonian multipliers associated with the private budget constraint and the dynamic accumulation equation of the time discount factor, representing the marginal utility of the accumulated assets and time discount rate, respectively; and η_t is the Lagrangian multiplier associated with the stock constraint, representing the marginal value of the stock asset.

The first-order conditions of this optimization are as follows:

$$U_c(c_t, l_t) = q_t, \quad (8)$$

$$U_l(c_t, l_t) = -q_t \bar{w}_t, \quad (9)$$

$$v'(m_t) = q_t(r_t + \pi_t), \quad (10)$$

$$\dot{q}_t = [\rho(\pi_t) - r_t]q_t. \quad (11)$$

The first two equations correspond to the familiar intratemporal first-order conditions for consumption and leisure. The third equation determines the optimal level of cash balances, and the fourth equation is the intertemporal condition of optimality.³

Using equations (8) and (9), c and l can be replaced as functions of q and $\bar{w}$:

$$c = c(q, \bar{w}), l = l(q, \bar{w}), \quad (12)$$

and hence

$$U(c, l) = u(q, \bar{w}). \quad (13)$$

From equation (10), the demand for cash balances depends only on q and the nominal interest rate i ($= r + \pi$), $m = \psi(q, i)$. Since the real interest rate r depends on the input levels k and l , and the labor supply l is a function of q and $\bar{w}$ in (12), the demand for real money balances can also be expressed as a function of k , q , $\bar{w}$, and π , namely,

$$m = \phi(k, q, \bar{w}, \pi), \quad (14)$$

which is the money demand function of the representative individual essentially. And it is easy to know that $\frac{\partial \psi(q, i)}{\partial i} = \frac{\partial \phi(k, q, \bar{w}, \pi)}{\partial \pi}$.⁴ For simplicity, it is assumed that the initial levels of the endogenous variables $P_0, M_0, k_0, b_0, q_0, \bar{w}_0$, and π_0 are exogenously given.

Differentiating equation (14) with respect to time t results in

$$(\theta - \pi_t)m_t = \dot{m}_t = \phi_k \dot{k} + \phi_q \dot{q} + \phi_{\bar{w}} \dot{\bar{w}} + \phi_{\pi} \dot{\pi}. \quad (15)$$

Equation (15) shows that there is a one-to-one relation between the growth rate of money and the inflation rate in the steady state. Hence, although the government controls θ , it is equivalent to assume that government chooses π or $\dot{\pi}$.

³In the following sections of the paper, whenever convenient, the time subscripts will be omitted.

⁴Note that $\psi(q, f_k(k, l) + \pi) = \varphi(k, q, \bar{w}, \pi)$. Taking the partial derivative with respect to π gives $\frac{\partial \psi(q, i)}{\partial i} = \frac{\partial \phi(k, q, \bar{w}, \pi)}{\partial \pi}$.

The second-best problem can now be formulated as follows:

$$\max \int_{t=0}^{\infty} e^{-\Delta t} \{u(q, \bar{w}) + v[\phi(k, q, \bar{w}, \pi)]\} dt,$$

subject to

$$\dot{k} = f(k, l(q, \bar{w})) - c(q, \bar{w}) - g, \quad (16)$$

$$\begin{aligned} \dot{b} = & f_k(k, l(q, \bar{w}))b - (f_l(k, l(q, \bar{w})) - \bar{w})l(q, \bar{w}) \\ & + g - \phi_k \dot{k} - \phi_q \dot{q} - \phi_{\bar{w}} \dot{\bar{w}} - \phi_{\pi} \dot{\pi} - \pi \phi(k, q, \bar{w}, \pi), \end{aligned} \quad (17)$$

$$\dot{q}_t = [\rho(\pi_t) - f_k(k, l(q, \bar{w}))]q_t, \quad (18)$$

$$\dot{\Delta}_t = \rho(\pi_t). \quad (19)$$

$$\dot{\bar{w}} = x, \quad (20)$$

$$\dot{\pi} = z. \quad (21)$$

Equation (16) is the resource constraint of the economy, which is derived from equations (4)-(7). Equation (17) comes from equations (5), (6), (12), (14), and (15). Equation (18) is essentially the intertemporal optimality condition of the private individual (11). Equations (20) and (21) are the dynamic equations of the net wage rate and inflation by definition. In the problem, the initial values of the state variables $k_0, b_0, q_0, \Delta_0, \bar{w}_0, \pi_0$ are exogenously given. The controls of the problem are the paths of x and z .

The optimal solution is determined by the present value Hamiltonian

$$H = e^{-\Delta t} \left\{ \begin{aligned} & u(q, \bar{w}) + v(\phi(k, q, \bar{w}, \pi)) + (\lambda + \mu \phi_k)[f(k, l(q, \bar{w})) - c(q, \bar{w}) - g] \\ & - \mu[f_k(k, l(q, \bar{w}))b - (f_l(k, l(q, \bar{w})) - \bar{w})l(q, \bar{w}) + g - \pi \phi(k, q, \bar{w}, \pi)] \\ & (\xi + \mu \phi_q)[\rho(\pi) - f_k(k, l(q, \bar{w}))]q_t + (\alpha + \mu \phi_{\bar{w}})x + (\beta + \mu \phi_{\pi})z - \gamma \rho(\pi_t) \end{aligned} \right\},$$

where $\lambda, -\mu, \xi, -\gamma, \alpha$, and β are the Hamiltonian multipliers (or co-state variables) associated with equations (2), (16)-(21), representing the shadow prices of the five state variables $k, b, q, \Delta, \bar{w}$, and π , respectively. The variable λ represents the social marginal value of the unique good in the economy. In the second-best problem, λ is in general different from the private marginal value of the good, q . The variable $-\mu$ represents the social marginal value of the public debt, which is also equal to the marginal excess burden of taxation. It is assumed that there is a unique dynamic path which satisfies the optimality conditions of the second best problem and converges to a steady state.⁵

⁵The proof of stability of the steady state is very complex, but similar to Chamley (1985b, 1986), Zou, Gong, and Zeng (2011), and Wang and Zou (2011).

2.3. The Optimal Inflation Tax

Among those dynamic equations which define implicitly the optimal solution to the problem of second best, four of them characterizes more specifically the optimal inflation rate:

$$z : H_z = e^{-\Delta t}(\beta + \mu\phi_\pi) = 0, \quad (22)$$

$$k : H_k = -\frac{d}{dt}(e^{-\Delta t}\lambda) = \rho(\pi_t)e^{-\Delta t}\lambda + e^{-\Delta t}\dot{\lambda}, \quad (23)$$

$$b : H_b = -\frac{d}{dt}(e^{-\Delta t}\mu) = \rho(\pi_t)e^{-\Delta t}\mu + e^{-\Delta t}\dot{\mu}, \quad (24)$$

$$\pi : H_\pi = -\frac{d}{dt}(e^{-\Delta t}\beta) = \rho(\pi_t)e^{-\Delta t}\beta + e^{-\Delta t}\dot{\beta}. \quad (25)$$

Equation (22) leads to

$$\beta = -\mu\phi_\pi(k, q, \bar{w}, \pi), \quad (26)$$

and

$$\dot{\beta} = -\dot{\mu}\phi_\pi - \mu\phi_{\pi k}\dot{k} - \mu\phi_{\pi q}\dot{q} - \mu\phi_{\pi \bar{w}}\dot{\bar{w}} - \mu\phi_{\pi \pi}\dot{\pi} \quad (27)$$

Equation (24) gives rise to

$$\dot{\mu} = [\rho(\pi) - f_k(k, l(q, \bar{w}))]\mu. \quad (28)$$

Together with equation (18), equation (28) shows that the marginal excess burden measured in units of private consumption is constant over time, namely,

$$v = \frac{\mu}{q}. \quad (29)$$

Substituting equations (10), (26), (27) and (28) into equation (25) results in

$$q(r + \pi)\phi_\pi + \mu m + (\xi + \mu\phi_q)\rho'(\pi)q - \gamma\rho'(\pi) = -(r + \pi)\mu\phi_\pi. \quad (30)$$

Multiplying both sides of equation (30) by $-\frac{1}{qm}$ gives rise to

$$\varepsilon = \frac{1}{1+v} \left\{ v + \frac{1}{qm} [q(\xi + \mu\phi_q) - \gamma]\rho'(\pi) \right\},^6 \quad (31)$$

where ε is the interest elasticity of the demand for money, where

$$\varepsilon = -\frac{i}{m} \frac{\partial \psi(q, i)}{\partial i} = -\frac{r + \pi}{m} \phi_\pi(k, q, \bar{w}, \pi).$$

⁶If $\rho'(\pi) = 0$ in equation (30) or (32), it is the Chamley (1985a) model.

In the steady state, we have $\dot{k} = \dot{q} = \dot{\lambda} = x = z = 0$, and $\rho(\pi) = r$. Equation (23) gives

$$\xi = \frac{1}{f_{kk}}[(1+v)(r+\pi)\phi_k - vf_{kk}b + vf_{lk}l - \mu\phi_q f_{kk}]. \quad (32)$$

Substituting equation (32), $f_{lk}l + f_{kk}k = 0$ ⁷, and $\rho'(\pi)\phi_k = f_{kk}\phi_\pi$ ⁸ into equation (30), rearranging, we obtain the simple formula that determines the optimal inflation rate in the steady state,

$$\varepsilon = \frac{1}{2(1+v)} \left\{ v \left[1 - \left(\frac{1-\omega}{\omega} \right) \rho'(\pi) \right] - \frac{\gamma \rho'(\pi)}{qm} \right\}, \quad (33)$$

where $\omega = \frac{m}{k+b+m}$ is the share of money in total financial wealth in the steady state. Equation (33) establishes

PROPOSITION 1. *In the dynamic second best economy populated by individuals with inflation aversion, the optimal inflation rate is determined by formula (33). If the marginal excess burden of other distorting taxes is zero, i.e., $v = 0$ (or $\mu = 0$), then Friedman's rule for optimum quantity of money is not optimal.*

Similar to Chamley (1985a), a simple rule for the optimal inflation rate is derived. Different from Chamley (1985a), two new items emerge in the formula: one is the share of money in the total wealth, the other is the “inflation aversion”. The formula involves more economic factors than the literature.

If the marginal excess burden of other distorting taxes approaches zero, the second-best problem degenerates to the first-best problem. Friedman (1969), Sidrauski (1967), and Chamley (1985a) show that when lump-sum taxation is feasible, or, the marginal excess burden of other distorting taxation is zero, i.e., $v = 0$, the nominal interest rate is equal to zero ($i = r + \pi = 0$) and hence the Friedman rule is optimal. Different from their studies, in this paper, if $v = 0$, equation (33) degenerates into

$$\varepsilon = -\frac{\gamma \rho'(\pi)}{2qm} < 0, \quad (34)$$

which is not equal to zero. Although the marginal excess burden of other taxes is very small, the Friedman rule is still not optimal. Moreover, the

⁷This equation comes from the property of constant return to scale of the production function.

⁸In the steady state, this equation holds. The proof is in appendix A.

negative nominal interest rate shows that the optimal inflation tax is negative. That is to say, the government should subsidy the individuals for their holdings of money. In other words, the optimal monetary policy is deflating more deeply than Friedman's rule, which can be seen from the transformation of formula (34), namely, $\pi = -[r - \frac{\gamma\rho'(\pi)}{2q\phi_\pi}] < -r$.⁹ In order to show the reason, by setting $r' = r + \frac{-\gamma\rho'(\pi)}{2qm\phi_\pi}$, we have $r' > r$, $f'(k') > f'(k)$, and hence $k' < k$, for the strict concavity of the production function. Then, the logic is clear. With the decreasing patience for inflation aversion, individuals consume more and save less. Hence, the steady-state levels of capital and consumption will be decreased in the long run, and the long-run level of the real interest rate will be higher, $[r - \frac{\gamma\rho'(\pi)}{2q\phi_\pi}] > r$. Therefore, the optimal inflation rate will be more negative than the Friedman rule, i.e., $-[r - \frac{\gamma\rho'(\pi)}{2q\phi_\pi}] < -r$.

PROPOSITION 2. *If the marginal excess burden of other distorting taxes is finite, i.e., $v \in (0, \infty)$, then the optimal inflation tax is indeterminate. Specifically,*

- (i) *If $v > [\frac{(1-\omega)v}{\omega} + \frac{\gamma}{qm}]\rho'(\pi)$, the optimal inflation tax is positive, i.e., $i > 0$;*
- (ii) *If $v < [\frac{(1-\omega)v}{\omega} + \frac{\gamma}{qm}]\rho'(\pi)$, the optimal inflation tax is negative, i.e., $i < 0$;*
- (iii) *If $v = [\frac{(1-\omega)v}{\omega} + \frac{\gamma}{qm}]\rho'(\pi)$, the optimal inflation tax is zero, i.e., $i = 0$.*

The proof of the proposition is straightforward. However, it provides a more general framework than the literature, in which the sign of the nominal interest rate is indeterminate. First of all, if the marginal excess burden of other distorting taxes is larger than the impatience effect of inflation, i.e., $v > [\frac{(1-\omega)v}{\omega} + \frac{\gamma}{qm}]\rho'(\pi)$, then the nominal interest rate (or the inflation tax) is positive, i.e., $i > 0$. This positive nominal interest rate result is consistent with Phelps (1973) and Chamley (1985a). However, in our opinion, a better explanation for the positive inflation tax could be stated: it is the tradeoff between two different kinds of distorting taxes (the inflation tax and the income tax) by the government, which determines the positive inflation tax, rather than the theorem of uniform commodity taxation. In order to decrease the distortions from income taxation, the government levies an inflation tax to some degree. Correspondingly, the optimal inflation rate or optimal monetary growth rate is larger than the negative value of the time preference rate in the steady state, $\pi = -r +$

⁹Totally differentiating equation (10) gives rise to $\frac{dm}{dk} = \frac{qf''(k)}{v''(m)} \geq 0$, and $\frac{dm}{d\pi} = \frac{q}{v''(m)} \leq 0$.

$\frac{m\varepsilon}{-\phi_\pi} > -\rho$, since the equilibrium time preference rate is equal to the real interest rate in the steady state. In particular, if $r = \frac{m\varepsilon}{-\phi_\pi}$, the optimal inflation rate could be zero or positive. Secondly, if the marginal excess burden of other distorting taxes is equal to the impatience effect of inflation, i.e., $v = [\frac{(1-\omega)v}{\omega} + \frac{\gamma}{qm}]\rho'(\pi)$, then the nominal interest rate is zero, $i = 0$. That is to say, when these two opposite effects are balanced, Friedman's rule for optimum quantity of money is optimal. Compared to Phelps (1973) and Chamley (1985a), the positive inflation tax on money is offsetted by the negative effect of inflation. Hence, the optimal inflation tax is zero in our model. Finally, if the marginal excess burden of other distorting taxes is less than the impatience effect of inflation, i.e., $v < [\frac{(1-\omega)v}{\omega} + \frac{\gamma}{qm}]\rho'(\pi)$, then the nominal interest rate is negative, i.e., $i < 0$. That is, compared to the distorting effects of other taxes, the dominating impatience effect of inflation determines in the end that if government prints too much papers, it is optimal for government to subsidize the consumers for their holdings of money, i.e., $i < 0$. Then, the optimal strategy of the government is to reduce the supply of money even more than Friedman's rule. Hence, the optimal inflation rate is less than the negative value of the time preference rate $\pi = -r + \frac{m\varepsilon}{-\phi_\pi} < -\rho$. Therefore, the result derived in Proposition 1 can be looked upon as an example of Proposition 2.

PROPOSITION 3. *Assume that the excess burden of other distorting taxes approaches infinite, i.e., $v \rightarrow \infty$, and the impatience effect of inflation is finite, i.e., $\frac{\gamma}{qm}\rho'(\pi) < \infty$. Then,*

- (i) *if $\frac{\omega}{1-\omega} > \rho'(\pi)$, the optimal inflation tax is positive, i.e., $i > 0$;*
- (ii) *if $\frac{\omega}{1-\omega} < \rho'(\pi)$, the optimal inflation tax is negative, i.e., $i < 0$;*
- (iii) *if $\frac{\omega}{1-\omega} = \rho'(\pi)$, the optimal inflation tax is zero. Especially, if $\omega = \frac{1}{2}$, and the time preference function is affine, i.e., $\rho(\pi) = \pi + a$, where a is an arbitrary constant, then the optimal inflation tax is zero. Hence, Friedman's rule for optimum quantity of money is optimal.*

Proof. The proof is in appendix B. ■

It is defined that ω is the proportion of money in total financial assets in the steady state, i.e., $\omega = \frac{m}{k+b+m}$. It is appropriate to think of ω as the relative demand for money, $1 - \omega$ as the relative demand for other financial assets, and hence $\frac{\omega}{1-\omega}$ as the optimal ratio of the proportions of money and other financial assets in total nonhuman wealth. Since money is in utility and $U_m > 0$, the level of $\frac{\omega}{1-\omega}$ can be looked upon as the utility effect of money. And the higher level of $\frac{\omega}{1-\omega}$ stands for a higher demand for money and a stronger utility effect of money. Naturally, $\rho'(\pi)$ stands for the impatience effect of inflation. Then, it is easy to explain the proposition. If the utility effect of money dominates the impatience effect of inflation,

i.e., $\frac{\omega}{1-\omega} > \rho'(\pi)$, then the nominal interest rate is positive, $i > 0$. That is, if the impatience effect is small and the utility of money is large, it is optimal for government to levy a positive inflation tax. To see this, setting $r' = r + \frac{1}{\phi_\pi} [1 - \frac{1-\omega}{\omega} \rho'(\pi)] < r$, we have $r' < r$, $f'(k') < f(k)$, and $k' > k$. Since the impatience effect of inflation is dominated by the utility effect of money, the demand for money increases. And more capital is accumulated since money and capital move in the same direction on the optimal path.¹⁰ Correspondingly, the optimal inflation rate is larger than the rate argued by Friedman and might be zero or positive. On the other hand, if the utility effect of money is dominated by the impatience effect of inflation, i.e., $\frac{\omega}{1-\omega} < \rho'(\pi)$, the nominal interest rate is negative, $i < 0$. In this case, the steady state levels of real balances and capital are both decreased. It is optimal for government to subsidize the consumers for their holdings of money. Hence, the optimal inflation tax is negative. Finally, if these two effects offset each other, the nominal interest rate is zero and the Friedman rule is optimal.

Two particular cases are presented as follows. Case 1, if the time preference function is affine, or $\rho'(\pi) = 1$, and the share of money in the total financial wealth is one half in the steady state, i.e., $\omega = \frac{1}{2}$, the Friedman rule is optimal, for $i = r + \pi = 0$. Case 2, if the share of money in total financial wealth is one in the steady state, i.e., $\omega = 1$ and the impatience effect of inflation is finite, $\frac{\gamma}{qm} \rho'(\pi) < \infty$, we have $\varepsilon = \frac{1}{2}$ by taking the limits on the both sides of equation (33) with respect to v . It is similar to Bailey (1956) and Chamley (1985a), which shows that when the excess burden of other taxes tends to infinity, the government maximizes the revenues from money creation and ε is equal to one.¹¹

3. GENERALIZATIONS TO NON-SEPARABLE UTILITY FUNCTION

The assumption of additive separability was introduced in the previous section for the sake of simplicity. It is now relaxed to the general concave function, $U = U(c, l, m)$. Then, the optimality conditions of the represen-

¹⁰See note 9.

¹¹The issue of time-consistency is similar to Chamley (1985a). If assuming that q_0 and P_0 are exogenously given and that the government is honest in the sense of Auerheimer (1974), it can be shown that there is no incentive for the government to change $\bar{w}_0$, π_0 .

tative individual are

$$U_c(c_t, l_t, m_t) = q_t, \quad (35)$$

$$U_l(c_t, l_t, m_t) = -q_t \bar{w}_t, \quad (36)$$

$$U_m(c_t, l_t, m_t) = q_t(r_t + \pi_t), \quad (37)$$

$$\dot{q}_t = [\rho(\pi_t) - r_t]q_t. \quad (38)$$

From equations (35) and (36), consumption and labor supply can be expressed as functions of q_t , $\bar{w}_t$, and m_t :

$$c_t = c(q_t, \bar{w}_t, m_t), l_t = l(q_t, \bar{w}_t, m_t). \quad (39)$$

Then, the optimality conditions of the firm turn into

$$r_t = f_k(k_t, l(q_t, \bar{w}_t, m_t)), w_t = f_l(k_t, l(q_t, \bar{w}_t, m_t)). \quad (40)$$

Substituting equations (39) and (40) into equation (37) gives us the money demand function, implicitly defined as a function of k_t , q_t , $\bar{w}_t$, and π_t , i.e., $m = \phi(k, q, \bar{w}, \pi)$. Taking derivatives with respect to t on both sides of the definition, we have

$$\dot{m}_t = \phi_k \dot{k} + \phi_q \dot{q} + \phi_{\bar{w}} \dot{\bar{w}} + \phi_{\pi} \dot{\pi}. \quad (41)$$

Similar to the case of additively separable utility, the Hamiltonian associated with the the optimization of government is

$$H = e^{-\Delta t} \left\{ \begin{aligned} & u(q, \bar{w}, \phi(k, q, \bar{w}, \pi)) + (\lambda + \mu \phi_k)[f(k, l(q, \bar{w}, \phi(k, q, \bar{w}, \pi))) - c(q, \bar{w}, \phi(k, q, \bar{w}, \pi)) - g] \\ & - \mu \left\{ \begin{aligned} & f_k(k, l(q, \bar{w}, \phi(k, q, \bar{w}, \pi)))b - \\ & [f_l(k, l(q, \bar{w}, \phi(k, q, \bar{w}, \pi))) - \bar{w}]l(q, \bar{w}, \phi(k, q, \bar{w}, \pi)) + g - \pi \phi \end{aligned} \right\} \\ & (\xi + \mu \phi_q)[\rho(\pi) - f_k(k, l(q, \bar{w}, \phi(k, q, \bar{w}, \pi)))]q_t + (\alpha + \mu \phi_{\bar{w}})x + (\beta + \mu \phi_{\pi})z - \gamma \rho(\pi_t) \end{aligned} \right\}$$

where

$$u(q, \bar{w}, \phi(k, q, \bar{w}, \pi)) = U(c(q_t, \bar{w}_t, \phi(k, q, \bar{w}, \pi)), l(q, \bar{w}, \phi(k, q, \bar{w}, \pi)), \phi(k, q, \bar{w}, \pi)). \quad (42)$$

The optimality conditions on the control variable z and the state variable b are analogous to the results of the case with separable utility function

$$\beta = -\mu \phi_{\pi}(k, q, \bar{w}, \pi), \quad (43)$$

$$\dot{\beta} = -\dot{\mu} \phi_{\pi} - \mu \phi_{\pi k} \dot{k} - \mu \phi_{\pi q} \dot{q} - \mu \phi_{\pi \bar{w}} \dot{\bar{w}} - \mu \phi_{\pi \pi} \dot{\pi}, \quad (44)$$

$$\dot{\mu} = \{\rho(\pi) - f_k[k, l(q, \bar{w}, \phi(k, q, \bar{w}, \pi))]\} \mu. \quad (45)$$

In the steady state (implying $r = \rho(\pi)$, and $\dot{k} = \dot{q} = \dot{\lambda} = x = z = 0$), the optimality conditions on the state variables π and k turn into

$$\phi_\pi A + \mu m - \gamma \rho'(\pi_t) = q(\xi + \mu \phi_q)[f_{kl} l_m \phi_\pi - \rho'(\pi)], \quad (46)$$

and

$$\phi_k A - \mu(f_{kk} b - f_{lk} l) = q(\xi + \mu \phi_q)[f_{kk} + f_{kl} l_m \phi_k], \quad (47)$$

respectively, where

$$A = \{u_m + \mu(r + \pi) + (\lambda + \mu \phi_k)(w l_m - c_m) - \mu l_m[f_{kl} b - f_{ul} l - (w - \bar{w})]\}.$$

Dividing equation (47) by equation (46) on both sides leads to

$$\frac{\phi_k A - \mu(f_{kk} b - f_{lk} l)}{\phi_\pi A + \mu m - \gamma \rho'(\pi_t)} = \frac{f_{kk} + f_{kl} l_m \phi_k}{f_{kl} l_m \phi_\pi - \rho'(\pi)},$$

which is equivalent to

$$\begin{aligned} & f_{kk}(\phi_\pi A + \mu m) + \mu m f_{kl} l_m \phi_k + \mu(f_{kk} b - f_{lk} l) f_{kl} l_m \phi_\pi \\ & + \rho'(\pi) \{[\phi_k A - \mu(f_{kk} b - f_{lk} l)] - \gamma(f_{kk} + f_{kl} l_m \phi_k)\} = 0. \end{aligned} \quad (48)$$

The property of constant return to scale of the production function results in

$$f_{kl} l + f_{kk} k = 0, f_{ul} l + f_{lk} k = 0. \quad (49)$$

Equations (35)-(37) and (42) establish

$$q c_m - q \bar{w} l_m + q(r + \pi) = u_m. \quad (50)$$

From equations (48), (49), and (50), it is easy to derive the formula for the optimal inflation rate as follows:

$$\varepsilon = \frac{(1 + \delta)}{[1 + \rho'(\pi)/f_{kk}]B} \left\{ v \left[1 + \frac{\rho'(\pi)(1 - \omega)}{\omega} \right] - \rho'(\pi)\gamma \right\}, \quad (51)$$

where

$$\begin{aligned} \delta &= \left(\frac{k}{m}\phi_k\right)\left(\frac{m}{l}l_m\right), \\ \omega &= \frac{m}{k + b + m}, \\ B &= \frac{1}{q(r + \pi)} \{(\lambda - q + \mu \phi_k)(w l_m - c_m) + (q + \mu)[r + \pi + (w - \bar{w})l_m]\}, \end{aligned}$$

Thus, we have

PROPOSITION 4. *In the framework of second best taxation with inflation aversion and nonseparable utility, the rule for the optimal inflation rate is given by equation (51). Similiar to the separable utility case, if the marginal efficiency cost of other distorting taxation is finite, i.e., $v < \infty$, the Friedman rule is not optimal, even the marginal efficiency cost of other taxes is zero.¹²*

4. CONCLUSION

The paper has analyzed the problem of the optimal inflation tax in a stylized dynamic model of second best with inflation aversion and derived interesting results different from the literature. The three propositions of section 2 present the main results. Firstly, when the marginal excess burden of other distortion taxes approaches zero, the paper shows that the optimal inflation tax is negative and Friedman's rule for optimum quantity of money is not optimal. Secondly, when the marginal excess burden of other distorting taxes is finite, the sign of the nominal interest rate relies mainly on the tradeoff of the marginal excess burden of other distorting taxes and the impatience effect of inflation. Specifically, if the marginal excess burden of other taxes dominates, then the nominal interest rate is positive; if the impatience effect of inflation dominates, then the nominal interest rate is negative; and if the two opposite effects offset each other, then the nominal interest rate is zero. Thirdly, when the marginal excess burden of other distorting taxes approaches infinite and the impatience effect is finite, the optimal inflation tax depends mainly on the tradeoffs between the utility effect of money and the impatience effect of inflation. If the utility effect of money dominates, the inflation tax is positive; if the impatience effect of inflation dominates, the inflation tax is negative; and if these two effects equal, then the inflation tax is zero.

APPENDIX A

Proof: Totally differentiating equation $m = \phi(k, q, \bar{w}, \pi)$ gives rise to

$$dm = \phi_k dk + \phi_q dq + \phi_{\bar{w}} d\bar{w} + \phi_{\pi} d\pi,$$

¹²When $v = 0$, equation (51) is simplified to $\varepsilon = -\rho'(\pi)\gamma(1 + \delta)\{[1 + \rho'(\pi)/f_{kk}]B\}^{-1} \neq 0$. Hence, different from Chamley (1985a), the Friedman rule is not optimal. But, if inflation aversion does not exit, i.e., $\rho'(\pi) = 0$, we return to the simple rule derived by Chamley (1985a), i.e., $\varepsilon = \frac{v(1+\delta)}{B}$.

which implies that $\frac{dm}{dk} = \phi_k$, $\frac{dm}{d\pi} = \phi_\pi$. Hence

$$\frac{d\pi}{dk} = \frac{\phi_k}{\phi_\pi}. \quad (\text{A.1})$$

Totally differentiating equation $\rho(\pi) = r = f_k(k, l(q, \bar{w}))$ in the steady state results in

$$\rho'(\pi)d\pi = f_{kk}dk + f_{kl}l_q dq + f_{kl}\bar{w}d\bar{w}.$$

Therefore,

$$\frac{d\pi}{dk} = \frac{f_{kk}}{\rho'(\pi)}. \quad (\text{A.2})$$

Then, equations (A.1) and (A.2) establish $f_{kk}\phi_\pi = \rho'(\pi)\phi_k$. ■

APPENDIX B

Proof: Setting $v \rightarrow +\infty$ and taking limits on both sides of equation (33) lead to

$$\begin{aligned} \lim_{v \rightarrow +\infty} \varepsilon &= \lim_{v \rightarrow +\infty} \frac{1}{2(1+v)} \left\{ v \left[1 - \left(\frac{1-\omega}{\omega} \right) \rho'(\pi) \right] - \frac{\gamma \rho'(\pi)}{qm} \right\} \\ &= \lim_{v \rightarrow +\infty} \frac{[1 - (\frac{1-\omega}{\omega}) \rho'(\pi)]}{2}, \quad (\text{where } \lim_{v \rightarrow +\infty} \frac{\gamma \rho'(\pi)}{qmv} = 0, \text{ for } \frac{\gamma \rho'(\pi)}{qm} \text{ is finite}) \\ &= \frac{[1 - (\frac{1-\omega}{\omega}) \rho'(\pi)]}{2}. \end{aligned}$$

Then, when the marginal excess burden of other distorting taxes approaches infinite, the result above can be written in the limit sense as $\varepsilon = \frac{[1 - (\frac{1-\omega}{\omega}) \rho'(\pi)]}{2}$. By the definition of the interest elasticity of the money demand, we have

$$i = \pi + r = \frac{1}{-2r\phi_\pi} \left[1 - \frac{1-\omega}{\omega} \rho'(\pi) \right]. \quad (\text{B.1})$$

where the item of $(-2\phi_\pi)$ is positive from note 9. Hence,

If $1 > \frac{1-\omega}{\omega} \rho'(\pi)$, i.e., $\frac{\omega}{1-\omega} > \rho'(\pi)$, then $i > 0$, and $\pi = -r - \frac{1}{2\phi_\pi} [1 - \frac{1-\omega}{\omega} \rho'(\pi)] > -r$;

If $1 < \frac{1-\omega}{\omega} \rho'(\pi)$, i.e., $\frac{\omega}{1-\omega} < \rho'(\pi)$, then $i < 0$, and $\pi = -r - \frac{1}{2\phi_\pi} [1 - \frac{1-\omega}{\omega} \rho'(\pi)] < -r$;

If $1 = \frac{1-\omega}{\omega} \rho'(\pi)$, i.e., $\frac{\omega}{1-\omega} = \rho'(\pi)$, then $i = 0$, and $\pi = -r$.

In particular, putting $\omega = \frac{1}{2}$ and $\rho'(\pi) = 1$ into equation (B.1) yields $i = 0$. ■

REFERENCES

- Atkinson, Anthony, and Joseph Stiglitz, 1972. The structure of indirect taxation and economic efficiency. *Journal of Public Economics* **1**, 97-119.
- Auernheimer, Leonardo, 1974. The honest government's guide to the revenue from the creation of money. *Journal of Political Economy* **82**, 598-606.
- Bailey, Martin, 1956. The welfare cost of inflationary finance. *Journal of Political Economy* **64**, 93-110.
- Blanchard, Olivier, and Stanley Fischer, 1990. Lectures on Macroeconomics. The MIT Press; Cambridge, Massachusetts; London, England.
- Bohm-Bawerk, Eugen, 1891. The Positive Theory of Capital. Macmillan, London.
- Burns, Arthur, 1978. The Condition of the American Economy. American Enterprise Institute for Public Research, Washington, D.C..
- Chamley, Christophe, 1985. On a simple rule for the optimal inflation rate in second best taxation. *Journal of Public Economics* **26**, 35-50.
- Chamley, Christophe, 1985. Efficient Taxation in a Stylized Model of Intertemporal General Equilibrium. *International Economic Review* **26**, 451-468.
- Chamley, Christophe, 1986. Optimal Taxation of Capital Income in General Equilibrium with Infinite Lives. *Econometrica* **54**, 607-622.
- Chari, Varadarajan, Lawrence Christiano, and Patrick Kehoe, 1996. Optimality of the Friedman Rule in Economies with Distorting Taxes. *Journal of Monetary Economics* **37**, 203-223.
- Chari, Varadarajan, and Patrick Kehoe, 1999. Optimal Fiscal and Monetary Policy. Handbook of Macroeconomics, Volume 1, Edited by J.B. Taylor and M. Woodford, Elsevier Science B.V.
- Correia, Isabel, and Pedro Teles, 1996. Is the Friedman Rule Optimal when Money is an Intermediate Good? *Journal of Monetary Economics* **38**, 223-244.
- Correia, Isabel, and Pedro Teles, 1999. The Optimal Inflation Tax. *Review of Economic Dynamics* **2**, 325-346.
- Da Costa, Carlos, and Ivan Werning, 2008. On the Optimality of the Friedman Rule with Heterogeneous Agents and Nonlinear Income Taxation. *Journal of Political Economy* **116**, 82-112.
- Drazen, Allan, 1979. The Optimal Rate of Inflation Revisited. *Journal of Monetary Economics* **5**, 231-248.
- Fabricant, Solomon, 1976. Economic Calculation under Inflation: the Problem in Perspective. In Economic Calculation under Inflation. Liberty Press.
- Friedman, Milton, 1969. The Optimum Quantity of Money, in: The Optimum Quantity of Money and other essays (Aldine Publishing Co., Chicago).
- Friedman, Milton, 1971. The Revenue from Inflation. *Journal of Political Economy* **79**, 846-856.
- Keynes, John, Maynard, 1936. The General Theory of Employment, Interest, and Money. A Harvest/HBJ Book, Harcourt Brace Jovanovich, Publishers.
- Katona, George, 1975. Psychological Economics. New York: Elsevier.

- Marty, Alvin, 1967. Growth and the Welfare Cost of Inflationary Finance. *Journal of Political Economy* **75**, 71-76.
- Marty, Alvin, 1973. Growth, satiety, and the tax revenue from money creation. *Journal of Political Economy* **81**, 1136-1152.
- Phelps, Edmund, 1972. *Inflation Policy and Unemployment Theory*. W.W. Norton and Co. New York.
- Phelps, Edmund, 1973. Inflation in the Theory of Public Finance. *The Swedish Journal of Economics* **75**, 67-82.
- Shiller, Robert, 1996. Why do people dislike inflation? NBER working paper.
- Sidrauski, Miguel, 1967. Rational Choice and Patterns of Growth in a Monetary Economy. *The American Economic Review* **57**, 534-544.
- Siegel, Jeremy, 1978. Notes on Optimal Taxation and the Optimal Rate of Inflation. *Journal of Monetary Economics* **4**, 297-305.
- Turnovsky, Stephen, and William Brock, 1980. Time Consistency and Optimal Government Policies in Perfect Foresight Equilibrium. *Journal of Public Economics* **13**, 183-212.
- Wang, Gaowang, and Heng-fu Zou, 2011. Inflation Aversion and Macroeconomic Policy in a Perfect Foresight Monetary Model. *Economic Modelling* **28**.
- Zou, Heng-fu, Liutang Gong, and Xinsheng Zeng, 2011. Inflation Aversion. *Annals of Economics and Finance* **12(1)**, 1-11.

Fiscal decentralization, revenue and expenditure assignments, and growth in China

Jing Jin ^{a,1}, Heng-fu Zou ^{b,c,*}

^a *Central University of Finance and Economics, 39, South College Road,
Haidian District, Beijing 100081, China*

^b *Peking University and Wuhan University, China*

^c *Research Department, The World Bank, MC3-639, 1818 H. St. NW,
Washington, DC 20433, USA*

Received 1 July 2005; received in revised form 15 September 2005; accepted 12 October 2005

Abstract

Theory suggests that a close match between revenue and expenditure assignments at sub-national levels benefits allocative efficiency, and hence economic growth. That is, a convergence of revenue and expenditure assignments at sub-national levels of government should, according to the theory, be positively associated with a higher growth rate. In the case of China, this paper shows, divergence, rather than convergence, in revenue and expenditures at the sub-national level of government is associated with higher rates of growth. A panel dataset for 30 provinces in China is used to examine the relationship between fiscal decentralization and economic growth over two phases of fiscal decentralization in China: (1) 1979–1993 under the fiscal contract system, and (2) 1994–1999 under the tax assignment system. The seeming contradiction between the theory and evidence in the China case is reconciled by taking into account the institutional arrangements that prevailed during the two phases of fiscal decentralization, in particular the inconsistency between the assumptions of the theory of fiscal decentralization and the institutional reality of China.

© 2005 Elsevier Inc. All rights reserved.

JEL classification: H2; H7; O1

Keywords: Fiscal decentralization; Economic growth; Revenue; Expenditure

* Corresponding author at: Research Department, The World Bank, MC3-639, 1818 H. St. NW, Washington, DC 20433, USA. Tel.: +1 202 473 7939; fax: +1 202 522 1154.

E-mail addresses: jjin68@hotmail.com (J. Jin), Hizou@worldbank.org (H.-f. Zou).

¹ Tel.: +86 137 175 41581; fax: +86 10 6228 9132.

1. Introduction

Since initiating economic reforms in 1978, fiscal decentralization has been a central component of China's economic policy at a time when China has experienced unprecedented economic growth. Although China remains a unitary political system, where sub-national government elections are virtually not exist, its fiscal system is nevertheless a decentralized one featured by a fiscal contract system (1980–1993)¹ and revenue assignment system (1994–present). Regardless of China's non-democratic institutions, the benefits of fiscal decentralization seem still applicable, according to Oates (1972, p. xvi), because “for an economist, however, constitutional and political structures are of less importance: What is crucial for him is simply that different levels of decision-making do exist, each of which determines levels of provision of particular public services in response largely to the interests of its geographical constituency. By this definition, practically any fiscal system is federal or at least possesses federal elements”.

The question of whether fiscal decentralization has contributed to China's economic success over the past 20 years is, however, open to debate. Some argue that fiscal decentralization has been fundamental to China's economic success (Oi, 1992; Qian, 1999; Qian & Weingast, 1997). It has been asserted that the fiscal contract system (1980–1993) provided material incentives that encouraged and rewarded sub-national governments to promote local economies (Oi, 1992; Qian, 1999). Secondly, Qian (1999) assumes that sub-national governments had less control over banks and therefore could not bail out their state-owned enterprises (SOEs) by extending credit to them as the central government did. Fiscal decentralization, they argue, hardened the budget constraints of sub-national governments' SOEs, and thus made these SOEs more efficient (Qian, 1999). The fiscal contract system, it is also asserted, allowed sub-national governments to conceal information about their financial position and enabled them to avoid revenue predation from the Center (Qian & Weingast, 1997), thus allowing them to retain the financial resources they needed for investments that promoted economic development.

Some studies have, however, offered evidence suggesting that fiscal decentralization fragmented the national market, and hence negatively affected economic growth. Instead of inducing jurisdictional competition that would have potentially enhanced allocative efficiency, decentralization, it is argued, created revenue incentives that encouraged sub-national governments to engage in protectionist behavior (Yang, 1997). Enterprise ownership by local governments provided an incentive to local governments to duplicate enterprises under their jurisdiction so as to capture the revenues that would have otherwise gone to the central coffers, leading to “backward specialization”,² as evidenced by the convergence of regional relative outputs and a divergence of regional relative factor allocations and labor productivities during the reform era (Young, 2000). As a result, the centrally controlled planned economy devolved, according to this argument, into one of many regional planned economies controlled by sub-national governments (Young, 2000). In addition, such ownership structure of SOEs enabled sub-national governments increasingly to mandate these firms to provide public goods—such as housing, healthcare, childcare, schooling and pension. Thus, it is argued that budget constraint on sub-national governments was effectively

¹ Under fiscal contract system, each subnational government level contracted with the next level up to meet certain revenue remittance targets, which could be either a nominal amount or a percentage share. The retention was at the sole discretion of subnational governments. For details, see Bahl and Wallich (1992), World Bank (1989, 1993).

² The term is used by Yang (1997).

softened by fiscal decentralization, since many local SOEs shared the spending responsibilities of local governments and became de facto government agencies and conduits for central-local financial transfers (Steinfeld, 1999).

The aim of this paper is to attempt to resolve and reconcile these outstanding issues, as well as to relate the Chinese experience to the orthodox theory of fiscal decentralization. Using panel dataset for 30 provinces from 1979 to 1993³ and 1994 to 1999, respectively, this paper investigates the relationship between the prevailing fiscal patterns, defined by both expenditure as well as revenue decentralization at the provincial level, and China's provincial economic growth. It further examines how the shift from the contracted revenue sharing (1980–1993) to tax assignment system (1994–1999) affected the relationship between fiscal decentralization and provincial economic growth. It aims to explain how intergovernmental fiscal relations under the two tax regimes affected growth.

Section 2 reviews the theoretical arguments and empirical studies on the relationship between fiscal decentralization and economic growth. Section 3 outlines the hypothesis, explanatory variables, and methodology used in this case study of fiscal decentralization in China. Section 4 reports the regression results and Section 5 summarizes the findings and conclusions.

2. Literature review

2.1. Theoretical considerations

It has long been held that, in theory, fiscal decentralization may be conducive to economic growth. If few public goods entail nationwide externalities, sub-national governments are likely to be more efficient in the production and delivery of public goods (Oates, 1972). It is also asserted that decision-making on expenditures at lower levels of government is more responsive to diversified local preferences and needs and, therefore, more conducive to allocative efficiency (Oates, 1972; Tiebout, 1956). Decentralizing revenue discretion to sub-national governments to match the spending assignments may also enhance accountability (Oates, 1972). It is held, therefore, that for a given level of government, revenue means should match expenditure needs as closely as possible, thereby (1) stimulating revenue mobilization from local sources, and improving a country's overall fiscal position; (2) improving accountability of sub-national governments; and (3) reducing the distorting effects of intergovernmental transfers (Shah, 1994).

Theorists of fiscal decentralization were inspired, for the most part, by their observations of the functioning of fiscal systems based in highly developed economies, like the United States (Brueckner, 2000). The implications of fiscal decentralization in the context of a developing country are, however, subject to various qualifications due to the divergence between the assumptions of orthodox theory and the institutional as well as economic realities in developing countries. As many have argued, if the standard assumptions of decentralization theory do not hold, the outcomes of fiscal decentralization may be detrimental to economic growth and efficiency (Oates, 1993; Prud'homme, 1995; Tanzi, 1996; Jin and Zou, 2003).

Prud'homme (1995) stresses, for example, that local provision of public goods may not be more cost-effective than at the national level because of economies of scale and economies of scope. It has also been suggested that assuming constituents universally can express their

³ Fiscal contracts between the Center and provinces were gradually introduced in 1980. This empirical study employed all available time series data since 1979.

preferences in their votes ignores (1) the patron–client relationships that define the local electoral behavior in developing countries, and (2) the usually vague and inconsistent electoral mandate of local elections in these countries (Prud'homme, 1995). In addition, even if local constituents can express preferences in their votes, and the elected officials want to satisfy the voters' needs, local bureaucracies that carry out the electoral mandate may be poorly motivated and/or qualified to carryout their responsibilities (Prud'homme, 1995).

Fiscal decentralization may also be conducive to corruption at local level because it confers discretion on local politicians and bureaucrats who are more susceptible and accessible to the demands of local interest groups (Prud'homme, 1995; Tanzi, 1996). Corruption at sub-national levels is likely to diminish, if not negate, the benefits that theory suggests fiscal decentralization brings to allocative efficiency and growth.

Moreover, in a non-democratic political system, the basic premise that sub-national governments have a stronger incentive to provided local public goods more efficiently may not apply (Tanzi, 1996). The principle-agent problem in a non-democratic political system may render fiscal decentralization as a tool to be used by sub-national authorities to exploit local constituents and the national treasury (for the case of China, please refer to Wong, 1991; Bahl & Wallich, 1992; Bahl, 1999).

2.2. Empirical evidence

The problem with the recent empirical studies can be summarized from the perspectives of (1) measurements of fiscal decentralization; (2) the relative relationship between expenditure and revenue decentralization; and (3) levels of government. Firstly, using expenditure shares alone to measure decentralization tends to produce a negative (for developing countries) or insignificant (for industrial countries) relationship between fiscal decentralization and economic growth (Davoodi & Zou, 1998; Xie, Zou, & Davoodi, 1999; Zhang & Zou, 1998).⁴ Using revenue shares alone to measure decentralization tends to give results suggesting a positive relationship with economic growth (Ebel & Yilmaz, 2001). What accounts for these fundamentally contradictory results? Perhaps the most important explanation is that expenditure in most of the countries is typically far more decentralized than revenue. For example, for the six-country sample data for 1999 used in Ebel and Yilmaz (2001), the mean of sub-national expenditure share in total government revenue is 22%, while the mean of sub-national own-taxes revenue share in total government revenue is only 6.2%. Since sub-national governments' own-taxes revenue share in total revenue is substantially lower than their expenditure share in total expenditure, it is therefore not surprising that using revenue shares alone to measure decentralization tends to give results to suggest that revenue decentralization (i.e., increasing the share of sub-national tax revenue share in total government revenue to meet the much larger spending assignments at the corresponding level) promotes economic growth. As such, neither the positive association between *revenue decentralization* and economic growth found in the study by Ebel and Yilmaz (2001) can undermine or refute the negative (or insignificant) findings between *expenditure decentralization* and growth found by Davoodi and Zou (1998), Xie et al. (1999), Zhang and Zou (1998), nor vice versa. Because the later use the expenditure shares, which are much more decentralized than revenue shares to assess the relationship between fiscal decentralization and economic growth,

⁴ Davoodi and Zou (1998) is a cross-country study, while Xie et al. (1999) is a case study on the US and Zhang and Zou (1998) is a case study on China.

and both studies consider only half of the story. Clearly what is necessary in analyzing the relationship between fiscal decentralization and economic growth is to test simultaneously the effect of the level of both expenditure and revenue decentralization, and the effect of the fiscal pattern they hence reveal (i.e., the extent to which expenditure and revenue decentralization converge or diverge), which is the approach taken in this case study of China. What should be note here is that such observations do not imply to detect the optimal level of expenditure or revenue decentralization, but about which directions they move. If the regression suggests that one should move closer to the other, then I call it a convergence. Otherwise divergence.

A second general observation on the recent empirical investigations is that when both decentralization measures are used, the results should be interpreted with respect not just to the coefficients of each measure but should also take into account the decentralization on the two sides of the government budget. In other words, the relationship between expenditure and revenue decentralization matters. For example, Akai and Sakata (2002) use both state expenditure and revenue share in total to proxy for fiscal decentralization. They conclude that “fiscal decentralization contributes to economic growth” because expenditure decentralization has a positive association with state GDP per capita growth rate in all equations. But a comparison of the relative levels of expenditure and revenue decentralization at US state level suggests a different conclusion. Specifically, since expenditure is 7.5% more decentralized than revenue at the US state level (see Table 3.A.2 in Akai & Sakata, 2002), to suggest that further expenditure decentralization promotes growth is to imply that expenditure and revenue assignments should diverge, rather than converge as the theoretical literature suggests is conducive to efficiency and growth.

Jin and Zou’s (1999) study, using both revenue and expenditure decentralization measures at both state/provincial and local levels⁵ find that a convergence of revenue and expenditure at state/provincial level and a divergence of them at local level promote growth. The finding that growth is promoted by the convergence of expenditure and revenue at the state/provincial level is consistent with the theoretical principle of fiscal federalism. However, the suggestion for divergence of the two – more expenditure assignments and fewer revenue assignments – at local level is not. The intuitive appeal of this result is that tax bases tend to be smaller and narrower at the local level than at the state/provincial level (Bird, 1992; Mello, 2000). Local governments simply do not have the social and economic endowments to generate the revenue required to finance their spending requirements.

The previous point leads to a third general observation: Most countries have three or more than three levels of government (federal, state and local), however the assignments of expenditure and revenue may have different implications at different sub-national levels of government (e.g., state or local). The question is, however, whether the results for one level of government can be generalized to another. It is well established, for example, that the revenue-generation capacity varies at different levels of government (Musgrave, 1983). The tax bases of local governments (vis-à-vis state/provincial level) are relatively narrow because of possible tax export, externalities in the public goods provision, factor mobility, and economies of scale (Mello, 2000). As such, decentralization of revenue assignments to match local expenditure assignments may not be efficient or growth promoting, as demonstrated in Jin and Zou (1999).

Finally, it is worth recalling that cross-country studies have the disadvantage of pooling countries with substantial differences in history, politics, institutions, and culture, which if not

⁵ A sample of 11 countries, 1970–1994.

taken into account in the analysis are likely to blur the true relationship between fiscal decentralization and growth (Akai & Sakata, 2002). Specific country studies, such as the present one, avoid this problem, though it may be argued that their results are less generalizable.

3. Hypothesis, explanatory variables and methodology

3.1. Hypothesis and explanatory variables

This study uses Chinese provincial panel data for two time periods, 1979–1993 and 1994–1999, to investigate the effect of fiscal reforms on provincial economic growth under the two fiscal regimes: the fiscal contract system and the tax assignment regime. The purpose of using two time periods is to focus on the effect of the policy change of tax structure and collection that brought by the 1994 reform. Therefore, instead of using tests of structural change to identify the break of time period, the second period is assigned from 1994, the year when the new tax assignment implemented with the split of tax collection between the Center and the provinces.

Fiscal decentralization is measured with respect to both expenditure and revenue assignments. Four fiscal decentralization measures are used. The two measures of expenditure decentralization are (1) provincial budgetary expenditure as a share in total budgetary expenditure, and (2) provincial extra-budgetary expenditure as a share in total extra-budgetary expenditure. The two measures of revenue decentralization are (1) provincial budgetary revenue as a share in total budgetary revenue, and (2) provincial extra-budgetary revenue as a share in total extra-budgetary revenue. Both provincial expenditure and revenue are expenditures spent and revenue collected at the provincial level. Using revenue collected at the provincial level as a share in total revenue to proxy the degree of revenue decentralization has the advantage of incorporating the tax collection aspect. More specifically, since China had a completely localized tax administration during the fiscal contract phase (1979–1993) – i.e., provinces collected taxes for the central government as its agents – provincial revenue share in total revenue should be, on average, larger than provincial expenditure share, the difference being the provinces' remittance to the Center (rather than central transfer to states/provinces, as is more typically the case).

The 1994 fiscal reform replaced localized tax administration by disaggregating tax collection into central and sub-national parts, with the central tax administration collecting central and shared taxes and sub-national bureaus collecting local taxes. Since 1995, central to provincial government transfers are recorded in the budget. In addition to the expenditure and revenue decentralization measures, the intergovernmental transfer, measured by central transfer to provinces as a percentage in total provincial expenditure, is taken into account in the analysis for the second phase of the fiscal reform so as to assess the potential distorting effect of such transfers at provincial level.

Conventional fiscal decentralization theory holds that the matching of revenue means and expenditure assignments at sub-national level promotes economic growth. Therefore, the signs on the coefficient of expenditure and revenue decentralization, taking into account the average levels of revenue and expenditure share, should indicate whether convergence or divergence of revenue and expenditure decentralization promotes growth.

Two tax variables are employed to examine the effects of distortion of taxes imposed by central and provincial governments. Tax rates are used as aggregate measures of distortion introduced by governments to finance their spending (Zhang & Zou, 1998). Specifically, the central tax rate, measured by central tax revenues as a percentage of total GDP, is used to capture the effect of distortion at the national level. The provincial tax rate, measured by provincial

(collected) tax revenue as a percentage of total provincial GDP, is used to capture the effect of distortion at the provincial level. It is expected that the higher the tax rate, the more the economy is distorted by the fiscal system (Barro, 1990).

While our main interest is the relationship between fiscal decentralization and economic growth, we must acknowledge that economic growth is subject to many influences beyond fiscal decentralization. In order to control for these influences we introduce a set of *control variables* to improve the robustness of the result. This set of control variables is consistent with the set of variables used in Zhang and Zou's (1998) case study of China, allowing their results to be compared to those presented in this study. The control variables used in this study includes: Physical and human capital investments, respectively measured by (1) the sum of gross investment (government and enterprises together) as a share in GDP at provincial level and (2) the growth rate of the provincial labor force.

Another important determinant of growth is openness to international trade, which is measured by the ratio of imports plus exports to GDP at provincial level. It is conventional to hypothesize a positive relation between openness and growth on grounds that international competition improves resource allocation via exports and more advance technology from industrial countries can be attained via imports (Feder, 1983 quoted in Zhang & Zou, 1998).

Finally, we allow for the potential effect of macroeconomic instability on economic growth, using the lagged inflation rate at the provincial level as a proxy for this variable. Inflation can have both a positive and negative effect on growth. The positive effect stems from the potential for inflation to promote savings and investment, as agents shift from financial wealth (money) to real assets (capital) to avoid the deleterious effects of inflation on real money balances (the Tobin portfolio-shift effect). On the other hand, inflation may dampen economic growth because it raises the transaction cost of economic activities (Zhang & Zou, 1998).

Data sources. The pre-1990 data are taken from Hseh, Li, and Liu (1993).⁶ The post 1989 data for the 30 provinces⁷ are from the *China Finance Statistical Yearbook* (various issues) and the *China Statistical Yearbook* (various issues). The panel datasets for thirty provinces cover 1979–1993 and 1994–1999 separately (using the same methodology for these two time periods).

A statistical summary of the key variables. Table 1 provides the statistics of annual budgetary expenditure and revenue shares in total government budgetary items across all provinces.

As indicated in Table 1, revenue is more decentralized than expenditure in every year.⁸ The difference between the expenditure and revenue shares was the provincial remittance to the Center. After introducing the fiscal contract system in 1980, the degree of revenue decentralization (averaged provincial revenue collection share in total government revenue) decreased from 2.6% in 1980 to 2.0% in 1985 while the degree of expenditure decentralization (averaged provincial

⁶ Hseh, Tien-tung, Li, Qiang, and Liu, Shucheng (Eds.), 1993. *China's Provincial Statistics 1949–89*. Boulder: Westview Press.

⁷ Chongqing became a direct municipal city in 1997. Since it is hard to disaggregate the statistics of Chongqing from that of Sichuan province before 1997, Chongqing's statistics are added back to that of Sichuan for the years after 1997.

⁸ Since the statistics in Table 1 are the means of each province's revenue and expenditure share in total. The mean discrepancy between revenue (2.4%) and expenditure decentralization (1.9%) therefore represents the share of a single province's revenue/expenditure in total government revenue/expenditure. To acquire an idea of the discrepancy between revenue and expenditure decentralization between central and provincial governments aggregated, the statistics in Table 1 should be multiplied by 29, which is the number of provinces (Tibet was dropped). For example, the discrepancy of revenue and expenditure decentralization between the Center and provincial level of government is average at $(2.4 - 1.9) \times 29 = 14.5\%$.

Table 1

Fiscal decentralization by year (1979–1993)

Expenditure decentralization by year (1979–1993)							Revenue decentralization by year (1979–1993)						
Year	Obs ^a	Mean	S.D.	Min	Max	Coefficient of variation	Year	Obs	Mean	S.D.	Min	Max	Coefficient of variation
1979	29	1.7	0.7	0.1	2.9	0.41	1979	29	2.7	3.0	0.1	15.1	1.09
1980	29	1.6	0.7	0.2	2.8	0.41	1980	29	2.6	3.0	0.1	15.1	1.14
1981	29	1.6	0.6	0.2	2.7	0.39	1981	29	2.5	2.9	0.1	14.8	1.15
1982	29	1.7	0.7	0.2	2.8	0.38	1982	29	2.5	2.7	0.1	13.9	1.10
1983	29	1.7	0.6	0.2	2.8	0.37	1983	29	2.2	2.2	0.1	11.4	1.01
1984	29	1.8	0.7	0.3	3.1	0.36	1984	29	2.0	2.0	0.1	10.0	0.96
1985	29	1.9	0.8	0.2	3.5	0.40	1985	29	2.0	1.8	0.1	9.2	0.89
1986	29	2.0	0.8	0.3	3.8	0.42	1986	29	2.2	1.7	0.1	8.5	0.81
1987	29	2.0	0.9	0.3	3.7	0.43	1987	29	2.3	1.7	0.1	7.7	0.73
1988	29	2.1	0.9	0.3	4.3	0.45	1988	29	2.3	1.6	0.2	6.9	0.68
1989	29	2.2	1.0	0.5	4.6	0.47	1989	29	2.4	1.6	0.2	6.3	0.64
1990	29	2.1	1.0	0.4	4.4	0.47	1990	29	2.3	1.4	0.2	5.8	0.62
1991	29	2.2	1.1	0.5	4.8	0.49	1991	29	2.5	1.5	0.2	5.6	0.61
1992	29	2.0	1.0	0.4	5.0	0.50	1992	29	2.4	1.5	0.2	6.4	0.63
1993	29	2.2	1.2	0.4	6.3	0.58	1993	29	2.6	1.8	0.2	8.0	0.72
		1.9	0.8	0.3	3.8				2.4	2.0	0.2	9.6	

^a Tibet is dropped due to unavailability of data.

expenditure share in total government expenditure) increased from 1.6% in 1980 to 1.9% in 1985. As a result, the gap between the average expenditure decentralization (across provinces) and revenue decentralization gradually narrowed.

After 1985, expenditure decentralization varied between 2.0 and 2.2% while revenue decentralization also started to increase from 2.0% in 1985 to 2.6% in 1993. As a result, the gap between revenue and expenditure assignments at provincial level steadily widened.

The relative change between expenditure and revenue decentralization across the first phase is also captured in Fig. 1.

In addition, the coefficients of variation of expenditure decentralization increased slightly from around 0.41 in the first half of 1980s to 0.58 in 1993 (Table 1 and Fig. 1). At the same time, the coefficients of variation of revenue decentralization decreased dramatically from 1.15 in 1981 to 0.61 in 1991 and then increased slightly to 0.72 in 1993. In other words, the degree of revenue decentralization across provinces converged through the 1980s. Such a convergence, in the context of a decreased revenue decentralization level (mean across provinces, as show by the decreasing mean statistic of the annual revenue decentralization across provinces shown in Fig. 1) in the first half of 1980s, indicates that the relatively wealthier provinces, with their lever of controlling tax collection, relaxed their revenue collections (to avoid sharing with the Center), hence their revenue decentralization level converged to the lower levels of the poorer ones.⁹

⁹ The coefficient of variation of revenue decentralization is the ratio of standard deviation of revenue decentralization across provinces divided by the mean of cross-province revenue decentralization. As the mean statistic (in the denominator) decreased before 1985, the standard deviation (in the numerator) decreased by more than the mean to yield a decreasing coefficient of variation. As such, the dispersion of revenue decentralization reduced significantly—i.e., the high degree of revenue decentralization of better off provinces reduced and converged to the lower level of revenue decentralization of poorer provinces.

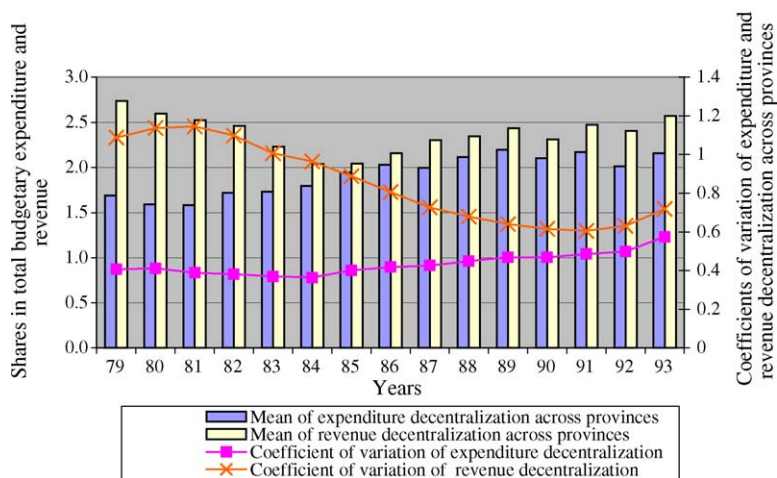


Fig. 1. Budgetary expenditure and revenue decentralization (1979–1993).

Furthermore, as noted before, the difference between expenditure and revenue decentralization is the provincial remittance to the Center. It is shown by the difference between revenue decentralization (white bars) and expenditure decentralization (darker bars) in Fig. 1. As demonstrated in Fig. 1, the provincial remittance to the Center (the excess part of white bars over darker bars) gradually declined since the implementation of fiscal contracts as the excess part of revenue over expenditure decentralization decreased.

From 1986 to 1991, the shares of provincial revenue in total government revenue further converged across provinces as shown by the further decrease of coefficient of variation of revenue decentralization. Table 1 shows that this is a result of a further decline in standard deviation of revenue decentralization across provinces at the time when the mean level of revenue decentralization increased. Such a convergence can be a result of either (1) the revenue collection of poor provinces grew faster than that of the wealthier provinces, or (2) the revenue collection of wealthier provinces grew relatively slowly because the contract regime incited them to engage in strategies that enabled them to accrue more of the revenue increments within their own jurisdictions (For details, see Wong, 1991; Bahl & Wallich, 1992; and Bahl, 1999), or both.

The coefficient of variation of revenue decentralization increased slightly in 1992 and sharply in 1993. The increase in 1992 was a result of the reduced cross-province mean of revenue decentralization (while the standard deviation was constant at 1991 level). The jump in 1993, however, was a result of an increase in the mean and a larger increase of standard deviation of revenue decentralization cross provinces. Such a change in 1993 was to a large extent a function of the design of 1994 fiscal reform. More specifically, the compromised plan of 1994 fiscal reform, under which provincial income level of 1994 was ensured by central “tax refund” up to their 1993 level, stimulated a sudden inflation of provincial revenue collection. Provinces thus attempted to boost the baseline of 1993 in order to entitle more “tax refund” from the Center in 1994 (for details, see Wang, 1997). Since wealthier provinces now switched strategy from hiding revenues to exhausting tax collections, this dramatically increased the dispersion in revenue decentralization across provinces.

Table 2 provides the statistics of annual budgetary expenditure and revenue share (mean across all provinces) in total government budgetary items during 1994–1999. Revenue

Table 2

Fiscal decentralization by year (1994–1999)

Expenditure decentralization by year (1994–1999)							Revenue decentralization by year (1994–1999)						
Year	Obs	Mean	S.D.	Min	Max	Coefficient of variation	Year	Obs	Mean	S.D.	Min	Max	Coefficient of variation
1994	29	2.4	1.4	0.3	7.2	1.52	1994	29	1.5	1.2	0.1	5.7	0.76
1995	30	2.4	1.5	0.3	7.7	1.60	1995	30	1.6	1.2	0.1	6.1	0.76
1996	30	2.4	1.5	0.4	7.6	1.69	1996	30	1.7	1.3	0.0	6.5	0.78
1997	30	2.4	1.5	0.4	7.4	1.64	1997	30	1.6	1.3	0.0	6.3	0.79
1998	30	2.4	1.5	0.4	7.6	1.68	1998	30	1.7	1.3	0.0	6.5	0.80
1999	30	2.3	1.4	0.4	7.3	1.63	1999	30	1.6	1.4	0.0	6.7	0.83
		2.4	1.5	0.4	7.5				1.6	1.3	0.1	6.3	

decentralization, which measured by the tax revenue collected at provincial level as a share in total revenue, therefore largely reduced. Now the degree of revenue and expenditure decentralization at the provincial level is reversed vis-à-vis the first phase of fiscal reform—unlike in the first phase, expenditure became more decentralized than revenue, with the average cross-province shares stabilized around 2.4 and 1.6%, respectively (Table 2). The gap, instead of representing provincial remittance to the Center, now reflects the Central transfer to the provinces.

3.2. Methodology

First, the regression analysis in this study uses the panel data econometric technique. Panel data sets combine time series and cross sections. They allow more flexibility in modeling. Time series data for each province in this cross-province regression analysis can better capture the dynamics of the relationship between fiscal decentralization and provincial economic growth. Second, all coefficients are estimated with fixed-effects with corrections for panel heteroskedasticity and panel serial correlation.

Using the panel data of 30 provinces, 1979–1993 and 1994–1999 separately, the following model is employed to examine how fiscal decentralization affects provincial growth:

$$\text{GDPgrowth}_{i,t} = \alpha_0 + \alpha_1 \text{FD}_{i,t} + \alpha_2 \text{TAX}_{i,t} + \alpha_3 \text{POLITICAL}_{i,t} + \alpha_4 \text{CONTROL}_{i,t} + \varepsilon_{i,t}$$

where $\text{GDPgrowth}_{i,t}$ represents real GDP growth rate of province i at time t . $\text{FD}_{i,t}$ is a set of vectors of fiscal decentralization measuring expenditure decentralization, revenue decentralization (both further disaggregated into budgetary and extra-budgetary terms), and intergovernmental transfers when applied to the time period of 1994–1999. $\text{TAX}_{i,t}$ is a set of vectors measuring the distorting effects of tax at both central and provincial level—central and provincial tax rates.

$\text{CONTROL}_{i,t}$ is a set of variables that control for provincial investment, labor force growth rate, the level of openness and provincial inflation (lagged).

4. Regression results

Table 3 reports the fixed- and random-effect results of how fiscal decentralization affected provincial economic reform for the time period 1979–1993. Table 4 reports the fixed and random-effect results for the time period 1994–1999.

Table 3
Fixed effects 1979–1993

Dependent variable	Provincial real GDP growth rate						
	1 ^a Coefficients (1979–1993)	2 ^a Coefficients (1979–1993)	3 ^a Coefficients (1979–1993)	4 ^a Coefficients (1982–1992)	5 ^a Coefficients (1982–1992)	6 ^a Coefficients (1982–1992)	7 ^a Coefficients (1982–1992)
Central tax rate	0.79 (3.47)***	0.80 (3.43)***	0.88 (3.83)***	0.97 (4.05)***	0.96 (3.81)**	1.10 (4.60)***	1.21 (4.80)***
Provincial tax rate	0.02 (0.28)	−0.12 (−0.94)	−0.21 (−1.66)*	−0.04 (−0.43)	−0.04 (−0.43)	−0.33 (−1.95)*	−0.24 (−1.33)
Expenditure decentralization (budgetary)	−2.89 (−3.11)***		−3.63 (−3.68)***			−5.53 (−4.36)***	−6.29 (−4.54)***
Revenue decentralization (budgetary)		0.54 (0.98)	1.25 (2.19)**			2.01 (2.50)**	1.57 (1.81)*
Expenditure decentralization (extra-budgetary)				−0.39 (−0.28)		0.86 (0.61)	−0.91 (−0.47)
Revenue decentralization (extra-budgetary)					−0.14 (−0.13)		2.43 (1.35)
Labor growth rate	0.21 (1.15)	0.24 (1.30)	0.21 (1.13)	0.47 (2.26)**	0.47 (2.24)**	0.46 (2.30)**	0.49 (2.42)**
Investment rate	0.26 (4.60)***	0.22 (3.90)***	0.26 (4.64)***	0.09 (1.37)	0.10 (1.35)	0.15 (2.25)**	0.11 (1.53)
Openness	0.09 (1.49)	−7.97E−03 (−0.14)	0.09 (1.41)	0.08 (1.07)	0.08 (1.04)	0.17 (2.32)**	0.17 (2.28)**
Provincial inflation (lagged)	2.85 (2.41)**	2.62 (2.17)**	2.48 (2.09)**	1.16 (0.89)	1.18 (0.90)	0.71 (0.56)	0.70 (0.55)
Constant	−2.04 (−0.63)	−4.94 (−1.59)	−0.57 (−0.17)	−1.36 (−0.31)	−1.81 (−0.45)	4.07 (0.92)	4.26 (0.97)
Number of observations	383	383	383	303	303	303	303
Number of groups	28	28	28	28	28	28	28
R ² within	0.15	0.13	0.16	0.13	0.13	0.19	0.194
R ² between	0.06	0.03	0.01	0.09	0.12	0.00	0.004
R ² overall	0.02	0.11	0.05	0.12	0.12	0.05	0.042

Note: The number in parentheses represents the *t*-statistic associated with each coefficient.

^a Independent variables.

* Indicates a significance level at 10%.

** Indicates a significance level at 5%.

*** Indicates a significance level at 1%.

Table 4
Fixed effects 1994–1999

Dependent variable	Provincial real GDP growth rate						
	1 ^a	2 ^a	3 ^a	4 ^a	5 ^a	6 ^a	7 ^a
	Coefficients (1994–1999)	Coefficients (1994–1999)	Coefficients (1995–1999)	Coefficients (1995–1999)	Coefficients (1995–1999)	Coefficients (1994–1998)	Coefficients (1994–1998)
Central tax rate	−1.10 (−3.68)***	−1.06 (−4.07)***	−1.43 (−4.10)***	−1.40 (−3.55)***	−1.32 (−3.78)***	−1.03 (−1.98)*	−0.97 (−1.87)*
Provincial tax rate	0.43 (1.26)	0.44 (1.44)	0.50 (1.25)	0.59 (1.33)	0.61 (1.51)	0.39 (0.92)	0.41 (0.97)
Expenditure decentralization (budgetary)	−1.06 (−1.00)			0.16 (0.13)	0.75 (0.58)	−1.08 (−0.90)	−1.18 (−0.98)
Revenue decentralization (budgetary)		−0.51 (−2.61)***		−0.36 (−2.19)**	−0.33 (−2.01)**	−0.49 (−2.39)**	−0.49 (−2.39)**
Central transfer			0.01 (0.26)	0.00 (−0.02)			
Provincial remittance					0.06 (1.28)		
Expenditure decentralization (extra-budgetary)						−0.92 (−1.26)	
Revenue decentralization (extra-budgetary)							−0.67 (−1.01)
Labor growth rate	0.09 (1.45)	0.11 (1.74)*	0.09 (1.58)	0.09 (1.71)*	0.10 (1.79)*	0.13 (1.66)*	0.12 (1.60)
Investment rate	0.12 (2.74)***	0.12 (2.76)***	0.06 (1.23)	0.07 (1.42)	0.03 (0.69)	0.10 (1.81)*	0.11 (1.84)*
Openness	0.01 (0.58)	0.01 (0.70)	0.00 (0.18)	0.01 (0.29)	0.01 (0.28)	0.01 (0.54)	0.01 (0.53)
Provincial inflation (lagged)	−2.27 (−7.83)***	−2.27 (−8.10)***	−1.59 (−3.68)***	−1.77 (−4.06)***	−1.47 (−3.52)***	−2.01 (−5.90)***	−2.06 (−6.13)***
Constant	20.99 (6.05)***	18.92 (9.47)***	19.19 (7.36)***	19.60 (4.82)***	16.77 (3.93)***	22.39 (4.61)***	22.09 (4.52)***
Number of observations	167	167	139	139	139	139	139
Number of groups	28	28	28	28	28	28	28
R ² within	0.5812	0.5987	0.6073	0.6251	0.631	0.5361	0.5335
R ² between	0.0767	0.0261	0.0183	0.026	0.0465	0.1634	0.1612
R ² overall	0.0737	0.1516	0.1965	0.1661	0.26	0.0003	0.0002

Note: The number in parentheses represents the *t*-statistic associated with each coefficient.

^a Independent variables.

* Indicates a significance level at 10%.

** Indicates a significance level at 5%.

*** Indicates a significance level at 1%.

4.1. First phase: 1979–1993

As Table 3 indicates, provincial economic growth is negatively associated with expenditure decentralization and positively associated with revenue decentralization. That is, further revenue decentralization and expenditure centralization promote growth. The negative association between expenditure decentralization and provincial real GDP growth rate contradicts the conventional wisdom of fiscal decentralization. It is, however, consistent with Zhang and Zou's (1998) result. Hence, their interpretation that "the central government may be in a better position to undertake public investment with nation-wide externalities in the early stages of economic development" is supported by this result. Second, the positive association between revenue decentralization and provincial real GDP growth rate supports the proponents of fiscal decentralization theories. In this case when revenue decentralization is measured by revenue collected at provincial level, this result specifically suggests that assigning more revenue collection to the sub-national levels leads to higher growth, because it stimulates revenue mobilization from local sources and improve overall fiscal position (Shah, 1994). In addition, central tax rate has a significant and consistent positive association with provincial economic growth. This result is counter intuitive, but may be reconciled by the observation that when central government revenue is low, countries are more prone to macroeconomic instability, which may deter growth. (Ahmad, Gao, & Tanzi, 1995; Yusuf, 1994).

If instead of analyzing expenditure and revenue decentralization along, we compare the results with the mean level of the degree of expenditure and revenue decentralization at China's provincial level, we reach rather different conclusions. Since in this phase, revenue is already on average more decentralized than expenditure (2.4% versus 1.9%), further decentralizing revenue and centralizing expenditure suggests a divergence of expenditure and revenue at provincial level is growth promoting (because further revenue decentralization means a higher share than 2.4% and further expenditure centralization means a lower share than 1.9%, hence the divergence). Without changing the fiscal contract regime, and assigning more revenue collection and less expenditure responsibilities to the provinces, allowing more revenue to be remitted to the Center appears to promote provincial economic growth.

In other words, provincial governments appear to be efficient in collecting money, while the central government appears to be more efficient in spending it. With an institutional setting of localized tax collection, the fiscal pattern suggested by the regression result, however, is consistent with the implementation of a fiscal contract system between central and provincial governments, under which the central government contracts tax collection out to its regional agents and claim a proportion of total revenue collected.

The control variables perform in the regression very much as expected, with provincial labor growth rate, investment rate, and openness all being positively associated with provincial economic growth. The lagged provincial inflation level also exhibits a positive association with provincial economic growth and the effect is statistically significant at 5% level when extra-budgetary expenditure and revenue share are absent from the equation (Equations 1, 2, and 3 in Table 3). This is not a surprise for a transition economy like China, which started its economic reform by liberalizing prices sector by sector. In addition, it may also suggest that inflation has a positive effect on growth by spurring investment in physical capital (Tobin portfolio-shift effect), overriding the negative effect of higher transaction costs on growth.

4.2. *Second phase: 1994–1999*

The overhaul of the fiscal contract system in 1994 substantially changed the relationship between provincial economic growth and the degree of fiscal decentralization. Table 4 presents the regression results testing the relationship between fiscal decentralization and growth for the period after 1994 when the tax assignment system was applied. Provincial economic growth rate is shown to have no statistically significant association with expenditure decentralization, and is negatively (rather than positively) associated with revenue decentralization, with a high level of consistency and statistical significance.

There is no significant association found between expenditure decentralization and provincial economic growth. Unlike in the earlier phase, revenue centralization is found to be positively associated with provincial economic growth, and the relationship is highly consistent and statistically significant (Table 4). Given that revenue is already more centralized than expenditure (averaged at 2.4% versus 1.6%) after the 1994 tax assignment reform,¹⁰ further revenue centralization suggests that a divergence of expenditure and revenue at provincial level – which should have implied that more transfers from the Center to provinces – promotes growth. Central transfers, however, are not found to be associated with higher growth in the regression results. A possible explanation lies in the compromise made between the Center and provinces at the onset of implementing tax assignment reform in 1994. That is, since the tax assignment reform would surely largely reduce the revenue collection at the provincial level, wealthier provinces that benefited the most from the contract regime resisted to comply. The Center compromised: for those provinces, whose own revenue would be reduced to lower than their 1993 level under the new tax assignment system, were entitled to a “tax refund” from the central government at a level that would ensure their revenues no lower than the 1993 level.¹¹ With an overwhelming proportion of “tax refund” in central transfers that are actually not at the discretion of the Center to serve macroeconomic stability, central transfers is found to have no significant positive association with provincial growth.

The control variables also performed as expected in the second phase. Provincial labor growth rate and investment rate show a positive signs with economic growth, but the association is weaker in both magnitude and statistical significance than in the earlier period. It is perhaps not surprising that physical inputs (investment and labor force growth) played a more significant role in the early years of the transition. However, as China moved into the 1990s, capital accumulation may have led to diminishing returns. Openness, although still positively associated with economic growth, is not a statistically significant explanatory variable in the second era. A possible explanation lies in the export VAT rebate implemented since 1994. Since the rebate falls solely on the central budget, provinces overstate exports in order to obtain more tax rebate. The degree of openness, measured by the total of exports and imports as a percentage in provincial GDP, may therefore be exaggerated.

Lagged provincial inflation, unlike in the first phase, is negatively related with provincial growth and the association is highly consistent and significant (most of the times at 1%

¹⁰ The split of central and subnational tax administration substantially increased revenue collected at central level and reduced revenue collected at provincial level, and hence reduced provincial revenue share in total revenue from its average level of 2.4% (1979–1993) to 1.7% (1994–1999). At the same time, expenditure were increasingly devolved to subnational levels, hence the degree of expenditure decentralization (provincial expenditure share in total government expenditure) increased from 1.9 in the (1979–1993) to 2.3 (1994–1999).

¹¹ For details, see Wang (1997).

level)—a sign of the overriding negative effect brought by the rise of transaction costs in the 1990s.

5. Summary and conclusion

This study attempts to examine how fiscal decentralization affected provincial economic growth in China. In addition, it investigates how the relationship between fiscal decentralization and provincial growth differed under the two different fiscal regimes that were adopted in China since 1980.

The conventional wisdom that fiscal decentralization – revenue means should match expenditure needs as close as possible at sub-national level – to improve allocative efficiency and promote economic growth does not apply in the case of China. Using a panel data set for China's 30 provinces for the time period from 1979 to 1993 and 1994 to 1999 separately, the results of this study suggest that in both time periods, expenditure and revenue decentralization levels should further diverge to benefit provincial growth.

For the revenue contract system (1979–1993), for example, tax collection was localized and the provinces collected taxes on the Center's behalf as its agents. Revenue decentralization, as measured by tax collection at each province as a percentage in total revenue, was therefore much more decentralized than expenditure because the provinces remitted a proportion (or a fixed amount plus a pre-determined growth rate) of the collected tax revenue to the Center and kept the rest (for detail, see [Ahmad et al., 1995](#); [Bahl & Wallich, 1992](#); [World Bank, 1993](#)). As such, the marginal budgetary revenue collection and marginal budgetary expenditure was highly correlated, and therefore suggests that more revenue decentralization spurs tax collection and allows for more spending (possibly by both central and provincial government) on investment ([Jin, Qian, & Weingast, 1999](#)). While this explanation supports the notion that revenue decentralization stimulates revenue mobilization from local sources ([Shah, 1994](#)), it also suggests that expenditure centralization promotes growth because the central government spends more efficiently than the provinces ([Zhang & Zou, 1998](#)).

The tax assignment reform in 1994 changed the tax administration with the establishment of central tax bureaus to collect central and shared-taxes, and sub-national tax bureaus to collect sub-national exclusive taxes. Revenue became more centralized than expenditure, with the expenditure gaps in provinces bridged by central transfers to the provinces. This is a fiscal pattern that is more comparable to other countries. The regression results suggest that at given level of expenditure decentralization, more revenue centralization contributes to growth. This finding supports the view that the Center is in a better position to allocate budgetary resources for horizontal balance, macroeconomic stability, and investment in projects of national significance.

This study adds to a growing body of evidence that under certain circumstances, fiscal decentralization can be detrimental to economic growth ([Steinfeld, 1999](#); [Yang, 1997](#); [Young, 2000](#); [Zhang & Zou, 1998](#)). The results of this study also underscore the fundamental proposition that institution matters. The effects of fiscal decentralization in any given case depend critically on the nature of the fiscal institutions and political system in place.

Acknowledgements

For comments and suggestion, we are grateful to Eiji Tajika and the participants of International Symposium on Tax Policy and Reform in Asia organized by Hitotsubashi University. We are responsible for all remaining errors.

Appendix A

See Tables A.1 and A.2.

Table A.1

Fiscal decentralization by province (1979–1993)

Expenditure decentralization by province (1979–1993)						Revenue decentralization by province (1979–1993)					
Province	Obs	Mean	S.D.	Min	Max	Provinces	Obs	Mean	S.D.	Min	Max
1 Beijing	15	1.7	0.2	1.2	2.0	Beijing	15	3.1	0.8	2.0	4.5
2 Tianjin	15	1.3	0.2	1.0	1.9	Tianjin	15	2.5	0.7	1.5	3.5
3 Hebei	15	2.4	0.2	2.1	2.9	Hebei	15	2.8	0.5	2.0	4.0
4 Shanxi	15	1.7	0.2	1.4	1.9	Shanxi	15	1.7	0.2	1.2	1.8
5 Neimeng	15	1.8	0.1	1.5	2.0	Neimeng	15	0.8	0.3	0.4	1.3
6 Liaoning	15	3.1	0.5	2.4	4.0	Liaoning	15	5.5	1.3	4.2	8.2
7 Jilin	15	1.8	0.3	1.4	2.3	Jilin	15	1.4	0.4	0.9	2.0
8 Heilongjiang	15	2.5	0.2	2.1	2.9	Heilongjiang	15	2.5	1.2	1.3	5.6
9 Shanghai	15	2.1	0.3	1.6	2.5	Shanghai	15	9.7	3.9	5.3	15.1
10 Jiangsu	15	2.7	0.3	2.1	3.1	Jiangsu	15	4.9	0.5	4.0	5.6
11 Zhejiang	15	2.0	0.4	1.4	2.5	Zhejiang	15	2.9	0.8	0.4	3.7
12 Anhui	15	1.7	0.2	1.4	2.1	Anhui	15	1.7	0.1	1.5	2.0
13 Fujian	15	1.6	0.3	1.2	2.2	Fujian	15	1.5	0.5	0.9	2.5
14 Jiangxi	15	1.5	0.1	1.3	1.6	Jiangxi	15	1.2	0.2	0.9	1.5
15 Shandong	15	3.0	0.5	2.3	3.7	Shandong	15	4.0	0.7	2.9	5.7
16 Henan	15	2.6	0.2	2.2	3.0	Henan	15	2.8	0.2	2.4	3.2
17 Hubei	15	2.4	0.2	2.1	2.7	Hubei	15	2.8	0.2	2.5	3.1
18 Hunan	15	2.2	0.2	1.9	2.5	Hunan	15	2.4	0.3	2.0	2.9
19 Guangdong	15	3.6	1.2	2.1	6.3	Guangdong	15	4.2	1.4	2.7	8.0
20 Guangxi	15	1.7	0.2	1.4	2.0	Guangxi	15	1.3	0.4	0.8	2.2
21 Hainan	15	0.3	0.2	0.1	0.7	Hainan	15	0.2	0.2	0.1	0.7
22 Sichuan	15	3.4	0.6	2.7	4.3	Sichuan	15	3.5	0.6	2.7	4.7
23 Guizhou	15	1.3	0.1	1.0	1.5	Guizhou	15	0.9	0.3	0.5	1.4
24 Yunnan	15	2.2	0.6	1.4	3.8	Yunnan	15	1.9	1.1	1.0	4.7
25 Tibet	0					Tibet	0				
26 Shaanxi	15	1.5	0.1	1.4	1.7	Shaanxi	15	1.3	0.2	0.9	1.7
27 Gansu	15	1.2	0.1	1.0	1.4	Gansu	15	1.1	0.3	0.8	1.8
28 Qinghai	15	0.5	0.1	0.4	0.6	Qinghai	15	0.2	0.1	0.1	0.3
29 Ningxia	15	0.5	0.1	0.4	0.6	Ningxia	15	0.2	0.0	0.1	0.3
30 Xinjiang	15	1.4	0.1	1.2	1.6	Xinjiang	15	0.5	0.2	0.1	0.8
		1.9	0.3	1.5	2.4			2.4	0.6	1.6	3.5

Table A.2

Fiscal decentralization by province (1994–1999)

Expenditure decentralization by province (1994–1999)						Revenue decentralization by province (1994–1999)					
Province	Obs	Mean	S.D.	Min	Max	Province	Obs	Mean	S.D.	Min	Max
1 Beijing	6	2.4	0.4	1.7	2.7	Beijing	6	1.9	0.6	0.9	2.5
2 Tianjin	6	1.3	0.1	1.2	1.4	Tianjin	6	1.0	0.0	1.0	1.1
3 Hebei	6	2.8	0.1	2.7	2.9	Hebei	6	2.0	0.1	1.8	2.1
4 Shanxi	6	1.6	0.1	1.4	1.7	Shanxi	6	1.1	0.1	1.0	1.2
5 Neimeng	6	1.6	0.0	1.5	1.6	Neimeng	6	0.7	0.0	0.7	0.8
6 Liaoning	6	3.8	0.2	3.5	4.0	Liaoning	6	2.8	0.2	2.4	3.0

Table A.2 (Continued)

Expenditure decentralization by province (1994–1999)							Revenue decentralization by province (1994–1999)					
	Province	Obs	Mean	S.D.	Min	Max	Province	Obs	Mean	S.D.	Min	Max
7	Jilin	6	1.8	0.0	1.8	1.8	Jilin	6	2.5	3.7	0.9	10.1
8	Heilong-jiang	6	2.5	0.1	2.4	2.6	Heilong-jiang	6	1.6	0.1	1.5	1.7
9	Shanghai	6	4.0	0.4	3.3	4.4	Shanghai	6	3.7	0.2	3.3	3.9
10	Jiangsu	6	3.9	0.2	3.7	4.3	Jiangsu	6	2.9	0.2	2.6	3.0
11	Zhejiang	6	2.6	0.0	2.6	2.7	Zhejiang	6	1.9	0.1	1.8	2.1
12	Anhui	6	2.1	0.3	1.6	2.3	Anhui	6	1.4	0.2	1.0	1.6
13	Fujian	6	2.4	0.1	2.1	2.5	Fujian	6	1.9	0.1	1.8	1.9
14	Jiangxi	6	1.6	0.0	1.6	1.7	Jiangxi	6	1.0	0.0	0.9	1.0
15	Shandong	6	4.2	0.3	3.8	4.5	Shandong	6	3.2	0.4	2.6	3.6
16	Henan	6	3.0	0.1	2.9	3.2	Henan	6	2.0	0.1	1.8	2.2
17	Hubei	6	2.5	0.1	2.4	2.6	Hubei	6	1.6	0.1	1.5	1.7
18	Hunan	6	2.6	0.1	2.4	2.7	Hunan	6	1.6	0.1	1.5	1.8
19	Guangdong	6	7.5	0.2	7.2	7.7	Guangdong	6	6.3	0.3	5.7	6.7
20	Guangxi	6	1.9	0.2	1.7	2.1	Guangxi	6	1.2	0.0	1.1	1.3
21	Hainan	6	0.6	0.1	0.4	0.7	Hainan	6	0.4	0.1	0.3	0.5
22	Sichuan	6	4.1	0.1	3.9	4.1	Sichuan	6	2.7	0.1	2.5	2.8
23	Guizhou	6	1.3	0.0	1.2	1.3	Guizhou	6	0.6	0.0	0.6	0.7
24	Yunnan	6	3.3	0.3	2.9	3.5	Yunnan	6	1.6	0.1	1.5	1.8
25	Tibet	5	0.4	0.0	0.4	0.5	Tibet	5	0.1	0.1	0.0	0.3
26	Shaanxi	6	1.5	0.0	1.5	1.6	Shaanxi	6	0.9	0.1	0.8	0.9
27	Gansu	6	1.2	0.0	1.1	1.2	Gansu	6	0.5	0.0	0.5	0.6
28	Qinghai	6	0.4	0.0	0.4	0.4	Qinghai	6	0.1	0.0	0.1	0.1
29	Ningxia	6	0.4	0.0	0.3	0.4	Ningxia	6	0.2	0.0	0.1	0.2
30	Xinjiang	6	1.3	0.1	1.2	1.4	Xinjiang	6	0.6	0.0	0.5	0.7
			2.3	0.1	2.2	2.5			1.7	0.2	1.4	2.1

References

- Ahmad, E., Gao, Q., & Tanzi, V. (Eds.). (1995). *Reforming China's public finances*. Washington, D.C. International Monetary Fund.
- Akai, N., & Sakata, M. (2002). Fiscal decentralization contributes to economic growth: Evidence from state-level cross-section data for the United States. *Journal of Urban Economics*, 52, 93–108.
- Bahl, R. (1999). *Fiscal policy in China: Taxation and intergovernmental fiscal relations*. The 1990 Institute.
- Bahl, R., Wallich, C. (1992). Intergovernmental fiscal relations in China. *Research Working Paper 863*. Washington, D.C.: Development Research Group, World Bank.
- Barro, R. (1990). Government spending in a simple model of endogenous growth. *Journal of Political Economy*, 98, 103–125.
- Bird, R. (1992). *Tax policy and economic development*. Baltimore: Johns Hopkins University Press.
- Brueckner, J. K. (2000). Fiscal decentralization in developing countries: The effects of local corruption and tax evasion. *Annals of Economics and Finance*, 1(1), 1–18.
- China Finance Statistical Yearbook. Various issues.
- China Statistical Yearbook. Various issues.
- Davoodi, H., & Zou, H. F. (1998). Fiscal decentralization and economic growth: A cross-country study. *Journal of Urban Economics*, 43, 244–257.
- Ebel, R., Yilmaz, S. (2001). On the measurement and impact of fiscal decentralization. *Policy Research Working Paper 2809*. Washington D.C.: The World Bank.
- Feder, G. (1983). On exports and economic growth. *Journal of Development Economics*, 12, 59–73.
- Hseh, T., Li, Q., & Liu, S. C. (Eds.). (1993). *China's provincial statistics 1949–89*. Boulder: Westview Press.
- Jin, H. H., Qian, Y. Y., & Weingast, B. R. (1999). *Regional decentralization and fiscal incentives: federalism, Chinese style*. Mimeo, Stanford University.

- Jin, J., & Zou, H. F. (1999). *Fiscal decentralization and economic growth*. Mimeo, Washington, D.C. World Bank.
- Jin, J., Zou, H. F. (2003). Soft budget constraints on local government in China. In Rodden, Eskeland, & Litvack (Eds.), *Fiscal decentralization and the challenge of hard budget constraints*. MIT Press.
- Mello, L. R. D. (2000). Fiscal decentralization and intergovernmental fiscal relations: A cross-country analysis. *World Development*, 28(2), 365–380.
- Musgrave, R. A. (1983). Whose should tax, where, and what? In C., McLure, Jr., (Ed.). *Tax assignment in federal countries*. Canberra: Australian National University Press.
- Oates, W. (1972). *Fiscal federalism*. New York: Harcourt Brace Jovanovich.
- Oates, W. (1993). Fiscal decentralization and economic development. *National Tax Journal*, XLVI, 237–243.
- Oi, J. C. (1992). Fiscal reform and the economic foundations of local state corporatism in China. *World Politics*, 45(1), 99–126.
- Prud'homme, R. (1995). On the dangers of decentralization. *The World Bank Research Observer*, 10(2), 201–220.
- Qian, Y. Y. (1999). The institutional foundations of China's market transition. *Annual World Bank Conference on Development Economics*.
- Qian, Y. Y., & Weingast, B. R. (1997). Federalism as a commitment to preserving market incentives. *The Journal of Economic Perspectives*, 11(4), 83–92.
- Shah, A. (1994). The reform of intergovernmental fiscal relations in developing and emerging market economies. *Policy and Research Series*, 23. The World Bank, Washington, D.C.
- Steinfeld, E. (1999). *Forging reform in China: The fate of state-owned industry*. Cambridge: Cambridge University Press.
- Tanzi, V. (1996). Fiscal federalism and decentralization: A review of some efficiency and macroeconomic aspects. *Annual World Bank Conference on Development Economics 1995*.
- Tiebout, C. (1956). A pure theory of local expenditures. *Journal of Political Economy*, 64, 416–424.
- Wang, S. G. (1997). China's fiscal reform in 1994: An initial assessment. *Asian Survey*, 37, 801–817.
- Wong, C. P. W. (1991). Central-local relations in an era of fiscal decline: The paradox of fiscal decentralization in post-Mao China. *The China Quarterly*, 128, 691–715.
- World Bank. (1989). China: Revenue mobilization and tax policy: issues and options. *Report No. 7605-CHA*. Washington, D.C.
- World Bank. (1993). China: Budgetary policy and intergovernmental fiscal relations. *Report No. 11094-CHA*. Washington, D.C.
- Xie, D. Y., Zou, H. F., & Davoodi, H. (1999). Fiscal decentralization and economic growth in the United States. *Journal of Urban Economics*, 45, 228–239.
- Yang, D. L. (1997). *Beyond Beijing: Liberalization and the regions in China*. London: Routledge.
- Young, A. (2000). The Razor's Edge: Distortions and incremental reform in the People's Republic of China. *NBER Working Paper* 7828.
- Yusuf, S. (1994). China's macroeconomic performance and management during transition. *Journal of Economic Perspective*, 8(2), 71–92.
- Zhang, T., & Zou, H. F. (1998). Fiscal decentralization, public spending and economic growth in China. *Journal of Public Economics*, 67, 221–240.

Finance and Income Inequality: What Do the Data Tell Us?

Author(s): George R. G. Clarke, Lixin Colin Xu, Heng-fu Zou

Source: *Southern Economic Journal*, Vol. 72, No. 3 (Jan., 2006), pp. 578-596

Published by: Southern Economic Association

Stable URL: <http://www.jstor.org/stable/20111834>

Accessed: 11/03/2010 11:19

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at <http://www.jstor.org/page/info/about/policies/terms.jsp>. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at <http://www.jstor.org/action/showPublisher?publisherCode=sea>.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.



Southern Economic Association is collaborating with JSTOR to digitize, preserve and extend access to *Southern Economic Journal*.

Finance and Income Inequality: What Do the Data Tell Us?

George R. G. Clarke,* Lixin Colin Xu,† and Heng-fu Zou‡

Although there are distinct conjectures about the relationship between finance and income inequality, little empirical research compares their explanatory power. We examine the relationship between finance and income inequality for 83 countries between 1960 and 1995. Because financial development might be endogenous, we use instruments from the literature on law, finance, and growth to control for this. Our results suggest that, in the long run, inequality is less when financial development is greater, consistent with Galor and Zeira (1993) and Banerjee and Newman (1993). Although the results also suggest that inequality might increase as financial sector development increases at very low levels of financial sector development, as suggested by Greenwood and Jovanovic (1990), this result is not robust. We reject the hypothesis that financial development benefits only the rich. Our results thus suggest that in addition to improving growth, financial development also reduces inequality.

JEL Classification: D3, G2, O1

1. Introduction

Recent studies have shown that financial sector development boosts economic growth (Levine 1997b).¹ But many people worry that financial development benefits only the rich and powerful. Because financial markets are fraught with adverse selection and moral hazard problems, borrowers need collateral. The poor, who do not have this, might, therefore, find it difficult to get loans even when financial markets are well developed. In contrast, the rich who do have property that can be used as collateral might benefit as the financial sector develops. If financial development improves access for the rich, but not the poor, it might worsen inequality.

But this might not be the case. As the financial sector grows, the poor, who were previously excluded from getting loans, might gain access to it. In this respect, finance might be an equalizer for people with talents, ambition, and persistence. Rajan and Zingales (2003, p. 92) argue that the revolution in financial markets is “opening the gates of the aristocratic clubs to everyone,” as

* Research Department, World Bank, 1818 H Street, NW, Washington, DC 20433; E-mail gclarke@worldbank.org; corresponding author.

† Research Department, World Bank, 1818 H Street, NW, Washington, DC 20433; Guanghua School of Management, Peking University, Beijing 100871, China; E-mail lxu1@worldbank.org.

‡ Research Department, World Bank, 1818 H Street, NW, Washington, DC 20433. E-mail hzou@worldbank.org.

We are grateful for suggestions of two anonymous referees. We thank Jerry Caprio, Robert Cull, Ross Levine, and Mattias Lundberg for comments. We are especially grateful to Thorsten Beck for early collaboration in this project, help in collecting and compiling data, and many useful discussions. The findings, interpretations, and conclusions expressed herein are those of the authors and do not necessarily reflect the views of the International Bank for Reconstruction and Development/The World Bank and its affiliated organizations, or those of the Executive Directors of The World Bank or the governments they represent. The World Bank does not guarantee the accuracy of the data included in this work.

Received October 2003; accepted June 2005.

¹ For the relationship between financial development and growth see, among others, Beck, Levine, and Loayza (2000), Levine, Loayza, and Beck (2000), and Rousseau and Wachtel (2000).

witnessed by the observation that, “in 1929, 70% of the income of the top 0.01% of income earners in the United States came from holding of capital In 1998, wages and entrepreneurial income made up 80% of the income of the top 0.01% of income earners in the United States, and only 20% came from capital.”

Consistent with the idea that financial development might benefit the poor, several theoretical models suggest that income inequality will be lower when financial markets are better developed (Banerjee and Newman 1993; Galor and Zeira 1993). These models show that when investments are indivisible, financial market imperfections perpetuate the initial wealth distribution, resulting in a negative relationship between financial development and income inequality even in the long run.

Although the relation between inequality and financial development could be linear, it is also possible that different mechanisms dominate at different levels of financial sector development, leading to a nonlinear relationship between financial sector development and inequality. Greenwood and Jovanovic (1990) show how financial and economic development might give rise to an inverted U-shaped relationship between income inequality and financial sector development. In their model, income inequality first rises as the financial sector develops but later declines as more people gain access to the system.

The relation between financial development and income distribution is important for policy makers—policy makers want to know how policies affect inequality as well as how they affect growth. Although recent work has established a robust link between financial sector development and economic growth (Levine 1997b), less work has focused on the relation between financial sector development and inequality. Understanding this relationship will allow policy makers to assess whether financial development will improve inequality and when it might be useful in doing so. Because different theoretical models give different predictions about the distributional impact of financial development on inequality, empirical investigation is needed to distinguish between the competing conjectures.²

This paper analyzes the relation between the distributional impact of financial intermediary development and income distribution using data from developing and developed countries from between 1960 and 1995. Specifically, we analyze whether financial intermediary development affects income inequality and whether the impact depends on the level of financial development. Because the different mechanisms might be more powerful at different levels of financial sector development, we allow the relationship to be nonlinear. Further, because causation could run either from financial sector development to inequality or from inequality to financial sector development, we control for endogeneity using instruments for financial sector development suggested in the financial sector development–growth literature (see, for example, Levine 1997a, 1999).

Our results show that inequality decreases as financial markets deepen, consistent with Galor and Zeira (1993) and Banerjee and Newman (1993). Although some weak evidence suggests that at low

² Li, Squire, and Zou (1998) and Li, Xu, and Zou (2000) include financial sector development in regressions looking at factors that affect income inequality. This paper, however, differs from Li, Squire, and Zou (1998) and Li, Xu, and Zou (2000) in several ways. First, neither of these earlier papers is primarily concerned with the impact of financial sector development on inequality. Li, Squire, and Zou (1998) focus on explaining international and intertemporal variations in income inequality, whereas Li, Xu, and Zou (2000) focus on the relationship between corruption and inequality (and growth). They do not try to distinguish the various hypotheses as we do here; that is, they assume a linear relationship, and given their focus, they do not run a battery of specifications to examine the robustness of their results. In addition, they do not deal with the endogeneity of financial development, use a different measure of financial sector development that measures financial development less precisely (M2 over GDP), and only include results from a pooled cross section.

levels of financial development inequality might increase as financial sector development increases, that is, that there is an inverted U-shaped relation between financial sector development and income inequality, as suggested by Greenwood and Jovanovic (1990), this second result is not highly robust. We strongly reject the hypothesis that financial development benefits only the rich: We do not find a positive and significant relation between financial development and inequality after controlling for the endogeneity of financial sector development.

In the next section, we briefly review the theoretical literature on the relation between income inequality and financial sector development. We then discuss the data that we use to test the theoretical hypotheses in Section 3. After discussing the empirical specification and some estimation issues in Section 4, we present empirical results in Section 5 and conclude in Section 6.

2. Theoretical Perspectives on Finance and Inequality

Although most economists would not expect financial development to widen income inequality in the long run, the popular press, some literature, and Marxist theory often depict financiers as greedy middlemen who serve only the interest of the rich and well connected. Indeed, these views are so common that the first chapter of a recent book defending the free-market system by two famous economists, Rajan and Zingales (2003), is entitled “Does finance benefit only the rich?”

One plausible reason why financial development might benefit the rich, especially when institutions are weak, is that the financial system might mainly channel money to the rich and well connected, who are able to offer collateral and who might be more likely to repay the loan, while excluding the poor.³ As financial sectors become more developed, they might lend more to rich households but continue to neglect the poor who are unable to provide collateral. As a result, even as the financial sector develops, the poor remain unable to migrate to urban areas, invest in education, or start new businesses. This tendency might be reinforced if the rich are able to prevent new firms from getting access to finance, preventing them from entering, and reducing the ability of the poor to improve their economic lot. If this were the case, we would expect to see a positive relation between financial development and income inequality—at least at some levels of financial development. We call this story the inequality-widening hypothesis of financial development.

Although the previous arguments suggest that high-income households might benefit more from financial sector development than low-income households, this is not necessarily the case. As financial markets become deeper, and access to finance improves, households that did not previously have access to finance might be the main beneficiaries. Because poor households cannot invest in human and physical capital or bear the start-up costs associated with starting a new business using only their own resources, they will be unable to do so unless they can borrow. In contrast, rich households are able to draw on their own resources for investment whatever the level of financial sector development. Therefore, capital constraints might be less binding for rich households at any level of financial sector development, and so they might gain less when these constraints are loosened.

Several recent theoretical models have formalized this intuition, suggesting that capital market imperfections might increase income inequality during economic development. Banerjee and Newman (1993) and Galor and Zeira (1993) suggest that capital market imperfections and

³ This paragraph mainly draws from Rajan and Zingales (2003), chapter 1.

indivisibilities in investment in human or physical capital may lead to divergence of income for the rich and the poor even in the long run. Further, depending on the initial wealth distribution, these imperfections might mean that income inequality persists even in the long run.

Galor and Zeira (1993) construct a two-sector model with bequests between generations, where agents who make an indivisible investment in human capital can work in a skill-intensive sector. However, given capital market imperfections, only individuals with bequests larger than the investment amount or who can borrow will be able to make this investment. This results in income inequality that is perpetuated through bequests to the next generation. In their model, an economy with capital market imperfections and an initially unequal distribution of wealth will maintain this inequality and grow more slowly than a similar economy with a more equitable initial distribution of wealth. Similarly, Banerjee and Newman (1993) construct a three-sector model, in which two of the technologies require indivisible investment. Because of capital market imperfections, only rich agents can borrow enough to run these indivisible, higher-return technologies. Once again, the initial distribution of wealth has long-run effects on income distribution and growth in the presence of capital market imperfections. With all else remaining equal, these models suggest that countries with larger capital market imperfections, that is, higher hurdles to borrow funds to finance indivisible investment, should have higher income inequality. Consequently, we should observe a negative relationship between financial development and income inequality. We call this hypothesis the inequality-narrowing hypothesis of financial development.

Offering a related, but different, perspective on these basic ideas, Greenwood and Jovanovic (1990) present a theoretical model that has elements of both ideas. In their model, agents operate the more profitable, but more risky, of two technologies only when they can diversify risk by investing in financial intermediary coalitions. However, the fixed costs (e.g., membership fees) associated with these coalitions prevent low-income individuals from joining them. Assuming that poor individuals save less and thus accumulate wealth more slowly, income differences between (high-income) members of intermediary coalitions and (low-income) outsiders will widen, resulting in an increase in income inequality. However, because the entrance fee is fixed, all agents eventually join these coalitions, resulting in an eventual reversal in the upward trend. Consequently, Greenwood and Jovanovic's (1990) model predicts a hump or inverted U-shaped relationship between income inequality and financial sector development, with income inequality first increasing and then decreasing before eventually stabilizing in the long run as more people join financial coalitions. We call this hypothesis the inverted U-shaped hypothesis of financial development.

There are, thus, quite different predictions about the relation between financial intermediaries and income inequality. Yet distinguishing among these three hypotheses is important. If the inequality-narrowing hypothesis is correct, improving the access to finance would reduce inequality and benefit low-income households in rich and in poor countries alike. In contrast, if the inverted U-shaped hypothesis is correct, improving the access to finance might initially worsen income inequality in poor countries, improving it only after the country has passed a certain stage of financial sector development. Finally, if the inequality-widening hypothesis is true, some countries might be trapped in a high-inequality world that would be only worsened by financial sector development. In what follows, we use data from a broad cross section of countries between 1960 and 1995 to assess the empirical validity of the different hypotheses.

It is perhaps useful to note that the inverted U-shaped hypothesis concerns a situation in which the empiricist observes the evolution of income inequality and financial development during the development process. Thus, the relationship would be most likely to show up in short- or medium-run time-series or panel data. In contrast, testing the inequality-widening and

inequality-narrowing hypotheses might require long-run data, such as cross-sectional data based on long time series.

3. Data

This section describes our indicators and data for financial intermediary development and income inequality as well as the set of conditioning information. Table 1 presents descriptive statistics and correlations.⁴ The income inequality data are based on a new data set of Gini coefficients compiled by Deininger and Squire (1996) and extended by Lundberg and Squire (2000). Although the original data set contained over 2600 observations, Deininger and Squire (1996) and Lundberg and Squire (2000) limited the data set by imposing several quality conditions. First, all observations had to be from national household surveys for expenditure or income. Second, the coverage had to be representative of the national population. Third, all sources of income and uses of expenditure had to be accounted for, including own consumption.⁵

To explore whether there is an inverted U-shaped relationship between economic development and income inequality, as proposed by Kuznets (1955), we regress the logarithm of the Gini coefficient on the log of real *per capita* GDP and its square. Figure 1 shows the result for the panel sample. The graph suggests the existence of an inverted U-shaped curve. However, this graph does not control for alternative explanations of income inequality, such as financial depth.

The recent literature on the relationship between financial intermediary development and economic growth has developed several indicators to proxy for the ability of financial intermediaries to identify profitable projects, monitor and control managers, ease risk management, and facilitate resource mobilization. We concentrate on credit to the private sector by financial intermediaries over GDP (private credit). This indicator, which comprises credit to private firms and households from banks and nonbank financial intermediaries (but which excludes central banks as lenders and government and state-owned enterprises as borrowers), seems a good proxy variable for the extent to which private sector agents have access to financial intermediation (as in Greenwood and Jovanovic 1990) or access to loans (as in Banerjee and Newman 1993, Galor and Zeira 1993). Many recent studies that have looked at the effect of financial sector development on economic growth have used this variable as a measure of financial sector development, showing that growth is faster in countries where private credit is higher (see, for example, Beck, Levine, and Loayza 2000; Levine, Loayza, and Beck 2000).

To assess the robustness of results, we use a second measure of financial development: claims on the nonfinancial domestic sector by deposit money banks divided by GDP (bank assets). In contrast to private credit, this measure excludes credits by nonbank financial intermediaries and includes credit to governments and state-owned enterprises.

⁴ The sample includes Algeria, Argentina, Australia, Austria, Belgium, Burkina Faso, the Bahamas, Bolivia, Botswana, Brazil, Cameroon, Canada, Chile, China, Colombia, Costa Rica, Cote d'Ivoire, Cyprus, Denmark, Dominican Republic, Ecuador, Egypt, El Salvador, Finland, France, Gabon, Gambia, Germany, Ghana, Greece, Guatemala, Guinea Bissau, Guyana, Honduras, Hong Kong (China), Indonesia, India, Ireland, Italy, Jamaica, Japan, Jordan, Kenya, Korea, Luxembourg, Madagascar, Malawi, Mali, Malaysia, Mexico, Morocco, Netherlands, New Zealand, Nicaragua, Niger, Nigeria, Norway, Pakistan, Panama, Paraguay, Peru, Philippines, Portugal, Senegal, Sierra Leone, Singapore, South Africa, Spain, Sri Lanka, Sudan, Sweden, Thailand, Trinidad and Tobago, Tunisia, Turkey, Uganda, United Kingdom, United States of America, Venezuela, Zambia, and Zimbabwe.

⁵ To account for different sampling methods, we adjust the data using a method suggested by Deininger and Squire (1996) and also applied by Li, Squire, and Zou (1998) and Lundberg and Squire (2000). Specifically, Deininger and Squire (1996) find a systematic difference of 6.6 points between the means of income-based and expenditure-based Gini coefficients. We, therefore, add 6.6 points to the expenditure-based Gini coefficients.

Table 1. Descriptive Statistics

	Gini Coef.	Private Credit	Bank Assets	Initial GDP <i>per Capita</i>	Risk of Exprop.	Ethno-linguistic Fract.	Gov't Cons.	Inflation Rate	Mod Sect. Val. Add.
Number of observations	205	205	205	205	205	163	205	205	205
Mean	38.4	44.9	42.4	5552	7.3	0.25	14.3	1.15	86.0
Minimum	22.4	1.6	2.5	160	3.3	0.00	5.6	1.00	43.3
Maximum	61.1	202.8	132.1	20,367	10	0.86	27.9	3.22	99.6
Gini coefficient	1.00								
Private credit	−0.38 (0.00)	1.00							
Bank assets	−0.48 (0.00)	0.86 (0.00)	1.00						
Initial GDP <i>per capita</i>	−0.59 (0.00)	0.69 (0.00)	0.61 (0.00)	1.00					
Risk of expropriation	−0.59 (0.00)	0.64 (0.00)	0.66 (0.00)	0.78 (0.00)	1.00				
Ethnolinguistic fractionalization	0.11 (0.15)	−0.38 (0.00)	−0.35 (0.00)	−0.45 (0.00)	−0.50 (0.00)	1.00			
Government consumption	−0.48 (0.00)	0.33 (0.00)	0.37 (0.00)	0.61 (0.00)	0.50 (0.00)	−0.36 (0.00)	1.00		
Inflation rate	0.32 (0.00)	−0.28 (0.00)	−0.30 (0.00)	−0.22 (0.00)	−0.21 (0.00)	−0.03 (0.72)	−0.21 (0.00)	1.00	
Modern sector value added/GDP	−0.25 (0.00)	0.55 (0.00)	0.54 (0.00)	0.68 (0.00)	0.67 (0.00)	−0.68 (0.00)	0.47 (0.00)	−0.06 (0.43)	1.00

Gini, measurement-adjusted Gini coefficient from Deininger and Squire (1996) and Lundberg and Squire (2003).

GDP *per capita*, real *per capita* GDP; Source: Loayza et al. (1999).

Private Credit, claims on the private sector by financial institutions divided by GDP. Source: Beck, Demirgüç-Kunt, and Levine (2000).

Bank Assets, claims on domestic nonfinancial sector by deposit money banks divided by GDP. Source: Demirgüç-Kunt and Levine (2000).

Risk of Expropriation, index indicating risk of expropriation through confiscation or forced nationalization. Higher values indicate that risk is lower. Source: PRS Group (2003).

Ethnolinguistic Fractionalization, average value of five indices of ethnolinguistic fractionalization, with values ranging from 0 to 1, with higher values indicating greater fractionalization. Source: Levine, Loayza, and Beck (2000).

Government Consumption, government consumption as share of GDP. Source: World Bank (2004).

Inflation Rate, log difference of Consumer Price Index. Source: International Monetary Fund (2002).

Modern Sector Value Added/GDP, value added of service and industrial sectors as share of GDP. Source: World Bank (2004).

We use private credit rather than the ratio of money and quasimoney (M2) to GDP (M2), a measure commonly used to measure financial sector development (King and Levine 1993; Levine and Zervos 1998), for several reasons. First, the ratio of M2 to GDP includes the liabilities of central banks in addition to banks and other financial intermediaries. Second, it includes credit to governments and state-owned enterprises. Because of this, it is a less clean measure of financial sector development than private credit.

Our sample shows a large variation in financial intermediary development. Private credit ranges from 2% of GDP in Uganda (1990–1995) to over 200% in Japan (1990–1995). The indicators of financial intermediary development are positively and significantly correlated (see Table 1). The pairwise correlations indicate that income inequality is lower in countries with deeper financial markets; financial sector development is significantly and negatively correlated with the Gini coefficient. Plotting the logarithm of the Gini coefficient and its fitted value (from the regression of the logarithm of the

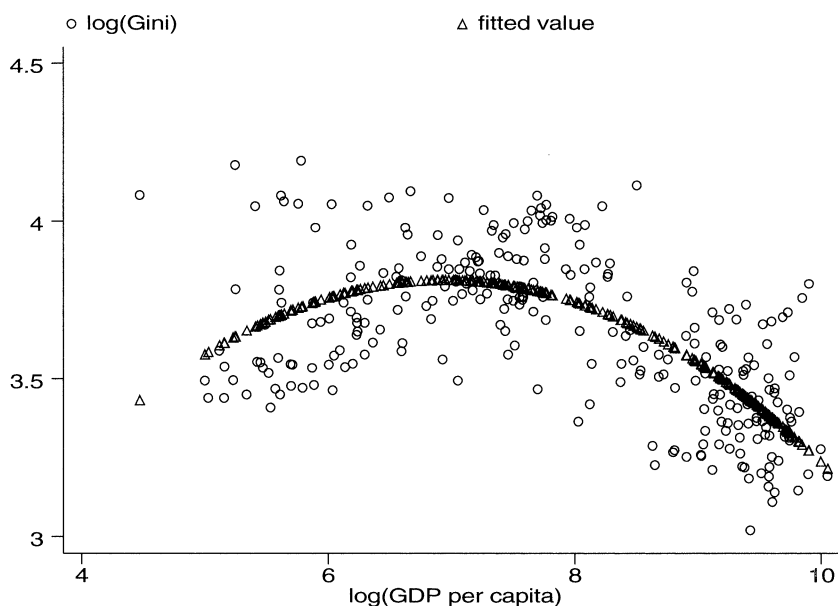


Figure 1. Log(Gini) and log(GDP *per capita*) in a panel of 91 countries. The fitted line is from a regression of log(Gini) on the log of real *per capita* GDP and its square. All data are averaged over seven 5-year periods between 1960 and 1995.

Gini coefficient on the logarithm of private credit) against the logarithm of private credit, Figure 2 suggests a negative, and possibly nonlinear, relation between the two.

4. Empirical Framework

To further explore the relationship between financial intermediary development and income inequality, we estimate the following regression:

$$\ln(\text{Gini Coef.}_{it}) = \alpha_0 + f(\text{Finance}_{it}) + \alpha_2 CV_{it} + \varepsilon_{it}. \quad (1)$$

As discussed previously, claims on the private sector by financial institutions as percentage of GDP (private credit) and claims on the nonfinancial domestic sector by deposit money banks divided by GDP (bank assets) are the measures of financial sector development. The focus of the analysis is $f(\text{Finance}_{it})$ which, based on earlier discussions, we assume has the following functional form:

$$\alpha_{11} \text{Finance}_{it} + \alpha_{12} \text{Finance}_{it}^2.$$

The inequality-narrowing hypothesis predicts $\alpha_{11} < 0$ and $\alpha_{12} = 0$, the inequality-widening hypothesis predicts $\alpha_{11} > 0$ and $\alpha_{12} = 0$, and the inverted U-shape hypothesis predicts $\alpha_{11} > 0$ and $\alpha_{12} < 0$.

In addition to the financial sector variables, we include several variables to control for other factors that might affect inequality. Specifically, we include linear and squared terms of the log of (initial) real *per capita* GDP to control for a direct “Kuznets effect” of economic development on income inequality that is independent of financial intermediary development. Once controlling for initial GDP, $f(\text{Finance}_{it})$ captures the effects of finance on steady-state inequality. If the real data do not reflect steady-state situations, initial GDP would capture whatever has been achieved by the force of convergence. However, because *per capita* GDP is highly correlated with financial sector development,

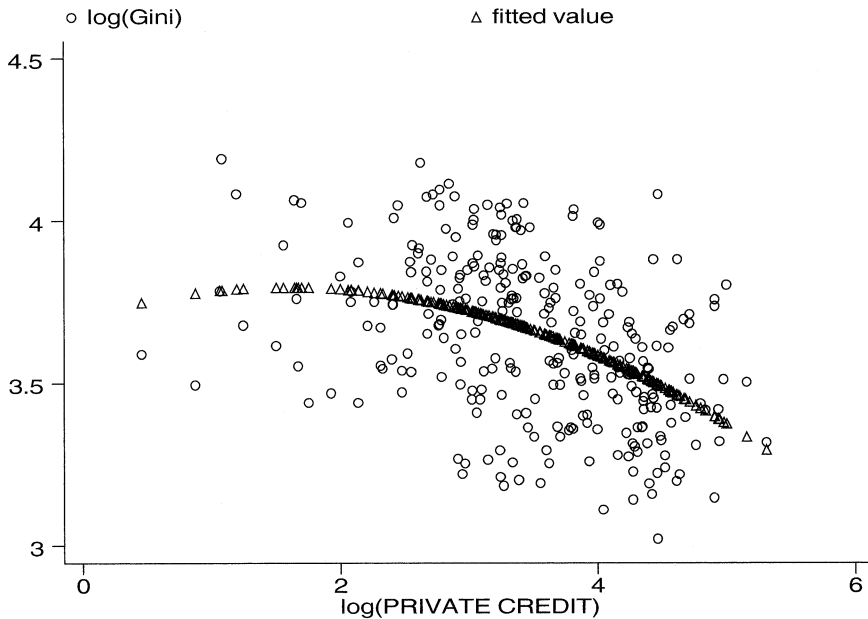


Figure 2. Log(Gini) against log(Private Credit) in a panel of 91 countries. The fitted line is from a regression of log(Gini) on the log of Private Credit and its square. All data are averaged over seven 5-year periods between 1960 and 1995.

to make sure that our test of the three hypotheses is robust with respect to the multicollinearity between GDP *per capita* and financial development, we also estimate the model omitting these variables.

In addition to these measures, we include several additional control variables. We include the inflation rate, conjecturing that monetary instability hurts the poor and the middle class relatively more than the rich because the latter have better access to financial instruments that allow them to hedge their exposure to inflation.⁶ We, therefore, expect inflation to have a positive coefficient.

Additionally, we include measures of government consumption, ethnolinguistic fractionalization, and a measure of the protection of property rights (the risk of expropriation). We might expect income inequality to be higher in countries where ethnic fractionalization is greater if, for example, people are averse to redistribution in countries where ethnic diversity is greater.⁷ This variable was not available across time, and, therefore, it is set equal to the same value for all periods.

It is less clear whether government consumption and property rights protection will increase or decrease income inequality. For example, although the protection of property rights might protect the rich against expropriation by the poor, it could also have the opposite effect, that is, protecting the poor against exploitation by the rich. Similarly, if most redistribution through the tax and transfer system is toward low-income groups, government consumption might result in greater equality. However, it could also have the opposite effect if rich households use their political power to exploit the poor.

⁶ See, for example, Easterly and Fischer (2001).

⁷ Consistent with this, Alesina, Baqir, and Easterly (1999) find that spending on productive public goods (e.g., on schools) is lower in U.S. cities where ethnic diversity is greater. Ethnolinguistic fractionalization was unavailable for many of the countries in our sample. To avoid excessive sample loss, we imputed values based on the other regressors for countries with missing data. Results were robust to using other imputation techniques, including hotdeck imputation (Mander and Clayton 1999) and multiple imputation (Royston 2004). Although the hotdeck approach was used for all regressions, the multiple imputation approach could be used only for OLS regressions. Results were similar in terms of size and statistical significance for the coefficients on the finance variables. Results were also similar for those coefficients when we simply dropped ethnolinguistic fractionalization from the estimation.

Kuznets (1955) suggests that income inequality might depend on the sectoral structure of an economy. Thus, we include a variable representing the share of value added accounted for by services and industry (as opposed to agriculture). The correlation of the modern, that is, nonagricultural sector, share of GDP and GDP per capita indicates that richer countries have larger modern sectors. Although the simple correlation between the modern sector's share of GDP and the Gini coefficient is negative, this appears to be because poorer countries have greater inequality and larger agricultural sectors. After controlling for *per capita* income, the partial correlation becomes positive and significant.

We conduct the analysis in two ways: a pure cross-sectional analysis, using data averaged over the entire period between 1960 and 1995, and a panel data analysis using five-year panels. The cross-sectional analysis might capture the long-term relationship between finance and inequality, offering a way of testing the long-term relationship featured in the inequality-narrowing and inequality-widening hypotheses. In contrast, the panel analysis might examine the *process* of comovement between finance and inequality and, therefore, might be a more appropriate setup in which to test the inverted U-shaped hypothesis.

Following the convention of most cross-country empirical panel studies, the panel analysis splits the sample period 1960 to 1995 into seven nonoverlapping 5-year periods. We use 5-year periods rather than shorter time spans because although financial intermediary data are available on a yearly basis for most countries in our sample, they might be subject to business cycle fluctuations that are controlled for by averaging over longer time periods. All panel regressions include time dummies to account for structural differences across periods. To take account of the panel structure of the data, we present results from random effects estimation.

Estimating Equation 1 using ordinary least squares (OLS) (or random effects) estimation might introduce bias because OLS does not allow for the possibility of reverse causality—that is, for the possibility that inequality affects the provision of financial services—something suggested in some of the theoretical models. For example, in Greenwood and Jovanovic's (1990) model, the initial distribution of wealth affects who is able to join financial intermediary coalitions and, therefore, might affect the size of the financial sector. Because we are primarily interested in the effect of financial sector development on income inequality, we use an instrumental variables approach, adopting instruments for financial sector development similar to the ones used in Levine (1997a, 1999), which assesses the exogenous impact of financial intermediary development on economic growth. The instruments are a set of dummy variables proposed by La Porta et al. (1998) that identify the origin of the country's legal system.⁸ We use the legal origin dummy variables, rather than the measures of creditor rights, also proposed by La Porta et al. (1998), because they are available for a wider sample of countries. Several papers have shown that differences in legal origin are significantly related to financial sector development, perhaps because different legal traditions put different levels of emphasis on the rights of property owners or because some systems are more adaptable to exogenous changes than others.⁹ In the empirical analysis, we examine the validity of the instruments using Hansen's J-test to test the overidentifying restrictions.¹⁰

⁸ The measures of legal origin were taken from the Global Development Network Growth Database produced by William Easterly and Mirvat Sewadeh (see Easterly 2001).

⁹ Beck, Demirgüç-Kunt, and Levine (2001) provide an excellent summary of much of the empirical and theoretical literature on this topic. La Porta et al. (1998) show that protection for corporate shareholders and creditors are strongest in common law countries and weakest in French Civil Law countries. La Porta et al. (1997) relate these variables to some measures of capital market development (external market capitalization over GDP, number of listed firms per capita, initial public offerings), showing that they are generally lower in civil law (especially French Civil Law) countries than in common law countries. Beck, Demirgüç-Kunt, and Levine (2001) show that private credit is lower in French Civil law countries than in German Civil Law and common law countries.

¹⁰ In similar regressions of financial sector development on economic growth, Levine (1997a, 1999) fails to reject the null hypothesis that the overidentifying restrictions are valid.

5. Empirical Results

Long-Term Relationship from Cross-Sectional Samples

To test the inequality-widening and the inequality-narrowing hypotheses, we regress the natural log of the Gini coefficient on linear terms for the measure of financial sector development (private credit) and the additional control variables. Before we control for the possible endogeneity of the measures of financial sector development, the coefficient on private credit is negative but statistically insignificant (see column 1 of Table 2). The coefficient on banks assets is also negative but is statistically significant, indicating that inequality is lower in countries where bank assets are greater as a share of GDP (see column 3 of Table 2). Results for both measures are qualitatively similar when we omit *per capita* GDP and *per capita* GDP squared from the regression (see Table 3). These results, as we discussed earlier, do not take into account the issue of the endogeneity of the finance variables.

After controlling for endogeneity using the indicators of legal origin as instruments, the coefficient on private credit remains negative but increases in size and becomes statistically significant (see column 5). Results are similar when bank assets are used as the measure of financial sector development (see column 7). This suggests that financial sector development reduces income inequality, supporting the inequality-narrowing hypothesis and rejecting the inequality-widening hypothesis of financial development. Hypothesis tests reject the null hypothesis that the financial variables are exogenous, favoring the results from the 2SLS regressions, consistent with the theoretical papers that view financial development as endogenous.¹¹ In addition, we are unable to reject the null hypothesis that the legal origin dummies are uncorrelated with the error term after controlling for the other variables, suggesting that they are appropriate instruments (see Hansen J-Statistics in the relevant tables). Based on the coefficient estimates in column 5, a 1% increase in private credit decreases the Gini coefficient by 0.31%. Results are similar when *per capita* GDP is omitted (see Table 3), with the point estimate of the parameter slightly smaller at -0.27 .

To test the inverted U-shape hypothesis of financial development, we include squared terms for the measures of financial sector development (see columns 6 and 8). Because the coefficient on the squared term is statistically insignificant in all model specifications, the results do not support this hypothesis. Although the coefficient on the linear term becomes statistically insignificant in both model specifications when the squared term is included, it is important to note that the coefficients on the linear and squared terms are jointly significant at a 1% level or higher when financial sector development is treated as endogenous. Thus, these regressions suggest that although financial sector development does affect inequality, it appears to do so in a roughly linear fashion. However, the panel data might provide a better way of testing the inverted U-shape hypothesis if panel data better capture short- or medium-run variations in the comovements of financial development and inequality.

After controlling for the endogeneity of the financial sector variables, many of the coefficients on the other control variables are statistically insignificant (see column 5 of the relevant tables). Although the coefficients on the linear and squared terms for initial GDP *per capita* are statistically insignificant, they are jointly significant in most model specifications.¹² The positive coefficient on the linear term and the negative coefficient on the squared term suggest an inverted U-shape, with income inequality

¹¹ When we perform a Durbin-Wu-Hausman test using an auxiliary regression (see Davidson and MacKinnon 1993), the null hypothesis that “private credit” is exogenous is rejected at a 1% significance level (p -value = 0.001). For “bank assets,” the null hypothesis that it is exogenous is rejected at a 5% significance level (p -value = 0.043).

¹² They are jointly significant at a 1% level or higher when bank assets are included, jointly significant at a 10% level when private credit is included linearly, and statistically insignificant when private credit is included linearly and in squared terms.

Table 2. Effect of Financial Intermediary Development on Income Inequality: Cross-Sectional Regressions

Dependent Variable Estimation Method	OLS					Gini Coefficient (natural log)					2SLS		
	1	2	3	4	5	6	7	8			6	7	8
Column	83	83	83	83	83	83	83	83			83	83	83
Observations													
Financial sector development													
Private credit (natural log, as share of GDP)	-0.0530 (1.10)	0.0188 (0.21)			-0.3133† (3.54)	-0.2626 (1.07)							
Private credit squared (square of natural log, share of GDP)		0.0307 (0.82)				0.0231 (0.21)							
Bank assets (natural log, as share of GDP)			-0.1650† (3.81)	-0.1843* (1.74)			-0.3072† (5.18)	-0.1022 (0.47)					
Bank assets squared (square of natural log, share of GDP)				-0.0078 (0.20)				0.0941 (0.95)					
Initial GDP													
Initial GDP per capita (natural log)	0.1486 (0.49)	0.2506 (0.81)	0.2182 (0.89)	0.1902 (0.67)	0.0272 (0.08)	0.1023 (0.20)	0.2568 (1.26)	0.6028 (1.58)					
Initial GDP per capita squared (square of natural log)	-0.0140 (0.73)	-0.0213 (1.09)	-0.0199 (1.28)	-0.0180 (0.99)	-0.0078 (0.35)	-0.0132 (0.38)	-0.0238* (1.83)	-0.0474* (1.86)					
Other controls													
Risk of expropriation (index-higher values mean lower risk)	-0.0467† (2.62)	-0.0455† (2.54)	-0.0311* (1.93)	-0.0313* (1.93)	-0.0254 (1.31)	-0.0241 (1.14)	-0.0139 (0.78)	-0.0084 (0.40)					
Ethnolinguistic fractionalization (higher values mean greater fractionalization)	0.1595* (1.79)	0.1715† (1.99)	0.1622† (2.12)	0.1583† (2.00)	0.0532 (0.43)	0.0610 (0.48)	0.1459† (1.99)	0.1893† (2.18)					
Government consumption (natural log, share of GDP)	-0.0176 (0.30)	0.0031 (0.05)	0.0014 (0.02)	-0.0036 (0.06)	-0.0302 (0.42)	-0.0148 (0.15)	0.0156 (0.23)	0.0794 (0.85)					
Inflation (natural log)	0.0152 (0.52)	0.0102 (0.35)	-0.0182 (0.62)	-0.0173 (0.59)	-0.0809* (1.75)	-0.0859* (1.71)	-0.0639* (1.78)	-0.0832† (2.06)					

Table 2. Continued

Dependent Variable	Gini Coefficient (natural log)									
Estimation Method	OLS					2SLS				
Column	1	2	3	4	5	6	7	8		
Observations	83	83	83	83	83	83	83	83		
Modern sector (services and industry)	0.0099† (2.82)	0.0105† (3.01)	0.0130† (3.75)	0.0129† (3.73)	0.0177† (3.48)	0.0182† (3.40)	0.0169† (4.15)	0.0181† (4.33)		
(share of GDP)										
Constant	2.8158† (2.48)	2.4690† (2.16)	2.1009† (2.37)	2.1905† (2.18)	2.0541* (1.83)	1.7831 (1.01)	1.3505† (1.97)	0.1189 (0.09)		
R-squared	0.43	0.44	0.52	0.52	0.32	0.19	0.45	0.34		
Significance of banking variables (significance level)	0.27	0.47	0.00†	0.00†	0.00†	0.00†	0.00†	0.00†		
Hansen's J-test (significance level)	—	—	—	—	0.95	0.90	0.75	0.97		

The instruments for the 2SLS analysis are dummies indicating legal origin. The null hypothesis for the Hansen's test is that the instruments are not correlated with the error term. The joint significance of banking variables is the joint test that the coefficients on the linear and squared terms are statistically significant.

†, †, * Denote statistical significance at 1%, 5%, and 10% levels. T-statistics are robust standard errors.

Table 3. Effect of Financial Intermediary Development on Income Inequality: Cross-Sectional Regressions Omitting Per Capita GDP

Dependent Variable	Gini Coefficient (natural log)							
	OLS				2SLS			
Estimation Method	1	2	3	4	5	6	7	8
Column								
Observations	83	83	83	83	83	83	83	83
Financial sector development								
Private credit (natural log, as share of GDP)	-0.0456 (0.90)	-0.0426 (0.53)			-0.2660† (4.13)	-0.2798 (1.21)		
Private credit squared (square of natural log, share of GDP)		0.0013 (0.04)				-0.0065 (0.06)		
Bank assets (natural log, as share of GDP)			-0.1365† (3.25)	-0.2137† (2.22)			-0.2140† (4.27)	-0.0761 (0.29)
Bank assets squared (square of natural log, share of GDP)				-0.0299 (0.88)				0.0659 (0.52)
Other controls								
Risk of expropriation (index-higher values mean lower risk)	-0.0689† (4.55)	-0.0691† (4.52)	-0.0623† (4.33)	-0.0577† (3.91)	-0.0521† (2.96)	-0.0511† (2.22)	-0.0566† (3.70)	-0.0643† (3.06)
Ethnolinguistic fractionalization (higher values mean greater fractionalization)	0.1632* (1.87)	0.1636* (1.88)	0.1655† (2.07)	0.1543* (1.89)	0.0797 (0.73)	0.0783 (0.70)	0.1571† (2.00)	0.1783† (2.01)
Government consumption (natural log, share of GDP)	-0.0294 (0.47)	-0.0288 (0.45)	-0.0177 (0.29)	-0.0309 (0.50)	-0.0359 (0.55)	-0.0389 (0.47)	-0.0118 (0.19)	0.0196 (0.23)
Inflation (natural log)	0.0248 (0.81)	0.0247 (0.80)	-0.0005 (0.02)	0.0018 (0.06)	-0.0537 (1.47)	-0.0524 (1.31)	-0.0241 (0.77)	-0.0390 (0.95)
Modern sector (services and industry) (share of GDP)	0.0074† (2.58)	0.0074† (2.59)	0.0094† (3.72)	0.0088† (3.39)	0.0120† (4.46)	0.0118† (3.09)	0.0112† (5.35)	0.0134† (3.52)
Constant	3.4889† (9.76)	3.4893† (9.70)	3.1316† (10.80)	3.1036† (10.67)	2.5589† (7.69)	2.5627† (7.81)	2.8196† (10.16)	2.7518† (10.59)
R-squared	0.40	0.40	0.46	0.47	0.21	0.22	0.44	0.33
Significance of banking variables (significance level)	0.37	0.66	0.00†	0.01†	0.00†	0.00†	0.00†	0.00†
Hansen's J-test (significance level)	—	—	—	—	0.83	0.65	0.33	0.20

The instruments for the 2SLS analysis are dummies indicating legal origin. The null hypothesis for the Hansen's test is that the instruments are not correlated with the error term. The joint significance of banking variables is the joint test that the coefficients on the linear and squared terms are statistically significant.
†, †, * Denote statistical significance at 1%, 5%, and 10% levels. T-statistics are robust standard errors.

increasing with income at low levels of income and decreasing at high levels. However, the turning point appears to be relatively low—it is less than \$600 in all regressions in Table 2 and is close to zero when private credit is used as the measure of financial sector development.

When bank assets are used as the measure of financial sector development, the coefficient on ethnolinguistic fractionalization is statistically significant and positive (columns 7 and 8), suggesting that income inequality might be higher in countries with greater fractionalization. This finding supports the conjecture that citizens prefer less redistribution when ethnolinguistic fractionalization is greater. However, this result is not robust because it does not hold when private credit is used as the measure of financial sector development. The negative coefficient on inflation suggests that inequality is generally lower in countries where inflation is greater.

After controlling for other factors that might affect income inequality, including *per capita* income and financial sector development, the coefficient on the share of the economy accounted for by services and industry, that is, sectors other than agriculture, is positive and statistically significant. This suggests that income inequality is lower in countries where agriculture accounts for a greater share of GDP.

Short- or Medium-Run Results from the Panel Sample

In addition to the cross-sectional results presented in the previous section, we present results from panel regressions. In addition to providing a useful robustness test, this might also provide better information on the short- and medium-run relationship between finance and inequality. As noted in the previous section, we divide the data up into seven five-year periods. To control for structural differences between periods, all regressions include period dummies.¹³ As a first exercise, we treat the financial variables as exogenous and estimate a random-effects regression.

When private credit is treated as exogenous, its coefficient is small, positive, and statistically insignificant (see column 1 of Table 4). Although the coefficient on bank assets is negative, its coefficient is also statistically insignificant at conventional significance levels (column 2). When squared terms are added to the base regressions, the linear and squared terms are statistically insignificant both singly and jointly.

A first question is whether the model should be estimated as a random-effects model or whether a fixed-effects estimator would be more appropriate. One concern with respect to the fixed-effects estimator is that any cross-sectional variation is removed when country dummies are added to the regression. This might be a problem because although inequality varies greatly between countries, it varies only modestly within countries over time. For example, Li, Squire, and Zou (1998) show that 90% of the variance in the Gini coefficient in their data (an updated version of which is used in this paper) is cross-country variation, compared with less than 1% from cross-time variation.¹⁴ In this respect, fixed effects will remove most of the variation in inequality that we are trying to explain. In addition, including fixed effects might exacerbate problems related to measurement error. This is a particular concern because income distribution is often measured poorly, and although inequality changes slowly over time, measurement error might be quite different in different periods.¹⁵ Hence,

¹³ When we added time dummies and tested the joint significance of these dummies, they were statistically significant at a 1% level or higher in all of the panel specifications in Table 4.

¹⁴ For example, Easterly (2002) suggests that it is unclear whether standard panel methods are appropriate given that income distribution is relatively stable over time.

¹⁵ See Griliches and Hausman (1986) for a discussion of errors in variables in panel regressions.

Table 4. Effect of Financial Intermediary Development on Income Inequality: Panel Regressions

Column	1	2	3	4	5	6	7	8
Estimation Method	Random Effects			Gini Coefficient		IV Random Effects		
Dependent Variable				(natural log)				
Observations	205	205	205	205	205	205	205	205
Financial sector development								
Private credit (natural log, as share of GDP)	0.0291 (1.46)	0.0160 (0.18)			-0.1114† (2.07)	0.2616 (0.92)		
Private credit squared (square of natural log, share of GDP)		0.0020 (0.15)				-0.0551 (1.36)		
Bank assets (natural log, as share of GDP)			-0.0157 (0.69)	0.0578 (0.57)			-0.1463‡ (2.81)	0.4246 (1.03)
Bank assets squared (square of natural log, share of GDP)				-0.0112 (0.74)				-0.0824 (1.39)
Initial GDP								
Initial GDP <i>per capita</i> (natural log)	0.6762‡ (4.36)	0.6864‡ (4.05)	0.7207‡ (4.62)	0.6878‡ (4.26)	0.8285‡ (4.60)	0.5147* (1.80)	0.8499‡ (4.92)	0.6187‡ (2.55)
Initial GDP <i>per capita</i> squared (square of natural log)	-0.0466‡ (5.04)	-0.0473‡ (4.56)	-0.0486‡ (5.23)	-0.0463‡ (4.74)	-0.0534‡ (5.04)	-0.0319* (1.74)	-0.0552‡ (5.39)	-0.0389‡ (2.46)
Other controls								
Risk of expropriation (index-higher values mean lower risk)	-0.0374‡ (3.56)	-0.0374‡ (3.54)	-0.0360‡ (3.40)	-0.0359‡ (3.38)	-0.0331‡ (2.76)	-0.0330‡ (2.63)	-0.0297‡ (2.53)	-0.0321‡ (2.67)
Ethnolinguistic fractionalization (higher values mean greater fractionalization)	-0.0578 (0.61)	-0.0584 (0.61)	-0.0840 (0.89)	-0.0871 (0.92)	-0.1500 (1.36)	-0.1442 (1.14)	-0.1197 (1.18)	-0.1288 (1.27)
Government consumption (natural log, share of GDP)	-0.1035‡ (2.47)	-0.1023‡ (2.39)	-0.1040‡ (2.45)	-0.1130‡ (2.55)	-0.1382‡ (2.88)	-0.1728‡ (3.02)	-0.0742 (1.54)	-0.1545‡ (2.05)
Inflation (natural log)	-0.0140 (0.23)	-0.0172 (0.26)	-0.0735 (1.26)	-0.0662 (1.12)	-0.2227‡ (2.19)	-0.1359 (1.01)	-0.2009‡ (2.55)	-0.1074 (1.04)
Modern sector (services and industry) (share of GDP)	-0.0015 (0.65)	-0.0014 (0.63)	-0.0015 (0.65)	-0.0017 (0.76)	-0.0017 (0.68)	-0.0030 (1.13)	-0.0014 (0.56)	-0.0030 (1.06)
Constant	1.8000‡ (3.26)	1.7793‡ (3.13)	1.7268‡ (3.11)	1.8075‡ (3.24)	1.6405‡ (2.63)	2.3422‡ (3.04)	1.4667‡ (2.42)	1.6788‡ (2.63)
R-squared	0.61	0.6	0.58	0.58	0.51	0.46	0.57	0.55
Significance of banking variables (significance level)	0.14	0.34	0.49	0.60	0.04‡	0.06*	0.00‡	0.01‡

All regressions include period dummies to control for differences between periods. The instruments for the 2SLS analysis are dummies indicating legal origin. The joint significance of banking variables is the joint test that the coefficients on the linear and squared terms are statistically significant. All regressions include period dummies to control for differences between periods. ‡, †, * Denote statistical significance at 1%, 5%, and 10% levels.

including fixed effects, which remove much of the variation in inequality, might leave us with a small amount of variation in inequality and a larger amount of variation in measurement error.

In practice, the results from the fixed- and random-effect regressions were qualitatively similar. When “private credit” is entered linearly, its coefficient is positive in both the fixed- and random-effects regressions—although the coefficient is statistically significant in the fixed-effects specification.¹⁶ A Hausman test fails to reject the null hypothesis that the coefficients from the two models with private credit are systematically different, favoring the random-effects specification for regressions including “private credit.”¹⁷ For bank assets, the results for the financial variables are similar in the fixed- and random-effects models—the coefficient is small and statistically insignificant in both specifications when entered linearly. However, a Hausman test rejects the null hypothesis that the coefficients are not systematically different for the models with “bank assets” included, favoring the fixed effects regressions.¹⁸ As in the cross-sectional regressions, there is no evidence of a non-linear effect—the coefficients on the squared terms are statistically insignificant in all model specifications.

The results from the cross-sectional regressions and *a priori* reasoning from existing theoretical models on finance and inequality all suggest that the financial variables might be endogenous.¹⁹ Therefore, we reestimate the panel regressions allowing for endogeneity. Because legal origin does not change over time, we are unable to estimate instrumental variable regressions that include fixed effects; that is, the instruments are collinear with the country dummies. Therefore, we reestimate the model using a random-effects instrumental variables model (see columns 5 to 8 in Table 4). When the financial variables are entered linearly, the coefficients on the financial sector variables are negative and statistically significant. This is consistent with the cross-sectional results.

When squared terms are included in the regression, the coefficients on the squared terms are statistically insignificant in both regressions for both financial variables. The results thus do not support the inverted U-shaped hypothesis of financial development. However, as before, the coefficients on the linear and squared terms are jointly significant (see final row of Table 4).

Although most of the control variables remain statistically insignificant, the panel results show greater evidence of an inverted U-shape with respect to initial GDP. The coefficients on initial GDP and initial GDP squared are statistically significant and indicate an inverted U-shape. In the IV random-effects model, the turning point is at about \$2300 in the private credit regression and \$2200 in the bank asset regressions. Another notable difference is that the coefficient on the risk of expropriation becomes statistically significant and negative, indicating that inequality is greater when the risk of expropriation is greater.²⁰

In contrast to the results for the cross section, results for the panel data are slightly different when initial GDP is dropped from the main regression (see Table 5). When the financial variables are treated as endogenous, the coefficients on the linear terms remain statistically significant and negative when

¹⁶ Results from fixed-effects regressions are available from authors on request.

¹⁷ The null hypothesis that the country effects are uncorrelated with the additional variables is not rejected at conventional significance levels (p -value = 0.20).

¹⁸ The null hypothesis is rejected at a 1% significance level (p -value = 0.00).

¹⁹ Although the Durbin-Wu-Hausman test is not available for the random-effects models, similar tests for the panel data when estimated using OLS and 2SLS, that is, as a pooled cross section, also strongly reject the null hypothesis that the financial variables are exogenous. With a Durbin-Hausman-Wu test, the null hypothesis is rejected at a 5% significance level for bank assets (p -value = 0.044) and at a 1% significance level for ‘private credit’ (p -value = 0.000). The coefficients from the two-stage least-squares model using the pooled cross section are similar in terms of size and statistical significance to the results from the IV random effects model.

²⁰ Recall that higher values on the index mean lower risk.

Table 5. Effect of Financial Intermediary Development on Income Inequality: Panel Regressions Omitting *Per Capita GDP*

Dependent Variable	Random Effects			IV Random Effects			
	Gini Coefficient (natural log)						
Column	1	2	3	4	5	6	7
Observations	205	205	205	205	205	205	205
Financial sector development							
Private credit (natural log, as share of GDP)	0.0190 (0.90)	0.2781† (3.36)			-0.1783† (2.53)	0.9578† (2.43)	
Private credit squared (square of natural log, share of GDP)		-0.0383† (3.23)				-0.1534† (2.94)	
Bank assets (natural log, as share of GDP)			-0.0127 (0.52)	0.2849† (2.91)			-0.2189† (2.68)
Bank assets squared (square of natural log, share of GDP)				-0.0455† (3.13)			1.2101† (2.28)
Other controls							
Risk of expropriation (index-higher values mean lower risk)	-0.0523† (4.77)	-0.0482† (4.47)	-0.0494† (4.49)	-0.0451† (4.16)	-0.0351† (2.48)	-0.0294† (2.02)	-0.0320† (2.30)
Ethnolinguistic fractionalization (higher values mean greater fractionalization)	0.0017 (0.02)	0.0011 (0.01)	-0.0260 (0.25)	-0.0550 (0.53)	-0.2232 (1.46)	-0.1546 (0.92)	-0.1487 (1.08)
Government consumption (natural log, share of GDP)	-0.1053† (2.27)	-0.1277† (2.79)	-0.1036† (2.21)	-0.1405† (2.97)	-0.1494† (2.58)	-0.2304† (3.47)	-0.0526 (0.89)
Inflation (natural log)	-0.0063 (0.09)	0.0508 (0.76)	-0.0505 (0.81)	-0.0258 (0.42)	-0.3169† (2.45)	0.0681 (0.36)	-0.2602† (2.48)
Modern sector (services and industry) (share of GDP)	0.0003 (0.15)	-0.0010 (0.50)	0.0006 (0.33)	-0.0005 (0.26)	0.0026 (1.05)	-0.0029 (0.94)	0.0026 (1.07)
Constant	4.2555† (20.79)	3.9774† (18.27)	4.3264† (21.50)	4.0391† (18.60)	4.8466† (15.15)	3.4492† (5.82)	4.7047† (16.93)
R-squared	0.45	0.46	0.46	0.47	0.30	0.33	0.36
Significance of banking variables (significance level)	0.36	0.00†	0.61	0.00†	0.00†	0.00†	0.00†

All regressions include period dummies to control for differences between periods. The instruments for the 2SLS analysis are dummies indicating legal origin. The joint significance of banking variables is the joint test that the coefficients on the linear and squared terms are statistically significant. All regressions include period dummies to control for differences between periods. †, ‡, * Denote statistical significance at 1%, 5%, and 10% percent levels.

entered linearly. However, the coefficient on the linear term is positive and statistically significant, and the coefficient on the squared term is negative and statistically significant when both linear and squared terms are included. This is broadly supportive of the inverted U-shape hypothesis with inequality first increasing as financial development increases and then decreasing. The point estimates suggest that the turning point is at about when private credit is equal to 22% of GDP. In 2003, private credit was equal to about 28% of GDP for low-income countries and 64% of GDP for middle income countries.²¹ Given the lack of support for the inverted U-shape hypothesis when initial GDP is included, and the support for the hypothesis without the inclusion of initial GDP, and in light of the fact of a close correlation between finance and initial GDP, we conclude that there is some weak support for the inverted U-shape hypothesis when short-term and medium-term variations are considered.

To summarize, after controlling for endogeneity, we find support for the inequality-narrowing hypothesis that inequality is lower in countries with better-developed financial sectors and reject the inequality-widening hypothesis in both panel and cross section. In contrast, we do not generally find strong support for the inverted U-shape hypothesis in the long-run cross-sectional data but do find some weak support for it with short- and medium-run panel data.

6. Conclusions

There has been little systematic empirical study on the relationship between finance and inequality. This paper attempts to examine this issue by testing empirically distinct predictions made by alternative theories. Specifically, Galor and Zeira (1993) and Banerjee and Newman (1993) predict a negative and linear relationship between finance and the Gini coefficient (the inequality-narrowing hypothesis), some popular press worry about the inequality-widening effects of financial development, while Greenwood and Jovanovic (1990) suggest a inverted U-shape relationship (the inverted U-shape hypothesis).

Exploring the link between indicators of financial intermediary development and the Gini coefficient in a large cross-country sample for the period 1960–1995, we experiment with both simple specifications and more sophisticated specifications that control for simultaneity. Overall, our results provide some support for the inequality-narrowing hypothesis. We find a significant negative coefficient on the measures of financial intermediary development once we control for endogeneity—and hypothesis tests suggest that this is important. In contrast, the results decisively reject the inequality-widening hypothesis. Moreover, while the cross-sectional (long-term) data do not provide much support for the inverted U-shaped hypothesis, the short- and medium-term panel data do provide some weak support for the inverted U-shape hypothesis. Overall, our results suggest that the growth-spurring effects of financial intermediary development are likely to be associated with positive effects on aggregate income distribution as well.²²

We recognize some limitation of our results, which stem mostly from the limitations of our measure of income inequality. Changes in the Gini coefficient can come about in different ways, by absolute and relative changes in one or several of the different income quintiles. We do not explore the impact that a higher level of financial intermediary development has on the income level of a specific quintile, for

²¹ Data from World Bank (2004).

²² See Levine (1997b) for a recent literature survey on this topic. See also Beck, Demirgüç-Kunt, and Levine (2001) for a discussion of more recent results.

instance the poor. Moreover, even results obtained by using quintile data have to be regarded with caution because they do not control for migration between the quintiles over the sample period. To analyze directly the effect of financial development on specific groups of the population, one would have to use disaggregated data, preferably at the household level. This poses new challenges for future research.

References

- Alesina, Alberto, Reza Baqir, and William Easterly. 1999. Public goods and ethnic divisions. *Quarterly Journal of Economics* 114:1243–84.
- Banerjee, Abhijit V., and Andrew F. Newman. 1993. Occupational choice and the process of development. *Journal of Political Economy* 101:274–98.
- Beck, Thorsten, Asli Demirgüç-Kunt, and Ross Levine. 2000. A new database on the structure and development of the financial sector. *World Bank Economic Review* 14:597–605.
- Beck, Thorsten, Ross Levine, and Norman Loayza. 2000. Finance and the sources of growth. *Journal of Financial Economics* 58:261–300.
- Beck, Thorsten, Asli Demirgüç-Kunt, and Ross Levine. 2001. Legal theories of financial development. *Oxford Review of Economic Policy* 17:483–501.
- Davidson, Russell, and James G. MacKinnon. 1993. *Estimation and inference in econometrics*. Oxford, UK: Oxford University Press.
- Deininger, Klaus, and Lyn Squire. 1996. A new data set measuring income inequality. *World Bank Economic Review* 10:565–91.
- Easterly, William. 2001. The lost decades: Developing countries' stagnation in spite of policy reform 1980–1998. *Journal of Economic Growth* 6:135–57.
- Easterly, William, and Stanley Fischer. 2001. Inflation and the poor. *Journal of Money, Credit and Banking* 33:160–78.
- Easterly, William. 2002. Inequality does cause underdevelopment. Unpublished paper, Center for Global Development, Washington, DC.
- Galor, Oded, and Joseph Zeira. 1993. Income distribution and macroeconomics. *Review of Economic Studies* 60:35–52.
- Greenwood, Jeremy, and Boyan Jovanovic. 1990. Financial development, growth, and the distribution of income. *Journal of Political Economy* 98:1076–107.
- Griliches, Zvi, and Jerry Hausman. 1986. Errors in variables in panel data. *Journal of Econometrics* 32:93–118.
- International Monetary Fund (IMF). 2002. *International finance statistics*. Washington DC: International Monetary Fund.
- King, Robert G., and Ross Levine. 1993. Finance, entrepreneurship, and growth: Theory and evidence. *Journal of Monetary Economics* 32:513–42.
- Kuznets, Simon. 1955. Economic growth and income inequality. *American Economic Review* 45:1–28.
- La Porta, Rafael, Florencio López-de-Silanes, Andrei Shleifer, and Robert W. Vishny. 1997. Legal determinants of external finance. *Journal of Finance* 52:1131–50.
- La Porta, Rafael, Florencio López-de-Silanes, Andrei Shleifer, and Robert W. Vishny. 1998. Law and finance. *Journal of Political Economy* 106:1113–55.
- Levine, Ross. 1997a. The legal environment, banks, and long-run growth. *Journal of Money, Credit and Banking* 30:596–613.
- Levine, Ross. 1997b. Financial development and economic growth: Views and agenda. *Journal of Economic Literature* 35:688–726.
- Levine, Ross, and Sara Zervos. 1998. Stock markets, banks, and economic growth. *American Economic Review* 88:537–58.
- Levine, Ross. 1999. Law, finance and economic growth. *Journal of Financial Intermediation* 8:8–35.
- Levine, Ross, Norman Loayza, and Thorsten Beck. 2000. Financial intermediation and growth: Causality and causes. *Journal of Monetary Economics* 46:31–77.
- Li, Hongyi, Lyn Squire, and Heng-fu Zou. 1998. Explaining international and intertemporal variations in income inequality. *Economic Journal* 108:26–43.
- Li, Hongyi, Lixin Colin Xu, and Heng-fu Zou. 2000. Corruption, income distribution, and growth. *Economics and Politics* 12:155–82.
- Loayza, Norman, Humberto Lopez, Klaus Schmidt-Hebbel, and Luis Servén. 1999. World saving database. Unpublished paper, World Bank.
- Lundberg, Mattias, and Lyn Squire. 2000. The simultaneous evolution of growth and inequality. Unpublished paper, World Bank.
- Lundberg, Mattias, and Lyn Squire. 2003. The simultaneous evolution of growth and inequality. *Economic Journal* 113:326–44.
- Mander, Adrian, and David Clayton. 1999. Hotdeck imputation. *Stata Technical Bulletin Reprints* 9:196–9.
- PRS Group. 2003. *International country risk guide*. East Syracuse, NY: PRS Group.
- Rajan, Raghuram G., and Luigi Zingales. 2003. *Saving capitalism from the capitalists*. New York: Crown Business.
- Rousseau, Peter L., and Paul Wachtel. 2000. Equity markets and growth: Cross-country evidence on timing and outcomes, 1980–1995. *Journal of Banking and Finance* 24:1933–57.
- Royston, Patrick. 2004. Multiple imputation of missing values. *Stata Journal* 4:227–41.
- World Bank. 2004. *World development indicators*. Washington DC: World Bank.

CHINA: REGIONAL DISPARITIES IN POVERTY DISTRIBUTION

CHUNLI SHEN

School of Public Policy, University of Maryland, College Park, MD 20742, USA

HENG-FU ZOU

CEMA, Central University of Finance and Economics, Beijing, China

China's remarkable poverty alleviation is quite uneven across regions in the last quarter of the century. It is important to explore why China has such huge disparity in poverty distribution in spite of overall dramatic economic growth and the vast improvement in per capita income. The aim of this paper is to fill the literature gap by focusing exclusively on the issue of regional disparities in poverty distribution in China. It finds an increasing concentration of the rural poor in south-western provinces and the urban poor in northern China. Behind the scene, political choices and public policies, particularly barriers restricting the flow of labor, and fiscal rules that provides the disadvantaged population and regions less access to the fruits of division of labour, have a critical impact on how the effects of endowment and geography play out in the country's poverty distribution. In efforts to fight against skewed poverty concentration and build a harmonious society, further policy actions are required to promote agricultural development and off-farm employment, enhance infrastructure investment in poor regions, lower fiscal disparities and promote equitable public services provision, and address the regressive inter-governmental fiscal system.

Keywords:

Poverty; income distribution; regional disparity; mobility; China; poverty

See <http://www.worldscinet.com/dltc/03/0301/S0219871108000343.html>

LOCAL GOVERNMENT TAX EFFORT IN CHINA: AN ANALYSIS OF PROVINCIAL TAX PERFORMANCE

Qian WANG*, Chunli SHEN** and Heng-fu ZOU***

Abstract - This paper aims to enhance the understanding of provincial tax performance in China, paying special attention to the recent fiscal reforms in the 1980s and in 1994. Using provincial panel data for the period 1986-2004, our analysis consists of two steps. First, a combined fixed time effects and random provincial effects model is used to analyze the statistical relationship between the tax share in GDP and economic and demographic variables. Results indicate that the decentralized fiscal system over the period 1986–1993 has had a positive impact on the tax share in GDP, whereas the recentralized fiscal system in the period 1994-2004 has had a negative impact. Second, provincial tax effort indices are calculated to estimate potential room for additional taxation. The findings from the analysis have important policy implications on the redistribution of fiscal resources as well as on the effectiveness of the tax administration.

Keywords : TAX EFFORT, TAX CAPACITY, FISCAL REFORMS, FISCAL DECENTRALIZATION,

JEL Classification: H20, H71, C23

* School of Public & Environmental Affairs (SPEA) Indiana University Bloomington, Indiana

** Central University of Finance and Economics, China. Email : shenchunli@gmail.com

*** Shenzhen University, Wuhan University and Central University of Finance and Economics, China

INTRODUCTION

Taxation is a major source of government revenue that finances essential public services such as education, health, infrastructure and environmental protection. In developing countries in particular, greater investment in public services is required to raise the standard of living and increase the pace of economic development. However, countries or localities that need tax revenues the most face more difficulty in raising tax revenues. Some studies in this area suggest that the root of the problem lies in the governments' inadequate efforts to collect tax revenues based on their tax bases. For example, Tanzi (1992) found that tax efforts in developing countries tend to be lower than in industrialized countries.

China has undergone several fiscal reforms with various forms of fiscal contracting systems (1978-1993) and later a tax sharing system (1994-present). This leads to the inevitable question: How have these fiscal reforms affected the tax performance of provincial governments? The objective of this paper is to answer this question using a detailed analysis of patterns in provincial finance. This paper predicts tax capacity and calculates tax effort indices using provincial data over the period 1986-2004.

This study contributes to the field of public finance in three important ways. First, the research findings contribute to a better understanding of the tax performance of provincial governments in China. Second, the study provides insights into how the major fiscal reforms in 1986 and 1994 have affected the tax performance of provincial governments. Third, the research has important policy implications for governments at different levels. For example, for the purpose of redistribution, the Chinese central government must know not only the provinces' capacity to raise tax revenues, but also to what extent they have exploited their tax capacity. Provincial governments with high tax effort should be rewarded for their performance. Otherwise, the risk exists that the central government might spoil those provinces with low tax effort by subsidizing their deficits that are supposed to be financed through provincial tax base. Finally, from the administrative perspective, since provincial governments do not have the right to change tax rates or decide what kind of tax they can levy, they have to focus on minimizing administrative costs. Therefore, this study provides the central government with a better understanding of the effectiveness of the current administrative reform that aims to reduce costs and improve administrative efficiency. The paper proceeds as follows. Section one discusses previous research and different methodologies that have been used to measure tax capacity and tax effort. Section two reviews tax reforms in China. Section three discusses the methodology this study employs and describes the data. Section four discusses the model selection and empirical results. Section five computes tax capacity and tax effort indices. Section six discusses policy implications.

1. TAX PERFORMANCE: WHAT IS KNOWN ?

Tax performance consists of two distinct measurements. One is tax capacity, the measurement of a government's hypothetical ability to raise revenue. The other is tax effort, which measures the extent to which a certain level of government actually has explored its available tax bases and utilized its tax capacity. Together, these two measurements of tax performance of a specific locality provide a picture of potential room for additional taxation for that place (Bahl, 1971; Bahl, 1972; Chelliah, 1971; Mertens, 2003; Tanzi, 1987; Tanzi, 1992). The literature defines tax effort by dividing the actual collected tax by the tax capacity. This section discusses two major methods the literature has used to link tax capacity and tax effort.

The first method is employed by the Advisory Commission on Intergovernmental Relations (ACIR, 1981, 1982, 1987, 1990 and 1993) in the U.S. ACIR uses the representative tax system (RTS) and representative revenue system (RRS) to measure the tax capacity and tax effort of all U.S. states. Although each state has different taxes, the RTS assumes a representative tax rate for every single tax across states, which is calculated by dividing the total actual revenues for a tax source from all states by the total estimated RTS/RRS base. Therefore, this methodology measures tax capacity by different taxes. For each revenue source, the tax capacity for every state is estimated by multiplying the RTS/RRS tax base by the representative tax rate. Correspondingly, a state's tax effort is calculated by dividing the actual tax collections by its capacity to collect taxes.

The other method in the literature to connect tax capacity and tax effort, which is widely used in OECD countries, uses a regression approach. Most of the OECD working papers regress tax capacity on explanatory variables that might affect a country's ability to raise tax revenues. In this literature, most studies employ the ratio of actually collected tax over GDP as a measurement of tax capacity (Tanzi, 1992). Ratios to GDP are used for the reason that "GDP includes income earned locally that accrues to non-residents and excludes income received from abroad by residents. Since local income accruing to non-residents typically is taxed while remittances from abroad are not, GDP produces a more accurate measure of taxable capacity" (Teera and Hudson, 2004). Therefore, the estimated tax share of GDP from such a regression is regarded as a measure of taxable capacity. Following this approach, tax effort is the ratio of actual tax share of GDP over estimated tax share of GDP (Mertens, 2003).

Since tax capacity is based on hypothetical calculations, different researchers focus on different sets of factors to capture such capacity. On the one hand, some studies emphasize economic and demographic variables, which are called "tax handles" (Musgrave, 1969), such as GDP per capita, population, and trade share of GDP (Ansari, 1982; Mertens, 2003; Sagbas, 2001; Teera and Hudson, 2004; Stotsky and Woldemariam, 1997). On the other hand, some studies focus on social and institutional factors, such as the administrative and political constraints on the fiscal system, attitudes toward the government, and

other government institutions (Eltony 2002; Teera and Hudson, 2004; Warner, 2001).

Tax capacity analysis has traditionally focused on economic and demographic characteristics. The literature suggests that a higher level of economic development reflects an increased demand for public expenditure and a greater taxable capacity to meet such demands, therefore a higher per capita income indicates a greater tax capacity (Teera and Hudson, 2004). Industry's share of GDP plays a positive role in generating tax revenue, as it is usually easier to collect tax from the industrial sector than from the agricultural sector given their relatively accurate accounting records of taxable resources (Bahl, 1971; Bahl, 1972; Chelliah, 1971; Mertens, 2003; Tanzi, 1968; Tanzi 1987; Tanzi, 1992). Moreover, there exist more public services and activities in urban areas than in rural areas. Therefore, the higher the agricultural share of GDP, the less public services are needed, and the less tax revenue needs to be generated. Tax capacity also depends on the volume of international trade, which measures the degree of openness. Stotsky and Woldemariam (1997) argue that the tax share is positively related to the degree of openness of the economy.

Other than the aforementioned variables, that are traditionally used to measure the tax capacity, more tax handle variables have been proposed to capture the determinants of tax capacity more precisely. Ansari (1982) argues that a high population density is assumed to be a negative indicator of tax capacity, because a high degree of congestion is considered to cause more problems of tax exemptions. However, Teera and Hudson (2004) argue that the tax collection cost will be reduced in a densely populated area, which is expected to encourage governments to collect tax revenues. In addition, Sagbas' results show that there is a strong positive relationship between tax capacity and expenditure trends (Sagbas, 2001).

Even though the literature emphasizes that the success of governments in exploiting tax potential and in attaining a taxation target depends to a large extent on their tax handles, the role of institutional factors has been widely discussed as well. Recent research suggests that institutional factors could also be significant predictors of tax performance. Teera and Hudson (2004) state that variables such as levels of literacy, the administrative and political constraints on the fiscal system, and social-political values, should also be taken into account to measure the overall willingness and ability of the government to raise taxes. In addition, Warner proposes that tax capacity is positively related to spatial effects, and it is negatively related to poverty and tax substitutes (e.g. state aid or federal aid) (Warner, 2001). Furthermore, Eltony (2002) argues that country-specific factors appear to be important determinants of tax share, e.g., the political system and other institutions of the government, and attitudes toward the government.

In China, uniform national tax laws are set by the central government, whereas provincial governments are responsible for tax administration and may give tax concessions to State Owned Enterprises (hereafter called SOEs). Therefore, it is important to analyze each province's hypothetical tax base and

the extent to which each provincial government exploits its tax base because this information will allow the central government to gain tighter control over the central and provincial tax systems. For example, the central government needs to inspire the revenue-raising incentives of provincial governments. In addition, it is better for the central government to have an overall picture of provincial tax collection, in case provincial governments offer their SOEs more tax concessions or tax holidays than necessary. However, few if any empirical studies have analyzed the tax performance of provincial governments in China. In their 1992 study, Bahl and Wallich used only two variables—per capita gross value of industrial output and the percentage of population living in urban areas—to estimate the tax capacity of provincial governments in China in one single year (1986). In this context, this present study employs the economic and demographic variables mentioned above to analyze the tax capacity and tax effort of provincial governments in China during 1986–2004. The reason why we chose this period is that there were two main fiscal reforms during the period. The first, the “Contracting System,” was introduced in 1986. The other, the “Tax Sharing System,” began in 1994. This paper will compare the different effects of the two fiscal reforms on tax collection.

2. FISCAL DECENTRALIZATION REFORM

Fiscal decentralization is widely recognized as an essential component in China’s transition to a market economy, and advocated by many for its contribution to the country’s remarkable economic performance over the last 25 years. The country has made substantial efforts to break down its highly centralized fiscal management system with various forms of fiscal contracting systems (1978-1993) and later a tax sharing system (1994-present) (Shen, 2008).

A fiscal revenue sharing system replaced the highly centralized system in 1980. From then on, the central and provincial governments each began to ‘eat in separate kitchens’, which provided sub-national governments with an incentive to collect revenues. Under this system, central-provincial sharing rules were established by the central government; provincial-municipal relations were governed by the province; and this principle extended to lower levels. There were three basic types of revenues under the reformed system: central-fixed revenues, local-fixed revenues, and shared revenues. During the period 1980–84, about 80 percent of the shared revenues were remitted to the central government and 20 percent were retained by local governments. The bases and rates of all the taxes, whether shared or fixed, were determined by the central government. Enterprises were supposed to pay taxes to the level of government they were subordinate to. Almost all revenues, except a few minor central-fixed revenues, were collected by local finance bureaus (Shen, 2008).

The uniform-sharing formula during the period 1980-1984 created undesired surpluses in affluent provinces and deficits in poor provinces, although the reform boosted more revenue collection in many localities. In 1985, the State Council redesigned revenue-sharing arrangements by varying schedules based on localities’ budget balances in the previous years. The

financially weak provinces were allowed to retain more revenues, but the wealthier regions, like Shanghai, Beijing, Tianjin, Liaoning, Jiangsu, and Zhejiang, were penalized by remitting more revenues to the center. As a consequence, the revenues from these regions generally grew more slowly than the national average since the high level of remittance curbed local enthusiasm for expanding their tax bases (Shen, 2008).

In the period 1988-1993, the “fiscal contracting system” was implemented. This system requires each level of government to contract with its subordinate level to meet certain revenue and expenditure targets. The central government signed contracts on a case-by-case basis with the provincial governments, specifying their remittance based on the profit of their enterprises. Six types of central-provincial revenue-sharing methods were adopted and each applied to some provinces¹. Consequently, all revenues were divided into two parts: the central fixed revenues and the local retained revenues. The provincial governments relied on their local retained revenues for their public expenditure requirements. In this case, to some extent, the provincial governments were self-financed. In other words, the responsibility for meeting the expenditure needs of provincial governments was decentralized (Bahl and Wallich, 1992).

Under this reform, the proportion of central revenue declined dramatically, causing a huge deficit at the central government level. In particular, certain categories of local revenues went to the “extra-budgetary fund” of the provincial government, which was not subject to sharing with the central government. Provincial governments tended to maximize their “extra-budgetary fund.” Consequently, two ratios (revenue/GDP and central/total revenue) eroded (see Figure 1), and the central government faced a huge deficit (Zhang and Zou, 1998). Therefore, in order to raise the ratio of central revenue over the total revenue, the central government introduced a new reform, the “Tax Sharing System,” in 1994.

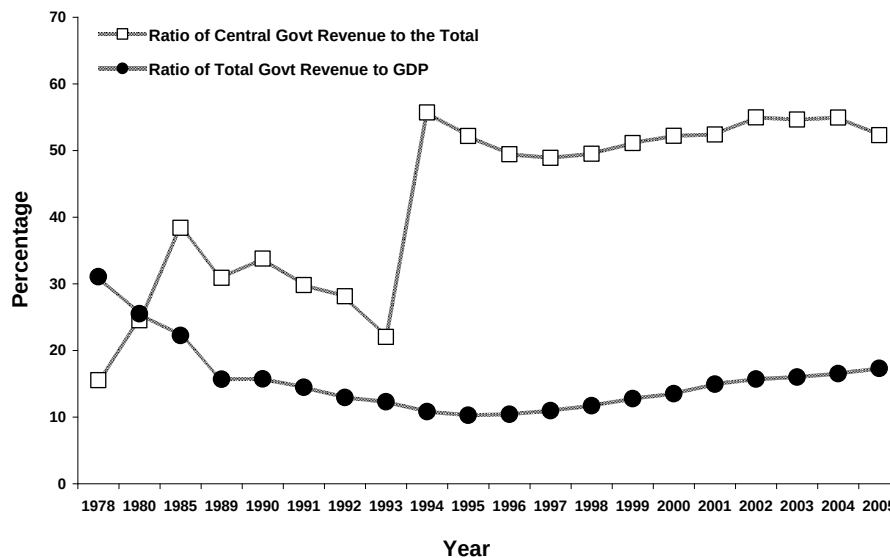
The 1994 fiscal reform was designed to base the fiscal relations between governments on the tax code: central, local, and shared taxes. Value-added tax, business tax, and several excise taxes were introduced both at the central governmental and the provincial level. The biggest tax is value-added tax, which is a shared tax. From value-added tax, the central government takes 75%, which accounts for a major portion of its fiscal revenue, and provincial governments retain only 25%. According to most scholars (Bahl and Wallich, 1992; Lin, Tao and Liu, 2003; Wong, 1998), the overall system reforms in China focused on the decentralization of economic management, which allowed the development of a greater autonomy for provinces and non-state sectors, but

¹ For example, one formula was “contracted sharing rate with fixed yearly growth rate of revenue”, which means the central-local revenue sharing rate and the yearly growth rate of local revenues were based on the revenue performance of the province over recent years and negotiated by the central and provincial governments. If the real growth rate was greater than the contracted rate, the province could keep all the surpluses. If the real growth rate was lower than the contracted rate, then the province had to make up the gap.

the 1994 fiscal reform actually recentralized the Chinese fiscal system. On the one hand, revenue is centralized under the tax sharing system because the central government takes a considerable amount of revenue.

On the other hand, since provinces keep only a small proportion of the total revenue, they need subsidies from the central government to meet their expenditure. In this respect, expenditure is also centralized.

Figure n°1 : The Two Ratios, 1978-2005

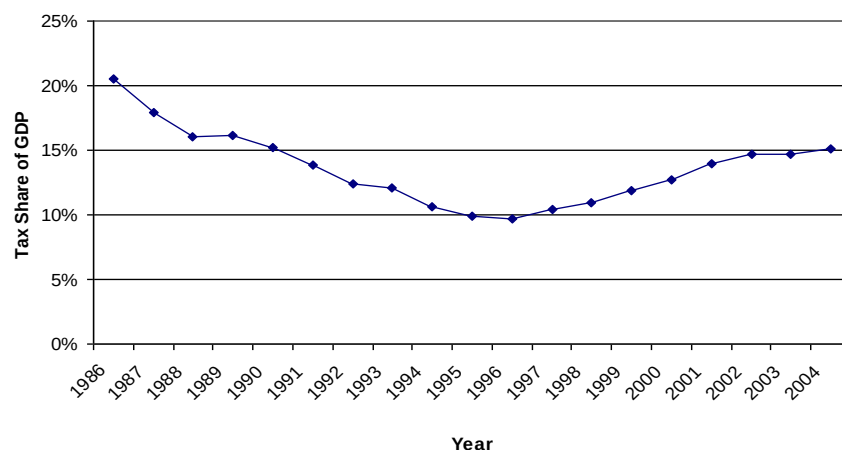


Source : Shen, 2008.

In terms of the outcome of these fiscal reforms, the 1980s fiscal reform led to a decreased overall tax share of GDP, while the 1994 fiscal reform resulted in a contrary outcome. As shown in Figure 1 below, the general trend over time is an increase in the tax share of GDP from 10% to 15% since 1996 (see Figure 2).

Such a trend suggests that China has enjoyed increases in tax shares, and hence, better overall tax-collection efforts in the past decade.

Theoretically, both the decentralization in 1980s and the recentralization in 1994 have had a significant impact on tax capacity and tax effort of provincial governments in China. In this study, we use panel data analysis to capture the policy reactions of local governments to the central government tax reforms.

Figure n°2 : The Ratio of Total Tax Revenue Share of GDP (% GDP)

Source: China Statistical Yearbooks (1986–2004).

3. METHODOLOGY

This study attempts to examine the effect of the tax base on the tax capacity of provincial governments in China, and therefore the dependent variable is the tax share of GDP actually collected and the independent variables are the agricultural share of GDP, industry's share of GDP, the trade share of GDP, and the population density. The relationship between the dependent and independent variables can be summarized by the equation below :

$$\text{Taxshare} = f(\text{GDPpc}, \text{Ind}, \text{Agri}, \text{Trade}, \text{PopDensi})$$

Where :

Taxshare = Tax to GDP ratio (% of GDP)

GDPpc = GDP per capita, in thousand yuan

Ind = the ratio of industry to GDP (% of GDP)

Agri = the ratio of agriculture to GDP (% of GDP)

Trade = the ratio of import and export to GDP (% of GDP)

PopDensi = population density (People per Sq.Km)

3.1. Tax share of GDP

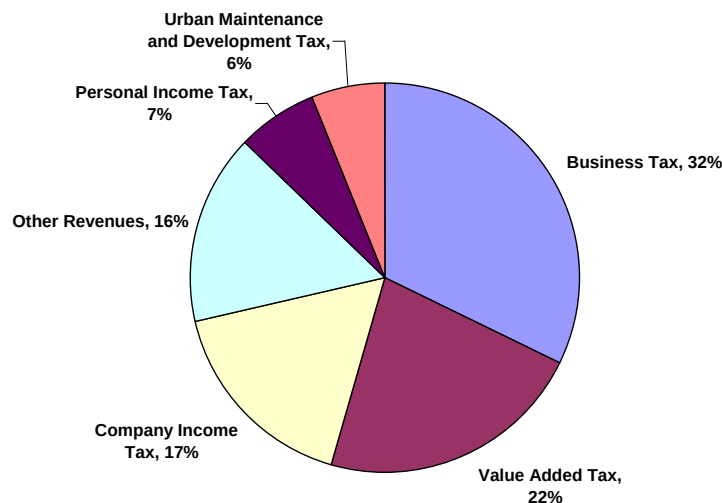
In this part, we will analyze the tax system of provincial governments before and after 1994, calculate the tax share of GDP as well as the tax growth rate, and discuss the tax buoyancy for each province.

In the pre-1994 fiscal system, the most important tax is “profits tax,” accompanied by value added tax, business tax, agricultural tax and so on. The central government stipulated a lump-sum tax obligation from provincial governments, based on their SOEs' profit for the current year. The tax

obligation will increase annually by an agreed rate if there are additional profits accruing to their SOEs. Usually, it is a fixed tax obligation for several years, but sometimes with an annual increment (Bahl and Wallich, 1992). Therefore, Wong (1992) argues that, under the fixed tax obligation, increased profits of SOEs will lead to a decreased representative tax rate.

In 1994, the tax sharing system consisted of central taxes, local taxes and shared taxes. Consumption taxes, tariffs and vehicle purchase taxes are all central taxes, while value added taxes, business income taxes, corporate income taxes and personal income taxes are shared taxes levied both at the central level and the local level. At the provincial level, China has introduced local taxes on very limited tax bases, including resource taxes, urban land use taxes, agriculture and related taxes, and taxes on contracts. Figure 3 uses 2005 data to demonstrate that value added tax and business tax usually comprise the largest share of taxes at the provincial level.

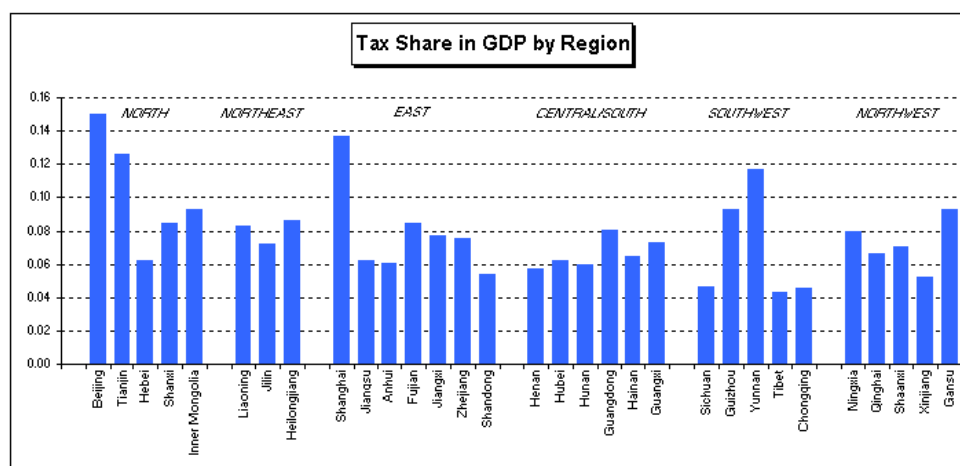
Figure n°3 : Main Tax Items of Provincial Governments in 2005



Source : Shen, 2008.

With respect to the average tax share of GDP for each province, as shown in Figure 4, Tibet has the lowest tax share of GDP (4.29%) followed by Chongqing, a new municipality entitled in 1996, and Sichuan, a province in the West region. Next comes Xinjiang, a minority province in the West region. Beijing has the highest share of GDP (14.97%), followed by the two municipalities of Shanghai and Tianjin. Surprisingly Shandong, which has a high GDP per capita and is located on the east coast, has the fifth lowest tax share of GDP, while, Yunnan, a minority province in the Southwest region, has the fourth highest tax share of GDP.

Figure n°4 : The Average Tax Share of GDP over the period of 1986–2004



Source : China Data Online (1986–2004).

3.2. GDP per capita

In Table 1 (see appendix), the provinces are ranked by average per capita GDP during the 1986–2004 period. The poorest provinces, which have the lowest average per capita GDP, are mostly inland provinces. On the contrary, the prosperous provinces with high per capita GDP are located in the coastal region.

Figure 5 in the appendix shows that Shanghai, the largest metropolitan city, had the highest average GDP per capita over the period 1986–2004. Beijing, the nation's capital, ranked number two in GDP per capita, followed by Tianjin, the third municipality after Shanghai and Beijing. Following the three municipalities rank the three coastal provinces of Zhejiang, Jiangsu and Guangdong. Guangdong is a coastal province favored by central government policies and was among the first to undertake economic reforms in 1978. Liaoning, one of China's heavy industrial centers, ranked number seven in per capita GDP. And Fujian ranks next, which has several special economic zones enjoying a special policy for the purpose of promoting economic development in that area. This can be accounted for by a special "open door" policy implemented in Guangdong and Fujian in 1978. Under this policy, four special economic zones in Guangdong and Fujian were established in 1980. In addition, 14 coastal cities were established as "coastal open cities" in 1984, which all work in favor of the coastal provinces, especially in the Southeast region. Due to the special "open door" policy, the coastal provinces were given not only special opportunity to develop their economy, but also special institutional environments and policies that grant them additional rights over

local economic activities beyond those of other provinces (Lin, Tao and Liu, 2003).

At the other extreme, Guizhou, a mountainous minority province in the Southwest, is the poorest province, followed by Gansu in the West region. Tibet, the minority province in the Southwest region, has very low per capita GDP. Sichuan, one of the most populous provinces in the West region, ranks the fourth poorest area, although its GDP is not among the lowest group. The Yunnan and Guangxi minority areas in the Southwest have very low per capita GDP.

Overall, most coastal provinces in the East region are rich provinces. While on the contrary, the minority provincial areas of the Southwest and the Northwest are among the poorer provinces.

3.3. Industry's share of GDP

“China's fiscal structure depends overwhelmingly on industry for the generation of revenues” (Wong, 1992). As Lin, Tao and Liu (2003) argue, since the first decentralization reform in 1957, the ownership of SOEs has been shifted from central government to provincial government. The tax revenue collection of provincial governments naturally fell on the shoulders of SOEs, since the provincial governments put their effort in revenue collection on the profits of SOEs. As Shen, Jin and Zou state in their report, provincial governments' revenue heavily rely upon their SOEs. Especially under the fiscal contracting system, the interests of the provincial governments are tightly linked with those of SOEs (Shen, Jin and Zou, 2006). SOEs and provincial governments have a strong connection not only because of the revenue collection, but also because SOEs provide their employees with basic services, which are otherwise supposed to be provided by provincial governments, such as education, health care, and pension services (Lin, Tao and Liu, 2003).

In the 1980s, over 80% of total local governmental revenues came from industry. The tax system remained narrowly focused on SOEs. However, Wong argues that this share has fallen with the Chinese fiscal reform, which has introduced a competitive market and declining profits of SOEs (Wong, 1992).

In the last decades, industry's share of GDP has grown in most of the provinces. Xinjiang (a minority province in the West region) has the highest growth rate of industry, followed by Hebei, an inland province in the North region. There are several exceptions, such as Tibet and Jiangxi, where industry's share of GDP has decreased by respectively 2.77% and 1.37% (see Table 1 in appendix).

The average share of industry of GDP during the period 1986 to 2004 is as low as 7.59% in Tibet, and as high as 51.94% in Shanghai (see Figure 6 in appendix). Shanghai is well known as a leading municipality in industrial and economic development. As provinces that strongly rely on heavy industry, Heilongjiang and Liaoning, in the Northeast region, have relatively high industry share of GDP. In most of the literature, industry's share of GDP

positively affects tax capacity, since urban areas need more public services than rural areas and in addition, there is a lower tax administrative cost in the industrial sector than in the agricultural sector (Bahl, 1971; Bahl, 1972; Chelliah, 1971; Mertens, 2003; Tanzi 1987; Tanzi, 1992). Therefore, we predict that industry's share of GDP is positively related to the tax share.

3.4. Agriculture's share of GDP

The values of the average agricultural share of GDP range from 2.63% in Shanghai to 39.27% in Tibet. Hainan, the island in the south of China, Guizhou, Guangxi and Jiangxi, in less developed, mostly inland areas — all of these provinces' revenues mainly stem from agricultural sources. By contrast, Shanghai, Beijing, Tianjin, Shanxi, Liaoning and Heilongjiang rely less on agriculture in their economy (See Figure 7 in appendix).

During the period 1986 to 2004, most of the provinces experienced a decrease in the agricultural sector, as a great amount of agricultural land was converted to industrial constructions for the purpose of urban development. Correspondingly, as shown in Table 1, agriculture's share of GDP has been diminishing. For example, Shanghai, Beijing and other eastern coastal provinces all decreased their agricultural share by more than 5%. However, agriculture's share of the GDP growth rate is 33.98% in Chongqing, 29.70% in Tibet, and 21.51% in Shanxi, where the largest number of coal mines are located (see Table 1 in appendix).

Agriculture is supposed to have a higher tax administrative cost than other sections, and rural areas enjoy fewer public services, which all make agriculture a negative factor that affects the tax capacity of a jurisdiction. However, some scholars (Lin, Tao and Liu, 2003) argue that in China, rural taxes and land requisition are charged excessively and abusively. Under this circumstance, one could expect agriculture to have a positive impact on the provincial governments' revenue.

3.5. Population Density

Tax capacity also depends on the population density. In China, the most populous area is Shanghai where the average population density over period 1986–2004 amounts to 2,100 inhabitants per Sq.Km, followed by the other two municipalities, Beijing and Tianjin. The least densely populated area is Tibet with only 2 inhabitants per Sq.Km, followed by Qinghai, Xinjiang and Inner Mongolia, all of which are minority areas (see Figure 8 in appendix). In China, the effect of the population density on tax capacity could lead to two diverging outcomes. On the one hand, a more populous area could result in a negative impact on tax capacity because of a high level of tax concession. On the other hand, a high density population area could play a positive role on the collection of tax revenue because of the reduced administrative cost.

3.6. Trade's share of GDP

Trade's share of GDP is used to measure the degree of openness, which is calculated by dividing the sum of imports and exports by GDP. The more open and the more developed the economies, the greater the tax bases. Figure 9 in the appendix shows that the most open area is Guangdong, which is also the first province that established a special economic zone and opened its door to the whole world. After this first open province rank Shanghai and Tianjin, the two coastal municipalities. Most of the coastal provinces, such as Fujian, Hainan and Jiangsu are very open too. The least open area is an inland province in the center, Henan, which is "a political and economic center of ancient China" (Zhang and Zou, 1998). Guizhou, Qinghai, Sichuan and Chongqing, in the Western inland area, rank second to fifth as the least open areas.

The effects of the share of industry, agriculture and trade, as well as the impact of the population density on provincial governments' tax revenue will be tested in the next section.

4. MODEL SELECTION AND EMPIRICAL RESULTS

This study uses panel data in order to allow for time and province heterogeneity. Without controlling the unobservable effects, the coefficients may be biased and inconsistent due to an omitted variable bias. For example, one such unobservable factor is the central policy. Both fixed effects and random effects can capture heterogeneity along both time and province dimensions. Specifically, three models are used:

$$\text{Pooled regression : } y_{it} = X'_{it}\beta + \alpha + \varepsilon_{it} \quad (1)$$

$$\text{Fixed effects : } y_{it} = X'_{it}\beta + \alpha_i + \varepsilon_{it} \quad (2)$$

$$\text{Random effects : } y_{it} = X'_{it}\beta + \alpha + \mu_i + \varepsilon_{it} \quad (3)$$

In equations (1), (2) and (3), i is the index for individual provinces and t denotes time or year. If there are no unobserved effects, equation (1), OLS is suitable to provide unbiased, consistent and efficient estimates. In equation (2), the fixed effects, α_i , capture the fixed individual effects. In equation (3), the province specific component in the error terms, μ_i is a group specific random element, which allows these unobservable effects to be randomly distributed across cross-sectional units.

When choosing between a fixed effects model and a random effects model for the time variable, we have chosen to use a fixed effects model. The reason is that we can only examine the effects of central fiscal on the tax share through the use of a fixed effects model. If a random effects model is used, the time variable, whose coefficient represents the effect of the central fiscal policy, would not enter the regression as an explanatory variable (it enters as one component of the error term).

For provincial effects, two tests are conducted to help choose a desirable model among pooled regression techniques, random effects and fixed effects. The first test is the Breusch and Pagan Lagrangian Multiplier test for random effects against pooled OLS. The LM test statistic is 1290.59 ($p < 0.01$). Hence we reject the null hypothesis that there are no such group-specific random elements. Then, we go on to use the Hausman test for a fixed effects model versus a random effects model. The Hausman test statistic is 23.63 ($p = 0.37$); on this basis, the null hypothesis cannot be rejected. In this case, the random effects model is consistent and efficient, but the fixed effects model is not efficient although still consistent. The reason is that there is no correlation between the included independent variables and the random effect (Greene, 2003). As a result, a provincial random effects model has been implemented. In this paper, we use fixed time effects and random provincial effects at the same time, denoted by:

$$y_{it} = X'_{it}\beta + \alpha_t + \mu_i + \varepsilon_{it} \quad (4)$$

μ_i is the unobservable province specific effect while α_t represents the fixed time effects that capture the impact of policy changes that affect all provinces each year. Estimations are carried out using the STATA statistical software package.

The tax capacity of a province is measured as a function of its GDP per head, the share of agriculture, trade and industry of GDP, and the population density. The model performs generally well with most of the variables significant at the .01 level. The signs of the coefficients are generally consistent with expectations. The results show that tax capacity is negatively, though not significantly, related to the level of per capita GDP. Also consistent with previous findings, industry's share plays a positive role in determining tax capacity. In other words, the higher the level of industrialization, the greater the capacity to raise taxes. Agriculture's and trade's share of GDP play a negative role in generating tax revenue. In addition, the importance of the population density as a major determinant of the level of tax capacity is not reliable, since it is not significant.

When comparing the effects of decentralization and centralization, the results of my panel analysis show that they have opposite effects on the level of the tax share. Decentralization had a positive and significant impact on the level of the tax share, while recentralization has had a negative impact. In the regression, the constant is dropped so that all of the time dummies can be included. The time effects are all significant except 1987. According to the Chow tests, the coefficients of the years after 1994 are significantly different from those before 1994. The results can be seen clearly in Figure 10. We plot the coefficients of the time dummies. There was clearly a slump in 1994.

Theoretically, there are two major views that can be used to explain the effect of decentralization on provincial tax capacity, i.e. Brennan and Buchanan's "Leviathan" model and Oates' model (Bird, Martinez-Vazquez and Benno, 2004).

Table n°2 : Estimation Results for the Determinants of Tax Share using a Fixed Time Effects and Random Provincial Effects Model

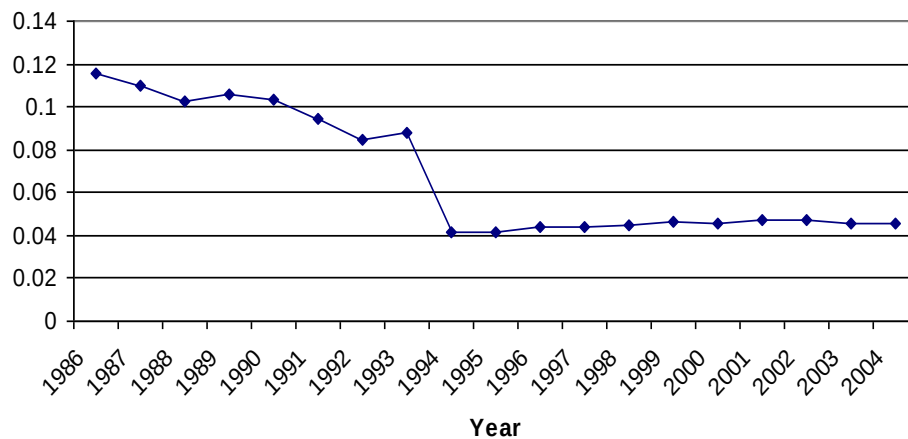
Independent Variable	Coefficients	Z Statistics
GDPpc	-0.0004	-1.42
Agrishare	-0.0950***	-3.41
Indushare	0.0811***	3.16
Trade	0.0110***	2.69
Popdensi	0.0279	0.29
year_1986	0.1152***	6.99
year_1987	0.1099	-1.22
year_1988	0.1025***	-2.92
year_1989	0.1061**	-2.07
year_1990	0.1031***	-2.76
year_1991	0.0948***	-4.55
year_1992	0.0845***	-6.58
year_1993	0.0877***	-5.70
year_1994	0.0417***	-15.43
year_1995	0.0414***	-15.22
year_1996	0.0441***	-14.47
year_1997	0.0438***	-13.88
year_1998	0.0451***	-13.08
year_1999	0.0465***	-12.04
year_2000	0.0458***	-11.67
year_2001	0.0475***	-10.96
year_2002	0.0476***	-10.64
year_2003	0.0454***	-10.59
year_2004	0.0456***	-9.51
R ² (Within)	0.7483	
R ² (Between)	0.3848	
R ² (Overall)	0.5760	
LM	1290.59	
H	23.63	

Notes : N=527. ** $p < 0.05$, *** $p < 0.01$. LM is the Breusch and Pagan Lagrangian multiplier test for random effects against pooled OLS. H is the Hausman test for Fixed versus random effects.

In Brennan and Buchanan's view, governments act as a Leviathan and seek to maximize their tax revenues through exploiting their tax base. Just like the private sector's desire to maximize profit, governments' rational behavior leads to increasing tax burdens and growing government size. They argue that the Leviathan behavior of governments can only be constrained by the

constitution that limits their access to tax and other fiscal instruments that encourage decentralization or federalism (Brennan and Buchanan, 1980). Based on this premise, Nelson (1986 and 1987) finds that decentralization divides a monolithic power into a number of relatively homogenous governmental units, and in turn this may result in competition among local governments in lowering taxes lest taxpayers vote with their feet or investments move to jurisdictions with lower tax rates. As a result, decentralization may serve as a constitutional constraint in limiting the revenue generating power of local governments. Marlow (1988) adds, “if greater decentralization in government increases competition in the public sector, then greater decentralization may lead to relatively low tax burdens.” Furthermore, Marlow’s study concludes that a decreasing federal share of total government could strengthen the importance of local governments in overall governmental activity. As a result, competition among local governments will cause them to lower tax shares. An alternate perspective offered by Oates contradicts these theories. Oates (1972 and 1985) argues that in a more decentralized system of government, local governments tend to increase public spending and the level of tax shares to meet their voters’ demands for government effectiveness and efficiency.

Figure n°10 : Coefficients of the Fixed Time Effects



The result reported above gainsays Brennan and Buchanan’s theory. This divergence can be explained by the lack of interjurisdictional mobility of people in China and the absence of explicit fiscal constraints on the taxing power of provincial governments. Consistent with Oates’ theory, under a decentralized system, provincial governments tend to increase tax revenues as a result of fiscal decentralization in order to meet their residents’ demands for public services. Therefore, the contract system can in theory provide incentives for provincial governments to collect revenue. In contrast, under the tax-sharing system, the high sharing rate with the central government may discourage tax collection at the provincial government level. As Bahl and Wallich (1992) state, if provinces are only able to keep a small proportion of what they collect, they

may not have the incentive to increase tax share. Therefore, under the recentralization of the tax system, provincial governments' tax collection performance is expected to decline.

5. TAX CAPACITY AND TAX EFFORT OF THE PROVINCIAL GOVERNMENT

In this section we will report the predicted tax capacity based on the data and model presented in earlier sections, discuss the provincial tax authority, and calculate tax effort indices for each province.

5.1. Tax Capacity

One can predict the tax share, or tax capacity, based on the level of per capita GDP, agriculture's and industry's share of GDP, and population density, using the coefficients generated in the earlier model. In this case the fixed time effects are not included in the tax capacity values because we do not want the provinces' tax capacities to be influenced by the national policies.

Figure n°11 : Average Tax Capacity over the period 1986–2004

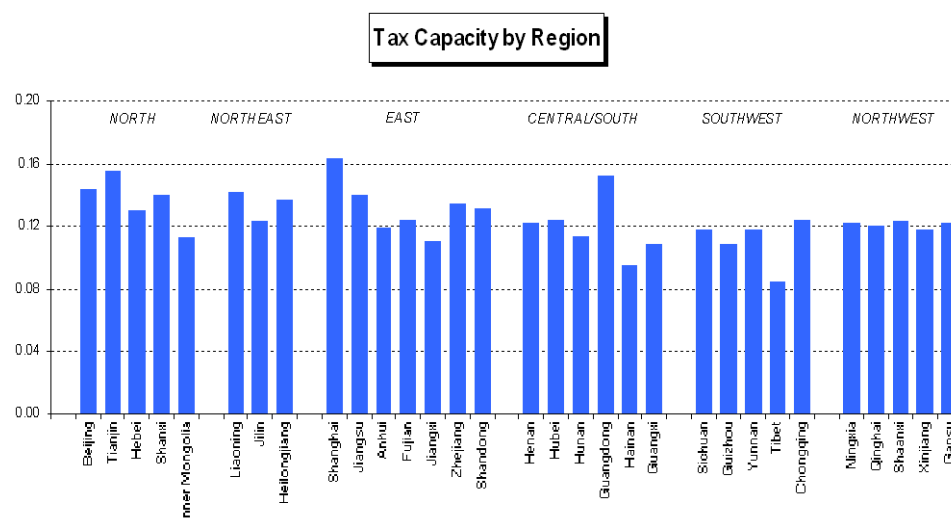


Figure 11 shows the average tax capacity of each province. Shanghai has the highest tax capacity and Tibet has the lowest tax capacity. Not surprisingly, the least developed provinces, such as Hainan, Guangxi, Guizhou, Jiangxi, have comparatively low abilities to raise taxes, while the most developed provinces have the highest abilities to levy taxes, including Tianjin, Guangdong, and Beijing. Liaoning, one of the heavy industrial centers in the northeast, ranks fifth in the level of tax capacity. Most of the top ten provinces are located in the east coast or northeast regions, except Hebei in the northern region. Facing their hypothetical capability to raise tax revenue, do provincial governments have the incentive or authority to control the extent to which they exploit their tax bases? The answer is yes.

5.2. Provincial Tax Authority

It is well known that China has a uniform tax system, under which tax rates and tax bases are determined by the central government. However, as Wong (1997) states, the Chinese tax system is ad hoc and negotiable, and provincial governments, to some extent, are entitled to change the de facto rate by offering special policies to their SOEs. Due to provincial protectionism, provincial governments are incited to award tax breaks to enterprises within their jurisdiction, which are called tax expenditures. Similarly, if provincial governments can get compensation from the central government through alternative sources, such as grants and subsidies, for inadequate local tax revenue, they have little incentive to collect the full tax from their tax base. To control the provincial governments' tax expenditures, and for redistributive purposes, the central government needs to know to what extent provincial governments are utilizing their tax capacity.

Tax expenditure

Even though tax rates are nominally set centrally, provincial governments still have an important impact on the amount of tax revenues raised within their jurisdictions. Provincial governments play the role of administering and collecting taxes and have a substantial degree of freedom to affect the level and composition of collected taxes, which determines the effective tax rate for their region. As Bahl and Wallich argue, "provincial governments have a surprising amount of discretion in granting tax relief," which is referred to as the policy of "stimulating enterprises through tax expenditures." Provincial governments, in most cases, award their SOEs tax concessions, which can substantially alter the effective tax rates paid by SOEs. Especially with the economic reform, or "open door" policy, markets in China tend to be increasingly competitive. Provincial governments are eager to attract additional investment from all over the world by offering special tax breaks, tax concessions and tax holidays (Bahl and Wallich, 1992). Nonetheless, this autonomy of provincial governments can result in serious problems if the tax concessions they offer are in conflict with the central governments' policy and can be detrimental to the fiscal environment as a whole.

Alternative Sources

Provincial governments have other revenue sources beyond their own tax revenue, including shared taxes with the central government, extrabudgetary funds, non-tax fees, tax rebates, earmarked grants, capital grants, and international aid (Bahl and Wallich, 1992; Zhang and Zou, 1998). These alternative sources have a significant impact on the tax effort, since provincial governments expect the central government to transfer grants so as to offset their deficit.

In addition, provincial governments can utilize "extra-budgetary revenues" to meet their expenditure needs, which includes "user charges of living infrastructure, various quasi-fiscal fees levied on provincial enterprises or

direct illegitimate fee charges on farmers by provincial governments who have almost all the autonomy of levying and spending the fees” (Wong, 1998). However, this extra-budget revenue falls outside the control of the central government. This lack of accountability may hurt the transparency of the fiscal system (Lin, Tao and Liu, 2003). Also, this provides opportunities for corruption.

In order to improve the transparency and accountability of provincial governments’ tax administration, it is necessary for the central government to have a clear idea of the extent to which provincial governments collect tax revenues from their own tax bases. With mandatory accounting practices, the tax effort could be better measured by the ratio of the actual tax share to the estimated tax share.

5.3. Tax Effort Indices

In this section, we employ the OECD method to calculate tax effort indices for each province by dividing the actual tax shares by the predicted tax shares. Tax effort indices suggest the willingness of provinces to use the available tax capacity to finance public expenditures. The higher a tax effort index, the greater the extent to which the province has exhausted its capacity for further taxation. This increases the likelihood that the province will have to explore other fiscal resources, such as central government subsidies and international aid. The national average of tax effort indices during the period 1986 to 2004 is 1.05, close to one, which suggests that the overall extent to which provinces utilize their tax capacity is close to the ideal one. It is noteworthy that the tax effort index of each province varies over time and the trends differ from each other.

Figure n°12 : Tax Effort Indices in the Northern Region

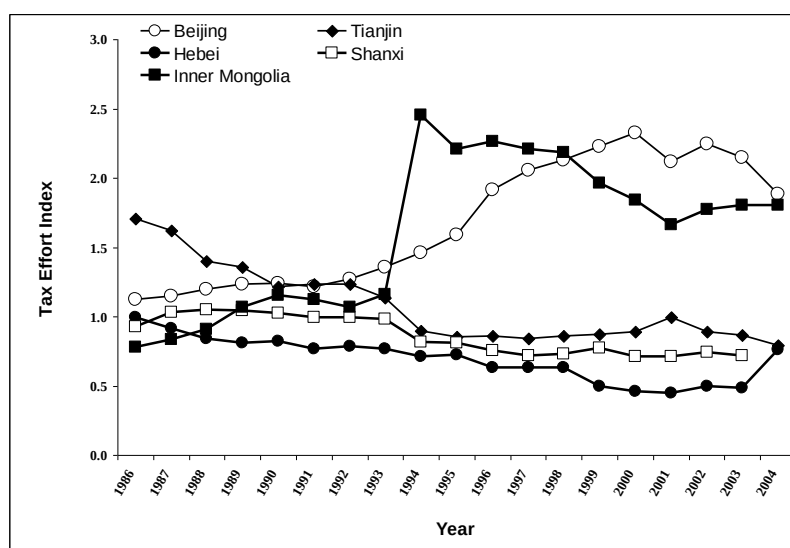
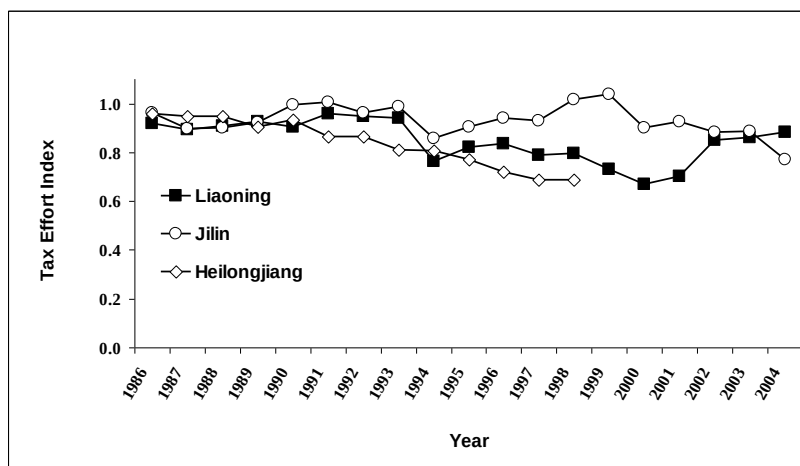
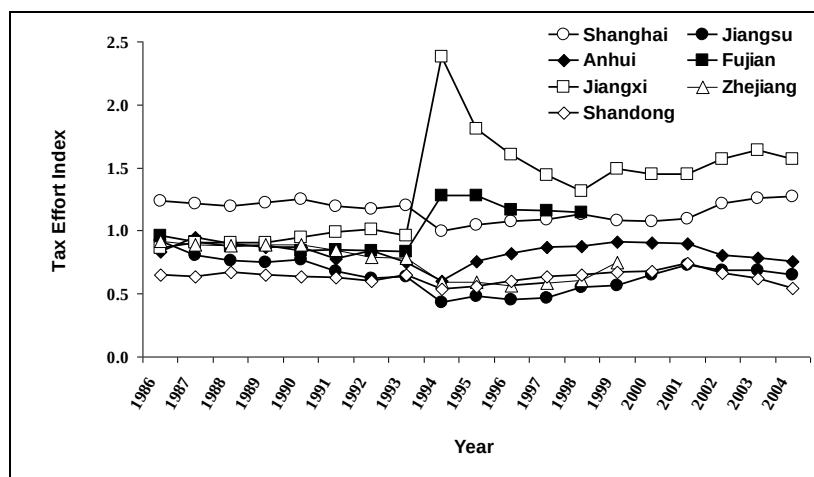


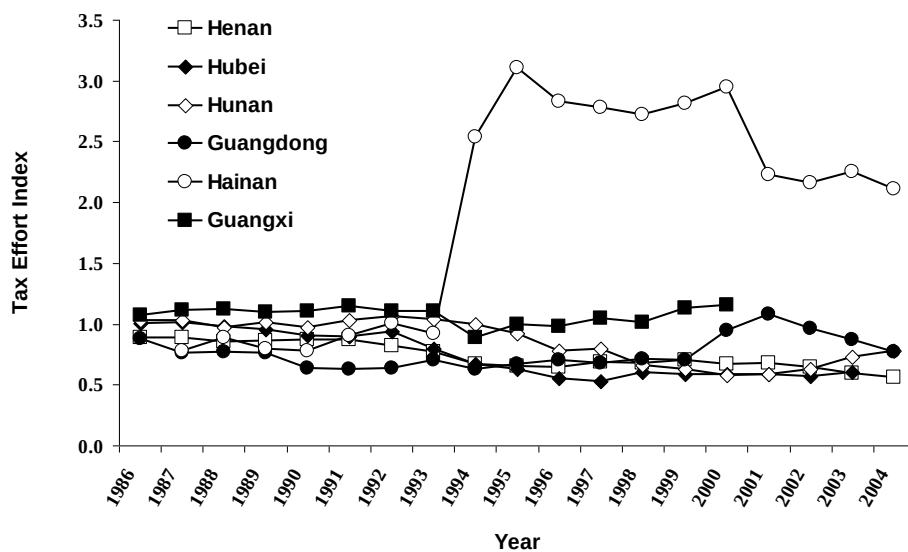
Figure n°13 : Tax Effort Indices in the Northeast Region**Figure n°14 : Tax Effort Indices in the Eastern Region**

In the Northeastern region, the three provinces, Liaoning, Jilin and Heilongjiang all have low tax effort indices. The trends are gradually decreasing with an average tax effort of approximately 0.8, which is below one. Since the three provinces are heavy industrial centers, the provincial governments are likely to offer tax breaks or tax holidays to their SOEs to help them in periods of hardship and promote their development. This would explain their low effective tax rates.

In the Eastern region, Anhui, Jiangsu, Zhejiang and Shandong are stable with mostly low tax effort indices, but all of them have a slump in 1994, which indicates the process of tax effort erosion accelerated under the economic reform. These provinces end up with indices averaging 0.75 in 2004. Jiangxi was stable before 1994, around one, but jumped to 2.38 in 1994 and then fell to 1.50 sharply in 1997. Shanghai's tax effort indices stay slightly above one.

In the Central South, most of the provinces' tax effort indices stay between 1 and 0.5, except Hainan. Hainan, an island in the South of China, which experiences a dramatic development in real estate and tourism after the open door policy. Hainan's tax effort indices soared sharply to 2.54 in 1994 and 3.11 in 1995, and fell slightly to 2.78 in 1997, but still maintain a high tax effort around 2.11 in 2004. This is partly because Hainan has taken advantage of the tax reform in 1994, which motivated efforts to raise tax revenue on commerce and services.

Figure n°15 : Tax Effort Indices in the Central/Southern Region



The tax effort indices of provinces in the Northwest were close to each other before the 1994 reform, but afterwards a divergent trend appears. Ningxia, Xinjiang and Gansu have average tax effort indices above one. In addition, Gansu's tax effort indices fell sharply after the 1994 reform from above 1.2 to around one.

In the Southwest, Guizhou and Yunnan (on the southwest border) have average tax efforts above one. Sichuan and Chongqing (located inside Sichuan) have average tax efforts close to one. Tibet witnessed a dramatic change from

1.00 in 1993 to 3.96 in 1994 and 4.68 in 1995, and then dropped to 1.13 in 2004.

Figure 16: Tax Effort Indices in the Northwestern Region

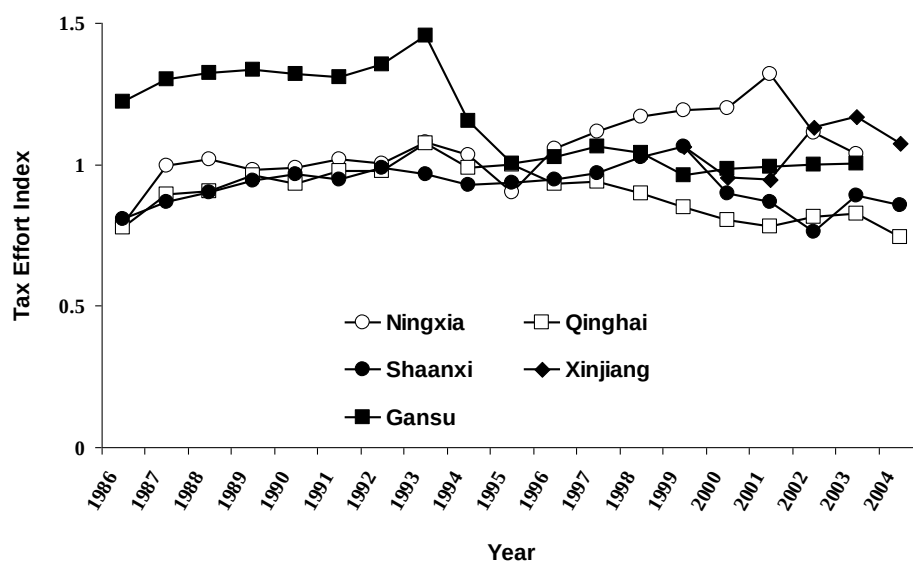


Figure n°17 : Tax Effort Indices in the Southwestern Region

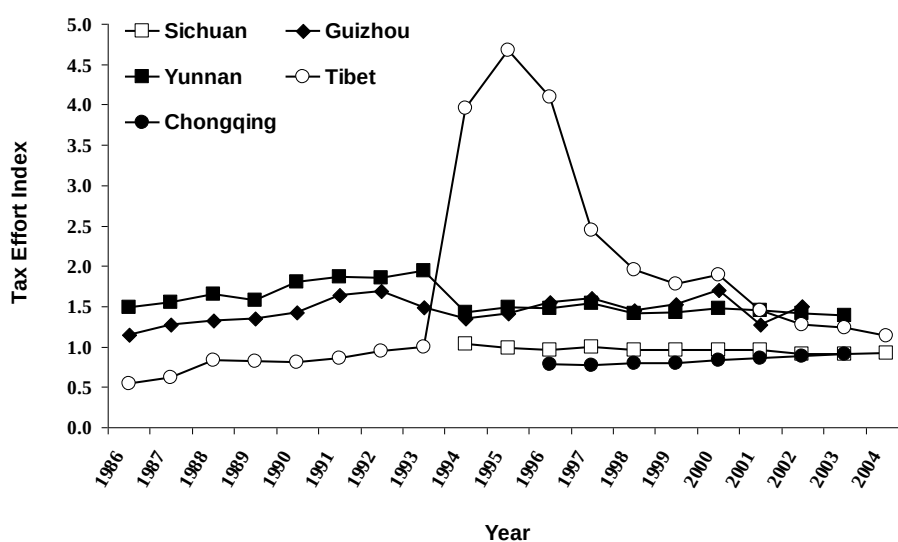
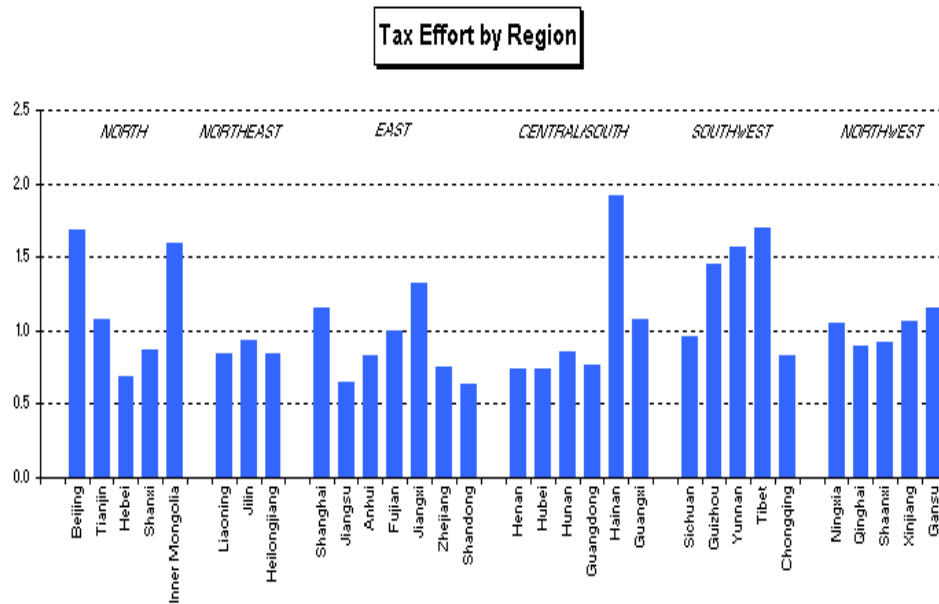


Figure n°18 : The Average Tax Effort Indices over the period 1986–2004

Nationwide, Hainan, Tibet and Inner Mongolia feature the highest tax effort, while, Shandong and Jiangsu, two coastal regions, have the lowest average tax effort. In general, tax collections are higher in provinces where per capita income is lower. In other words, some low-income provinces collect more taxes than might be predicted by their tax capacity. Poorer provinces, such as Inner Mongolia, Gansu, Guizhou, Yunnan, and Tibet, have above-average tax efforts, especially Inner Mongolia and Tibet, 60% above average, which indicate a regressive tax effect. They may wish to look for alternative financial resources, such as grants from the central government or international agencies, because there is limited room for them to further utilize their tax bases in order to meet expenditure needs.

At the other end of the spectrum, many of the higher income provinces appear to exert a lower level of tax effort: Jiangsu, Zhejiang, Shandong and Guangdong all register at 80% of the average tax effort. These provinces are not limited by a low capacity to generate tax revenues. Rather, for different reasons, they have problems with exploring their potential to collect taxes. For example, they may need to consider lowering administrative costs.

The results indicate that provinces in the coastal region generally have relatively low indices of tax effort. In addition, some provinces have substantially increased their tax efforts in recent years while others have experienced marked declines. The results suggest that most provinces, such as Yunnan, Guizhou, Sichuan, and Shanxi, are relatively stable in their tax effort

indices over the period 1986–2004 and are close to one. However, some provinces experience a dramatic change during the same time period; for example, Tianjin and Qinghai experience a downward trend.

5.4. Coefficient of Variation

From Figure 19, the tax capacity across provinces shows a higher degree of divergence after 1994 than before 1994. The coefficient of variation increased by 80% after 1994. The increased variation in tax capacity provides further evidence for increased disparities in the ability to collect tax revenue since 1994. But the variation coefficients go slightly down after 1995 and then go up again after 2002. This could be explained by the increasing divergence of economies across regions since the economic reform.

The variation coefficient of tax efforts has grown from 0.24 to 0.45, indicating growing dispersion of tax effort among the provinces during the period 1986 to 2004. After the reform in 1986, the coefficient of variation for tax effort has risen slowly. From 1986 to 1993, the coefficient of variation ranged from 0.25 to 0.29. But in 1994, the coefficient of variation jumped to 0.66. Therefore, inequity in the tax effort appears to have widened probably due to disparities of tax administration and tax expenditures in 1994. However, the variation coefficient for tax effort has fallen since 1995 (following the financial reforms in 1994) from 0.73 to reach 0.45 by 2004. There has been a converging trend of tax effort among provinces since 1995.

Figure n°19 : Variation Coefficient of Tax Capacity

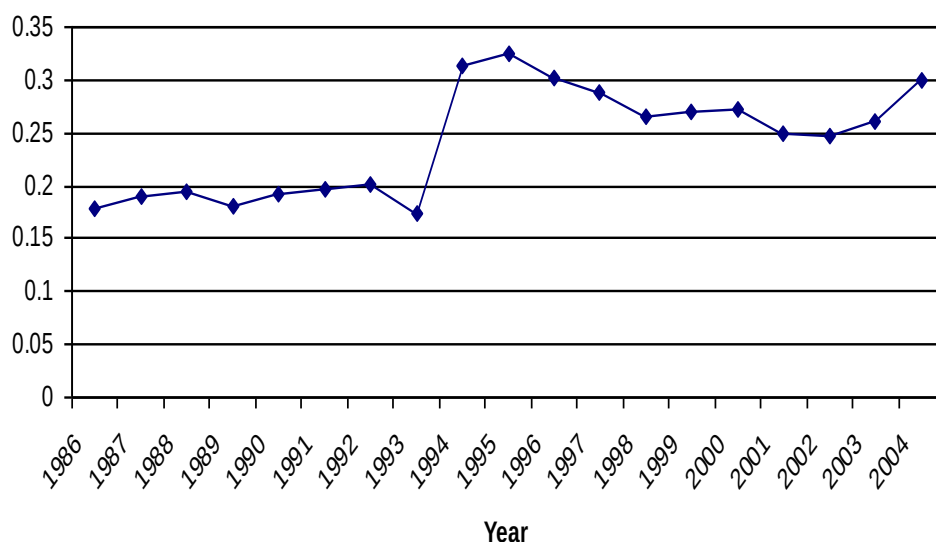
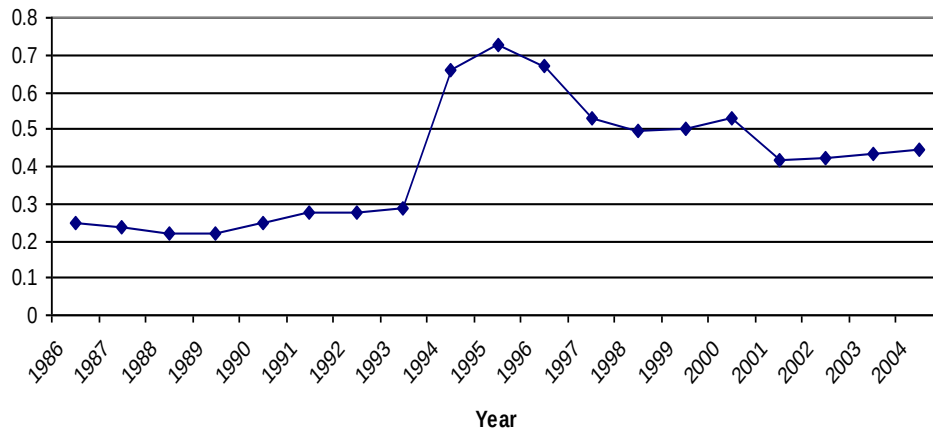


Figure n°20 : Variation Coefficient of Tax Effort

Overall, the fiscal reform in 1994 has led to highly differentiated tax capacities across regions as well as a high heterogeneity of tax effort.

6. POLICY IMPLICATIONS

This study provides two major policy implications regarding tax capacity and tax effort. The first is related to the redistribution issue. Tax capacity and tax effort indices could help determine the amount of resources that should be allocated to each provincial government. The second is related to the tax administration of the provincial governments.

6.1. Redistribution

The central government could use an approach based on tax capacity and tax effort indices for allocating grants and subsidies among its provinces. Before going any further, it is necessary to review both the past and the current redistribution and transfer of shared taxes, tax rebates, grants, and subsidies in China. Tremendous changes have occurred after 1994.

In the 1980s, for the total amount of tax subject to sharing, the shared rate was the result of negotiations between the central government and provincial governments. According to Bahl and Wallich's (1992) sharing formula, the shared rate was determined by combining the original amounts of tax collections and negotiation. Historically, the redistribution of shared taxes, grants, and subsidies to the provinces was determined by the ratio of the actual amount of "allowable" provincial government expenditures over the actual amount of provincial fixed plus shared revenues collected. Usually the least developed and minority provinces received a deficit subsidy. The other approach was a fixed tax quota contracted with the central government. To get a desirable contract, provinces bargained with the central government for an ad hoc tax quota. In the bargaining process, the prosperous and high-yield provinces, such as Jiangsu, Zhejiang, Beijing, Guangdong, and Shanghai,

typically took advantage of their special economic development policies when negotiating with the central government for greater subsidies (Bahl and Wallich, 1992; Zhang and Zou, 1998).

After the 1994 reform, among all the grants and subsidies transferred from the central government to the provincial governments, tax rebates became the largest subsidy. The size of the tax rebate is highly correlated with the income level, which is a regressive effect. Therefore, this method of redistribution has little equalizing effects because the coastal provinces and the most developed regions are favored (Lin, Tao and Liu, 2003). The earmarked grants, the second largest transfer item, were designed as subsidies for food and other consumer goods, which favor urban areas, and which also still have the problem of a regressive effect (Wong, 1997). To compensate for this inequality problem, in 1996, the government introduced an equalizing transfer to aid poor regions. The transfer is based on variables from both the supply side, such as GDP, and the demand side, such as student-teacher ratios, the number of civil servants, and the population density (Lin, Tao and Liu, 2003). To a certain extent, this approach could be considered as a redistribution based on the tax capacity.

Additionally, the tax effort of a government is viewed by some political entities as an indicator of the desirability for allocating further resources to that government. For example, international lending agencies use measures of tax efforts as a basis for allocating grants, thus favoring high tax effort countries (Leuthold, 1991). Similarly, in some countries, the central government uses the capability of local governments to generate tax revenues as the basis for judging their performance, and in turn allocates its grants to each local government accordingly. In addition, in countries including Canada, Australia, Germany, and Denmark, the redistribution system is based on tax capacity equalization. The equalization transfers are designed to offset tax capacity differentials (Ahmad and Craig, 1997).

Furthermore, both the tax capacity and tax effort should be taken into account when considering the redistribution. The reason for using tax effort indices to determine the redistribution rate is to give provincial governments greater incentives to exploit their own tax base. Of equal or greater importance, using the tax effort will, to some extent, offset the aforementioned problem of a regressive effect. As shown in the previous section, some provinces with low per capita GDP have tax effort indices far above one, for example, Inner Mongolia, Gansu, Guizhou, Yunnan, and Tibet. Grants and subsidies should be distributed to them since they are limited in their potential to utilize tax bases to meet their expenditure needs.

Ideally speaking, according to Boadway (2001), under the equalization redistribution, each provincial government with a comparable level of tax effort should be provided with a comparable tax capacity to make a uniform set of public services available. Therefore, redistribution should reflect all three factors: differences in hypothetical tax bases, the extent to which the provincial governments utilize their tax bases, and differences in need across provinces.

6.2. Administrative reform

The final amount of tax revenue at each government level depends not only on the tax base and the tax rate, but also on the tax administration of governments. A low efficient administration with high administrative costs (defined as the cost for government agencies to collect tax) will decrease the tax effort index significantly. In this regard, low tax effort indices might be seen as reflecting administrative problems of provincial governments, such as the failure to reform public administration and the inefficiencies introduced by under-qualified government officials and by the intervention of enterprises in the provincial administration.

The administrative reform, which aims at both lowering the administrative cost and improving administrative efficiency, has been carried out from the central government to all levels of subnational governments. Before 1994, it was the provincial tax administration's responsibility to collect tax revenue and submit it to the central government. In order to improve administrative efficiency and to limit the provincial tax administration power, in 1994, the central government split the tax administration into two parts, namely the national tax administration and provincial tax administration. The former is in charge of collecting central taxes and shared taxes. The latter is responsible for local taxes only.

The central government can judge the achievement of provincial government officials by examining their tax effort indices. For the provinces with tax effort indices situated significantly below one, such as the two coastal provinces Shandong and Jiangsu, such a judgment may introduce a potentially serious problem concerning the officials' efficiency and special relationships with enterprises, in that the most profitable enterprises may end up paying less tax than they should.

CONCLUSION

This study has shed light on the tax performance of provincial governments in China. The analysis carried out in this study comprises two steps. First, we use a fixed time effects and random provincial effects model to analyze the statistical relationship between tax shares and economic and demographic variables, including per capita GDP, the share of agriculture, industry and trade, and the population density. In general, the decentralized fiscal system over the period of 1986 to 1993 had a positive impact on the tax share of GDP, whereas the recentralized system over the period from 1994 to 2004 dramatically decreased the tax share of GDP.

Secondly, we employ the estimated coefficients from the model to calculate tax capacity and tax effort indices for each province in China. Tax effort indices for each province vary over time and provincial trends show significant differences. The results suggest that some prosperous and coastal provinces, such as Shandong, Jiangsu, and Guangdong, which have a high tax capacity, show relatively low tax efforts. These provinces may consider placing

greater emphasis on administrative reforms as a means to increase local tax revenues and therefore reduce their reliance on other funding resources. On the other hand, some poorer inland provinces, such as Guizhou, Gansu, and Tibet, have a low tax capacity and a high tax effort. They may wish to look for alternative financial resources because there is limited room for them to exploit their tax bases to meet expenditure needs. The results for the variation coefficient of tax capacity and tax effort indices indicate that the fiscal reform in 1994 has led to greater differentiation of tax capacities across provinces and also to more divergence in tax efforts.

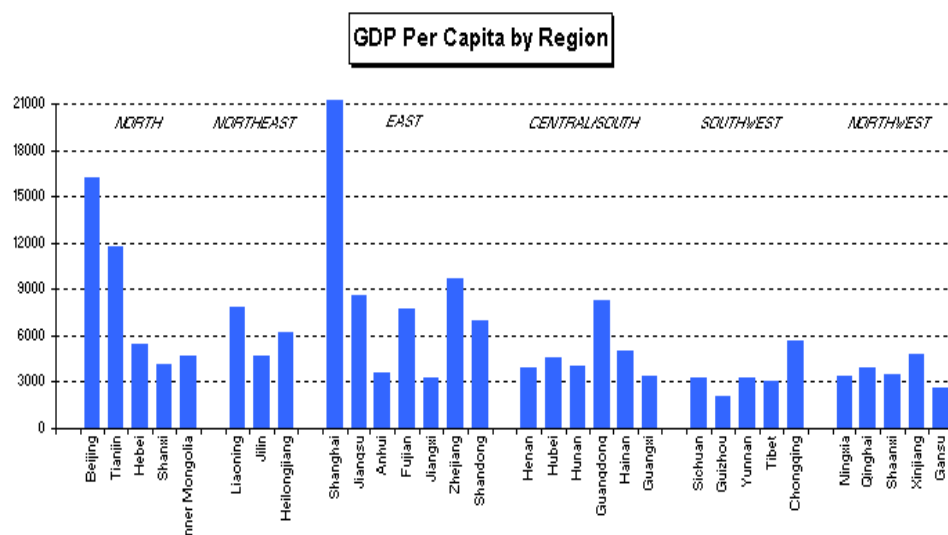
The findings from this study have important policy implications. First, tax capacity and tax effort indices could help the central government to redistribute grants and subsidies to each province. Second, these measures can also help to make judgments about the administrative efficiency of provincial governments. Along with information on expenditure needs, alternative financial sources, and political and cultural differences among provinces, measuring tax capacity and effort can provide valuable, and somewhat objective, information on the levels of tax utilization in individual provinces.

APPENDIX

Table n°1 : Average values over the period 1986–2004.

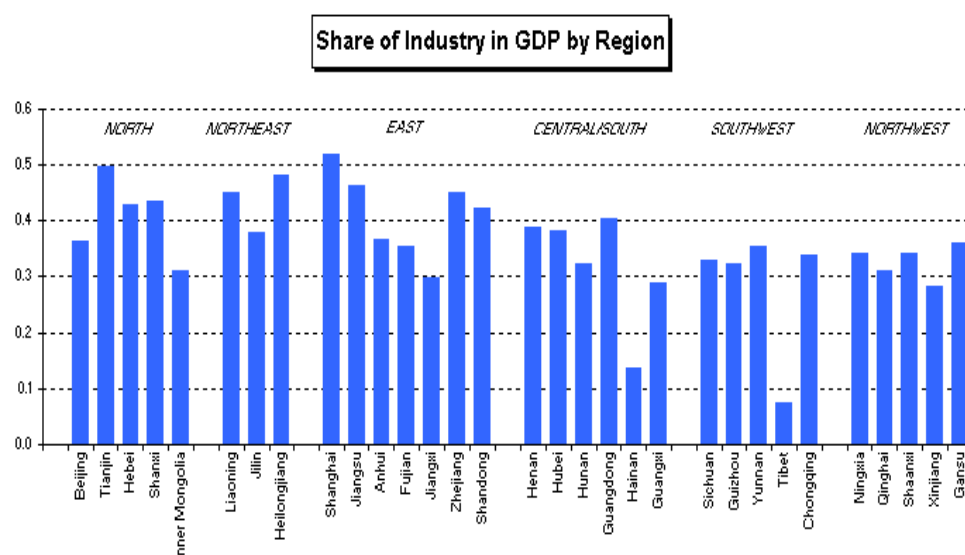
Province	GDPPC	GDPpc growth	Tax share	Tax growth	Buoyancy	Agri share	Agri growth	Indu share	Indu growth	Popden	Tax Capacity	Tax effort
Shanghai	21478.32	15.96%	13.71%	12.97%	0.98	2.63%	-10.94%	51.94%	0.93%	2100	0.16	1.16
Beijing	16229.70	15.55%	14.97%	15.97%	1.17	5.42%	-10.40%	36.31%	-1.31%	695	0.14	1.68
Tianjin	11775.42	15.61%	12.59%	8.55%	0.66	6.32%	-8.40%	49.79%	3.17%	787	0.16	1.08
Zhejiang	9645.20	18.17%	7.55%	15.57%	1.23	16.25%	-6.69%	45.14%	1.56%	431	0.13	0.76
Jiangsu	8555.75	17.68%	6.23%	15.41%	1.30	16.98%	-6.05%	46.46%	2.94%	683	0.14	0.65
Guangdong	8315.00	17.40%	8.01%	18.42%	1.32	16.70%	-5.30%	40.27%	1.82%	366	0.15	0.76
Liaoning	7877.70	13.98%	8.26%	10.61%	0.89	13.19%	-3.36%	45.06%	0.36%	276	0.14	0.85
Fujian	7673.95	18.50%	8.45%	20.76%	0.88	21.58%	-0.89%	35.35%	1.40%	264	0.12	1.00
Shandong	6927.15	17.63%	5.36%	15.57%	1.11	20.70%	-2.44%	42.40%	2.17%	566	0.13	0.63
Heilongjiang	6155.60	14.28%	8.65%	8.80%	0.71	16.22%	-1.34%	48.31%	0.79%	78	0.14	0.84
Chongqing	5637.67	10.27%	4.57%	14.40%	1.79	18.69%	33.98%	34.03%	-0.45%	376	0.12	0.83
Hebei	5419.70	17.04%	6.24%	10.09%	0.67	19.95%	-4.42%	42.78%	10.80%	336	0.13	0.69
Hainan	4983.25	15.39%	6.47%	16.90%	0.97	38.83%	1.08%	13.79%	-1.09%	209	0.10	1.92
Xinjiang	4811.53	15.11%	5.26%	16.93%	1.95	27.90%	-3.15%	28.32%	15.66%	10	0.12	1.06
Inner Mongolia	4695.35	16.80%	9.30%	17.95%	1.27	27.58%	-2.74%	31.16%	0.78%	19	0.11	1.59
Jilin	4647.89	14.64%	7.20%	10.59%	1.02	24.32%	-1.61%	37.97%	1.72%	135	0.12	0.93
Hubei	4551.95	14.94%	6.22%	10.15%	1.04	24.31%	-1.67%	38.13%	0.36%	304	0.12	0.74
Shanxi	4128.60	16.44%	8.43%	11.07%	0.87	13.04%	21.51%	43.42%	0.05%	197	0.14	0.86
Hunan	4010.10	15.49%	5.99%	11.10%	0.95	29.11%	-0.08%	32.20%	-0.59%	302	0.11	0.85
Henan	3940.95	16.64%	5.74%	11.66%	0.92	26.29%	2.69%	38.95%	0.24%	543	0.12	0.74
Qinghai	3917.40	13.57%	6.63%	12.14%	1.08	19.97%	-2.63%	31.02%	0.71%	7	0.12	0.90
Anhui	3609.90	14.28%	6.02%	12.00%	1.29	28.05%	-3.82%	36.76%	0.73%	429	0.12	0.83
Shaanxi	3460.15	15.19%	7.03%	13.21%	1.10	20.80%	-2.25%	34.20%	0.66%	169	0.12	0.92
Ningxia	3369.37	13.57%	7.96%	14.82%	1.39	21.45%	3.03%	34.28%	0.03%	77	0.12	1.06
Guangxi	3321.40	16.34%	7.32%	13.65%	0.98	31.23%	4.55%	28.85%	-0.77%	191	0.11	1.07
Yunnan	3314.20	15.52%	11.66%	14.00%	1.00	26.77%	-1.37%	35.43%	1.19%	102	0.12	1.57
Jiangxi	3258.63	15.37%	7.70%	15.78%	0.81	30.28%	3.71%	29.84%	-1.37%	240	0.11	1.33
Sichuan	3222.37	15.68%	4.61%	13.47%	0.67	28.87%	6.91%	33.04%	-1.06%	167	0.12	0.96
Tibet	3068.11	13.28%	4.29%	14.19%	1.42	39.27%	29.70%	7.59%	-2.77%	2	0.08	1.70
Gansu	2605.68	12.93%	9.25%	9.12%	0.90	23.09%	-0.42%	36.15%	0.37%	53	0.12	1.16
Guizhou	2018.05	13.39%	9.26%	13.67%	1.47	32.26%	-2.23%	32.26%	1.29%	200	0.11	1.45

Figure n°5 : Average per Capita GDP (in thousand RMB Yuan) over the period 1986–2004



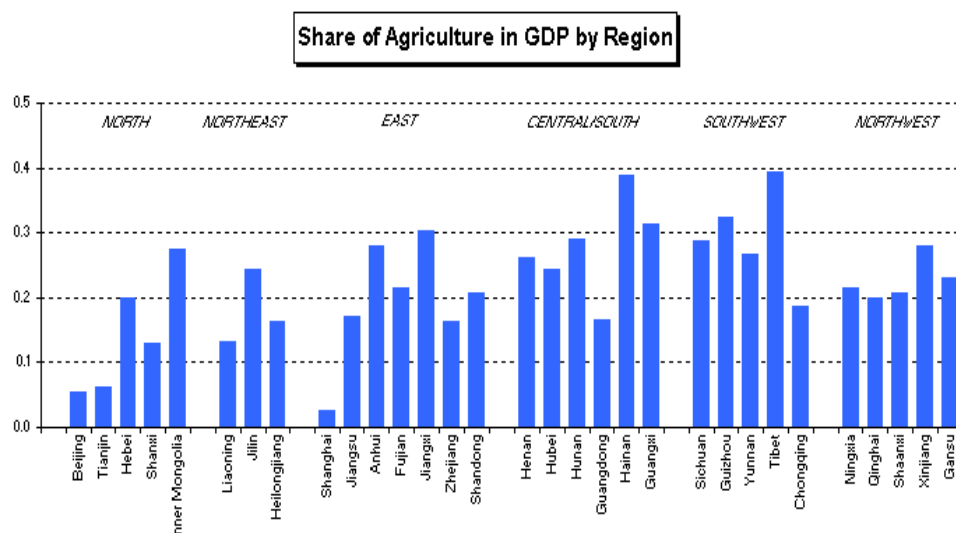
Source : China Data Online (1986–2004).

Figure n°6 : Average Industrial Share of GDP over the period 1986–2004



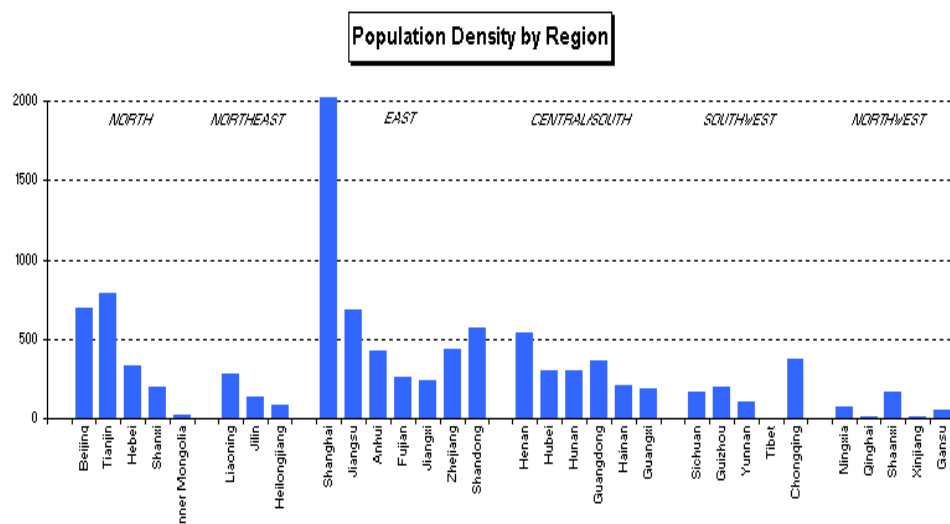
Source : China Data Online (1986–2004).

**Figure n°7 : Average Agricultural Share of GDP
over the period 1986–2004**



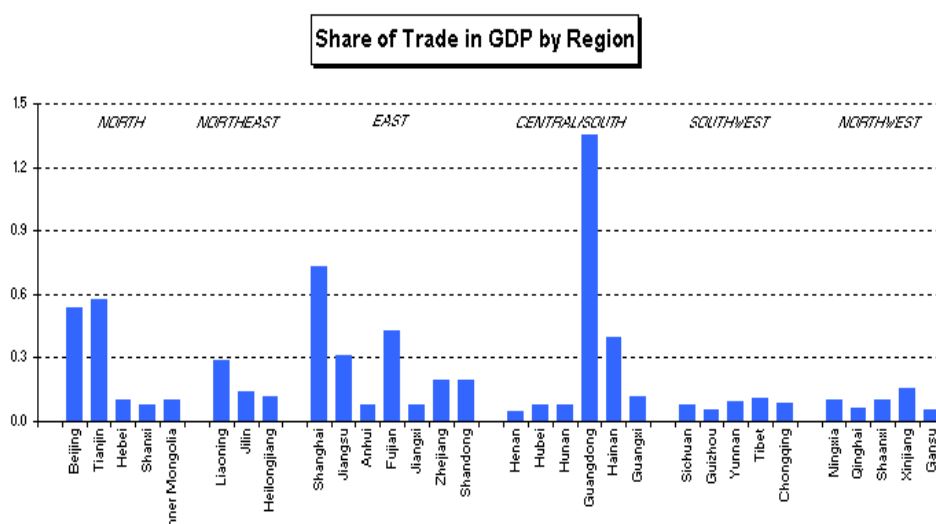
Source : China Data Online (1986–2004).

**Figure n°8 : Average Population Density (Inhabitants per Sq.Km)
over the period 1986–2004**



Source : China Data Online (1986–2004).

**Figure n°9 : Average Trade Share of GDP (%GDP)
over the period 1986–2004**



Source : China Data Online (1986–2004).

REFERENCES

- Advisory Commission on Intergovernmental Relations (ACIR), 1981, *Measuring the Fiscal Capacity and Effort of State and Local Areas*. Washington, D.C. : U.S. Government Printing Office.
- Advisory Commission on Intergovernmental Relations (ACIR), 1982, *Tax Capacity of the Fifty States : Methodology and Estimates*. Washington, D.C.: U.S. Government Printing Office.
- Advisory Commission on Intergovernmental Relations (ACIR), 1987, *Measuring State Fiscal Capacity*. Washington, D.C. : U.S. Government Printing Office.
- Advisory Commission on Intergovernmental Relations (ACIR), 1990, *State Fiscal Capacity and Effort*. Washington, D.C. : U.S. Government Printing Office.
- Advisory Commission on Intergovernmental Relations (ACIR), 1993, *Significant Features of Fiscal Federalism*. Vol. 2. Washington, D.C. : U.S. Government Printing Office.
- Advisory Commission on Intergovernmental Relations (ACIR), 1993, *State Fiscal Capacity and Tax Effort*. Washington, D.C.

- Ahmad Ehtisham and John Craig, 1997, "Intergovernmental Transfers", in Teresa TerMinassian (ed.), *Fiscal Federalism in Theory and Practice*, Washington, D.C. : International Monetary Fund, 73-105.
- Ansari M, 1982, "Determinants of Tax Ratio: a cross-country analysis." *Economic and Political Weekly*, 19, 1035-1042.
- Bahl, Roy W, 1971, "A Regression Approach to Tax Effort and Tax Ratio Analysis." *IMF Staff Papers*, November, 570-608.
- Bahl Roy W, 1972, "A Representative Tax System Approach to Measuring Tax Effort in Developing Countries" *IMF Staff Papers* 19, 87-124.
- Bahl Roy and Christine Wallich, 1992, "Intergovernmental Fiscal Relations in China," Working Papers, Country Economics Department, World Bank, WPS 863.
- Bird R.M., Jorge Martinez-Vazquez and Torgler Benno, 2004, "Societal Institutions and Tax Effort in Developing Countries." ITP paper 04011, International Tax Program, Rotman School of Management, University of Toronto, October.
- Boadway R. W., 2001, "Inter-Governmental Fiscal Relations: The Facilitator of Fiscal Decentralization." *Constitutional Political Economy*, 12, 93-121.
- Brenna G. and J. Buchanan, 1980, *The Power to Tax: Analytical Foundations of a Fiscal Constitution*. Cambridge : Cambridge University Press.
- Chelliah Raja J., 1971, "Trends in Taxation in Developing Countries." *IMF Staff Papers*, July, 254-327.
- Eltony M. Nagy, 2002, "Measuring Tax Effort in Arab Countries." Arab Planning Institute (API) Working Paper, September.
- Greene William, 2003, *Econometric Analysis*, Prentice Hall, 5th Edition.
- Leuthold Jane, 1991, "Tax Shares in Developing Economies." *Journal of Development Economics*, Vol. 35, 173-185.
- Lin Yifu, Ran Tao, Mingxing Liu, 2003, "Decentralization and Local governance in China's Economic Transition" *Conference: The Rise of Local Governments in Developing Countries*, London School of Economics, May.
- Marlow Michael L, 1988, "Fiscal decentralization and government size" *Public Choice*, Vol. 56, 259-269.
- Mertens Jo Beth, 2003, "Measuring Tax Effort In Central and Eastern Europe." *Public Finance and Management*, 3(4), 530-563.
- Musgrave Richard A, 1969, *Fiscal Systems*. New Haven : Yale University Press.
- Nelson Michael A, 1986, "An empirical analysis of state and local tax structure in the context of the Leviathan model of government" *Public Choice*, Vol. 49, 283-294.

- Nelson Michael A, 1987, "Searching for Leviathan : Comment and Extension" *The American Economic Review*, Vol. 77, No. 1, 198-204.
- Oates W.E., 1972, *Fiscal Federalism*. New York : Harcourt Brace Jovanovich.
- Oates Wallace E., 1985, "Searching for Leviathan : An Empirical Study" *The American Economic Review*, Vol. 75, 748-757.
- Sagbas Isa, 2001, "An Econometric Analysis of Local Fiscal Response to Revenue Sharing in Turkey." *Environment and Planning C: Government and Policy*. Vol.19, 85- 101.
- Shen Chunli, 2008, "Bridging Fiscal Divergence in China." Dissertation, School of Public Policy, University of Maryland.
- Shen, Chunli, Jing Jin, and Heng-fu Zou, 2006, "Fiscal Decentralization in China Revisited." Working Paper, Word Bank.
- Stotsky Janet G. and Asegedech Woldemariam, 1997, "Tax Effort in Sub-Saharan Africa." *IMF Working Paper*. No.97/107.
- Tanzi Vito, 1968, "Comparing International Tax 'Burdens': a suggested Method." *The Journal of Political Economy*, Vol. 76, No. 5, 1078-1084.
- Tanzi Vito, 1987, "Quantitative Characteristics of the Tax Systems of Developing Countries", in Newbery D, Stern N. (eds), *Theory of Taxation for Developing Countries*, Oxford University Press: New York, 205-241.
- Tanzi Vito, 1992, "Structural factors and tax revenue in developing countries: a decade of evidence", in Ian Goldin and L. Alan Winters (eds.), *Open Economies: Structural Adjustment and Agriculture*, Cambridge: Cambridge University Press, 267-81.
- Teera Joweria M. and Hudson John, 2004, "Tax Performance: A Comparative Study." *Journal of International Development*, 16, 785-802.
- Warner Mildred, 2001, "State policy under devolution: Redistribution and Centralization." *National Tax Journal*, Vol. 54, No. 3, 541-556.
- Wong Christine P.W., 1992, "Fiscal Reform and Local Industrialization: The Problematic Sequencing of Reform in Post-Mao China" *Modern China*, Vol. 18, No. 2, 197-227.
- Wong Christine P.W. (ed.), 1997, *Financing Local Government in the People's Republic of China*. Hong Kong: Oxford University Press.
- Wong Christine P.W., 1998, "Fiscal Dualism in China: Gradualist Reform and the Growth of Off-Budget Finance", in Dohand J.S. Brean (ed.), *Taxation in Modern China*, New York and London: Routledge.
- Zhang Tao and Heng-fu Zou, 1998, "Fiscal decentralization, public spending, and economic growth in China" *Journal of Public Economics*, Vol. 67, 221-240.

**LA POLITIQUE FISCALE DES GOUVERNEMENTS LOCAUX
EN CHINE : UNE ANALYSE DE LA PERFORMANCE
FISCALE DES PROVINCES**

Résumé – Cet article cherche à comparer la performance des politiques fiscales régionales entre 1986 et 2004 en Chine en mettant plus particulièrement l'accent sur les effets des réformes majeures effectuées dans les années 1980 et en 1994. Dans un premier temps, en s'appuyant sur des données de panel, nous utilisons un modèle hybride qui combine effets temporels fixes et effets régionaux aléatoires, permettant d'analyser la relation entre le poids de la fiscalité dans le PIB et les variables économiques et démographiques régionales. Les résultats obtenus montrent que le système décentralisé de la période 1986-1993 a eu des effets plus bénéfiques sur la croissance du PIB que le système, plus centralisé, de la période 1994-2004. Dans un deuxième temps, nous proposons la construction d'indicateurs régionaux représentant le potentiel de recettes fiscales additionnelles et permettant d'analyser l'impact d'une modification des politiques fiscales locales en matière de redistribution et de croissance.