Tax competition, Tax policy, and Innovation

DISSERTATION

Presented in Partial Fulfillment of the Requirements for the Degree Doctor of Philosophy
in the Graduate School of The Ohio State University

By

Yuan Wang, M.A.

Graduate Program in Economics

The Ohio State University

2017

Dissertation Committee:

Pok-Sang Lam, Advisor

Lucia Dunn

Stephen Cosslett

Abstract

My research concerns how governments make economic decisions and interact with other governments, to increase social welfare; in particular, my focus lies in the area of taxation and technological innovation.

In a globalized economy with mobile capital, increasing interest has been paid to capital tax policy. My research is among the first to examine empirically and explain theoretically the tax competition among states in the U.S. Moreover, I also study how state governments set their tax rates using historical data and explain why the pattern observed is different from the zero-tax theory.

Due to the absence of state-level average capital tax rate data, I first construct a panel dataset of average capital income tax rates at the state level for the period 1958-2007 for the capital taxation studies.

In Chapter 1, I analyze the tax policy of each individual state government. Empirical evidence implies that tax rates are history-dependent. I provide an alternative explanation for nonzero capital tax rate, reexamining Ramsey's (1927) rule. With a lack of commitment power from government, households form adaptive expectations on capital tax rates. The equilibrium capital tax rate is thus history-dependent with a balanced-budget requirement on state governments. The investment decision combines income and substitution effects, and the U.S. states differ on investment sensitivity to capital tax

rates. I provide empirical findings on investment sensitivity for each state, and then a structural model is applied to replicate the empirical.

In Chapter 2, I analyze the pattern of strategic interaction on capital tax rates among states in the U.S. This paper is the first to apply MLE estimation of the SAR panel data model with fixed-effects to study tax competition behavior. Through a joint investigation into both tax competition behavior and capital allocation decision, I demonstrate the existence of capital tax competition among states in the South and West, but competition is less significant in the Midwest and Northeast. I then apply a high-order SAR panel data estimation with fixed-effects to study the impact of population growth on tax competition, and results suggest that faster population growth significantly relates to stronger reaction to changes in neighbors' tax policy. I also apply two weighting schemes of neighbors to validate the findings. A two-period structural model with a saving decision is developed to explain this result. The model features a capital dilution effect which is also tested empirically.

In Chapter 3, a quality ladder model is developed in which the technology gap between the North and the South is endogenously determined. A stronger intellectual property rights (IPR) in the South discourages imitation and reduces the FDI cycle length. The optimal IPR strength balances two effects, long-run and short-run effects, and it is non-monotonic in the market size and increasing in the number of imitating firms. The social welfare of the South is decreasing in the FDI cycle length, but is decreasing in IPR strength given cycle length.

Acknowledgments

I would like to express my sincere gratitude to my advisor Prof. Pok-Sang Lam for his

great help and persistent support during my Ph.D. study, and for his encouragement,

guidance and care.

I would like to express the deepest appreciation to my committee members Prof. Lucia

Dunn and Prof. Stephen Cosslett, for their insightful feedbacks, encouraging comments,

great patience and support.

In addition, I would like to thank Prof. Paul Evans, who helped me to start my research in

my Ph.D. study, with his immense knowledge and motivating advice.

Vita

Fields of Study

Major Field: Economics

Table of Contents

List of Tables

List of Figures

Chapter 1: History-Dependent Capital Taxation

1 Introduction

Ramsey's (1927) seminal contribution on zero capital taxation states that in order to ensure a bounded future implicit consumption tax and avoid capital accumulation distortion, it is optimal to levy zero taxation on capital investment in the long run in an infinitely-lived household model. Further studies including Chamley (1986) and Judd (1985) validated this result under different economic environments. Based on these, a lot of literature was extended in various directions and gave stories that optimal capital tax should not be zero, including different discount factors across individuals (Diamond and Spinnewijn, 2011); financial market failures (Glenn and Judd,1986; Aiyagari, 1995); life-cycle models (Erosa and Gervais, 2002; Garriga, 2003); nonseparable utility (Kuhn and Koehne, 2013) and so on.

After observing data on capital tax rates all over the world, the gap between tax rates in reality and that suggested by Ramsey's theory is evident: some countries impose relatively high taxation on capital, whereas some other countries levy no tax on capital.

In the US, apart from federal capital taxation, each state imposes its own tax rate

of capital. Thus I continue to check historical capital taxation across 50 states in the US, and found that capital tax rates are different across states in terms of both levels and historical patterns, which can be decreasing, increasing or oscillating. Furthermore, the ratio of capital to personal income(inflation-adjusted) shows a fluctuating but convergent pattern in most states.

This paper provides an explanation to capture these empirical facts across states in the US, based on observations above.

The time lag between individuals' investment and return realizations leads to uncertainty about future returns when households invest. If government has no commitment power on capital taxation, households form their own expectations on future tax rate and invest accordingly. In the next period, capital tax rate is realized based on existing capital stock, and households update their belief according to new information.

Economists have been debating over the assumptions of rational expectation and adaptive expectation when it comes to study of economic behavior. The rational expectation hypothesis is argued to be a possible source of the Lucas Critique, and is thus supported and applied widely. However,this possibility does not validate rational expectations due to a lack of empirical support, as suggested by Chow (2011). Furthermore, Chow (2011) and Chow (1988) presented strong statistical and econometric evidence for adaptive expectation. Logical argument is also provided for using adaptive expectation as a better proxy for psychological expectation. Hence, house-

holds in my model update their belief using adaptive expectations. The expected tax rate is a weighted sum of past information with geometrically declining weights with respect to time.

All but one U.S states are required to expend no more than the revenue they can raise[1]. States start with different initial beliefs on tax rates, which lead to different historical patterns of tax rates. With low initial expectation on capital tax rate, households invest a lot and increases the tax base, government only needs to levy low tax rate, which confirms households' initial belief; while with high initial capital tax expectation, households reduce investment sufficiently, which forces government to tax heavily on capital return to meet government budget, thus government can do nothing to revert households' belief back to low level, and get "stuck" in the high tax equilibrium. With balanced budget requirement, the government cannot borrow to alter households' belief; nor can this be achieved with no commitment power from the government. Thus, the existence of multiple equilibria is possible, suggesting that given different levels of initial capital stock or initial expectation, each state will end up at different steady states.

Moreover, the equilibrium with higher capital tax rate is associated with lower capital stock, which matches the findings of empirical work that the capital level as well as investment is negatively correlated with capital tax rate (Knight, 2000).

The historical path of capital tax rates is also determined by the elasticity of

---

[1]The National Conference of State Legislatures (NCSL) has traditionally reported that 49 states must balance their budgets, with Vermont being the exception.

investment to changes in tax rates. I empirically estimate the investment sensitivity to tax rates for each state. This investment decision rule is a combination of substitution and income effects, which offset with each other: with an increasing expected capital tax rate, the substitution effect decreases capital investment, while income effect increases capital investment. As states differ in industry structure, productivity level, education level and degree of economic inequality, it's natural to observe different investment behavior empirically. Each state's specific investment curve and starting belief characterize the pattern of tax rates evolution. The simulated sequence of tax rates qualitatively fits the observed data.

The structure of this paper is as follows: Section 2 introduces dataset and provides empirical findings, and Section 3 introduces the empirical model. Section 4 summarizes the result and Section 5 introduces the potential theoretical model. Lastly, Section 6 concludes.

2   Data

2.1   Average Capital Gains Tax Rates

The officially available data on capital taxation includes marginal capital gains tax rates, brackets and so on. These information have been used to calculate effective capital tax rates. Pomerleau (2013) shows a wide range of effective rates across states.

In most theoretical models, return from capital investment is taxed proportionally

and thus the tax rate is simplified as an average tax rate. However, insufficient empirical work has been done to obtain average capital tax rates in the US or in the states. In order to be consistent with theoretical models, I obtained my own series of average capital tax rates for each state [2] from 1958 to 2008.

From US Census Bureau, I first summed up three sources of revenue to account for capital tax revenue: property tax, corporate net income tax, death and gift tax [3]. And I used the data "dividends, interest and rent" for taxable capital income. Then I divided total capital tax revenue by taxable capital income to get the average capital tax rates.

The calculated tax rates range from 0 to 0.25, and most of them fall into the range of 0 to 0.1. States start with different initial values of rates in 1958, which I summarize in Appendix A. The paths of historical rates can be categorized into three patterns: decreasing (e.g. Texas), oscillating (e.g. North Dakota), and increasing (e.g. New Hampshire). The figures of these three states' patterns are displayed in Figure 1.

---

[2] I included District of Columbia for analysis as well. And there exist some errors in data for several years of Alaska, which is a fiscal and geographic outlier in the US. The observations of Alaska could thus be dropped.

[3] Death and Gift tax is the tax imposed on transfer of property at death, in contemplation of death, or as a gift.
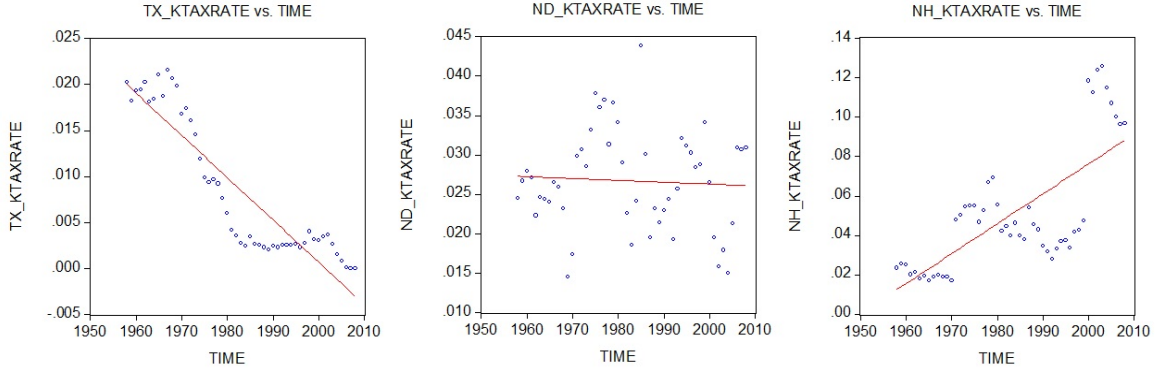
Figure 1: Capital tax rates in Texas, North Dakota and New Hampshire

2.2 Detrended Capital Stocks

To estimate investment decision empirically, I obtain the data on capital stock from the database created by Garofalo and Yamarik(2002) and Yamarik(2012). As capital grows over time in each state, I first detrend capital by dividing the level of capital in each state by the aggregate level of capital in the United States in each year. The detrended capital stock thus represents the percentage of each state's capital level in the US. And these data will be used for empirical regression, with the details presented in section 3.

The correlation between detrended capital and capital tax rates is either negative or positive, which implies different investment behaviors across states. This will be discussed in next section.

The patterns of detrended capital of the three states mentioned in last section are displayed in Figure 2.

6
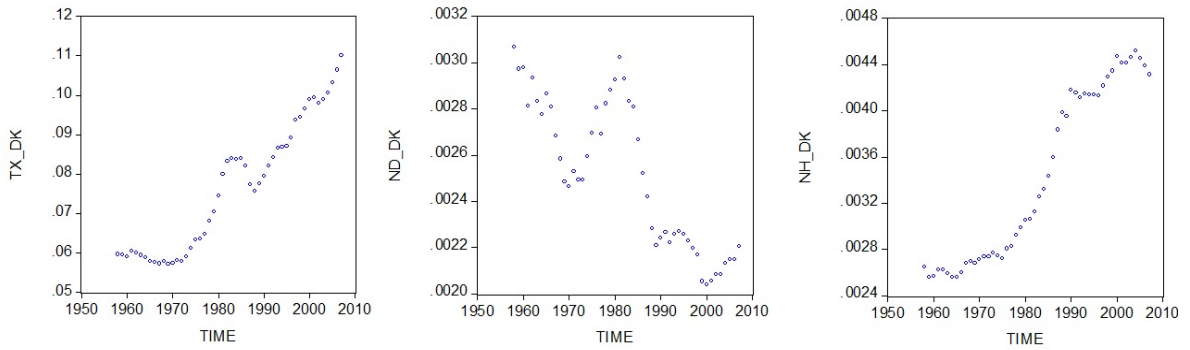
Figure 2: Detrended capital stock in Texas, North Dakota and New Hampshire

2.3 Government spending

State government is not encouraged to borrow to meet their expenditure, so the major source of raising revenue is from taxes.

In a growing economy, both capital tax revenue and state government expenditure show an upward sloping trend in each state. To detrend government spending and obtain the portion of state government expenditure covered by capital taxation, I divided capital tax revenue by total government expenditure. This is consistent with the way capital is detrended.

3 Empirical Model

3.1 Investment Sensitivity

The tax sensitivity of investment has important implications for analyzing historical pattern of capital taxation. The investment decision is influenced by a combination of income and substitution effects, and the total effect can vary at different values of the capital tax rate. With an increase in expected capital tax rate, the expected net return decreases and households reduce investment, which characterizes the substitution effect. A higher expected capital tax rate also reduces expected income next period, and in order to ensure a certain level of consumption, households increase investment. This income effect offsets the substitution effect.

Households' responses to changes in taxation at different levels of rates generate an investment decision curve over the whole range of capital tax rates. Each state has its own feature of industrial structure, productivity level, education level and degree of economic inequality, which altogether produce a specific investment decision curve for each state.

As argued in Young (1988), many factors influence the decision on investment, with some of them difficult to be quantified. In this paper, the effects of tax rates are isolated to be analyzed.

To estimate the elasticity of investment for each state, I run a regression of the capital level on a polynomial of tax rates as in Equation (1), which is based on the Hartman (1985) model. I use annual data on detrended capital and contemporary

capital tax rate [4] for each state. And the regression result can be linear, quadratic or cubic, the most significant one of which is chosen as the investment decision curve for each state, so c_2, c_3 can be zero [5].

$$dk_t = c_0 + c_1\theta_t + c_2\theta_t^2 + c_3\theta_t^3 + \epsilon_t \tag{1}$$

Estimation results indicate that states are distributed approximately equally into four main patterns of investment behavior: decreasing, increasing, U-shaped and inverse U-shaped, with a few left maintaining a polynomial of cubic or even higher degree. Estimation results of representative state in each pattern are displayed in Table 1.

---

[4]There is no significant difference in regression results when lagged capital tax rates are used as explanatory variables, from a randomly selected sample of states.

[5]Only a few states have regression results with a polynomial of degree higher than 3.

Dependent variable: DetrendedCapital

Sensitivity Patterns	Decreasing	Increasing	U-shape	Inverse U-shape	Cubic
(State)	(Alabama)	(Iowa)	(Virginia)	(New Mexico)	(Illinois)
c_0	0.013***	0.008***	0.025***	0.003***	0.097***
	(0.0002)	(0.0009)	(0.0005)	(0.0005)	(0.0063)
c_1	-0.026***	0.119***	-0.228***	0.083***	-5.387***
	(0.005)	(0.037)	(0.0408)	(0.0282)	(1.0480)
c_2	0	0	1.382**	-0.927**	185.828***
			(0.6219)	(0.34860)	(43.1458)
c_3	0	0	0	0	-1894.047***
					(508.7329)

Note: These are least squares estimates of the parameters in Eq. (1).

Robust standard errors in parentheses. ***$p<0.01$, **$p<0.05$, *$p<0.1$.

Table 1: Tax Sensitivity Estimation of Representative States

And the corresponding investment curves for those representative states are depicted in Figure 3, where I denote t as capital tax rate and dkfcst as the estimation forecast value of detrended capital.

Different curves of investment decision imply different combinations of substitution and income effects, with the main patterns summarized below:

Case 1 *Decreasing pattern*

Substitution effect dominates income effect in the whole range of tax rates. As the expected capital tax rate increases, the return of investment decreases and thus investment decreases.

Case 2 *Increasing pattern*

The income effect dominates the substitution effect in the whole range of tax rates. As the expected capital tax rate increases, income from investment decreases and thus investment increases to compensate for the loss in expected income.

Case 3 *U-shaped pattern*

The substitution effect dominates in the lower range of tax rates while the income effect dominates in the higher range. As the expected capital tax rate rises in the higher range, low income further decreases, which households are very sensitive to, and investment increases accordingly.

Case 4 *Inverse U-shaped pattern*

The income effect dominates in the lower range of tax rates while the substitution effect dominates in the higher range. Households are more sensitive to tax rate changes at higher level of capital tax rates; while at lower level of tax rates, households care more about the income loss.
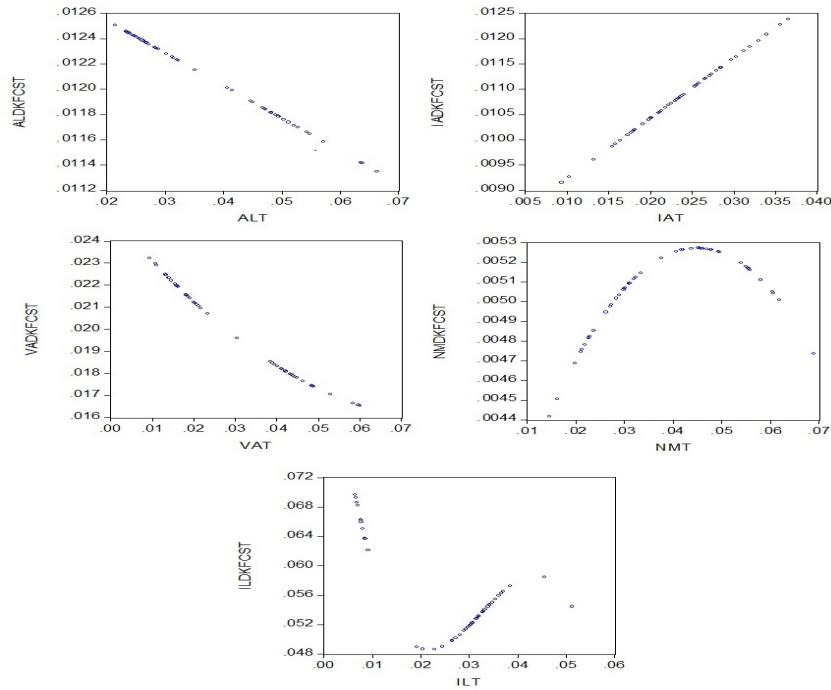
Figure 3: Investment Decision

States differ in terms of investment patterns, which are summarized in Appendix B. Further analysis to account for their patterns are in Section 5.

12

3.2 Adaptive Expectation

Economists have been studying the hypotheses of Rational Expectation (RE) and Adaptive Expectation (AE), by testing for the empirical validity of each. Campbell and Shiller (1987), Poterba and Summers (1987), Fama and French (1988) and West (1988) realized the inconsistency with data from the assumption of Rational Expectation in present-value models. Moreover, Chow (1988) found that by replacing Rational Expectation hypothesis with Adaptive Expectation, the performance of present-value models in explaining data improves.

Chow (2011) summarizes how these two competing hypotheses can be effectively assumed and provide further econometric support for Adaptive Expectation. Though Rational Expectation has been long accepted for its potential to serve as a source of the Lucas critique, this alone does not rationalize the use of Rational Expectation as the empirical economic hypothesis over Adaptive Expectation, with insufficient evidence supporting Rational Expectation.

Moreover, a large body of research has been testing on the RE hypothesis using survey data of inflation expectations, including Bonham and Cohen (2001), Bonham and Dacy (1991) and Croushore (1997). All of them failed to empirically justify the RE assumption. Similarly, literature such as Frankel and Froot (1987b, 1990a), Froot (1989), Friedman (1990) and Jeong and Maddala (1996) applied the survey data of interest rate forecasts from foreign exchange markets and found that the traders' behaviors display behavioral instead of rational patterns. Thus, these findings also

rejected the RE hypothesis and motivated economists to search for alternative models to match the survey data. Markiewicz and Pick (2013) is one of these contributions to support the approach of adaptive learning.

Furthermore, a brief observation of the patterns of historical capital tax rates (as shown in Figure 1) suggests the use of Adaptive Expectation hypothesis. Rather than jumping into the steady state equilibrium immediately which is implied by Rational Expectation hypothesis, tax rates gradually converge or oscillate around.

Hence, I assume Adaptive Expectation in what follows, which also makes logical sense as households form their expectations by averaging past information with geometrically declining weights.

Assumption: Adaptive Expectation Denote θ_t^e as the expected tax rate for period t, and θ_t as the realized capital tax rate set by the government at period t, households form expectation according to:

$$\theta_t^e = \lambda \theta_{t-1} + (1 - \lambda)\theta_{t-1}^e, 0 \leq \lambda \leq 1 \tag{2}$$

Households update their belief on capital tax rates using newly realized tax rates weighted λ. When λ=1, households fully utilize new information and believe that government will set the same capital tax policy next year. When λ=0, households insist on their initial belief and consider the change in realized capital tax rate merely as a perturbation. A higher λ implies more weight on new information.

3.3 The Model

3.3.1 Households

In this section, households in State i invest according to a reduced form investment function [6], denoted by

$$k_t = f_i(\theta_t^e) \tag{3}$$

which is empirically estimated in Section 3.1.

Starting with an initial belief θ_0^e, capital level next period is determined and capital tax rate is realized. Households update their expected tax rate according to Adaptive Expectation Assumption applying new information obtained.

3.3.2 Government

Investment is an intertemporal decision, but government lacks commitment power setting tax policy. Suppose government announces zero capital tax for next year, households who believe it will invest largely for the high return. If faced with a positive spending shock when it comes to next year, government has an incentive to deviate by setting a slightly higher than zero capital tax rate on a large capital

---

[6]The structural model is presented in Section 5

stock, in order to acquire revenue and meet the budget requirement. Foreseeing this, commitment from government is not credible to households, which motivates households to form their own beliefs on policy.

Government collects revenue from tax collection. State government is faced with State Balanced-Budget Provision, and cannot borrow to cover expenditure.

Assuming Cobb-Douglas production function, the return from capital is $r_t = A\alpha k_t^{\alpha-1} N^{1-\alpha} - \delta$. Thus the balanced budget equation of government is[7]

$$G_t = k_t r_t \theta_t \tag{4}$$

Government determines capital tax rate each period from this equation.

3.4   Patterns of Tax Evolution

This section describes how different patterns of historical capital tax rates are generated.

To simplify the analysis using figures, I assume $\lambda=1$ in equation of Adaptive Expectation.

$$\theta_t^e = \theta_{t-1} \tag{5}$$

---

[7]I do not include other sources of tax revenue, as the data I use for simulation is the percentage of government expenditure collected from capital tax revenue. Results still hold qualitatively in an alternative setting with other taxes included.

And I fix government spending level constant for better illustration, so curve (represented by Equation (3)) does not shift around over time.

Investment decision rule (Equation 3), government balanced budget (Equation 4) and adaptive expectation (Equation 5) simultaneously determine a path of capital tax rates, given an initial expected capital tax rate.

Many possible patterns can be resulted in, which are summarized below.

Pattern 1

Figure 4 depicts the case with decreasing investment decision curve. Starting from a low initial expected capital tax rate for period 1 at θ_1^e, households invest and capital level is k_1 at period 1, determined by the investment curve. Given k_1, government chooses capital tax rate at θ_1, determined by government revenue curve. Then households update their belief on capital tax rate for period 2 by $\theta_2^e = \theta_1$. With capital tax rate expected at θ_2^e, households invest up to capital level at k_1. Capital tax rates increase monotonely and converge to a stable steady state equilibrium with positive tax rate. This sequence is summarized as pattern 1 in the right graph.
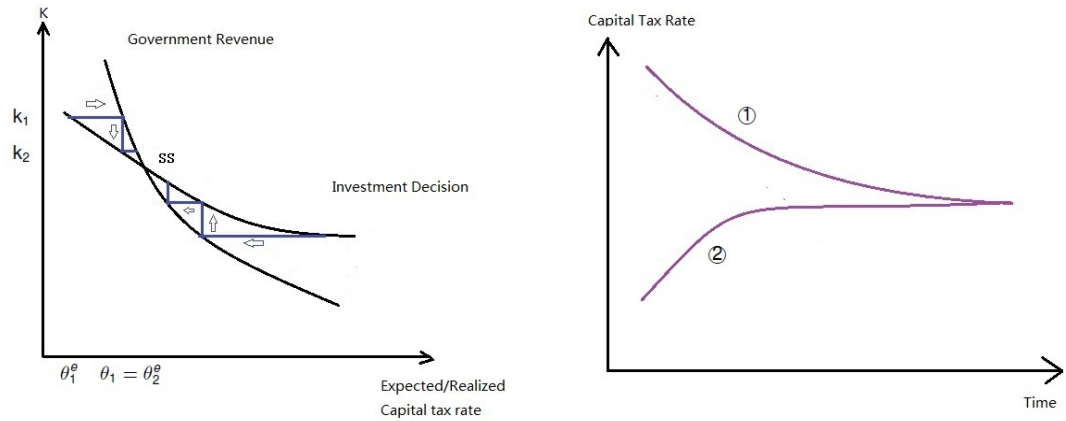
Figure 4: Pattern 1

Similarly, starting from a high initial expected capital tax rate, tax rates monotonely decrease over time and converge to the same stable steady state equilibrium.

Evidently, the Ramsey result does not hold here, as the capital tax rates converge to a positive value rather than zero. With limited commitment power, government cannot alter households' expectation by announcement. Furthermore, with balanced budget requirement, government cannot borrow to set a low rate permanently to enforce a low belief. Hence, households invest according to their initial belief, which is reinforced gradually by government action until the equilibrium is reached.

Moreover, initial belief held by the households matters for the pattern of convergence. A low initial belief gives rise to an increasing pattern, whereas a high initial belief produces a decreasing pattern.

Pattern 2

18

Here is another case with decreasing investment decision curve as depicted in Figure 5, which is however more steeply sloped than that in Pattern 1.

Different from previous case, if households start with a relatively low expected capital tax rate, the economy converges to a zero capital tax rate, which is consistent with the Ramsey's result. If starting from a relatively high rate, however, the tax rates diverge. Apparently, there does not exist any stable steady state equilibrium in this case.

A more steeply sloped investment curve suggests more elastic response to tax rates by households. At low values of capital tax rates, a reduction in tax rates significantly increases investment, which allows government to further reduce tax rates, and ultimately leads to the convergence to zero tax rate. This is beneficial to the economy, with lower degree of distortion and higher level of capital stock.

At high values of capital tax rates, however, an increase in tax rate reduces investment greatly due to higher sensitivity to tax changes. With big drops in the capital base, the government has to further increase tax rates to meet the budget. This is devastating to the economy with escalating capital tax rates over time.

Initial belief is also crucial here: a low starting belief combined with a sensitive investment curve leads to a decreasing capital tax rate to zero; while a high starting belief combined with the same investment curve "traps" the government in this worsening situation and collapses the economy.
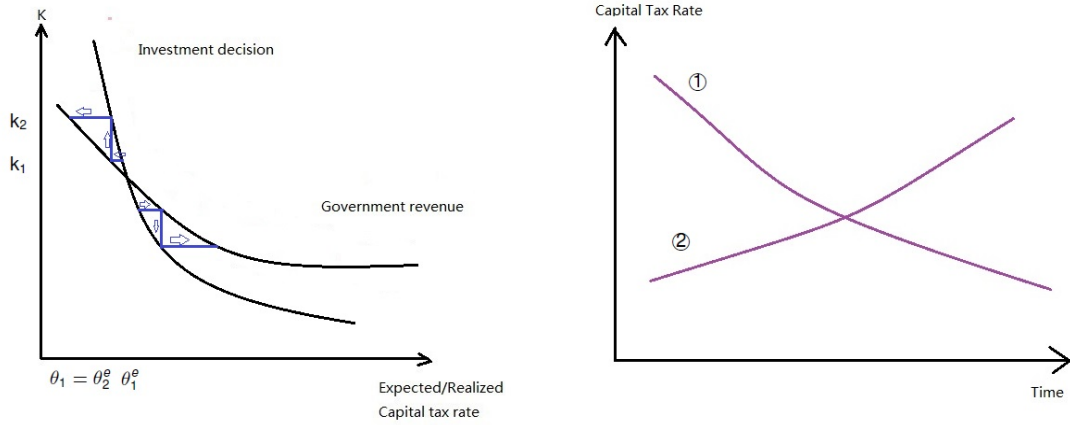
Figure 5: Pattern 2

Pattern 3

Figure 6 shows the case with U-shaped investment decision curve. Starting from any expected rate, tax rates oscillate until reaching the stable steady state equilibrium and stay there. The tax rate at the steady state is positive.

This deviates from Ramsey's rule due to a less elastic investment decision. In the range of high tax rates, as investment decision dominates, households invest a lot and bring down the tax rate to the low range of tax rates. Therefore, no matter where the economy starts, it ends up at the same steady state equilibrium.

Similarly, the specific pattern of tax rates path depends on the initial belief.
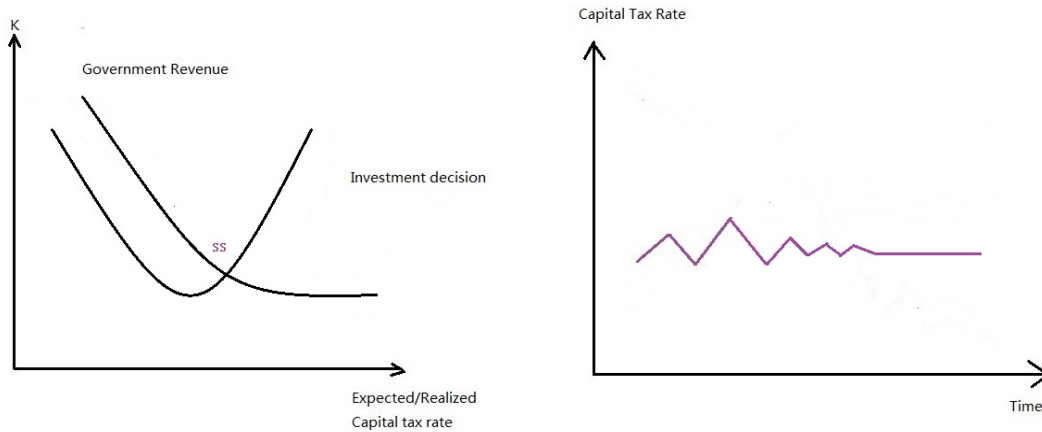
Figure 6: Pattern 3

Pattern 4

Similar to Pattern 3, investment decision is U-shaped but with higher substitution effect in the low range.

There are two equilibria in this economy, and neither of them is stable. Starting from any belief other than these two points, this economy converges to zero capital tax, coinciding with Ramsey's rule. Even if the economy starts with high belief, households invest significantly due to strong income effect, and this brings down the rate to the low range. Then the strong substitution effect comes into play and leads the economy to a decreasing tax rate and increasing capital level. This pattern is also beneficial to the economy.
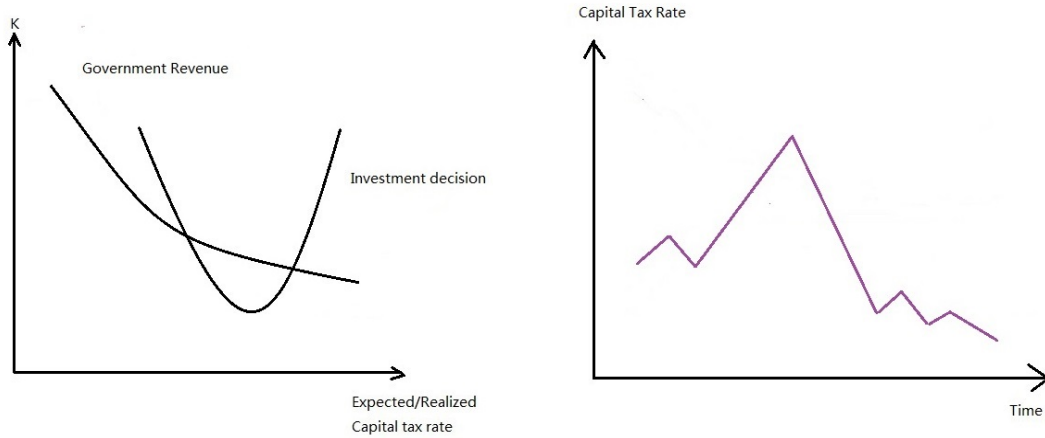
Figure 7: Pattern 4

Pattern 5

In the case with inverse U-shaped investment decision curve, there exist two equilibria and only one of them can be stable. Starting from any initial belief below a threshold level (denoted t in the figure), tax rates either oscillate around point O in Figure 8 or converge to it as a steady state. If the initial value is greater than t, tax rates diverge up to 1.

As income effect dominates at the low range, when tax rate increases, households have more incentive to invest, which in turn could bring down the tax rate. Thus tax rates alternate between high and low values, or ultimately converge to the steady state depending on the degree of income effect. At the high range, however, substitution effect dominates with households investing less with tax increase, which deteriorates the situation. Thus, tax rates diverge from a high value.

22
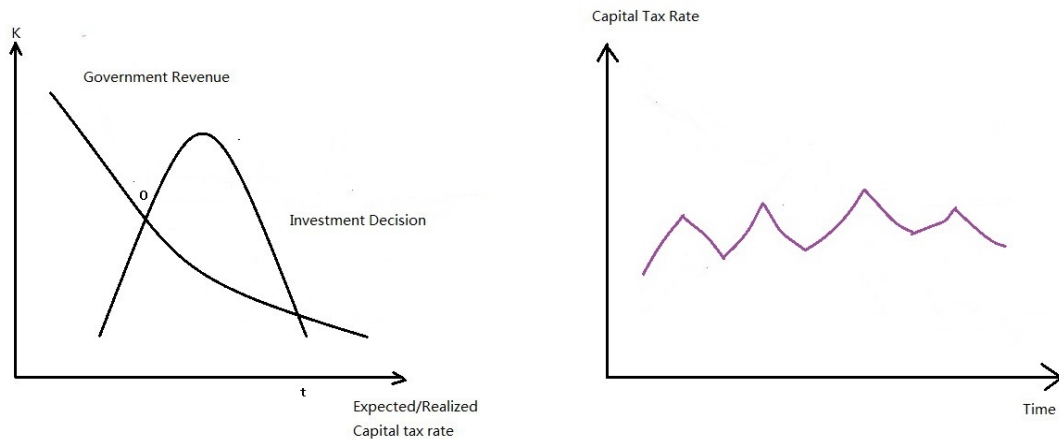
Figure 8: Pattern 5

Pattern 6

The case with cubic or higher degree polynomial investment curve is more complicated.

In the graph on the left of Figure 9, there exist multiple steady state equilibria but no stable one; and in the graph on the right, there exist two stable steady state equilibria and another unstable one.

The evolution path of capital tax rates can be various depending on the starting point and curvature of investment curve.

23
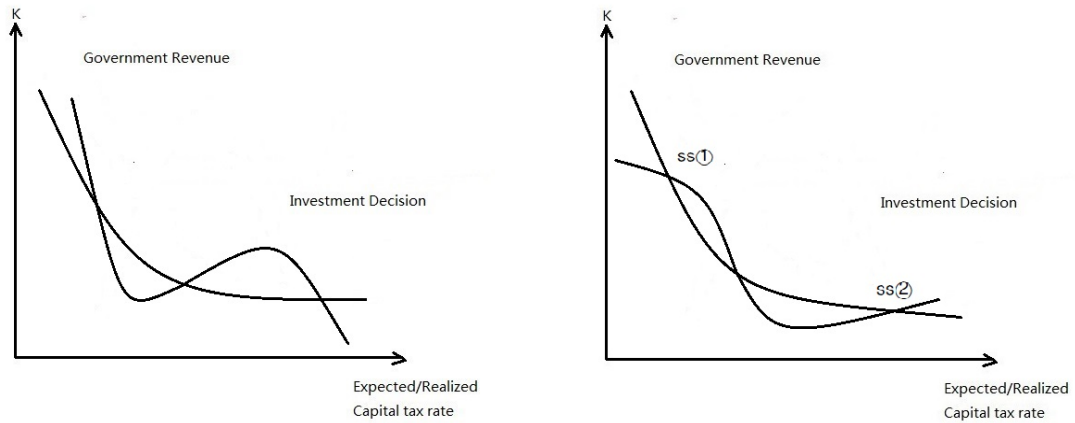
Figure 9: Pattern 6

4 Results

This section evaluates the model's performance to account for the empirical data observed empirically across 50 states.

4.1 Preliminary Test

The existence of multiple equilibria is tested. The two stable steady state equilibria in Figure 9, for instance, suggest that states with a low starting belief converge to an equilibrium with low rate while states with a high starting belief converge to an equilibrium with high rate.

In order to test this theory, I divided 51 states (DC included) into two subgroups by tax rates observed in year 1958. I assume $\theta^e_{1959} = \theta_{1958}$ as the starting belief of each

state for year 1959. These two groups are named low-initial group and high-initial group, separated by the median value of initial beliefs.

I run AR(1) regression to obtain the steady state value. Then I use panel regression to test fixed effects across two groups, which is shown in Table 2.

Dependent Variable: TR
Method: Pooled Least Squares

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.006087	0.000831	7.321916	0.0000
LTR	0.866378	0.010162	85.25481	0.0000
Fixed Effects (Cross)				
_HIGH--C	0.001787			
_LOW--C	-0.001787			

Effects Specification

Cross-section fixed (dummy variables)

R-squared	0.755105	Mean dependent var	0.046272	
Adjusted R-squared	0.754909	S.D. dependent var	0.069102	
S.E. of regression	0.034210	Akaike info criterion	-3.911390	
Sum squared resid	2.917637	Schwarz criterion	-3.904392	
Log likelihood	4884.415	F-statistic	3843.436	
Durbin-Watson stat	1.837470	Prob(F-statistic)	0.000000	

Table 2: Fixed-Effect Test for Multiple Equilibria

I denote TR as capital tax rates and LTR as lagged capital tax rates. The hypothesis of different steady state levels across two groups is not rejected, and high-initial group does converge to a higher level of equilibrium rate. This preliminary test supports the model.

4.2   Simulation Results

I do simulations based on theoretical model to generate a sequence of capital tax rates for each state, and compare it with real data.

For adaptive expectation, I choose $\lambda = 0.9$ [8].

For the government revenue equation, I first normalize $N_t=1$ and set $A_t=25$ [9] for each state.

I calibrate the capital share α_t by $1-\frac{TotalWage}{GSP}$ for each state and year.

As mentioned in previous section, the investment decision curve $k_t = f_i(\theta^e_t)$ for state i is chosen as the polynomial with the most significant regression result.

Feeding in $\theta^e_{1959} = \theta_{1958}$, the economy starts and a sequence of detrended capital chosen by households is generated, together with the sequence of realized capital tax rates.

Pick Alabama to interpret Case 1 of the theory.

---

[8] This weight on new information is chosen such that the simulated result fits the data well. Comparative analysis on this weight is provided in later section.

[9] I choose the same A for states so as to fix the effects from A and isolate the effects of the model. Note that A captures factors more than Total Factor Productivity, such as composition of government tax revenue.
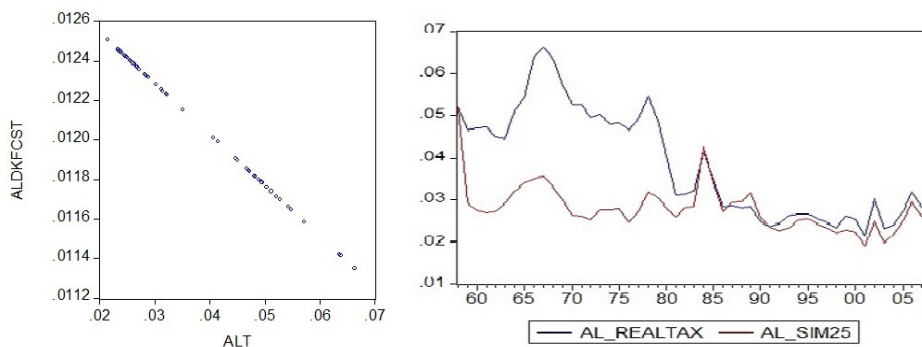
Figure 10: Simulation Results of Alabama

Graph on the left in Figure 10 shows that the investment decision in Alabama follows a decreasing pattern. Alabama starts with an initial tax rate equal to 0.05, which is relatively high. Then the pattern generated is a decreasing sequence, which matches the implication of the model. Moreover, the simulated data fits the real data after about 1980 [10].

Iowa represents the case with increasing investment decision curve, as shown in Figure 11. Following the implication of Inverse U-shaped case, tax rates will fluctuate around a steady state value if the economy does not start with a too high initial belief. Iowa's initial belief is 0.023, which is in the middle of tax range, and the simulation produces a fluctuating sequence.

---

[10]The generated pattern has small fluctuations over the decreasing trend rather than monotonely decreases, due to the fact that real data on government spending is not constant as in the model.

27
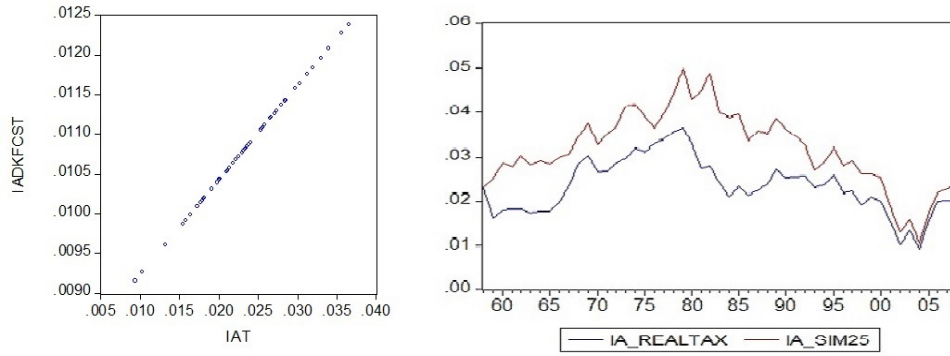
Figure 11: Simulation Results of Iowa

Similarly, New Mexico with an Inverse U-shaped investment curve also generates an oscillating path of capital tax rates, as is shown in Figure 12.
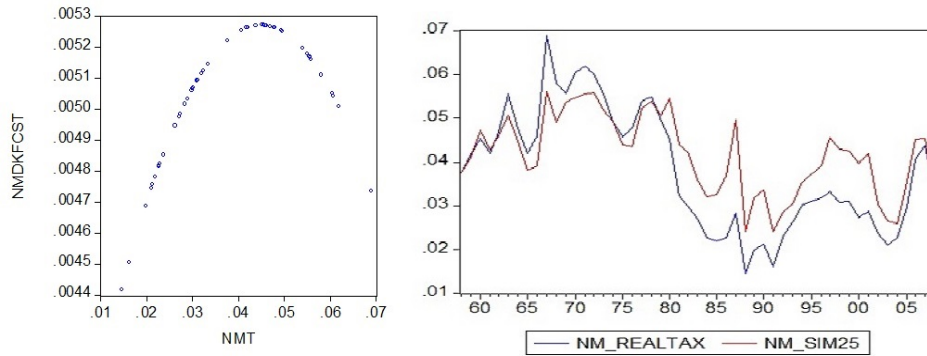
Figure 12: Simulation Results of New Mexico

Simulation result also produces a convergent-to-zero path, as suggested by the model. Virginia maintains a U-shaped investment curve and starts with a high belief at 0.06. The generated sequence fluctuates around a decreasing pattern and converges to 0, which also matches real data quantitatively.
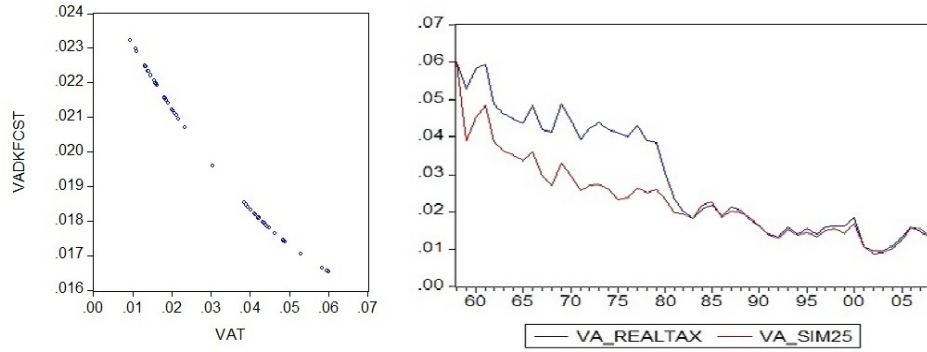
Figure 13: Simulation Results of Virginia

Illinois and Minnesota are two examples with cubic polynomial investment curve. Illinois starts with a low belief at 0.0069 and Minnesota starts with 0.056. Though the model suggests no uniform pattern in this more complicated case, the model can still generate data which captures the observed level and pattern in data, as depicted in Figure 14 and 15.
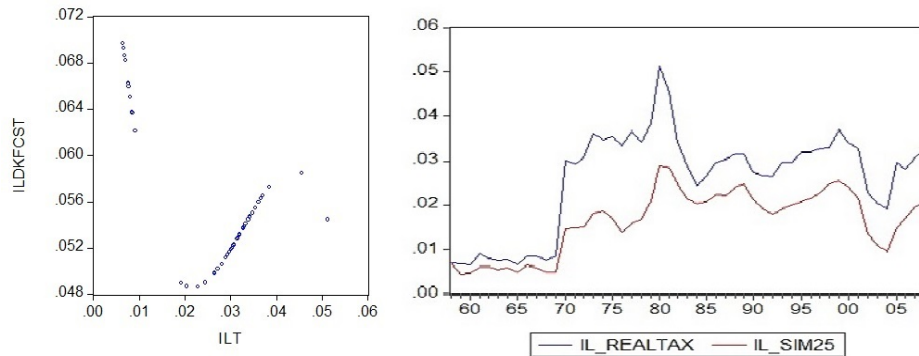
Figure 14: Simulation Results of Illinois
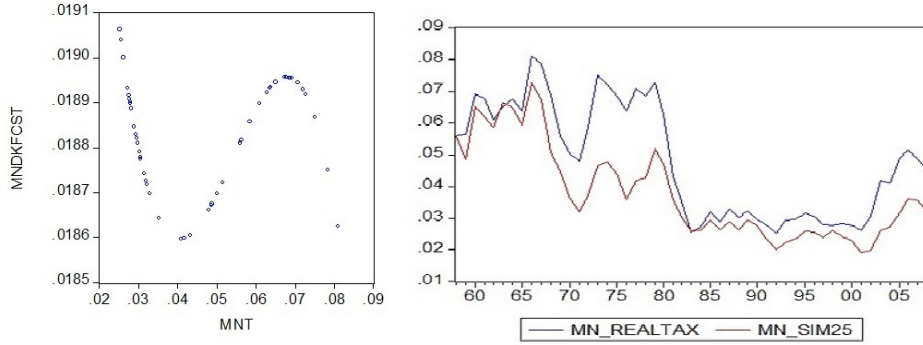
Figure 15: Simulation Results of Minnesota

4.3 Comparative Analysis

Comparative analysis is done regarding initial value and weight in Adaptive Expectation.

Firstly, I use Arizona data for comparative analysis on initial belief. Arizona follows a decreasing pattern of investment curve and according to theory, if there exist a positive steady state equilibrium which is what observed in data, then tax rates should increase monotonely toward it if starting from a low value.

The graph on the left of Figure 16 shows the simulation result as well as real data from the starting belief at a relatively high value, 0.079, which is the real tax rate in 1958. Both sequences display a decreasing pattern. The one on the right, however, feeds in a low initial belief to the economy. Consistent with the theory, tax rates increase to the steady state.
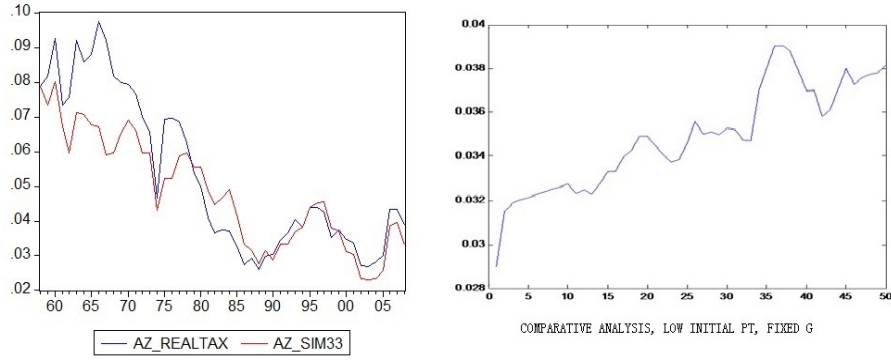
Figure 16: Comparative Analysis on initial belief

I use South Dakota to investigate the effects of belief updating process, which is shown in Figure 17. Recall that λ is the weight households place on new information to form expectation. The left figure presents the case with $\lambda = 1$, and the right one with $\lambda = 0.1$. Though tax rates follow the same pattern, tax rate sequence with $\lambda = 1$ has many spikes, while that under $\lambda = 0.1$ is more smoothed out.

With $\lambda = 1$, households completely rely on new information to update their belief, and with $\lambda = 0.1$, they gradually update their belief and the investment is smoothed out and so is the realized tax rate sequence. Putting a lower weight on previous expectation leads to more jumps of capital levels and thus of tax rates, when government spending level is not stable.
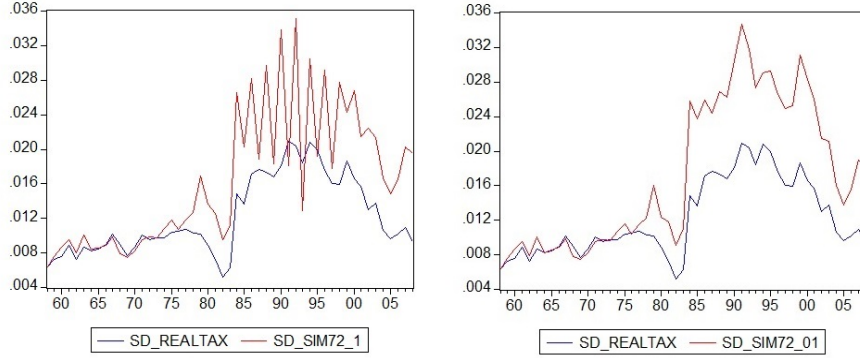
Figure 17: Comparative Analysis on belief updating process

5   Theoretical Model

This section bridges the gap between empirical model and real data with a theoretical model. The theoretical model targets the reduced-form investment decision curve, which is a key ingredient in the empirical model in the previous section.

5.1   Empirical Facts

States fall into different patterns of investment decision, as summarized in Appendix B. Each state maintains its own feature of geographic, economic and educational condition, and it is vital to discover the common traits in each group to explain states' different investment behaviors.

Two empirical facts on common traits are found[11]. Firstly, states with a decreasing

---

[11]The differences in these two common traits are not significant in the groups of states with U-shaped or inverse U-shaped investment curves.

investment curve are tested to have a higher average Gini coefficient than those with an increasing investment curve[12], with details in Appendix C. Households in a more equal economy tend to increase their investment as the expected tax rate increases, while households in a more polarized economy decrease aggregate investment as the expected tax rate increases.

Secondly, states with a decreasing investment curve are tested to have a higher level of education than those with an increasing investment curve[13], with details in Appendix C.

An overlapping generation model generates these two empirical facts.

5.2 Overlapping Generation Model with two types of households

Erosa and Gervais (2002) apply an Overlapping Generation (OLG) Model to present a reason for the nonzero capital tax if tax rates cannot be conditional on age. Garriga (2003) also theoretically analyzes the nonzero capital taxation under the framework of OLG model for a large class of preferences. In the finitely-lived household model, the distortion from capital taxation is much smaller than that in an infinitely-lived household model, and the consumption across the lifecycle of a

---

[12]The data on Gini coefficient in 2010 was obtained by U.S Census Bureau.
[13]I use Bachelor Degree Attainment from IMF as a measure to represent the level of education in each state.

household's life is not constant.

Following Swarbrik (2012), I introduce a two-agent two-period OLG Model with "wealthy" and "poor" households. Each household lives for two stages, young and old. Only "wealthy" households invest in capital when they are young and get capital income at the "old" stage. "Poor" households have no savings and consumes all their income each period. The "wealthy" makes up a portion of γ in the population[14] whilst the "poor" makes up the remaining $1 - \gamma$ of the population. A superscript w denotes variables for the "wealthy" and p for the "poor".

Households obtain pension transfers from the government when they are old.

Households work and receive income from providing labor when they are young. "Wealthy" households can also invest in capital when they are young, and as they age, they have income sources from both capital returns and government pension transfers.

Assume utility function satisfies Inada conditions, and wealthy household chooses consumption for both periods as well as investment[15] to maximize

$$u(c_t^{y,w}) + \beta E_{\theta_{t+1}} u(c_{t+1}^{o,w}) \tag{6}$$

---

[14] I assume $\gamma \epsilon(\frac{1}{2}, 1)$ to analyze the effect of equality in the economy, which ensures sufficient capital for production.

[15] To isolate the analyze on investment behavior, labor is normalized to $n_t = 1$ for both wealthy and poor for now.

subject to the budget constraints for both periods:

$$c_t^{y,w} + k_{t+1} \leq w_t n_t \tag{7}$$

$$c_{t+1}^{o,w} \leq (1 + r_{t+1}(1 - \theta_{t+1}))k_{t+1} + T_{t+1}^w \tag{8}$$

Poor households without investing in capital can only consume with income from government pension transfers at the old stage. They only choose consumption in both periods to maximize:

$$u(c_t^{y,p}) + \beta E_{\theta_{t+1}} u(c_{t+1}^{o,p}) \tag{9}$$

subject to the budget constraints for both periods:

$$c_t^{y,p} \leq w_t n_t \tag{10}$$

$$c_{t+1}^{o,p} \leq T_{t+1}^p \tag{11}$$

Solving the problem of the "wealthy" gives the intertemporal Euler Equation with a complete global insurance company[16]:

$$u'(c_t^{y,w}) = \beta u'(c_{t+1}^{o,w}(\theta_{t+1}^e))(1 + r_{t+1}(1 - \theta_{t+1}^e)) \tag{12}$$

---

[16]The completeness in the global insurance company enables households to fully insure their consumption against capital tax uncertainty.

With a lack of commitment power from the government, households update their beliefs on future tax rates by Adaptive Expectation:

$$\theta_t^e = \lambda \theta_{t-1} + (1 - \lambda)\theta_{t-1}^e \qquad (13)$$

A representative firm produces consumption goods with rented capital and employed labor with Cobb-Douglas production function:

$$Y_t = AK_t^\alpha N_t^{1-\alpha} \qquad (14)$$

where aggregate capital $K_t = \gamma k_t$ and aggregate labor $N_t = 1$.

Rental rate and wage rate are:

$$r_t = A\alpha K_t^{\alpha-1} N_t^{1-\alpha} - \delta \qquad (15)$$

$$w_t = A(1 - \alpha)K_t^\alpha N_t^{-\alpha} \qquad (16)$$

Government's budget constraint determines the capital tax rate for each period t, given G_t, T_t^w and T_t^p.

$$G_t + \gamma T_t^w + (1 - \gamma)T_t^p = K_t r_t \theta_t \qquad (17)$$

The free parameters A and γ in the model capture the level of education and degree of inequality in the economy.

Moretti (2004) provides evidence that human capital is positively correlated with productivity due to externalities. He calculates the fraction of college-educated workers among all to index the level of human capital, which is consistent with the empirical data of Bachelor Degree Attainment I use for each state. Moretti (2004) finds that with human capital spillover, cities with a larger stock of human capital are more productive than those with a smaller stock. Supported by Moretti (2004)'s work, it is legitimate to capture A in the model by Bachelor Degree Attainment in the state. As for how γ in the model captures the degree of equality in an economy, Gini coefficient is calculated with the illustration of Figure 18.
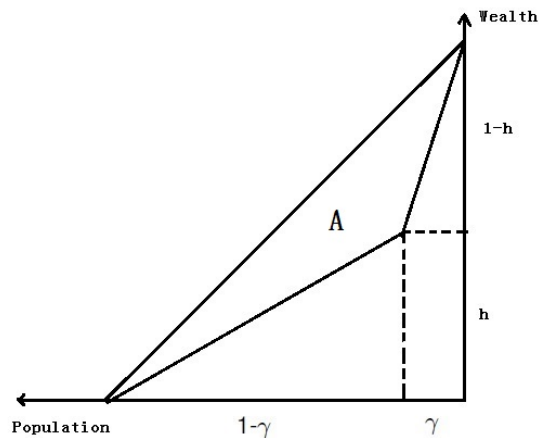
Figure 18: Gini Coefficient

In the economy with two types of households holding wealth w_1, w_2 respectively, where $w_1 < w_2$. Gini coefficient (G) equals $\frac{A}{\frac{1}{2}} = 1 - \gamma - h$, where $h = \frac{(1-\gamma)w_1}{(1-\gamma)w_1+\gamma w_2}$. Plugging in gives $G = \frac{w_2-w_1}{\frac{w_1}{\gamma}+\frac{w_2}{1-\gamma}}$. $\frac{w_1}{\gamma} + \frac{w_2}{1-\gamma}$ decreases in $\gamma \in (0, \gamma^*)$ and increases in $\gamma \in (\gamma^*, 1)$, where $\gamma^* < \frac{1}{2}$. Thus the economy is most unequal at γ^* and becomes more equal as γ approaches to either 0 or 1.

Now take differentiation on Intertemporal Euler Equation (11) with respect to θ_{t+1}^e and k_{t+1} to obtain the curve of investment decision.

Thus,

$$\frac{dk_{t+1}}{d\theta_{t+1}^e} =$$

$$\frac{\beta r_{t+1}\Psi}{\beta\Psi(1 - \theta_{t+1}^e)\frac{dr_{t+1}}{dk_{t+1}} + \beta u''[(1 + r_{t+1}(1 - \theta_{t+1}^e))k_{t+1} + T_{t+1}^w](1 + r_{t+1}(1 - \theta_{t+1}^e))^2 + u''(w_t - k_{t+1})}$$

$$(18)$$

where

$$\Psi = u''[(1+r_{t+1}(1-\theta_{t+1}^e))k_{t+1}+T_{t+1}^w](1+r_{t+1}(1-\theta_{t+1}^e))k_{t+1}+u'[(1+r_{t+1}(1-\theta_{t+1}^e))k_{t+1}+T_{t+1}^w]$$

and $\frac{dr_{t+1}}{dk_{t+1}} = A\alpha(\alpha - 1)\gamma^{\alpha-1}k_{t+1}^{\alpha-2} < 0$.

Suppose $\Psi > 0$, then the denominator is negative and the investment decision curve is downward sloping.

As A decreases and γ increases, Ψ decreases with a utility function satisfying certain conditions characterized as follows.

Condition 5 *The utility function holds the following condition:*

$u(c)$ has the Elasticity of Intertemporal Substitution, $-\frac{u'(c)}{u''(c)c}$, below one at lower range of c and above one at higher range of c such that $u''(c)c + u'(c) < 0$ at low c and > 0 at high c.

Ψ can decrease to be negative. With a smaller A and a bigger γ, $\left| \frac{dr_{t+1}}{dk_{t+1}} \right|$ decreases. Hence, with $\Psi < 0$, the denominator can remain negative and the total effect is positive, which implies that the investment decision curve is upward sloping.

The intuition is as follows: A lower TFP value decreases households' income at each period. With a strong consumption smoothing effect at a low consumption value as suggested by Condition 1, the income effect becomes stronger when the income is lower. Moreover, a higher γ increases the portion of wealthy households who invest in capital which in turn reduces the return of aggregate investment and alleviates the substitution effect. Thus in a more equalized economy with more investors and a lower productivity level, income effect dominates substitution effect and the investment decision curve is sloped upwards.

6 Conclusion

This paper provides an alternative explanation for the possibility of nonzero capital taxation in the economy. Government's lack of commitment power forces households to form their own expectations on tax rates. Furthermore, the balanced budget constraint disenables the government to freely set the capital tax rates in order to alter households' belief. Consequently, capital tax rates are history-dependent.

The pattern of historic capital rate development depends on two factors: initial belief on the capital tax rate and state-specific investment behavior. The overall education level as well as the degree of equality in the economy determines the investment decision curve for each state, according to empirical observations. An overlapping generation model with two heterogeneous agents can produce this result as long as the utility function satisfies certain conditions.

A more equalized economy with a lower productivity level increases income effect and decreases substitution effect, which generates an increasing investment decision curve.

For some cases of this theory, the economy will converge to a zero capital tax rate, which coincides with Ramsey's result. With an elastic investment curve and a low enough initial tax belief, capital tax rates converge to zero in the long run.

This paper simulates capital tax rate patterns which match the real data. The future study is possibly to extend the existing model to further rationalize the U-shaped and inverse U-shaped investment decision curves.

Chapter 2: Tax Competition with Population Growth

1   Introduction

A long line of literature has been focusing on interaction among governments. One source of interaction is the mobile capital moving across jurisdictions, which leads to the series of theoretical literature on tax competition. Bucovetsky (1991) is among the pioneering literature which presents that strategic interaction leads to underprovision of public goods as each jurisdiction sets a tax rate so low to preserve the tax base. Kanbur and Keen (1993), together with Bucovetsky (1991), provide models with unequal population size and conclude that the equilibrium tax rates are higher in more populated areas. Pi and Zhou (2013) consider all-purpose public goods, which increase private firms' productivity through provision of infrastructure, and demonstrate that tax competition does not necessarily lead to inefficient outcomes.

The main strand of empirical literature tests the presence of strategic interaction among governments, through estimating reaction functions, and tax competition framework represents the best known example of the resource-flow models. Brueckner (2003) provides an overview of related empirical studies, which summarizes papers estimating tax reaction functions in Boston metropolitan areas (Brueckner and Saave-

dra, 2001), in Canada (Brett and Pinkse, 2000; Hayashi and Boadway, 2001) and etc. Almost all the empirical results confirm a positive presence of strategic interaction, implying that the decision variables are "strategic complements".

In the U.S., competition over capital can also be a potential issue among states. In 2005, Intel company, originated in California, decided to establish their multi-billion chip-making factory in Arizona, due to the more favorable corporate income tax environment there. In 2015, General Electric warned their 42-year-old home state Conneticut of their rising corporate income tax rate, before actually leaving for Boston. Besides all these facts of firms making business decisions based on capital income tax system, there also seems to exist capital tax policy interaction among some states. New Mexico has started a schedule of cutting state corporate income tax rate from 6.9 in 2016 to 6.6 in 2017 and to a target of 5.9 in 2018; its neighbor Arizona, meanwhile, has reduced its corporate income tax rate from 6.0 in 2015 to 5.5 in 2016, and has planned to keep this falling trend to 2017 and 2018.(Walczak, Drenkard, and Henchman, 2016)

Utilizing a panel data of average capital tax rates from 1958 to 2007 at the state-level in the US, this paper verifies the existence of capital tax competition. Besides OLS panel regression, I apply spatial autoregressive (SAR) panel estimation proposed by Elhorst (2003) to avoid the potential endogeneity problem of regressors. The results of SAR estimation are qualitatively consistent with those of OLS estimation.

Moreover, this paper is the first to uncover the difference in competition pat-

terns among states in the South and West, with that in the Midwest and Northeast. Furthermore, it is also the first to explore the underlying reason for this difference, utilizing high-order SAR panel estimation with fixed-effects. Controlling for macro-economic and political environment features of each state, the effect of population growth rate on the reaction coefficient is positive and significant. Faster population growth induces stronger tax competition behavior.

To support the argument that tax competition explains the interaction of tax rates, the relationship between tax base and its own and neighbors' tax rates is estimated. As expected, capital is negatively related to own tax rate and positively related to neighbors' tax rates, which further confirms the view of states having a tax cut to fight over the tax base.

In contrast to the result in Chirinko and Wilson (2013), the response coefficients obtained in this paper are positive and siginificant in the South and West. They esti-mate the tax competition pattern among states in the U.S. using data on investment tax credits (ITC) and corporate income tax (CIT), but fail to show the existence of tax competition empirically. Compared to their study, this paper applies a new data series of average capital tax rates with a longer timespan and also deals with specific features in different areas of the U.S.

The theoretical literature, however, has been silent regarding the slope of the reac-tion function. This paper studies the behavior of tax competition and its relationship with population growth.

Keen and Kotsogiannis (2002) explore vertical and horizontal tax externalities with a saving model. Population growth is introduced in this paper based on their framework, to study the interaction between population growth and tax competition pattern.

Many researchers are concerned that a faster population growth brings cost to a society by reducing natural resources, physical and human capital per worker, which is widely known as "dilution effect". Apart from many theoretical support (Samuelson, 1975; Deardoff, 1976; Galor and Weil, 1996), Mankiw, Romer and Weil (1992) examine a sample that includes almost all countries[1] between 1960-1985 and provide empirical evidence that population growth rate has important effect on per capita income quantitatively. A higher population growth rate spreads capital and other resources more thinly such that capital per cap is lower, while a lower rate increases capital intensity in the economy.

In this paper, faster population growth leads to a larger gap between the number of people who save and people who share the increased capital, and thus capital is more diluted in the second period. Any tax cut attracts less inflow of capital per worker in the area with faster population growth. An additional effect is that given any tax cut from neighboring state, the effect of capital outflow is more severe since the compensated capital from saving is more spread out in the second period. Hence, states compete in a more fierce manner due to the dilution effect.

Eakin (1994) investigates the role of public infrastructure on private firms' produc-

---

[1]Central-planned countries are excluded.

tion at state-level in the US, and shows that public good has little effect on private firms' production possibilities, while private capital has effect on its productivity. This empirical evidence motivates the form of production function applied in the theoretical model, which is different from Pi and Zhou (2013).

There is no ambiguity regarding the effect of population growth rate on the degree of inefficiency. In particular, tax competition is more damaging when competing states have higher population growth rates. This paper analyzes the welfare implications of tax competition from another point of view, based on Keen and Kotsogiannis (2004).

The structure of this paper is as follows: Section 2 introduces the dataset, and Section 3 provides empirical findings. Section 4 presents the theoretical model with population growth. Lastly, Section 5 concludes.

2 U.S. State-Level Panel Data

The estimation of tax competition is based on the estimated coefficients of capital-tax reaction function in different areas of the U.S. The U.S. state-level panel data is for the period 1958-2007. I estimate the results for the four areas of the U.S.: Midwest, South, West and Northeast[2]. The analysis is on how the capital tax rate of one state is determined by the capital tax rates of its neighbors within the same area. Each area has its own specific growth rate of population for the past half century, and the

---

[2]A list of states in this four areas is included is in Appendix D.

study focuses on the relationship between the degree of capital tax competition and how fast population grows. Details about data sources and variable definitions are presented in Appendix E.

A. Capital tax rate

The officially available data on capital taxation includes marginal capital gains tax rates, brackets and so on. These information have been used to calculate effective capital tax rates. In most theoretical models, return from capital investment is taxed proportionally and thus the tax rate is simplified as an average tax rate. Moreover, average capital tax rates can combine the effects of different categories of capital taxation into one index, which allows for the fact that states might use different tax instruments to attract business. However, insufficient empirical work has been done to obtain average capital tax rates at the state-level in the U.S. Thus, I obtained my own series of average capital tax rates for each state.

B. Control Variables

Capital is not only taxed at the state-level, but also at the federal-level in the U.S. Thus, the first control variable is the federal effective capital gains tax rate at each year, which is common to all the states. The influence from capital tax rate at the federal-level on the tax rate at the state-level can be examined.

I also account for macro-economic condition and political environment.

Personal income at the state-level is applied to represent macro-economic condition in each state for each year.

Political environment is hardly observed, and electoral outcome serves as a good proxy. I apply the series of data on legislature's party of each state. I measure three alternatives: the fraction of State House that is Democrat, the fraction of State Senate that is Democrat and a dummy variable representing whether the majority of State House and Senate are Democrat.

C. Weighting Scheme

There are many possible schemes for econometricians to describe a neighbor and assign the weights. The notion of close proximity can refer to closedness of geopraphic location or similarity of industrial environment.

Pinkse, Slade and Brett (2002) investigate the nature of competition with measures including nearest neighbors geographically, sharing markets with common boundaries and located a certain Euclidean distance apart. They find that the competition is highly localized and rivalry decays abruptly with geographic distance.

Moreover, physical capital is imperfectly mobile across states, with cost of moving and adjusting to new social, cultural and political environment. Thus, it is natural to start with a geographic-based weighting scheme, following many empirical literature including Brueckner and Saavedra (2001), Chirinko and Wilson (2013), Buettner (2003), Brett and Pinkse (2000).

The weighting matrix W can be time-invariant or time-variant. I first consider the case with time-invariant W, such that $W = I_T \otimes W_n$.

The first scheme assigns equal weights to all contiguous states[3], so $w_{ij} = 1$ if states i and j share the same border geographically.

Equally weighted scheme, however, is insufficient to discriminate among all the contiguous neighbors in the same area. The second scheme is to combine both geographic and economic distance, where w_{ij} is adjusted by population size for each contiguous state, assuming a bigger influence from a more populous neighbor. I take the time-average population size for each state first, so W is time-invariant.

Scheme 1	Contiguous neighbors only, equally weighted.
Scheme 2	Contiguous neighbors only, weighted by time-average population size.

Table 3: Weighting Schemes of SAR Panel Estimation

D. Population growth rates

This paper examines how population growth rate influences the degree of capital tax competition in each area.

The major four areas in the U.S. have different population growth rates. Population has been growing much faster in the South and West, compared to Midwest and Northeast. I obtain the series of state population data from 1958 to 2007 and calculate the time-average population growth rates for these 50 years for each state. The time-average growth rates are summarized in Appendix E.

---

[3]To focus on the pattern of tax competition in each area, I confine the pool of neighbors as all the states in that area only.

The time-average population data is also used to assign weights in the spatial estimation, as above in section 2.C.

And the series of historic population data for each state and each year is also used in the capital response regression.

3   Empirical Findings

This paper analyzes the patterns of capital tax competition among states in the South, Midwest, West and Northeast. Southern states such as Alabama and Georgia are known to have higher population growth rates, compared to Midwestern states like Michigan. The main goal in this section is to first examine the existence of capital tax competition in these four areas of the U.S., and then to test whether the tax competition is stronger among states with faster population growth, through estimating the tax reaction function.

The estimation starts with OLS estimation and proceeds to spatial autoregressive (SAR) panel estimation.

3.1   Empirics on tax competition pattern

The basic estimated reaction equation takes the form:

$$OTR_{st} = \beta \cdot TN_{st} + \gamma \cdot TF_t + \mathbf{X}_{st} \cdot \tau + u_s + \epsilon_{st} \tag{1}$$

where OTR_{st} is the own capital tax rate of state s at year t, TN_{st} is the average neighbors' capital tax rates of state s at year t, and thus β captures the degree of capital tax competition. TF_t is the federal capital tax rate at year t. $\mathbf{X}_{st}$ is a row vector of exogenous explanatory variables, with macroeconomic and political environments included in this paper. PI_{st} is personal income level as an explanatory variable to account for the macro-economic characteristic in state s at time t. Policy makers' preferences are largely involved in the tax setting process. To account for political environment, I add legislature's party as another explanatory variable. D_H_{st} and D_S_{st} are the fraction of state house that is democrat, the fraction of state senate that is democrat respectively, which represent the political environment in state s at time t. d_{st} captures whether democrat is majority in state house and state senate. Details of variables are presented in Appendix E. As unobservable individual features of each state, including historical or institutional factors, may influence policy on capital taxation, u_s captures the fixed-effects. ϵ_{st} is a random error term.

All variables are summarized in Table 4.

OTR_{st}	own capital tax rate of state s at year t
TN_{st}	average neighbors' capital tax rates of state s at year t
TF_t	federal capital tax rate at year t
g_s	time-average population growth rate of state s
g_{st}	population growth rate of state s at year t
X_{st}	exogeneous features of state s at year t
PI_{st}	personal income of state s at year t
$\mathrm{D\_H}_{st}$	fraction of state house that is democrat of state s at year t
$\mathrm{D\_S}_{st}$	fraction of state senate that is democrat of state s at year t
k_{st}	capital per cap of state s at time t
d_{st}	whether democrat is majority in state house and senate of state s at time t

Table 4: Abbreviations of variables

Figure 19 and 20 contrast the pattern of capital tax rates between Alabama and Michigan with their neighbors' average[4]. Tax rates in Alabama and its neighbors closely follow each other; while in Michigan, no such pattern exists.
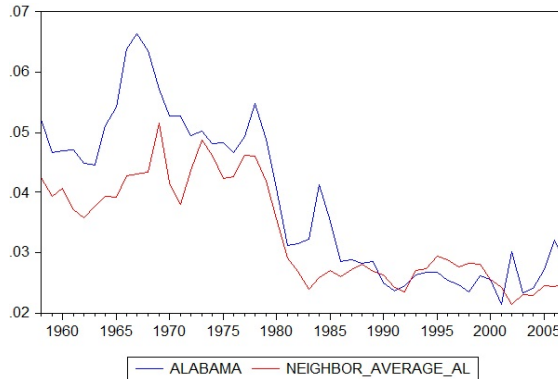
Figure 19: Capital tax rates of Alabama and its neighbors' average.
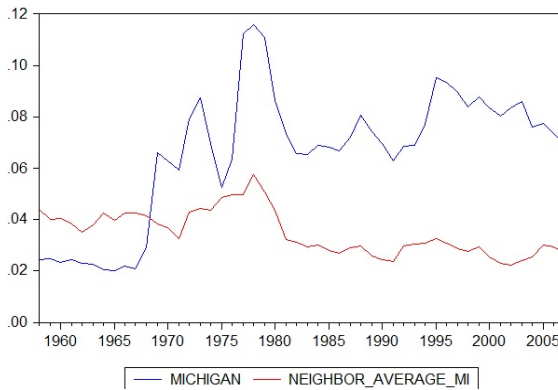
Figure 20: Capital tax rates of Michigan and its neighbors' average.

Moreover, Table 5 displays a preliminary comparison between Alabama and Michigan, where own state tax rates respond much stronger to neighbors' tax change in Alabama than that in Michigan.

[4]Scheme 1 of defining neighbors is applied in this estimation, with every contiguous neighbor being equally weighted.

Dependent variable: OwnTaxRate

Explanatory Variables	Alabama	Michigan
TaxNeighbor	1.184***	0.157
	(0.090)	(0.424)
TaxFed	-0.095***	0.371***
	(0.022)	(0.104)
Constant	0.016**	-0.005
	(0.006)	(0.028)

Note: These are least squares estimates of the parameters in Eq. (1).

Robust standard errors in parentheses. ***p<0.01, **p<0.05, *p<0.1.

Table 5: Tax competition regressions

To investigate tax competition patterns in the South, Midwest, West and Northeast, I run individual panel regressions with fixed effect[5] for each of these four areas. Specifications with different explanatory variables are estimated[6].

Results of both neighboring schemes show that capital taxes compete in a stronger and more significant manner in the South and West, compared to that in the Midwest and Northeast. Moreover, adjusting the weights by population size magnifies these differences.

In addition, state capital tax rates respond to federal capital tax rates negatively

---

[5]Hausman test suggests that fixed-effect estimator is preferred.

[6]Results of OLS estimation are robust to those of SAR panel estimation and thus omitted in this paper.

yet insignificantly in most specifications.

To avoid the endogeneity problem of the regressors, SAR estimation is adopted in next subsection.

Alternatively, Chirinko and Wilson (2013) suggest estimating with political preference as an instrumental variable can take care of the endogeneity problem.

3.2 Spatial Autoregressive Panel Estimation

To avoid simultaneity problem of regressors from OLS estimation, I use a spatial autoregressive (SAR) panel model to estimate the effect of neighbors' tax rates on own state tax rates.

$$Y = \lambda WY + X\beta + l_T \otimes u_n + \epsilon \tag{2}$$

where Y is an $nT \times 1$ vector of own state tax rates, W is an $nT \times nT$ weighting matrix, X is an $nT \times k$ vector of exogenous variables, β is a $k \times 1$ vector of parameters, u_n is an $n \times 1$ vector of fixed-effect errors, and ϵ is an $nT \times 1$ vector of random errors. n is the number of states in one area, T is the number of years and k is the number of exogeneous state-dependent exogeneous variables included. And λ captures the degree of capital tax competition.

For South, Midwest, West and Northeast, I run the SAR Panel estimation with spatial fixed effect for each specification and each neighboring scheme, with the results summarized below.

	Scheme 1				
	South	South	South	South	South
$WTaxrate$	0.314***	0.235***	0.280***	0.313***	0.274***
	(0.037)	(0.039)	(0.037)	(0.037)	(0.037)
TaxFed	-0.050***	-0.033***	-0.044***	-0.052***	-0.031**
	(0.013)	(0.012)	(0.012)	(0.013)	(0.013)
Personal Income	-3.1e-08***	2.2e-09	1.8e-08***	-2.8e-08***	1.3e-08**
	(4.5e-09)	(6.1e-09)	(6.2e-09)	(5.8e-09)	(6.5e-09)
Democrat_House		0.036***			0.005
		(0.004)			(0.006)
Democrat_Senate			0.045***		0.046***
			(0.004)		(0.006)
Political Dummy				0.001	-0.004***
				(0.001)	(0.001)
Log-likelihood	2414.284	2446.208	2470.462	2414.522	2475.196

Note: These are SAR Panel estimates with fixed-effect of the parameters in Eq. (2).

Robust standard errors in parentheses. ***p<0.01, **p<0.05, *p<0.1.

Table 6: Tax competition regressions, SAR Panel with spatial fixed effect South 1

	Scheme 2				
	South	South	South	South	South
$WTaxrate$	0.348***	0.297***	0.320***	0.350***	0.315***
	(0.034)	(0.035)	(0.034)	(0.034)	(0.034)
TaxFed	-0.045***	-0.027**	-0.039***	-0.047***	-0.026**
	(0.013)	(0.012)	(0.012)	(0.013)	(0.013)
Personal Income	-2.6e-08***	6.0e-09	2.2e-08***	-2.4e-08***	1.6e-08***
	(4.5e-09)	(6.0e-09)	(6.1e-09)	(5.7e-09)	(6.5e-09)
Democrat_House		0.034***			0.005
		(0.004)			(0.006)
Democrat_Senate			0.046***		0.046***
			(0.004)		(0.006)
Political Dummy				0.001	-0.004***
				(0.001)	(0.001)
Log-likelihood	2419.263	2450.126	2474.039	2419.437	2478.969

Note: These are SAR Panel estimates with fixed-effect of the parameters in Eq. (2).

Robust standard errors in parentheses. ***$p<0.01$, **$p<0.05$, *$p<0.1$.

Table 7: Tax competition regressions, SAR Panel with spatial fixed effect South 2

	Scheme 1				
	Midwest	Midwest	Midwest	Midwest	Midwest
$WTaxrate$	0.082	0.049	0.071	0.056	0.051
	(0.051)	(0.051)	(0.050)	(0.050)	(0.050)
TaxFed	-0.012	-0.015	-0.023	-0.021	-0.021
	(0.017)	(0.016)	(0.017)	(0.016)	(0.017)
Personal Income	1.3e-08	2.0e-08***	2.1e-08***	2.0e-08***	2.1e-08***
	(7.7e-09)	(7.6e-09)	(7.7e-09)	(7.6e-09)	(7.6e-09)
Democrat_House		0.024***			0.010
		(0.006)			(0.008)
Democrat_Senate			0.019***		2.9e-04
			(0.006)		(0.008)
Political Dummy				0.004***	0.003**
				(8.4e-04)	(0.001)
Missing_Political		0.032***	0.030***	0.025***	0.029***
		(0.005)	(0.005)	(0.004)	(0.006)
Log-likelihood	1738.740	1756.306	1754.422	1759.054	1759.899

Note: These are SAR Panel estimates with fixed-effect of the parameters in Eq. (2).

Robust standard errors in parentheses. ***p<0.01, **p<0.05, *p<0.1.

Table 8: Tax competition regressions, SAR Panel with spatial fixed effect Midwest 1

	Scheme 2				
	Midwest	Midwest	Midwest	Midwest	Midwest
$WTaxrate$	0.109**	0.047	0.070	0.052	0.045
	(0.048)	(0.048)	(0.048)	(0.047)	(0.048)
TaxFed	-0.012	-0.015	-0.023	-0.021	-0.021
	(0.017)	(0.016)	(0.017)	(0.016)	(0.017)
Personal Income	1.2e-8	2.0e-08***	2.0e-08***	2.0e-08	2.0e-08***
	(7.7e-09)	(7.6e-09)	(7.7e-09)	(7.6e-09)	(7.6e-09)
Democrat_House		0.023***			0.010
		(0.006)			(0.008)
Democrat_Senate			0.019***		3.5e-04
			(0.006)		(0.008)
Political Dummy				0.004***	0.003**
				(8.4e-04)	(0.001)
Missing_Political		0.031***	0.029***	0.024***	0.029***
		(0.005)	(0.005)	(0.004)	(0.006)
Log-likelihood	1740.178	1756.349	1754.613	1759.106	1759.921

Note: These are SAR Panel estimates with fixed-effect of the parameters in Eq. (2).

Robust standard errors in parentheses. ***p<0.01, **p<0.05, *p<0.1.

Table 9: Tax competition regressions, SAR Panel with spatial fixed effect Midwest 2

	Scheme 1				
	West	West	West	West	West
$W'Taxrate$	0.373***	0.282***	0.293***	0.333***	0.257***
	(0.047)	(0.047)	(0.048)	(0.047)	(0.048)
TaxFed	-0.028*	-0.016	-0.009	-0.011	-0.009
	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)
Personal Income	-1.02e-08***	-1.19e-08***	-1.18e-08***	-1.20e-08***	-1.25e-08***
	(3.94e-09)	(3.80e-09)	(3.86e-09)	(3.88e-09)	(3.80e-09)
Democrat_House		0.043***			0.032***
		(0.006)			(0.007)
Democrat_Senate			0.033***		0.018***
			(0.005)		(0.006)
Political Dummy				0.005***	1.21e-04
				(8.94e-04)	(0.001)
Log-likelihood	1605.688	1633.698	1627.627	1620.589	1638.636

Note: These are SAR Panel estimates with fixed-effect of the parameters in Eq. (2).

Robust standard errors in parentheses. ***p<0.01, **p<0.05, *p<0.1.

Table 10: Tax competition regressions, SAR Panel with spatial fixed effect West 1

	Scheme 2				
	West	West	West	West	West
$WTaxrate$	0.333***	0.234***	0.245***	0.287***	0.210***
	(0.043)	(0.044)	(0.044)	(0.043)	(0.044)
TaxFed	-0.039**	-0.025	-0.019	-0.022	-0.018
	(0.017)	(0.016)	(0.016)	(0.017)	(0.016)
Personal Income	-1.00e-08**	-1.22e-08***	-1.22e-08***	-1.22e-08***	-1.29e-08***
	(4.01e-09)	(3.85e-09)	(3.91e-09)	(3.95e-09)	(3.85e-09)
Democrat_House		0.045***			-0.033***
		(0.006)			(0.007)
Democrat_Senate			0.034***		-0.018***
			(0.005)		(0.007)
Political Dummy				0.005***	5.31e-05
				(9.01e-04)	(0.001)
Log-likelihood	1599.890	1629.410	1622.410	1615.070	1634.353

Note: These are SAR Panel estimates with fixed-effect of the parameters in Eq. (2).

Robust standard errors in parentheses. ***p<0.01, **p<0.05, *p<0.1.

Table 11: Tax competition regressions, SAR Panel with spatial fixed effect West 2

	Scheme 1				
	Northeast	Northeast	Northeast	Northeast	Northeast
$WTaxrate$	0.290***	0.256***	0.247***	0.248***	0.245***
	(0.047)	(0.047)	(0.047)	(0.047)	(0.047)
TaxFed	-0.013	-0.041	-0.048	-0.041	-0.048
	(0.033)	(0.033)	(0.032)	(0.032)	(0.032)
Personal Income	-3.35e-08***	-4.24e-08***	-3.20e-08***	-3.61e-08***	-3.11e-08***
	(1.03e-08)	(1.04e-08)	(9.96e-09)	(1.00e-08)	(1.04e-08)
Democrat_House		0.048***			-0.017
		(0.011)			(0.017)
Democrat_Senate			0.052***		0.038***
			(0.009)		(0.013)
Political Dummy				0.010***	0.007***
				(0.002)	(0.003)
Log-likelihood	1049.748	1058.676	1066.667	1065.917	1070.420

Note: These are SAR Panel estimates with fixed-effect of the parameters in Eq. (2).

Robust standard errors in parentheses. ***$p<0.01$, **$p<0.05$, *$p<0.1$.

Table 12: Tax competition regressions, SAR Panel with spatial fixed effect Northeast 1

	Scheme 2				
	Northeast	Northeast	Northeast	Northeast	Northeast
$WTaxrate$	0.129***	0.085*	0.092*	0.079	0.078
	(0.049)	(0.050)	(0.049)	(0.049)	(0.049)
TaxFed	-0.027	-0.059*	-0.065*	-0.056*	-0.066**
	(0.034)	(0.034)	(0.033)	(0.033)	(0.033)
Personal Income	-2.81e-08***	-4.07e-08***	-2.78e-08***	-3.29e-08***	-2.91e-08***
	(1.08e-08)	(1.07e-08)	(1.03e-08)	(1.04e-08)	(1.08e-08)
Democrat_House		0.059***			-0.009
		(0.011)			(0.018)
Democrat_Senate			0.060***		0.041***
			(0.009)		(0.013)
Political Dummy				0.012***	0.007***
				(0.002)	(0.003)
Log-likelihood	1035.784	1048.550	1056.610	1055.343	1060.551

Note: These are SAR Panel estimates with fixed-effect of the parameters in Eq. (2).

Robust standard errors in parentheses. ***p<0.01, **p<0.05, *p<0.1.

Table 13: Tax competition regressions, SAR Panel with spatial fixed effect Northeast 2

Estimation results are not only consistent with those from OLS regression, but also show a sharper contrast in the degrees of capital tax competition.

Both neighboring schemes' results suggest the following: there exists a significant pattern of tax competition in the South and West under all specifications. Tax competition is positive but insignificant in the Midwest under most specifications, except one estimation with population adjusted neighbors. For the Northeast, no pattern of tax competition exists under all specifications.

To summarise, state capital tax competition is much stronger as well as more significant in the South and West where population have been growing faster, than that in the Midwest and Northeast with lower popoulation growth rates.

Moreover, the SAR panel estimation is more efficient, compared to OLS estimation.

3.3   Effect of population growth on capital tax competition

States in the South and West have been experiencing faster population growth than states in the Midwest and Northeast[7]. I examine whether the higher population growth rate induces stronger tax competition.

I use a high-order spatial autoregressive (SAR) panel model with fixed effects to estimate the effect of population growth rate on the degree of tax competition.

---

[7]Statistics of population growth rates are summarized in Appendix E.

$$Y_{nt} = \lambda_1 W_{1n} Y_{nt} + \lambda_2 W_{2n} Y_{nt} + X_{nt}\beta + u_n + \epsilon_{nt} \tag{3}$$

$$t = 1, 2, ..., T, \tag{4}$$

where $Y_{nt} = (y_{1t}, y_{2t}, ..., y_{nt})'$ is an $n \times 1$ vector of own state tax rates, W_{1n} is an $n \times n$ nonstochastic weighting matrix, $W_{2n} = G_n W_{1n}$, G_n is an $n \times n$ matrix with diagonal entries equal to the time-averaged population growth rates of each state. X_{nt} is an $n \times k$ vector of exogenous time varying variables, β is a $k \times 1$ vector of parameters, u_n is an $n \times 1$ vector of fixed-effect errors, and $\epsilon_{nt} = (\epsilon_{1t}, \epsilon_{2t}, ..., \epsilon_{nt})'$ is an $n \times 1$ vector of random errors. n is the number of states in one area, T is the number of years and k is the number of exogeneous state-dependent exogenous variables included.

Thus, $\lambda_1 + \lambda_2 G_n$ represents the degree of tax competition and the coefficient λ_2 captures how population growth rate affects the degree of tax competition.

Followed by Lee and Yu (2014), GMM estimation is applied through a transformation approach to take account of the fixed effects. $[F_{T,T-1}, \frac{1}{\sqrt{T}} l_T]$ is the orthonormal matrix of eigenvectors of $J_T = (I_T - \frac{1}{T} l_T l_T')$, and $F_{T,T-1}$ is composed of the eigenvectors corresponding to all eigenvalues equal to one, so $F_{T,T-1}$ is $T \times (T-1)$. The variables are transformed as follows: $[Y_{n1}^*, Y_{n2}^*, ..., Y_{nT-1}^*] = [Y_{n1}, Y_{n2}, ..., Y_{nT}]F_{T,T-1}$, $[X_{n1}^*, X_{n2}^*, ..., X_{nT-1}^*] = [X_{n1}, X_{n2}, ..., X_{nT}]F_{T,T-1}$, and $[\epsilon_{n1}^*, \epsilon_{n2}^*, ..., \epsilon_{nT-1}^*] = [\epsilon_{n1}, \epsilon_{n2}, ..., \epsilon_{nT}]F_{T,T-1}$.

With the fixed effects eliminated, the estimated equation becomes:

$$Y_{nt}^* = \lambda_1 W_{1n} Y_{nt}^* + \lambda_2 W_{2n} Y_{nt}^* + X_{nt}^* \beta + \epsilon_{nt}^* \tag{5}$$

$$t = 1, 2, ..., T - 1, \tag{6}$$

I then apply 2SLS estimation with optimum instrumental variables (IVs) chosen as suggested by Kelejian and Prucha (1998)[8]. The estimated results are summarized below.

	Scheme 1			
$WTaxrate$	0.006	0.021	-0.044	0.012
	(0.170)	(0.166)	(0.179)	(0.169)
$G * WTaxrate$	18.979*	14.770	22.864*	14.812
	(11.303)	(10.292)	(11.917)	(9.811)
TaxFed	-0.024	-0.033*	-0.034**	-0.032*
	(0.017)	(0.017)	(0.017)	(0.017)
Personal Income	-0.000	-0.000	-0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Democrat_House	0.031***			0.009
	(0.007)			(0.009)
Democrat_Senate		0.035***		0.026***
		(0.006)		(0.007)
Political Dummy			0.005***	0.001
			(0.001)	(0.001)
Missing_Political	0.033***	0.037***	0.023***	0.038***
	(0.006)	(0.006)	(0.004)	(0.006)

Note: These are high-order SAR estimates with fixed-effect of the parameters in Eq. (5).

Robust standard errors in parentheses. ***p<0.01, **p<0.05, *p<0.1.

Table 14: Tax competition regressions, high-order SAR Panel with spatial fixed effect Scheme 1

---

[8] The procedure of choosing optimum IVs are included in Appendix F.

	Scheme 2			
$WTaxrate$	-0.061	-0.081	-0.167	0.077
	(0.193)	(0.186)	(0.220)	(0.190)
$G*WTaxrate$	34.432**	31.984**	42.868***	21.784*
	(13.632)	(14.054)	(16.163)	(12.387)
TaxFed	-0.024	-0.032*	-0.035*	-0.029
	(0.018)	(0.018)	(0.018)	(0.018)
Personal Income	0.000	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Democrat_House	0.028***			0.006
	(0.008)			(0.009)
Democrat_Senate		0.032***		0.025***
		(0.007)		(0.007)
Political Dummy			0.005***	0.001
			(0.001)	(0.001)
Missing_Political	0.028***	0.032***	0.019***	0.033***
	(0.006)	(0.006)	(0.004)	(0.006)

Note: These are high-order SAR estimates with fixed-effect of the parameters in Eq. (5).

Robust standard errors in parentheses. ***p<0.01, **p<0.05, *p<0.1.

Table 15: Tax competition regressions, high-order SAR Panel with spatial fixed effect Scheme 2

As shown in Table 14 and 15, there exists tax competition $(\lambda_1 + \lambda_2 \cdot G > 0)$. Moreover, specifications under both neighboring schemes show that λ_2 is positive and significant. With the population weighted neighboring scheme, this result is more significant. States with higher population growth rates compete in a much stronger manner than those with lower population growth rates[9].

This can be also shown in Figure 21 where representative states are marked.

---

[9]For robustness check, I run estimation with a different G as the matrix of time-averaged growth rate of per capita income for each state. The result shows that G has insignificant effect on the degree of tax competition for most specifications in the first neighboring scheme, and the significance levels are lower than those with G as population growth rate for most specifications in the second neighboring scheme.
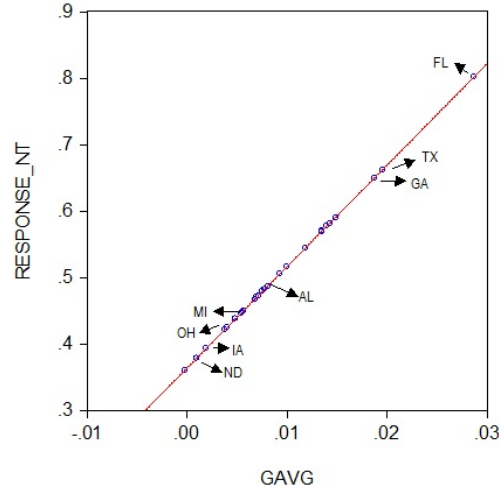
Figure 21: Response Coefficients and Population Growth Rates

Tax rates at federal level affect own state tax rates negatively in many estimation results.

3.4   Response of capital level to taxes

Another key empirical finding is on capital allocation among competing states, and how it is affected by capital tax rates and population growth rates.

To show that competition over capital leads to the observed patterns of tax interactions among states, I estimate an equation relating tax base to tax rates in own state and neighboring states, as in Brett and Pinkse (2000). This is also vital in explaining the theoretical channel in Section 4.

I run one panel of 48 states in Midwest, South, West and Northeast with fixed

66

effects:

$$\log(k_{st}) = \alpha + \beta_1 \cdot OTR_{st} + \beta_2 \cdot (g_{st} \cdot OTR_{st}) + \lambda_1 \cdot TN_{st} + \lambda_2 \cdot (g_{st} \cdot TN_{st}) + \gamma \cdot TF_t + \mathbf{X}_{st} \cdot \tau + u_s + \epsilon_{st}$$

(7)

where k_{st} denotes capital per cap in state s at time t, OTR_{st}, TN_{st}, and TF_t are capital tax rates of own state, neighbors' average, and federal government, respectively. g_{st} is the population growth rate in state s at time t, and $\mathbf{X}$ is a row vector of exogenous explanatory variables, with macroeconomic and political variables previously defined. u_s is the fixed effect.

This estimates how own state capital level responds to changes in own state capital tax rates and neighbor states' capital tax rates. Moreover, the influence of population growth rates on the response of capital levels to tax rates is also examined. I first isolate the response to own tax rates, including the interacting term $g_{st} \cdot OTR_{st}$ and focusing on whether a faster population growth affects the response of capital allocation to changes in own state tax rates[10]. I continue to isolate the response to neighbors' tax rates, including the interacting term $g_{st} \cdot TN_{st}$ and testing whether faster population growth affects how much capital responding to changes in neighbors' tax rates[11]. Then I combine the responses to own state and neighbors' tax rates and include both interacting terms $g_{st} \cdot OTR_{st}$ and $g_{st} \cdot TN_{st}$ in the regression.

Both neighboring schemes are applied in the estimation.

Details about data source and variable definition of capital level are presented in

---

[10]Results for this part is presented in Appendix G.

[11]Results for this part is presented in Appendix G.

E.

The results are qualitatively identical and lead to the same conclusions whether the responses are isolated or not. Thus, I present below the result of combined responses to own state and neighbors' tax rates, and how the responses are influenced by population growth rates.

Dependent variables: log capital per cap

Specifications	Scheme 1			Scheme 2		
	II	III	IV	II	III	IV
Explanatory Variables						
OwnTax	-1.899***	-1.916***	-1.925***	-2.245***	-2.256***	-2.267***
	(0.188)	(0.188)	(0.188)	(0.191)	(0.191)	(0.190)
Growth×OwnTax	173.486***	172.769***	171.508***	231.332***	230.867***	228.910***
	(15.330)	(15.355)	(15.341)	(14.861)	(14.872)	(14.872)
TaxNeighbor	-1.639***	-1.685***	-1.710***	-1.575***	-1.585***	-1.613***
	(0.261)	(0.259)	(0.258)	(0.270)	(0.269)	(0.269)
Growth×TaxNeighbor	251.014***	251.030***	250.503***	176.163***	175.173***	173.904***
	(18.849)	(18.852)	(18.842)	(20.358)	(20.397)	(20.322)
TaxFed	0.108**	0.099*	0.090*	0.079	0.075	0.064
	(0.054)	(0.054)	(0.054)	(0.055)	(0.056)	(0.055)
log Personal Income	0.269***	0.270***	0.270***	0.266***	0.267***	0.267***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Democrat_House	-0.014			-0.007		
	(0.016)			(0.016)		
Democrat_Senate		0.005			0.004	
		(0.014)			(0.015)	
Political Dummy			0.005*			0.006**
			(0.003)			(0.003)
Missing_political	-0.044*	-0.032	-0.026	-0.048*	-0.040	-0.032
	(0.026)	(0.026)	(0.024)	(0.026)	(0.026)	(0.025)
Constant	7.273***	7.260***	7.260***	7.318***	7.311***	7.310***
	(0.022)	(0.026)	(0.020)	(0.023)	(0.023)	(0.021)

Note: These are least squares estimates with fixed-effect of the parameters in Eq. (7).

Robust standard errors in parentheses. ***$p<0.01$, **$p<0.05$, *$p<0.1$.

Table 16: Capital allocation regression

Note that $\beta_1 + \beta_2 \cdot g < 0$ and $\lambda_1 + \lambda_2 \cdot g > 0$ when evaluated at the sample mean of time-average population growth rate. This shows that $\frac{\partial k_i}{\partial t_i} < 0$ and $\frac{\partial k_i}{\partial t_j} > 0, i \neq j$, which means that own state capital level responds negatively to a change in its own tax rate, while positively to a change in its neighbors' average tax rate[12].

Moreover, the degree of how much tax rates can influence capital allocation depends on population growth rate. As $\beta_2 > 0$ and $\lambda_2 > 0$, a higher population growth rate reduces the magnitude of own tax rate's effect on own capital level, while increases the magnitude of neighbors' tax rates' effect on own capital level.

Federal tax rates' effect on capital allocation is significantly positive in most results but insignificant in some.

3.5   Industrial Neighbors

In almost all existing empirical literature[13], the existence of capital tax competition is examined among geographic neighbors in different areas all over the world. The neighboring schemes include contiguous jurisdictions, sharing markets with common boundaries, located with certain Euclidean distance apart, weighted by the inverse of the distance and so on. Overall, neighbors are defined according to their geographic locations.

However, instead of geographic neighbors only, jurisdictions may also compete

---

[12] Consistent with the finding in Buettner (2003), the impact of local tax rate has a negative effect on tax base, while the average tax rate of adjacent neighboring jurisdictions has a positive effect if interacted with population size.

[13] Brueckner (2003), Brueckner and Saavedra (2001), Brett and Pinkse (2000), Hayashi and Boadway (2001), and Chirinko and Wilson (2013).

with other "neighbors" who share similarities in certain aspects. For instance, the state of California may find itself competing with the state of New York in finance and real estate industry, probably in a more fierec manner than its contiguous neighbor Oregon whose major industry is durable goods manufacturing.

In this paper, capital tax competition among U.S. states is examined under different definitions of "neighbors". I utilize the panel dataset of 48 states[14] from 1958 to 2007. Geographicallly contiguous neighboring scheme is first examined as a benchmark. Then I continue to examine the existence and degree of capital tax competition among "industrial neighbors".

I use spatial autoregressive (SAR) panel estimation as follows:

$$Y = \lambda WY + X\beta + l_T \otimes u_n + \epsilon \qquad (8)$$

W is the weighting matrix which defines neighbors. The coefficient λ captures the strength of capital tax competition among states in the U.S.

3.5.1 Contiguous neighbors only, equally weighted, all 48 states

I first run estimation on geographic neighbors as a benchmark. For each state, I equally weigh its neighbors who share the same border. And the results are summarized in Table 17.

There exists significant capital tax competition among geographic neighbors, even though the competition is insignificant in Midwest and Northeast areas (Wang, 2016).

---

[14]Alaska and Hawaii are excluded.

And the overall degree of capital tax competition is close to those in South and West

areas.

$WTaxrate$					
	0.347***	0.285***	0.289***	0.314***	0.279***
	(0.024)	(0.025)	(0.024)	(0.024)	(0.025)
TaxFed	-0.029***	-0.025***	-0.032***	-0.035***	-0.032***
	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)
Personal Income	1.7e-8***	1.1e-08***	-7.2e-09**	1.4e-08***	-7.4e-09**
	(3.0e-09)	(3.0e-09)	(3.0e-09)	(3.0e-09)	(3.0e-09)
Democrat_House		0.031***			0.007*
		(0.003)			(0.004)
Democrat_Senate			0.034***		0.026***
			(0.003)		(0.004)
Political Dummy				0.005***	0.001
				(5.4e-04)	(0.001)
Missing_Political		0.032***	0.035***	0.022***	0.036***
		(0.005)	(0.005)	(0.005)	(0.005)

Note: These are SAR Panel estimates with fixed-effect of the parameters in Eq. (1).

Robust standard errors in parentheses. ***p<0.01, **p<0.05, *p<0.1.

Table 17: Tax competition regressions, SAR Panel with spatial fixed effect-contiguous equal

3.5.2 Same leading industry, equally weighted, all 48 states

For the second neighboring scheme, I define neighbors as "industrial neighbors", who share the same leading industry. For each state, I equally weigh its "industrial neighbors".

I obtain the series of GDP data by state and by industry from Bureau of Economic Analysis. The series of data is from 1963 to 2007. I caluculate the portion of six major industries out of total state GDP for each year: Agriculture, forestry and fishing; Mining; Durable goods manufacturing; Nondurable goods manufacturing; Finance, insurance and real estate; Services. Then the average portion of each industry over the time span 1963 to 2007 is calculated for each state, and the industry with the highest average portion is the leading industry for each state.

Estimation results are summarized in Table 18. There exists significant capital tax competition among "industrial neighbors", and strength of competition is lower than that among contiguous neighbors.

$WTaxrate$					
	0.284***	0.230***	0.244***	0.268***	0.230***
	(0.030)	(0.030)	(0.030)	(0.030)	(0.030)
TaxFed	-0.035***	-0.029***	-0.036***	-0.041***	-0.036***
	(0.010)	(0.009)	(0.009)	(0.009)	(0.009)
Personal Income	-2.0e-8***	-1.2e-08***	-8.0e-09***	-1.6e-08***	-8.1e-09***
	(3.1e-09)	(3.1e-09)	(3.1e-09)	(3.1e-09)	(3.1e-09)
Democrat_House		0.036***			0.011**
		(0.003)			(0.004)
Democrat_Senate			0.039***		0.028***
			(0.003)		(0.004)
Political Dummy				0.006***	0.001
				(5.5e-04)	(0.001)
Missing_Political		0.035***	0.039***	0.024***	0.040***
		(0.005)	(0.005)	(0.005)	(0.005)

Note: These are SAR Panel estimates with fixed-effect of the parameters in Eq. (1).

Robust standard errors in parentheses. ***$p<0.01$, **$p<0.05$, *$p<0.1$.

Table 18: Tax competition regressions, SAR Panel with spatial fixed effect-industrial equal

3.5.3 Same leading industry, weighted by inverse distance, all 48 states

I continue to apply industrial neighboring scheme. Instead of assigning equal weights to each "industrial neighbor", I weigh them by the inverse distance between pairs of states, which automatically assigning higher weights to neighbors located closer.

Estimation results are shown in Table 19. With distance adjusted weighting scheme, capital tax competition is slightly stronger than that with equal weights in Section 3.5.2.

$WTaxrate$					
	0.302***	0.232***	0.236***	0.273***	0.225***
	(0.026)	(0.027)	(0.027)	(0.027)	(0.027)
TaxFed	-0.036***	-0.031***	-0.038***	-0.042***	-0.038***
	(0.010)	(0.009)	(0.009)	(0.009)	(0.009)
Personal Income	-2.0e-8***	-1.2e-08***	-8.9e-09***	-1.6e-08***	-9.0e-09***
	(3.1e-09)	(3.1e-09)	(3.1e-09)	(3.0e-09)	(3.1e-09)
Democrat_House		0.034***			0.010**
		(0.003)			(0.004)
Democrat_Senate			0.037***		0.026***
			(0.003)		(0.004)
Political Dummy				0.006***	0.001*
				(5.5e-04)	(0.001)
Missing_Political		0.035***	0.038***	0.024***	0.040***
		(0.005)	(0.005)	(0.005)	(0.005)

Note: These are SAR Panel estimates with fixed-effect of the parameters in Eq. (1).

Robust standard errors in parentheses. ***p<0.01, **p<0.05, *p<0.1.

Table 19: Tax competition regressions, SAR Panel with spatial fixed effect-industrial distance weighted

3.5.4 Contiguous neighbors AND same leading industry, equally weighted, all 48 states

In this neighboring scheme, I combine the concepts of both geographic and industrial neighbors. For each state, I assign weight to a neighbor only if the neighbor both shares contiguous border and the same leading industry.

Estimation results are provided in Table 20. There exists capital tax competition, but the degree of tax competition is smaller than other neighboring schemes.

$WTaxrate$	0.182***	0.141***	0.136***	0.163***	0.130***
	(0.020)	(0.020)	(0.020)	(0.020)	(0.020)
TaxFed	-0.044***	-0.037***	-0.045***	-0.050***	-0.044***
	(0.010)	(0.009)	(0.009)	(0.009)	(0.009)
Personal Income	-2.1e-8***	-1.3e-08***	-9.6e-09***	-1.7e-08***	-9.5e-09***
	(3.1e-09)	(3.1e-09)	(3.1e-09)	(3.1e-09)	(3.1e-09)
Democrat_House	0.036***	0.036***			0.012***
	(0.003)	(0.003)			(0.004)
Democrat_Senate			0.037***		0.026***
			(0.003)		(0.004)
Political Dummy				0.006***	0.001
				(5.5e-04)	(0.001)
Missing_Political	0.036***	0.036***	0.039***	0.025***	0.041***
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)

Note: These are SAR Panel estimates with fixed-effect of the parameters in Eq. (1).

Robust standard errors in parentheses. ***p<0.01, **p<0.05, *p<0.1.

Table 20: Tax competition regressions, SAR Panel with spatial fixed effect-contiguous industrial equal

4   Benchmark Model

4.1   Tax Competition

There are two periods in the model. A nation is divided into two states, each of which is populated by a large number of identical residents in each period. Labor is immobile and grows at the same rate g in each state. Capital is perfectly mobile between states. Using labor and capital in the same production function, a single homogeneous good is produced in each state.

Each household in both states is endowed with income e in the first period, and saves for period 2. In the second period, each household earns labor income and receives the return from saving. Denote K_i, L_i as the aggregate level of capital and labor located in state i at period 2, $i = 1, 2$. The production function $F(K_i, L_i)$ has constant returns to scale, is concave in both inputs and twice continuously differentiable. The production function can be written in intensive form $f(k_i)$, where k_i is capital per worker at period 2.

Normalizing the price of the private good to one. Capital is taxed in each state with the unit tax rate t_i, $i = 1, 2$. Due to mobility of capital, net-of-tax returns are equalized between jurisdictions:

$$f'(k_1) - t_1 = f'(k_2) - t_2 = \rho \tag{9}$$

where ρ denotes this uniform net return. These non-arbitrage conditions define the demand for capital in each state $k_i = k(\rho + t_i)$, with $k'(\rho + t_i) = \frac{1}{f''(k_i)} < 0$.

Residents in each state get utility from consuming both private goods and public goods in two periods, with total utility $u_i(x_i^1, z_i^1) + \beta u_i(x_i^2, z_i^2)$ where β is the discount factor, x_i^t and z_i^t are levels of private goods and public goods consumed by residents in state i at period t.

All tax revenue collected by the government in each state is spent on public goods. As capital is accumulated only in the second period, government provides public goods only at period 2. This public good can be either excludable $(z_i = t_i k_i)$ or non-excludable shared by all $(z_i = t_i k_i L_i)$.

Households choose saving s to maximize

$$u_i(e - s_i) + \beta v_i((1 + \rho)s_i, z_i) \tag{10}$$

Following Keen and Kotsogiannis (2002), the representative household acts as both worker and investor, and utility function is assumed to be

$$u(e - s_i) + f(k_i) - k_i \cdot f'(k_i) + (1 + \rho)s_i + \Gamma(z_i) \tag{11}$$

Assume utility functions are identical in two states.

First-order condition describes saving behavior $s_i(\rho, t_1, t_2)$, $i = 1, 2$, where $u'(e - s_i) = (1 + \rho)$. Assume saving only depends on net return $s(\rho)$ with $s'(\rho) \geq 0$.

Suppose states start with same population of labor L in the first period, and given that the growth rate of population is g for both states, the following market-clearing

condition holds:

$$(1+g)\sum k_i = \sum s(\rho) \tag{12}$$

So,

$$\frac{\partial \rho}{\partial t_i} = \frac{(1+g)k'(\rho + t_i)}{\sum s'(\rho) - (1+g)\sum k'(\rho + t_i)} \tag{13}$$

Assuming $f''(k_i) = \gamma < 0^{15}$, then

$$\frac{\partial \rho}{\partial t_i} = -\frac{1}{2 + \frac{(-\gamma)\sum s'(\rho)}{1+g}} \in (-\frac{1}{2}, 0) \tag{14}$$

Compared to one-period model where total capital is fixed[16], a one unit change in tax rate affects net return by less with saving in this two-period model.

There are two effects associated with tax change in this model: capital reallocation effect and saving effect. As one state cuts tax, more capital inflow is attracted. In addition, the return of investing in that state is higher which stimulates more saving nationwide. This saving effect drives up total capital, and reduces $f'(k_i)$, thus the net effect on $\rho = f'(k_i) - t_i$ is less since change in tax not only reallocates capital between states but also affects total saving.

When $s'(\rho) = 0$, $\frac{\partial \rho}{\partial t_i} = -\frac{1}{2}$, same as the result when total capital is exogenously fixed, since saving is independent of ρ and there is capital reallocation effect only.

---

[15] A standard assumption on production function with one example being quadratic production form, which is also assumed in Brueckner and Saavedra (2001).

[16] The results obtained in Hoyt (1989), Bucovetsky (1991) show that $\frac{\partial \rho}{\partial t_i} = -\frac{1}{N}$ where N is the number of total states. Only reallocation effect exists.

With a higher population growth rate, a certain amount of increased saving needs to be shared with more people, which is known as "dilution effect". Thus, each household has a lower increase in k, leading to a smaller drop in $f'(k_i)$ and a bigger effect on ρ.

Lemma 1 *The magnitude of $\frac{\partial \rho}{\partial t_i}$, $\left| \frac{\partial \rho}{\partial t_i} \right|$, is positively dependent on g. Given a same amount of tax cut, net return on capital increases more in the states with a higher population growth rate.*

In the extreme case where $g \to +\infty$, $\frac{\partial \rho}{\partial t_i} = -\frac{1}{2}$, same result as when $s'(\rho) = 0$. Only allocation effect remains when population grows too rapidly. Given any amount of total saving, capital is thinly spread out and each resident gets an insignificant share, the change in tax rate only affects the allocation of capital between the two states. To summarize, a higher g reduces saving effect.

Utilizing equations (8) and (13),

$$\frac{\partial k_i}{\partial t_i} = \frac{1}{\gamma} \cdot \frac{1 + \frac{(-\gamma)\sum s'(\rho)}{1+g}}{2 + \frac{(-\gamma)\sum s'(\rho)}{1+g}} < 0 \tag{15}$$

$$\frac{\partial k_i}{\partial t_j} = \frac{1}{-\gamma} \cdot \frac{1}{2 + \frac{(-\gamma)\sum s'(\rho)}{1+g}} > 0, i \neq j \tag{16}$$

Different from one-period model[17], there is asymmetric effect on capital from own tax cut and neighbor's tax cut, where $\left| \frac{\partial k_i}{\partial t_i} \right| \geq \frac{\partial k_i}{\partial t_j}$. And if $s'(\rho) > 0$, $\left| \frac{\partial k_i}{\partial t_i} \right| > \frac{\partial k_i}{\partial t_j}$. A change in own tax rate affects capital by more in magnitude than a neighbor's cut

---

[17]In Hoyt (1989), for instance, $\frac{\partial k_i}{\partial t_i} = -\frac{\partial k_i}{\partial t_j}$.

tax.

There are two effects associated with a tax cut in a state: reallocation effect transferring capital from the state to the tax-cut state; and saving effect which increases total capital stock nationwide. Obviously, reallocation effect increases k in one state (the state which initiates a tax cut), and reduces k in the other state by the same amount if total capital is fixed, which is the result obtained in one-period model. The saving effect, however, increases k in both states, leading to a further increase in k of the tax-cut state, and compensating some loss in k of the other state. Therefore, $\left|\frac{\partial k_i}{\partial t_i}\right|$ is higher than $\left|\frac{\partial k_i}{\partial t_j}\right|$ with saving in the model.

Lemma 2 $\left|\frac{\partial k_i}{\partial t_i}\right|$ *is negatively related to g, while* $\left|\frac{\partial k_i}{\partial t_j}\right|$ *is positively related to g.*

As a higher g results in a bigger increase in ρ given the same amount of tax cut (Lemma 1), k increases by less in the state which initiates the tax cut, as more people have to share the total capital. Similarly, as the new equilibrium net return ends up at a higher value, k in the other state drops by more. This is due to the increased total saving has to be shared by more, so each gets compensated by less.

This theoretical result is consistent with the empirical finding in Section 3, and the dilution effect is verified both empirically and theoretically.

Each state government plays Nash with its neighboring state. Starting with the case of excludable public goods, government is benevolent and chooses its own capital tax rate t_1 to maximize aggregate utility in two periods.

Each government solves:

$$\max_{t_i} \; u(e - s_i(\rho(t_1, t_2))) + f(k_i(t_1, t_2)) - k_i(t_1, t_2) \cdot f'(k_i(t_1, t_2))$$

$$+(1 + \rho(t_1, t_2))s_i(\rho(t_1, t_2)) + \Gamma(t_i \cdot k_i(t_1, t_2)) \tag{17}$$

Taking FOC,

$$s_i \cdot \frac{\partial \rho}{\partial t_i} - f''(k_i) \cdot k_i \frac{\partial k_i}{\partial t_i} + \Gamma_z \cdot (k_i + t_i \frac{\partial k_i}{\partial t_i}) = 0 \tag{18}$$

Suppose $\Gamma_z = \eta > 0$[18] and $s'(\rho)$ is not a function of ρ. Then utilizing equations (13),

(14) and (15), the response action is:

$$\frac{\partial t_1}{\partial t_2} = -1 + \frac{\eta}{\gamma} \cdot \frac{2 - A + \frac{A-1}{\eta}}{A^2 s'(\rho) + \frac{2\eta}{\gamma} + \frac{2\eta}{-\gamma}A - \frac{(1-A)^2}{\gamma}} \tag{19}$$

where $A = \frac{1}{2 + \frac{(-\gamma)\sum s'(\rho)}{1+g}} \; \epsilon \; (0, \frac{1}{2})$ and $\frac{\partial A}{\partial g} > 0$.

Proposition 1 *As long as the value on public goods is high enough, i.e. $\eta > \underline{\eta}$, there exists tax competition where $\frac{\partial t_1}{\partial t_2} > 0$.*

From equation (17), first-order condition $\eta \cdot k_i = \eta \cdot (-t_i \frac{\partial k_i}{\partial t_i}) + s \cdot (-\frac{\partial \rho}{\partial t_i}) + \gamma \cdot k_i \frac{\partial k_i}{\partial t_i}$

implies that given neighbor's tax rate t_2, own tax rate is chosen by equalizing the cost

of a tax cut and the benefit of a tax cut. The cost of a tax cut is the loss in utility

---

[18] As is assumed in Brueckner and Saavedra (2001).

from public goods, as the government collects less revenue from each unit of capital, while the benefit combines an increased tax base from capital inflow contributing to higher utility from public goods, a higher return from saving made by households, with an extra benefit of a tax cut, increasing wage income by attracting more capital to production.

The threshold value is $\underline{\eta} = \max(1, \frac{A^2 s'(\rho)(-\gamma)}{2(1-A)} + \frac{1-A}{2})$. As $s'(\rho)$ increases, the value governments impose on public goods needs to be higher to initiate tax competition. From cost-benefit analysis, when neighbor cuts tax, there is reallocation effect resulting in capital outflow. Hence, at the previously chosen tax rate, the cost of a tax cut drops with a lower level of capital base, which means the potential revenue loss from cutting tax reduces. On the other hand, whenever $s'(\rho) > 0$, saving nationwide increases after a tax cut in neighboring state, which increases the benefit of cutting own tax as consumption increases with a higher return from savings. The benefit from higher wage income, however, decreases with a lower capital base, as wage positively depends on capital level. And the benefit from higher public goods remains unchanged. And own state should compete with neighbor by cutting own tax rate, as long as reducing tax brings net benefit. As the net change in benefit is ambiguous, it can be negative, and if η is too small, the drop in benefit might even exceed that in cost, leading to cost of tax cut higher than its benefit, and own tax rate is raised as a result of neighboring tax cut.

Moreover, marginal utility (MU) of a tax increase equals $\eta \cdot k_i - \eta \cdot (-t_i \frac{\partial k_i}{\partial t_i}) - s \cdot$

$(-\frac{\partial \rho}{\partial t_i}) - \gamma \cdot k_i \frac{\partial k_i}{\partial t_i}$, and $\frac{\partial MU}{\partial t} = \frac{2\eta(1-A)-(1-A)^2}{\gamma} + A^2 s'(\rho)$. Interior solution is attained

whenever $\frac{\partial MU}{\partial t} < 0$. In the case of $s'(\rho) > 0$, however, $\frac{\partial MU}{\partial t}$ can be positive if η is

relatively small compared to the value of $s'(\rho)$, then utility function is convex with

corner solution. The intuition is when the government values little on public goods,

the consumption from saving as well as wage income is valued more. The government

tends to reduce the tax rate to the bottom such that the return gained from saving

increases to the utmost, without much loss in public goods. Thus, only the case where

$\eta > \underline{\eta}$ is considered, so that the interior solution is obtained.

As shown from the results in Section II, there exists tax competition in areas in the

U.S., which implies that the value on public goods by the government is sufficiently

high.

Furthermore, from the equation of $\underline{\eta}$, this threshold value increases with popula-

tion growth rate g when $s'(\rho) > 0$, i.e. the value on public goods needs to be higher

to induce tax competition. From Lemma 1, a higher population growth rate leads to

a stronger effect on net return from any tax change, implying that cutting tax brings

higher return on saving. Combined with Lemma 2, a higher population growth rate

results in a smaller effect on own capital level after own tax change, implying that the

degree of capital inflow is lessened even with tax cut, leading to a lower level of po-

tential gain in public goods provision. Unless the value on public goods is sufficiently

high, $\frac{\partial MU}{\partial t}$ is positive. Otherwise, states are better off benefiting from a higher saving

return from neighbors' cutting tax.

Proposition 2 *A higher population growth rate g leads to stronger tax competition whenever there exists tax competition, i.e.* $\frac{\partial(\frac{\partial t_1}{\partial t_2})}{\partial g} > 0$ *whenever* $\eta > \underline{\eta}$.

Applying equation (17) again, after neighbor's cutting tax rate, there is capital outflow leading to a reduction in the cost of own tax cut. Moreover, if $s'(\rho) > 0$, saving increases after a tax cut, leading to an increase in the benefit of own tax cut on the right-hand side. Since benefit exceeds cost, own state needs to cut tax rate.

From Lemma 1 and Lemma 2, a higher g leads to a stronger response in both net return and reduction in capital after neighbor's cutting tax, widening the gap between the benefit and the cost of tax cut. In addition, a higher population growth rate results in a smaller increase on own capital level after own tax cut. Thus, states compete in a more fierce method.

As a higher population growth rate reduces saving effect, residents obtain only a smaller share of increased saving. Hence, with a smaller "pie" for each resident, governments compete more strongly for the mobile capital.

4.2   Social Planner

One question is whether inefficiency arises from tax competition, and how population growth rate g affects the magnitude of the inefficiency.

Consider a social planner's problem maximizing the total welfare of two states'

residents:

$$\max_{t} \ u(e - s(\rho(t))) + f(k(t)) - k(t) \cdot f'(k(t)) + (1 + \rho(t))s(\rho(t)) + \Gamma(t \cdot k(t)) \quad (20)$$

and a coordinated change in both states' tax rates affects net return by:

$$\frac{\partial \rho}{\partial t} = \frac{\partial \rho}{\partial t_i} \cdot 2 = -\frac{1}{1 + \frac{(-\gamma)s'(\rho)}{1+g}} \in (-1, 0) \quad (21)$$

The effect of a coordinated increase in both state taxes gives $s \cdot \frac{\partial \rho}{\partial t} + \Gamma_z \cdot (k + t\frac{\partial k}{\partial t}) - kf''(k)\frac{\partial k}{\partial t}$, comparing with the result of corresponding symmetric equilibrium by substracting equation (17) from it: $s \cdot (\frac{\partial \rho}{\partial t} - \frac{\partial \rho}{\partial t_i}) + (\Gamma_z \cdot t - kf''(k)) \cdot (\frac{\partial k}{\partial t} - \frac{\partial k_i}{\partial t_i}) = (s + \frac{\Gamma_z \cdot t}{\gamma} - k) \cdot (\frac{\partial \rho}{\partial t} - \frac{\partial \rho}{\partial t_i})$. And from equation (20), $\frac{\partial \rho}{\partial t} - \frac{\partial \rho}{\partial t_i} = -\frac{1}{2 + \frac{2(-\gamma)s'(\rho)}{1+g}} < 0$.

Proposition 3 *The tax competition allocation is efficient if and only if $s + \frac{\Gamma_z \cdot t}{\gamma} - k = 0$. Tax rates from uncoordinated tax setting are too high if $s + \frac{\Gamma_z \cdot t}{\gamma} - k > 0$; and tax rates from competition are too low if $s + \frac{\Gamma_z \cdot t}{\gamma} - k < 0$.*

Consistent with Keen and Kotsogiannis (2004), whether uncoordinated chosen tax rates in a free market are too high or too low depends on the elasticities of the demand for capital and the value on public goods. When $|\gamma|$ is too small, marginal productivity is less sensitive to change in k, implying that after neighbor cutting tax, capital outflow would be more significant. Together with a higher value on public goods, own state has to fight much strongly in order to provide public goods, resulting in efficiency loss from too much competition.

89

Proposition 4 *The degree of inefficiency is higher when the population growth rate g is higher.*

Whether inefficiency arises from a tax rate that is too high or too low than socially optimal, a higher population growth g widens the gap between the effect on net return from a competitive and a coordinated tax cut.

When value on public goods is not large enough to initiate competition, states take advantage of higher net return from neighbors' lower tax and do not lower taxes accordingly. Each state ignore the positive externality it confers on its neighbors' net return by cutting tax, and as a higher population growth rate g magnifies the effect on net return from one unit tax cut, the loss in efficiency is bigger.

Whenever there exists tax competition, while competing over capital pool to provide public goods, each state ignores the negative externality imposed on its neighbors' capital level. As a higher population growth increases this externality, neighbors are forced to fight stronger.

5 Conclusion

The empirical contribution of this paper is to first quantify the degree of tax competition among states in the US, applying MLE estimation of the SAR panel data model with fixed-effects. Another empirical finding is that states in the South and West compete in setting capital tax rate much more strongly than states in Midwest and Northeast. One explanation, which is empirically tested in this paper applying

a high-order SAR panel data estimation with fixed-effects, is that population growth rates are much higher in the South and West than the growth rates in the Midwest and Northeast.

The supporting related empirical finding is that capital allocation is affected by tax rates in own state and neighboring states. Amount of capital inflow to own state negatively depends on own tax rate while positively depends on neighbors' tax rates. Moreover, the magnitude of tax rates' effect on capital allocation significantly depends on population growth rates. This verifies the "capital dilution" effect among the states in the US.

Capital tax competition not only exists among geographically contiguous states in the U.S., but also among states who share the same leading industries. Although states may not share a border, but they still compete with states who share the same leading industry by offering tax incentives to attract firms' investment.

Distance-adjusted industrial neighbors' competition is slightly stronger than that of equally weighted industrial neighbors. It may suggest that distance is one factor for firms to make investment decision, provided states have the same leading industry.

A model with intertemporal saving decision can account for these empirical facts. Different from most tax competition literature, the pool of total capital that states compete over is not fixed. Whenever one state's tax cutting increases the net return of saving, households save more for the second period. A high population growth rate increases the gap between the two periods' population, people who save and

people who share the savings. Thus, faster population growth dilutes the increase in capital by more, leading to a lower increase in capital per cap for the tax-cut state and a bigger loss in capital per cap in its neighborhood. The same unit tax cut brings benefit by less and cost by more, leading to states cutting tax more fiercely and stronger strategic interaction among each other.

Regarding social efficiency, a higher population growth rate leads to greater inefficiency cost. The policy implication is whenever tax competition is observed, it is of more importance to regulate those states experiencing faster population growth.

Chapter 3: Foreign Direct Investment Cycles and Intellectual Property Rights in Developing Countries

1 Introduction

There has been an ongoing debate as to whether developing countries should strengthen their intellectual property rights (IPR). Opponents argue that strong IPR regimes reduce consumer wlefare by prolonging innovators' monopoly power, and slow down the technology progress of developing countries by discouraging imitation. Proponents counteract with the argument that strong IPR regimes encourage foreign direct investment (FDI) from developed countries (North) to developing countries (South), which benefits the South.

The existence of FDI cycle means that firms in developed countries will not always do FDI in the developing countries.[1] Anecdotal evidence suggests that FDI cycle exists in some industries. For instance, Volkswagen Passat B2 was introduced in Europe in 1981. Its variant Santana has been produced in China since 1986, and

---

[1] Consider HD TV as an example, and suppose the North keeps innovating new generations of TVs with higher resolutions. An FDI cycle length of 3 means that the current technology leaders in the North will do FDI in every 3 generations, say generations of 3, 6, 9 and so forth. And the current technology leaders of other generations will not do FDI.

another variant Quantum was produced in Brazil from 1985 to 2002. In the late 1980s, new generations of Passat, B3 and B4, were introduced in Europe (1988) and North American (1990). But they were never produced in China, and Volkswagen started to produce them in South America only after 1995. The newer generation of Passat, B5, was introduced in 1996, shortly after it was produced in China.[1]

In this model, innovations only occur in the North, and the rate of innovation is exogenously given. There is no international trade, which means that in order to sell in the South, North firms have to do FDI. FDI is assumed to be costly: each FDI has to incur some fixed (sunk) cost. There are a fixed number of firms in the South, who are active in imitation. The South firms can imitate the products of FDI, but they are not able to imitate the products of the North firms who only produce in the North. Imitation is costly and imitation intensity is endogenous. FDI and South firms engage in Bertrand price competition. IPR strength is modelled as the probability that a successful imitation is ruled illegal. The timing is that the South government chooses and commits to an IPR strength in the very beginning.

In equilibrium FDI occurs cyclically, due to the fact that FDI entails a fixed cost. Due to Bertrand competition, the price that a new FDI charges increases in the technology gap between the current leaders in the North and in the South. If the technology gap is not big enough, then new FDI by the current leader in the North will not be profitable, as it can only charge a lower price and is not able to cover

---

[1]Source: http://en.wikipedia.org/wiki/Volkswagen_Passat
http://en.wikipedia.org/wiki/Volkswagen_Santana
http://en.wikipedia.org/wiki/Shanghai_Volkswagen_Automotive

its fixed cost. Every new generation of FDI faces the threat of imitation. Once a South firm successfully imitates and is ruled legal, that firm replaces the FDI serving the South market as South firms have a lower production cost. The equilibrium intensity of imitation decreases over a FDI cycle. As the technology frontier of the North advances, the next FDI will render the current FDI product profitless and become more imminent. As a result, the expected duration of monopoly of successful imitation decreases, and South firms' incentive to imitate decreases as well.

With a higher South IPR strength, the equilibrium FDI cycle length decreases or FDI occurs more frequently, as it reduces the prospect of successful imitation thus reduces the equilibrium imitation intensity. This increases the profitability of FDI and reduces the equilibrium cycle length.

This paper continues to study the optimal IPR strength that maximizes the discounted social welfare of the South. Two effects are identified. The first effect is the "free upgrade" effect. Across cycles, South consumers pays the same price for higher quality products. A smaller FDI cycle length brings more frequent upgrades to consumers. The second effect is the "imitation effect." Within each cycle, successful imitation always increases South welfare as it reduces the price consumers pay and brings profit to South firms. Thus, given the cycle length, the South government tends to induce the highest possible aggregate imitation intensity. Free upgrade effect implies that the IPR strength should be high in order to reduce the cycle length. However, imitation effect implies that the IPR strength should be set as low as possi-

ble given any cycle length. The optimal IPR strength balances these two effects, and they imply that the optimal IPR strength will not be the zero strength or the full strength. Essentially, the imitation effect captures the short-run (within a cycle) benefit of lowering IPR strength as it encourages imitation, while the free upgrade effect reflects the long-run (across cycles) cost as it makes FDI less frequent and enlarges the technology gap between the North and South.

The rest of the paper is organized as follows. Section 2 sets up the model. Section 3 characterizes the equilibrium FDI cycle and investigates how the cycle length depends on various factors. Section 4 studies the optimal IPR strength that maximizes the discounted social welfare, and how it is affected by the market size, imitation costs and the number of firms in the South. Concluding remarks are offered in Section 5. All the proofs can be found in the Appendix H.

2   The Model

There's an economy with two countries: South and North. There is a single industry in both countries, which produces a single good. The good can be of different qualities on a quality ladder, a la Grossman and Helpman (1991a). Denote the quality level of the good as $q_j \in \{0, 1, 2, 3, ...\}$. The quality improvement between any two adjacent quality levels is $\lambda > 1$. That is, if a consumer consumes a good of quality q_j, he derives a gross utility of λ^{q_j}. In each country, there exists a maximum technologically feasible quality level, which can be improved through either innovation

or imitation. Time is continuous, with each agent's discount rate being ρ.

Two countries differ in their innovating abilities. Only the North innovates, while the South only imitates. The industry in the North is perfectly competitive. Denote the aggregate R&D intensity in the North as ι, which is exogenously given. That is, at each instant with a Possion arrival rate ι some North firm(s) improves the maximum quality level by one step. Note that North firms have to climb up the quality ladder step by step (no skipping). The quality level of the leading North firm is publicly observable. North firms might sell in the South market. There is no international trade. In order to sell in the South market, North firms have to go through foreign direct investment (FDI).[2] Each FDI requires a sunk cost $F > 0$, and the marginal cost of production for each FDI in the South (regardless of quality) is $\xi \in (1, \lambda)$.

In the South, the industry has $N \geq 1$ firms. The South firms can improve quality level only through imitating products sold by the North FDI firm in the South, and they are not able to imitate the products produced only in the North.[3] This means that FDI is the only channel of international technology transfer, and without it the technology progress in the South will be stagnant. Denote the quality level of the up-to-date FDI as q_{FDI}, and the technology level of the leading South firm(s) as q_S. Let $\Delta \equiv q_{FDI} - q_S \in \{0, 1, 2, ...\}$ be the step difference between q_{FDI} and q_S. By

---

[2]An alternative setting is that there is international trade, but the South imposes tariff. Now the gross benefit of FDI is the save of tariff, instead of the access of the South market in the basic model. Our main results still hold qualitatively in this alternative setting.

[3]FDI creates local knowledge spillovers, hence it is much easier for South firms to imitate. Glass and Saggi (2002) make a similar but weaker assumption: it is less costly to imitate an FDI's product than to imitate a North firm's product.

imitation, q_S can jump to the level of q_{FDI} directly, and this is the only source of technology progress in the South. Let μ_{Ij} be South firm j's imitation intensity. With a Poisson arrival rate μ_{Ij} South firm j will successfully imitate the product of the FDI firm at each instant. Immitation is costly. By choosing imitation intensity μ_{Ij}, at each instant dt firm j incurs a cost of $a_I(\Delta)\mu_{Ij}^2 dt$.[4] The cost $a_I(\Delta)$ is (weakly) increasing in technogical difference Δ, which captures the fact that it is more difficult to imitate more advanced technology.

There are a mass of M_S consumers in the South market. The marginal cost of production of any South firm is (normalized to) 1 regardless of the quality level. Note that South firms have advantages in production relative to North FDI firms, $\xi > 1$, which reflects the fact that FDI firms are operating in an unfamiliar environment. In the product market, firms compete in prices.

When a North firm with a quality level more advanced than any operating firms in the South does FDI, it acquires a patent from the South government. The patent potentially prevents the South firms from imitating products of the FDI. The patent law and enforcement in the South, however, is not perfect. In particular, the IPR strength in the South is captured by one parameter $p \in [0, 1]$. If a South firm successfully imitates the FDI's product, with probability p the imitating product is ruled illegal and cannot be sold in the market, and with probability $1 - p$ it is ruled

---

[4]The quadratic cost function is not essential to our qualitative results, but it does simplify our computation. Any increasing and convex cost function would work, at the expense of more complicated algebra.

legal and sold in the market.[5] The probability p is i.i.d. across firms and across different tries of the same firm. In case that a South firm's successful imitation is ruled illegal, the successfully imitated product is discarded, and this firm has to start imitation from scratch.[6] This assumption ensures that all South firms are symmetric in terms of immitating, regardless whether a firm has successfully come up with imitations but rule illegal beforehand.

The timing is as follows. First, the South government chooses the IPR strength p at time 0 and commits to it afterwards. Then, after observing p, all firms play their parts over time. At time 0, the North and the South both are at the lowest quality ladder $q_j = 0$.

3   Equilibrium FDI Cycles

3.1   Preliminary Analysis

In any instant of time, in the South either a single firm or two firms have the leading technology (will be verified later). If a single firm has the leading technology, then it must be an FDI firm. If a South firm sucessfully imitates and is ruled legal, then the FDI and the South firm both have the leading technology. Suppose only an FDI has the leading technology. In this case, there is no successful imitation product

---

[5]Our modelling of the IPR strength is rather abstract. For the details of patent law, eg. patent leading breadth, lagging breadth, and patentibility, see O'Donoghue (1998).

[6]One can think that each successful imitation produces a slightly different version. If one version happens to be ruled illegal, then this version will never be ruled legal in future tries.

on the market, thus the technology gap between the FDI and South firms is $\overline{\Delta}$. Since firms are engaging in Bertrand competition and $\xi < \lambda$, the FDI is a monopoly in the South market, and it charges a "limit" price $\lambda^{\overline{\Delta}}$. Now suppose one South firm successfully imitates the FDI's product and is ruled legal.[7] In this case, two firms' products have the same quality. But since the South firm has cost advantage, the South firm becomes a monopoly in the South market and charges a "limit" price ξ. Once a South firm successfully imitates and is ruled legal, all other South firms will stop imitating. This is because even if another South firm successfully imitates and is ruled legal, due to Bertrand competition with the existing South firm, it would have earned a zero profit as two firms have the same quality level and the same production cost.

Let Δ_{NS} be the quality difference between the leading North firm's quality level and the leading technology in the South (including the most recent FDI).[8] Recall that the North industry is perfectly competitive. This means that on the quality ladder different North firms will be the leading firm at different times. Moreover, only one firm will be the leading firm in the North. The leading firm in the North has an incentive to do FDI in the South market, in order to earn extra profit. But due to the sunk cost F, it might not be profitable due to the imitation threat of the South firms and the future FDI of the future North leading firms, which makes the current

---

[7]At each instant, the probability that more than one South firms come up with successful imitation is negligible, relative to the probability that a single South firm successfully imitates.

[8]The leading technology in the South is just the quality level of the most recent FDI. This is because successful imitation will not advance the quality level in the South beyond that of the most recent FDI.

FDI obsolete. Therefore, new FDI will occur only if Δ_{NS} reaches some threshold level. Denote the expected gross discounted payoff of a new FDI as $V_F(\Delta_{NS})$. $\overline{\Delta}$ is the smallest Δ_{NS} such that $V_F(\Delta_{NS}) \geq F$. This is because once Δ_{NS} reaches $\overline{\Delta}$, new FDI becomes profitable and the current leading firm in the North will immediately carry out FDI. On the other hand, if the North-South technology gap is smaller than $\overline{\Delta}$, the current leading firm in the North will not carry out FDI as it is not profitable by the definition of $\overline{\Delta}$. Therefore, FDI must occur periodically or cyclically in equilibrium, with the cycle length being $\overline{\Delta}$.

To summarize, FDI occurs cyclically. Once the technology gap between North and South reaches $\overline{\Delta}$, the leading North firm at that moment immediately does FDI in the South, which starts a new cycle. At that moment, the technology gap between North and South becomes 0. Within a cycle, the most recent FDI first has the monopoly in the South market, and South firms try to imitate the product of the FDI. Once imitation is successful and ruled legal, a South firm replaces the FDI as the monopoly of the South market. At the same time, the technology frontier in the North is advacing stochastically, which means that the technology gap between North and South is widening. Once the technology gap between North and South reaches $\overline{\Delta}$ again, the leading North firm at that moment immediately does FDI, which ends the current cycle (regardless whether imitation has succeeded or not) and starts a new cycle.

3.2 Imitation

In this subsection, South firms' incentive to imitate is investigated. Recall that Δ_{NS} is the quality difference between the leading North firm's quality level and the leading technology in the South. In particular, $\Delta_{NS} = \{0, 1, ..., \overline{\Delta}-1\}$, which indicates the phases of the cycle. Note that Δ_{NS} affects South firms' incentive to imitate. This is because a bigger Δ_{NS} implying a shorter remaining length of the current cycle, as new FDI will arrive sooner in expectation, which will render successful imitation obsolete.

Denote $V_I(i, \overline{\Delta})$ as the discounted expected payoff of a successful imitator in the current cycle given $\Delta_{NS} = i$ and the cycle length being $\overline{\Delta}$, and $V_F(i, \overline{\Delta})$ as the discounted payoff of the most recent FDI. Let $\mu_{Ij}(i, \overline{\Delta})$ be firm j's imitation intensity in state i. The aggregate imitation intensity is $\mu(i, \overline{\Delta}) = \sum_{j=1}^{N} \mu_{Ij}(i, \overline{\Delta})$. The value function of $V_I(i)$ can be written as:

$$\rho V_I(\overline{\Delta} - 1) = M_S(\xi - 1) + \iota[0 - V_I(\overline{\Delta} - 1)],$$

$$\rho V_I(i) = M_S(\xi - 1) + \iota[V_I(i + 1) - V_I(i)] \text{ for } i \leq \overline{\Delta} - 2.$$

The above value functions can be interpreted as follows. The flow payoff of holding the asset $V_I(i)$ equals to the instantaneous profit $M_S(\xi - 1)$, plus the change in asset value: with probability ι the leading technology in the North advances by one step, hence $V_I(i)$ is changed to $V_I(i + 1)$. Note that $V_I(\overline{\Delta}) = 0$, since when $\Delta_{NS} = \overline{\Delta}$

new FDI will arrive and the South firm loses the market. Solving the value functions recursively,

$$V_I(i) = M_S(\xi - 1)\frac{1 - (\frac{\iota}{\rho + \iota})^{\overline{\Delta} - i}}{\rho}. \tag{1}$$

Observing (1), $V_I(i)$ is decresing in i. Intuitively, a bigger i implies a shorter (expected) remaing length of the current cycle, which further implies a smaller value of successful imitation.

Now South firms' equilibrium imitation intensity is derived. Symmetric equilibrium is considered in which all firms choose the same (individual) imitation intensity $\mu_I^*(i)$. Thus the equilibrium aggregate imitation intensity is $\mu^*(i) = N\mu_I^*(i)$. To derive the symmetric equilibrium, suppose all other South firms choose $\mu_I^*(i)$, and consider firm j. Firm j's discounted payoff of imitation (before any successful imitation occurs), $w_j(\mu_{Ij}, i)$, can be written as (suppress argument i):

$$\rho w_j(\mu_{Ij}) = -a_I\mu_{Ij}^2 + \mu_{Ij}(1 - p)[(V_I - w_j) + (N - 1)\mu_I^*(1 - p)(0 - w_j).$$

In the above expression, $-a_I\mu_{Ij}^2$ is the instantaneous payoff. With probability $\mu_{Ij}(1 - p)$ firm j successfully imitates and its product is ruled legal, in which case firm j collects V_I. With probability $(N - 1)\mu_I^*(1 - p)$, one of the other firms successfully imitates and its product is ruled legal, in which case firm j's payoff becomes 0.

Solving for $w_j(\mu_{Ij})$ from the above expression,

$$w_j(\mu_{Ij}) = \frac{\mu_{Ij}(1-p)V_I - a_I\mu_{Ij}^2}{\rho + (N-1)\mu_I^*(1-p) + \mu_{Ij}(1-p)}. \tag{2}$$

Firm j chooses μ_{Ij} to maximize w_j. Taking partial derivative,

$$\frac{\partial w_j}{\partial \mu_{Ij}} \propto [(1-p)V_I - 2a_I\mu_{Ij}][\rho + (N-1)\mu_I^*(1-p)] - a_I(1-p)\mu_{Ij}^2.$$

Since the above expression is strictly decreasing in μ_{Ij}, the FOC is necessary and sufficient in characterizing the equilibrium μ_{Ij}^*. After imposing symmetry, $\mu_{Ij}^* = \mu_I^*$, the equation characterizing μ_I^* becomes

$$a_I(2N-1)\mu_I^{*2} + [2a_I\frac{\rho}{1-p} - (N-1)(1-p)V_I]\mu_I^* - V_I\rho = 0. \tag{3}$$

Lemma 1 *(i) There is a unique symmetric equilibrium, with $\mu_I^* \in (\frac{(N-1)(1-p)V_I}{(2N-1)a_I}, \frac{(1-p)V_I}{2a_I})$.*

(ii) Both μ_I^* and μ^* are increasing in V_I, decreasing in p and a_I, and increasing in N.

Part (ii) of Lemma 1 shows that the equilibrium (both the individual and aggregate) imitation intensities are increasing in the value of imitation "prize" V_I, decreasing in the IPR strength p and imitation cost a_I, and increasing in the number of South firms N. The first three properties are easy to understand. As to the last property, note that an increase in the number of firms means that, other things equal,

it becomes more likely that one of the other firms will succeed first in imitation. This increases the effective discount rate for each individual firm (see equation (2)). As a result, each individual firm increases imitation intensity in order to speed up its own imitation. Another property worth mentioning is that, as the IPR strength p increases, the equilibrium aggregate imitation cost $a_I(\mu^*)^2/N$ would decrease since μ^* decreases. This is different from Glass and Saggi (2002), in which an increase in IPR, though reduces equilibrium imitation intensity, leads to a higher aggregate imitation cost and more recources being devoted to imitation.

For simplicity, $\mu^*(i)$ is written as $\mu(i)$, with the understanding that it denotes for equilibrium aggregate imitation intensity. Following part (ii) of Lemma 1, $\mu(i)$ is decreasing in i. That is, the intensity of imitation is monotonically decreasing over a cycle. This is because the prize of successful imitation, $V_I(i)$, is decresing over a cycle.

3.3   Incentive to FDI

For $i \leq \overline{\Delta} - 2$, the value function of $V_F(i)$ can be written as:

$$
\begin{aligned}
\rho V_F(\overline{\Delta} - 1) &= M_S(\lambda^{\overline{\Delta}} - \xi) + [-(1-p)\mu(\overline{\Delta} - 1)V_F(\overline{\Delta} - 1) - \iota V_F(\overline{\Delta} - 1)] \\
\rho V_F(i) &= M_S(\lambda^{\overline{\Delta}} - \xi) + [-(1-p)\mu(i)V_F(i) + \iota(V_F(i+1) - V_F(i))] \quad (4)
\end{aligned}
$$

The RHS of the above equations has two terms. The first term is the FDI's instantaneous profit, before any successful imitation occurs. The second term is the

change in asset value. With intensity $(1-p)\mu(i)$ imitation is successful, and the value changes from $V_F(i)$ to 0; with intensity ι the leading technology in the North advances by one step, hence $V_F(i)$ is changed to $V_F(i+1)$. Solving the value functions (4) recursively,

$$V_F(i) = M_S(\lambda^{\overline{\Delta}} - \xi)\sum_{j=i}^{\overline{\Delta}-1}\frac{1}{\rho + (1-p)\mu(j) + \iota}\prod_{k=i}^{j-1}\frac{\iota}{\rho + (1-p)\mu(k) + \iota}, \quad (5)$$

$$V_F(0) = M_S(\lambda^{\overline{\Delta}} - \xi)\sum_{j=0}^{\overline{\Delta}-1}\frac{1}{\rho + (1-p)\mu(j) + \iota}\prod_{k=0}^{j-1}\frac{\iota}{\rho + (1-p)\mu(k) + \iota}. \quad (6)$$

Generally, whether $V_F(i)$ is decreasing in i (or over the cycle) cannot be determined with certainty. This is because an increase in Δ_{NS} has two effects on $V_F(i)$. A decrease in the remaining length of the current cycle implies that the FDI's expected time length as a monopoly in the current cycle (if imitation is not successful) is reduced. This tends to reduce $V_F(i)$. The other effect is that an increase in Δ_{NS} reduces the imitation intensity $\mu(i)$ as mentioned earlier, which reduces the possibility of successful imitation and thus tends to increase $V_F(i)$. Depending on which effect is stronger, $V_F(i)$ could either decrease or increase over the cycle.

Observing (1), we see that V_I has a lower bound $\underline{V}_I = V_I(\overline{\Delta} - 1) = M_S(\xi - 1)/(\rho + \iota)$ and an upper bound $\overline{V}_I = \lim_{\overline{\Delta}\to\infty} V_I(0, \overline{\Delta}) = M_S(\xi - 1)/\rho$. For analytical convenience, following conditions are assumed to hold regarding the analytical results throughout the paper.

Condition 2 *The following condition holds:*

$$\lambda \iota < \rho + \iota \Leftrightarrow (\lambda - 1)\iota < \rho.$$

Condition 3 *The following condition holds*

$$\lambda \rho > \rho + \iota \Leftrightarrow (\lambda - 1)\rho > \iota.$$

Condition 2 basically says that the speed of technology progress ($\lambda \iota$) in the North is lower than the discount rate, which ensures that consumers' discounted utility will not explode but converge to some well defined limit. Condition 3 requires that λ and ρ are big enough relative to ι. This condition ensures that the changes in imitation intensity over the cycle is relatively small, as will be shown later. Combining conditions 2 and 3, $\frac{\iota}{\rho} < \lambda - 1 < \frac{\rho}{\iota}$. Thus, ρ should be relatively big, ι should be realtively small, and λ should lie in between.

Lemma 4 *(i) For all $\overline{\Delta}$, $\mu(0, \overline{\Delta}) \leq \lambda \mu(\overline{\Delta} - 1, \overline{\Delta})$. (ii) $V_F(0, \overline{\Delta})$ is strictly increasing in $\overline{\Delta}$.*

The intuition for Lemma 4 is as follows. As the cycle length increases from $\overline{\Delta}$ to $\overline{\Delta} + 1$, there are four effects on the leading North firms' incentive to do FDI, $V_F(0)$. First, it implies that each FDI can charge a higher price ($\lambda^{\overline{\Delta}+1}$ instead of $\lambda^{\overline{\Delta}}$) in the case of no successful imitation. Second, an increase in the cycle length increases the expected length of the monopoly of each FDI. Both effects tend to increase $V_F(0)$.

Third, an increase in the cycle length also implies that successful imitator will now enjoy a longer expected length of monopoly as well, and this effect tends to increase the intensity of imitation and reduce $V_F(0)$. Finally, a bigger $\overline{\Delta}$ weakly increases the cost of imitation, which tends to reduce the intensity of imitation and increase $V_F(0)$. Condition 3 ensures that the third effect is weaker than the first two effects, so that $V_F(0)$ is strictly increasing in the cycle length. In particular, Condition 3 implies that, the increase in the intensity of imitation due to an increase in cycle length is small enough relative to the first effect. Note that even with the imitation cost a_I being constant in the cycle length $\overline{\Delta}$, $V_F(0)$ could be strictly increasing in $\overline{\Delta}$.

3.4 Equilibrium cycle length

Let Λ be the set of $\overline{\Delta}$s such that $V_F(0, \overline{\Delta}) \geq F$. That is, $\Lambda = \{\overline{\Delta}^E : V_F(0, \overline{\Delta}^E) \geq F\}$. As shown earlier, the equilibrium cycle length $\overline{\Delta}^*$ is the smallest $\overline{\Delta}^E$ among all $\overline{\Delta}^E \in \Lambda$. Note that, if Λ is nonempty, then $\overline{\Delta}^*$ is unique. Since, by part (ii) of Lemma 4, $V_F(0)$ is strictly increasing in the cycle length, the equilibrium cycle length satisfies the following conditions: $V_F(0, \overline{\Delta}^* - 1) < F$ and $V_F(0, \overline{\Delta}^*) \geq F$.

Proposition 1 *Suppose Λ is nonempty. Given the South's IPR strength p, FDI occurs periodically or cyclically: new FDI occurs when the technology gap exactly reaches $\overline{\Delta}^*$. The equilibrium length of the FDI cycle is unique and satsifies: $V_F(0, \overline{\Delta}^* - 1) < F$ and $V_F(0, \overline{\Delta}^*) \geq F$.*

Note that Condition 3 is sufficient but not necessary for $V_F(0,\overline{\Delta})$ being strictly increasing in $\overline{\Delta}$ and the characterization of equilibrium cycle length $\overline{\Delta}^*$ in Proposition 1. For instance, for the characterization in Proposition 1 to be valid, Condition 3 can be weakened. In particular, part (i) of Lemma 4 does not need to hold for all $\overline{\Delta}$, but only for the relevent range of $\overline{\Delta}$ such that $\overline{\Delta} \leq \overline{\Delta}^*$.[9] In numerical simulations Condition 3 is not imposed.

Specifically, the price charged by every generation of FDI is always $\lambda^{\overline{\Delta}^*}$, and the price charged by each generation of successful imitator is always ξ. Moreover, the imitation intenisities of the same phase (the same Δ_{NS}) across different cycles are always the same. What is different across cycles is that technology advances, and hence consumers' gross utilities increase, across cycles. Thus the South consumers are also benefiting from the technology advancement of the North as well.

Proposition 2 *The equilibrium FDI cycle length, $\overline{\Delta}^*$, is: (i) weakly decreasing in p, the IPR strength of the South, and weakly decreasing in a_I, the cost of imitation; (ii) weakly increasing in ξ, the cost of FDI production; (iii) weakly decreasing in λ, the size of each step in the quality ladder; (iv) weakly increasing in N, the number of South firms; (v) weakly decreasing in M_S, the size of the South market.*

The intuition for the comparative statics in Proposition 2 is as follows. When the IPR strength p increases, the probability of successful imitation is directly reduced

---

[9]Specifically, part (i) of Lemma 4 only requires $\mu(0,\overline{\Delta}^*) \leq \lambda\mu(\overline{\Delta}^* - 1, \overline{\Delta}^*)$, which requires a weaker condition than Condition 3 as $V_I(0,\overline{\Delta}^*) < \lim_{\overline{\Delta}\to\infty} V_I(0,\overline{\Delta})$.

and the incentive of imitation is indirectly dampend. Both tend to increase the profitability of FDI and reduce the equilibrium cycle length. Similarly, an increase in the imitation cost a_I reduces imitation intensity and weakly reduces the equilibrium cycle length. When the cost of FDI production ξ increases, the instantaneous profit of FDI decreases, and the instantaneous profit of successful imitation increases, which increases the intensity of innovation. Both effects reduce the profitability of FDI, and thus the equilibrium cycle length will increase.

When the step size of the quality ladder λ increases, it tends to increase the price charged by the FDI, while the incentive of imitation is not affected. As a result, the equilibrium cycle length will decrease. The comparative statics regarding the North innovation rate ι is ambiguous. As ι increases, for any given $\overline{\Delta}$ it reduces the expected time length of the monopoly of the existing FDI. But at the same time, it reduces the expected time length of the monopoly of the successful imitator as well, which dampens the incentive to imitate and increases the value of FDI. Either effect could dominate. One can think that the step size of the quality ladder λ depends on the patent policy in the North. In particular, a more stringent patentability requirement of the North implies a bigger λ. Thus part (iii) of Proposition 2 implies that a more stringent patentability requirement in the North leads to a smaller FDI cycle length, or FDI will occur more frequently.[10]

---

[10]This implication should be viewed with caution. Given our assumption that the innovation rate in the North is exogenous, roughly $\lambda\iota$ is constant. As λ increases, ι will decrease correspondingly. But given that the impact of changes in ι on the equilibrium cycle length is ambiguous, a more stringent patentability requirement in the North will likely reduce the equilibrium cycle length.

As the number of South firms N increases, the aggregate imitation intensity increases, which reduces the profitability of FDI and increases the equilibrium cycle length. This implies that if the South industry is more competitive, then the FDI cycle will be longer or FDI occurs less frequently.

As the size of the South market M_S increases, it increases the instantaneous profit of existing FDI. But at the same time, it increases the instantaneous profit of the sucessful imitator as well, which increases imitation intensity and reduces the value of FDI. However, the first effect dominates and the equilibrum cycle length will weakly decrease. Intuitively, FDI benefits more from an increase in the market size than successful imitators do. This is because the profit margin of the FDI is $\lambda^{\overline{\Delta}} - \xi$, which is bigger than $\xi - 1$, the profit margin of successful imitators. Moreover, given Conditions 2 and 3, the changes in equilibrium imitation intensities are not that sensitive to changes in the prize of successful imitation V_I. The implication of this result is that, other things equal, a bigger South market will have a shorter FDI cycle length. More specifically, FDI should occur more frequently in South countries with bigger markets, such as China, than in South countries with smaller markets, such as Thailand.

Our model can be easily extended to the case with many symmetric industries. Although all the industries have the same equilibrium cycle length, the stochastic nature of innovation means that in real physical time the phases of cycles of different industries are in general stagggered. Thus, the FDI cycle length in our model can be

interpreted as the volume of FDI: longer cycle length means a smaller volume. With this interpretation, previous results imply that South countries with a bigger market size should have a bigger FDI volume.

By part (i) of Proposition 2, the equilibrium cycle length, $\overline{\Delta}^*$, is weakly decreasing in p. Note that $p \in [0, 1]$. Correspondingly, fixing other parameter values, the equilibrium cycle length has a lower bound and an upper bound, which we call the minimum cycle length ($p = 1$) and maximum cycle length ($p = 0$), and denote as $\overline{\Delta}^*_{\text{min}}$ and $\overline{\Delta}^*_{\text{max}}$, respectively. Specifically, when $p = 1$, $\mu(\cdot) = 0$, and $V_F(0, 1, \overline{\Delta})$ (where the second argument denotes $p = 1$) becomes

$$V_F(0, 1, \overline{\Delta}) = \frac{M_S(\lambda^{\overline{\Delta}} - \xi)}{\rho}[1 - (\frac{\iota}{\rho + \iota})^{\overline{\Delta}}].$$

The minimum cycle length $\overline{\Delta}^*_{\text{min}}$ satisifes $V_F(0, 1, \overline{\Delta}^*_{\text{min}}) \geq F$ and $V_F(0, 1, \overline{\Delta}^*_{\text{min}} - 1) < F$. Similarly, the maximum cycle length $\overline{\Delta}^*_{\text{max}}$ satisifes $V_F(0, 0, \overline{\Delta}^*_{\text{max}}) \geq F$ and $V_F(0, 0, \overline{\Delta}^*_{\text{max}} - 1) < F$. Note that $\overline{\Delta}^*_{\text{max}}$ is bounded, since $\mu(\cdot)$ is bounded even if $p = 0$ as imitation is costly. To summarize, the number of possible equilibrium cycle length is finite: $\overline{\Delta}^* \in \{\overline{\Delta}^*_{\text{min}}, \overline{\Delta}^*_{\text{min}} + 1, ..., \overline{\Delta}^*_{\text{max}}\} \equiv \Gamma$.

4   The Optimal IPR Strength

Analysis in the previous section shows that, given the IPR strength of the South, p, the pattern of equilibrium FDI cycles is uniquely determined. The South government

tries to maximize its discounted social welfare.

4.1 Discounted social welfare

Let CS_t and PS_t be the consumer surplus and producer surplus of the South at time t. The South goverment's social welfare at time t, w_t, is the sum of CS_t and PS_t. In particular, $w_t = CS_t + PS_t$. Let $W(p)$ be the (expected) discounted social welfare of the South in equilibrium, given p, evaluated at the very beginning of the time, time 0. Let $W_k(0)$ be the equilibrium discounted social welfare of the South, starting in kth cycle with state $i = 0$. Note that $W = (\frac{\iota}{\rho+\iota})^{\overline{\Delta}^*} W_1(0)$, since it takes $\overline{\Delta}^*$ steps for the first FDI to occur and the first FDI cycle to start. Denote $w_k(i, J)$ as the instantaneous social welfare of the South in the kth cycle with state i, where $J = F, S$. In particular, F stands for the case that an FDI is serving the South market (no successful imitation) and S stands for the case that a south firm is serving the market (after successful imitation). More explicitly

$$w_k(i, F) = (M_s \lambda^{k\overline{\Delta}^*}) + [-M_s \lambda^{\overline{\Delta}^*} - a_I(\overline{\Delta}^*)\mu^2(i)/N],$$

$$w_k(i, S) = (M_s \lambda^{k\overline{\Delta}^*}) + [-M_s].$$

To understand the above expressions, note that when an FDI is serving the market, the consumers of the South gets a flow (gross) utility of $M_s \lambda^{k\overline{\Delta}^*}$, and they pay a total price of $M_s \lambda^{\overline{\Delta}^*}$, and South firms in total incur a flow imitation cost of $a_I(\overline{\Delta}^*)\mu^2(i)/N$. When a South firm is serving the market, again the consumers of the South get a

flow (gross) utility of $M_s\lambda^{k\overline{\Delta}^*}$ and pay a total price of $M_s\xi$. The South firms in total have a profit of $M_s(\xi - 1)$, and they no longer incur imitation costs.

Observing the two above expressions, we see that the first term $M_s\lambda^{k\overline{\Delta}^*}$ (consumers' gross utility) is increasing across cycles, but remains constant within cycles. The second terms are changing within cycles (depending on state i), but are isomorphic across cycles. Based on these observations, $W_1(0)$ which is a discounted sum of instantaneous social welfare, is decomposed into two components. The first one is the discounted sum of consumers' gross utility, which is growing due to periodoic free upgrade across cycles. We denote this term (starting at the beginning of kth cycle) as C_k. The second one is the discounted sum of the remaining terms (could be interpreted as changes in social welfare due to imitation within cycles). This term is denoted as $R(i, J)$, where $J = F, S$. Recall that this term is stationary across cycles. In short, $W_1(0) = C_1 + R(0, F)$ and $W_k(0) = C_k + R(0, F)$, since every cycle starts with a new FDI who initially serves the South market.

$$C_k = \frac{M_s\lambda^{k\overline{\Delta}^*}}{\rho}[1 - (\frac{\iota}{\rho + \iota})^{\overline{\Delta}^*}] + (\frac{\iota}{\rho + \iota})^{\overline{\Delta}^*}C_{k+1}.$$

In the above expression, the first term is the discounted consumer gross utility within the current cycle, while the second term is continuation payoff. Solving recursively,

$$C_1 = \frac{M_s\lambda^{\overline{\Delta}^*}}{\rho}\frac{1 - (\frac{\iota}{\rho + \iota})^{\overline{\Delta}^*}}{1 - (\frac{\lambda\iota}{\rho + \iota})^{\overline{\Delta}^*}}. \tag{7}$$

By equation (7), C_1 is bounded due to Condition 2, $\lambda\iota < \rho + \iota$, which essentially means that the rate of upgrade is less than the discount rate. One can think of C_1 as the increase in South welfare due to periodic free upgrades of products resulting from technology growth in the North. In other words, C_1 captures the trend of growth across cycles. Removing C_1 makes each cycle stationary.

Since $R(i, F)$ is stationary across cycles, it can be written as

$$
\begin{aligned}
\rho R(i, F) &= -M_s \lambda^{\overline{\Delta}^*} - a_I(\overline{\Delta}^*)\mu^2(i)/N + (1-p)\mu(i)[R(i, S) - R(i, F)] \\
&\quad + \iota[R(i+1, F) - R(i, F)], \text{ for } i \leq \overline{\Delta}^* - 2, \\
\rho R(\overline{\Delta}^* - 1, F) &= -M_s \lambda^{\overline{\Delta}^*} - a_I(\overline{\Delta}^*)\mu^2(\overline{\Delta}^* - 1)/N + \\
&\quad (1-p)\mu(\overline{\Delta}^* - 1)[R(\overline{\Delta}^* - 1, S) - R(\overline{\Delta}^* - 1, F)] \\
&\quad + \iota[R(0, F) - R(\overline{\Delta}^* - 1, F)].
\end{aligned}
$$

In the above expressions, the first two terms are the instantaneous payoff, while the next two terms are the changes in the values. Similarly, $R(i, S)$ can be expressed as

$$
\begin{aligned}
\rho R(i, S) &= -M_s + \iota[R(i+1, S) - R(i, S)], \text{ for } i \leq \overline{\Delta}^* - 2, \\
\rho R(\overline{\Delta}^* - 1, S) &= -M_s + \iota[R(0, F) - R(\overline{\Delta}^* - 1, S)].
\end{aligned}
$$

Solving the value functions recursively,

$$R(0, F) = \frac{1}{1 - (\frac{\iota}{\rho+\iota})^{\overline{\Delta}^*}} \times \tag{8}$$

$$\{\sum_{j=0}^{\overline{\Delta}^*-1}[\frac{-M_s\lambda^{\overline{\Delta}^*} - a_I(\overline{\Delta}^*)\mu^2(j)/N - (1-p)\mu(j)\frac{M_s}{\rho}[1 - (\frac{\iota}{\rho+\iota})^{\overline{\Delta}^*-j}]}{\rho + (1-p)\mu(j) + \iota}$$

$$\prod_{z=0}^{j-1}\frac{\iota}{\rho + (1-p)\mu(z) + \iota}]\} \tag{9}$$

Combining euqations (7) and (8),

$$W(p) = (\frac{\iota}{\rho+\iota})^{\overline{\Delta}^*}[C_1 + R(0, F)]$$

$$= \frac{M_s}{\rho}\frac{(\frac{\lambda\iota}{\rho+\iota})^{\overline{\Delta}^*}}{1 - (\frac{\lambda\iota}{\rho+\iota})^{\overline{\Delta}^*}}[1 - (\frac{\iota}{\rho+\iota})^{\overline{\Delta}^*}] - \frac{(\frac{\iota}{\rho+\iota})^{\overline{\Delta}^*}}{1 - (\frac{\iota}{\rho+\iota})^{\overline{\Delta}^*}} \tag{10}$$

$$\times\{\sum_{j=0}^{\overline{\Delta}^*-1}[\frac{M_s\lambda^{\overline{\Delta}^*} + a_I(\overline{\Delta}^*)\mu^2(j)/N + (1-p)\mu(j)\frac{M_s}{\rho}[1 - (\frac{\iota}{\rho+\iota})^{\overline{\Delta}^*-j}]}{\rho + (1-p)\mu(j) + \iota}$$

$$\prod_{z=0}^{j-1}\frac{\iota}{\rho + (1-p)\mu(z) + \iota}]\} \tag{11}$$

4.2   Two major effects

Recall that, by previous results, the number of possible equilibrium cycle length $\overline{\Delta}^*$ is finite. Given that p is continuous, for each possible $\overline{\Delta}^*$ there is a range of p such that all p in this range induce the same equilibrium cycle length $\overline{\Delta}^*$. Formally, let $P(\overline{\Delta}^*) \equiv \{p$: the equilibrium cycle length is $\overline{\Delta}^*$ under $p\}$. Since $\overline{\Delta}^*$is weakly decreasing in p, $P(\overline{\Delta}^*)$ is an interval : $P(\overline{\Delta}^*) = [\underline{p}(\overline{\Delta}^*), \overline{p}(\overline{\Delta}^*)]$, where $\underline{p}(\overline{\Delta}^*)$ is the

lower bound and $\overline{p}(\overline{\Delta}^*)$ is the upper bound. To abuse terminology, the equilibrium cycle length induced by the optimal IPR strength p^o is called the optimal cycle length, with label $\overline{\Delta}^o$.

Proposition 3 *Among all IPR strengths that induce the same equilibrium cycle length* $\overline{\Delta}^*$, $p \in P(\overline{\Delta}^*)$, *the discounted social welfare is the highest under the smallest IPR strength* $\underline{p}(\overline{\Delta}^*)$.

Proposition 3 implies that among all IPR strengths inducing the same equilibrium cycle length, the South goverment always tries to choose the lowest IPR strength. Denote p^o as the optimal p that maximizes the South welfare, and $\underline{P} \equiv \{\underline{p}: \underline{p} = \underline{p}(\overline{\Delta}^*)$ for some $\overline{\Delta}^* \in \Gamma\}$. Then by Proposition 3 $p^o \in \underline{P}$, and it must be the case that $V_F(0, p^o, \overline{\Delta}^*) = F$; that is, every generation of FDI earns zero expected profit.[11] This means that generically, $p^o < 1$, or the optimal IPR protection of the South is never perfect.

The underlying reason for Proposition 3 is that, compared to social optimum, at each instant of any state i South firms underinvest in imitation. This is because the South's (flow) social gain of imitation at each instant is $M_S(\lambda^{\overline{\Delta}^*} - 1)$, while the flow profit for successful imitation is only $M_S(\xi - 1)$. Moreover, the expected time length to enjoy social gain and to enjoy private profit of a successful imitation is the same. As a result, the South government wants to reduce p as much as possible in order to induce more intensive imitation. This effect is the "imitation" effect.

---

[11] This implies that the possibility of doing FDI in the South does not affect North firms' incentive to innovate.

It is surprising that this result holds regardless of the number of South firms, N. Recall that as N increases, the aggregate imitation intensity increases. One would think that as N becomes large enough, the aggregate equilibrium imitation intensity might be bigger than the socially optimal one. And as a result, the South government might want to increase the IPR strength to reduce the aggreagte imitation intensity. However, this will never happen, and the reason is as follows. Recall that as N goes to infinity, the individual equilibrium imitation intensity converges to $\frac{(1-p)V_I}{2a_I}$. Since the private prize of successful imitation, V_I, is less than the social gain of sucessful imitation, each firm still underinvests in imitation even as N goes to infinity. In terms of aggregate imitation intensity, note that the total costs of imitation across different firms are additive, instead of being increasing and convex in aggregate imitation intensity. This implies that if each South firm underinvests in imitation, then aggregately firms underinvest in imitation as well, regardless of the number of firms N.

Proposition 4 *The discounted gross utility of the consumers of the South, $(\frac{\iota}{\rho+\iota})^{\overline{\Delta}^*}C_1$, is strictly decreasing in equilibrium cycle length, $\overline{\Delta}^*$.*

The result of Proposition 4 is quite intuitive. As the equilibrium cycle length increases, consumers will get free upgrades less often, which decreases the discounted value of consumers' gross utility (free upgrades). To illustrate this point more clearly, consider two equilibrium cycle lengths $\overline{\Delta}^*$ and $\overline{\Delta}^{*\prime}$, with $\overline{\Delta}^{*\prime} > \overline{\Delta}^*$. Now consider a grand cycle with length $\overline{\Delta}^*\overline{\Delta}^{*\prime}$. Note that the comparsion of the two equilibrium

paths are the same across different grand cycles. Now inspect a grand cycle. The first observation is that the overall upgrades within the grand cycle are the same under two different cycle lengths. However, with equilibrium length $\overline{\Delta}^*$ consumers will get more frequent free upgrades (a total number of $\overline{\Delta}^{*\prime}$) of smaller steps (each upgrade has $\overline{\Delta}^*$ steps), while with equilibrium length $\overline{\Delta}^{*\prime}$ consumers will get less frequent free upgrades (a total number of $\overline{\Delta}^*$) of bigger steps (each upgrade has $\overline{\Delta}^{*\prime}$ steps). Because of discounting, more frequent upgrades lead to a higher discounted value.

Given that the equilibrium cycle length is weakly decreasing in IPR strength p, Proposition 4 reveals the downside of a weak IPR strength: it will increase the equilibrium cycle length and consumers will get free upgrades less frequently. Because of this "free upgrade" effect, the South government has an incentive to implement a higher IPR strength.

To summarize, the imitation effect identified in Proposition 3 means that the South government tends to choose an IPR strength as small as possible in order to speed up imitation (within each cycle). On the other hand, the free of upgrade effect identified in Proposition 4 implies that the South government tends to choose a high IPR strength, in order to reduce the equilibrium cycle length and get more frequent upgrades. The optimal IPR strength p^o tries to balance these two opposite effects. Of course, the optimal IPR strength also depends on effects other than the two just mentioned. For instance, an increase in equilibrium cycle length also leads to a higher price charged by FDI, which tends to reduce the social welfare of the South. This

"price" effect means that the South government tends to choose a high IPR strength. Moreover, an increase in equilibrium cycle length might increase the cost of imitation a_I, which again tends to reduce the social welfare of the South. This "imitation cost" effect implies that the South government tends to choose a high IPR strength as well.

Example 5 $\lambda = 1.105$, $\rho = 0.05$, $\iota = 0.04$, $\xi = 1.1$, $M_S = 10$, $F = 15$, $a_I = 3$, $n = 0$, $N = 5$.[12] *The minimum cycle length is* 2, *and the maximum cycle length is* 32. *The optimal cycle length is* 3 *and the optimal IPR strength is* 0.9089.

Example 6 $\lambda = 1.21$, $\rho = 0.05$, $\iota = 0.04$, $\xi = 1.2$, $M_S = 10$, $F = 1$, $a_I = 3$, $n = 0$, $N = 2$. *The minimum cycle length is* 1, *and the maximum cycle length is* 4. *The optimal cycle length is* 2 *and the optimal IPR strength is* 0.3994.

The above examples share a common feature: the optimal cycle length is very close to the minimum cycle length. Actually, in both examples the optimal cycle length is just one step longer than the minimum cycle length. In all the numerical simulations, the optimal cycle length either coincides with the minimum cycle length or is one step longer than the minimum cycle length. This indicates that, quantitatively, the free upgrade effect is the dominant effect.

---

[12] *In the numerical examples, we assume that the cost of imitation has the following form: $a_I(\overline{\Delta})^n$. Thus $n = 0$ means that imitation cost is independent of cycle length, and $n = 1$ implies that imitation cost is linear in cycle length.*

Note that Condition 2 does not hold for these examples.

4.3 Comparative statics

Proposition 5 *Suppose the optimal cycle length $\overline{\Delta}^o$ does not change. (i) When the step size of the quality ladder, λ, increases, the optimal IPR strength p^o decreases, and the equilibrium imitation intensity increases. (ii) When the marginal cost of FDI production, ξ, increases, the optimal IPR strength p^o increases. (iii) When the size of the South market, M_S, increases, the optimal IPR strength p^o decreases, and the equilibrium imitation intensity increases. (iv) When the number of South firms, N, increases, the optimal IPR strength p^o increases.*

The results of Proposition 5 are straightforward given the comparative statics results in Proposition 2. For example, as the size of the South market increases, any FDI becomes more profitable. Given that the optimal cycle length does not change, the South government now can reduce its IPR strength. A lower IPR strength and a bigger market imply a higher aggregate imitation intensity. This result implies that countries with bigger markets, such as China, tend to have lower IPR strength and higher aggregate imitation intensity. Part (iv) of Proposition 5 implies that countries with a more competitive industry tends to have higher IPR strength.
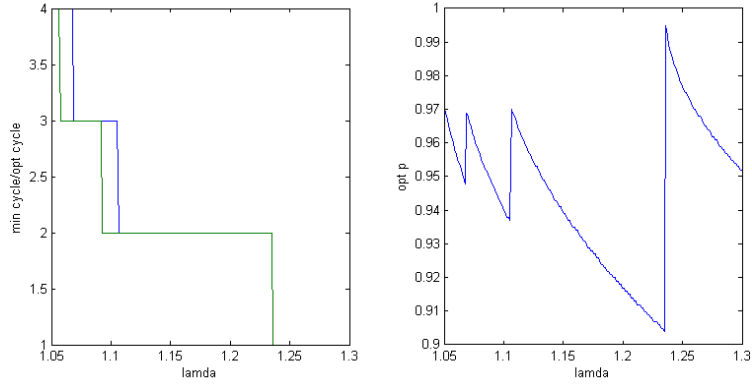
Figure 22: Optimal Cycle Length and IPR as λ Changes-1
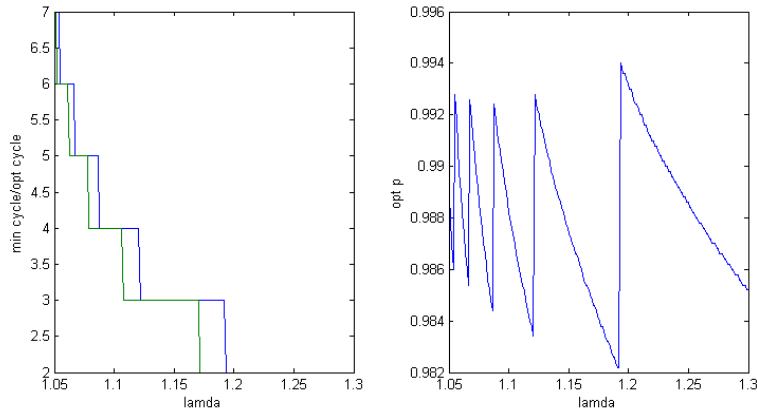
Figure 23: Optimal Cycle Length and IPR as λ Changes-2

Figure 22 and Figure 23 illustreate how the optimal cycle length (the left panels) and the optimal IPR strength (the right panels) change as λ, the step size of the quality ladder, varies. The parameter values for Figure 22 are: $\rho = 0.05$, $\iota = 0.04$, $\xi = 1.1$, $M_S = 10$, $F = 15$, $a_I = 3$, $n = 0$, and $N = 10$. And those for Figure 23 are: $\rho = 0.05$, $\iota = 0.02$, $\xi = 1.1$, $M_S = 10$, $F = 50$, $a_I = 0.3$, $n = 1$, $N = 20$.[13] Two

---

[13] In the examples, $\lambda < \xi$ for some value of λ. But $\lambda^{\overline{\Delta}} > \xi$ always hold for all λ, which means that the priced charged by FDI is always higher than that of a successful imitator.

122

figures exhibit the same pattern. As λ increases, both the minimum cycle length and the optimal cycle length weakly decrease. Moreover, the optimal cycle length either coincides or is one step bigger than the minimum cycle length. As to the optimal IPR strength, it monotonically decreases as λ increases when the optimal cycle length remains the same; and it jumps up discretely when an increase in λ causes a decrease in the optimal cycle length.

Recall that an increase in λ directly increases the price charged by the FDI. Thus the minimum cycle length is decreasing in λ. Another pattern worth mentioning is that, as λ increases, when the minimum cycle length decreases the optimal cycle length does not decreases immediately. It decreases to the minimum cycle length only when λ increases further by some amount. In other words, the decreases in the optimal cycle length "lag behind" those of the minmum cycle length. To understand this pattern, note that when the minimum cycle length decreases by one, at that particular λ for the optimal cycle length to match the minimum cycle length (decreases by one as well) the IPR strength p has to be 1 (by the definition of the minimum cycle length). But $p = 1$ implies a complete shutdown of imitation. However, since the cost of imitation is convex in imitation intensity, the cost of imitation approches 0 faster than the social return does. This implies that the social return of imitation is very big relative to the cost of imitation when the imitation intensity goes to 0. As a result, in the neighborhood of that particular λ the social planner would optimally choose not to increase p all the way to 1, though a smaller cycle length is feasible.

In other words, the imitation effect outweighs the free upgrade effect, leading to a one-step gap between the optimal and the minimum cycle length.

Figure 24: Optimal Cycle Length and IPR as ξ Changes-1
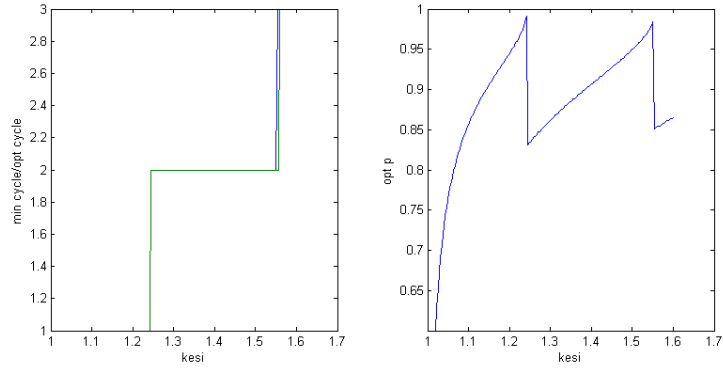
Figure 25: Optimal Cycle Length and IPR as ξ Changes-2

Figure 24 and Figure 25 illustrate the impacts of changes in FDI's production cost ξ. The parameter values for Figure 24 are: $\lambda = 1.2$, $\rho = 0.05$, $\iota = 0.04$, $M_S = 10$, $F = 50$, $a_I = 1$, $n = 0$, and $N = 10$. And those for Figure 25 are: $\lambda = 1.25$, $\rho = 0.05$, $\iota = 0.02$, $M_S = 10$, $F = 1$, $a_I = 2$, $n = 1$, $N = 20$. Two figures exhibit the same

pattern. As ξ increases, both the minimum cycle length and the optimal cycle length weakly increase, and they are at most one step apart. As to the optimal IPR strength, it monotonically increases as ξ increases when the optimal cycle length remains the same; and it jumps down discretely when an increase in ξ causes an increase in the optimal cycle length.

It is easy to understand why an increase in ξ leads the minimum cycle length to increase. This is because, as pointed out ealier, an increase in ξ directly reduces the profitability of FDI. Another pattern worth mentioning is that, as ξ increases, the increases in the optimal cycle length "precede" those of the minmum cycle length. Again, the underlying reason is that the social return of imitation is very big relative to the cost of imitation when the imitation intensity is close enough to 0. When ξ is close enough to the next "jump" point of ξ at which the minimum cycle length jumps up, to match the equilibrium cycle length to the minimum cycle length the IPR strength has to be very close to 1. Since the soical return of imtation is relatively very high when μ is close to 0, the optimal cycle length jumps up "earlier" than the minimum cycle length does to ensure that the imitation intensity is not close to 0.
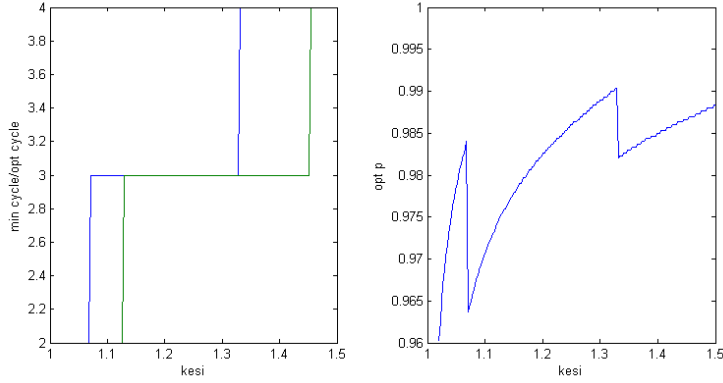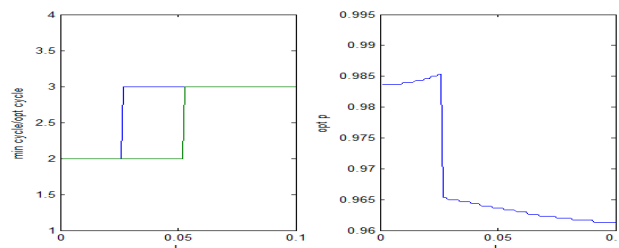
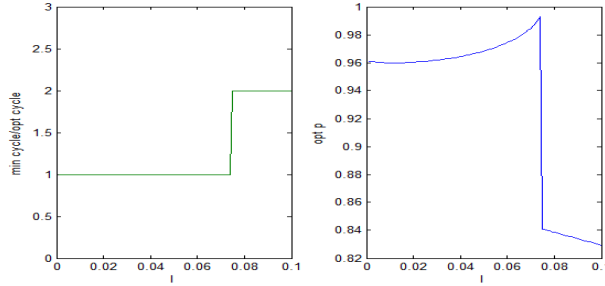

Figure 26: Optimal Cycle Length and IPR as ι Changes-1

125

Figure 27: Optimal Cycle Length and IPR as ι Changes-2

Figure 26 and Figure 27 illustrate the impacts of changes in ι, the North innovate rate. The parameter values for Figure 26 are: $\lambda = 1.2$, $\rho = 0.05$, $\xi = 1.1$, $M_S = 10$, $F = 50$, $a_I = 3$, $n = 0$, and $N = 20$. And those for Figure 27 are: $\lambda = 1.2$, $\rho = 0.05$, $\xi = 1.1$, $M_S = 10$, $F = 8$, $a_I = 2$, $n = 1$, $N = 10$. As ι increases, both the minimum cycle length and the optimal cycle length weakly increase, and they are at most one step apart. Moreover, the optimal cycle length either coincides or is one step bigger than the minimum cycle length (in the second figure, they always coincide). As to the optimal IPR strength, when the optimal cycle length remains the same it is increasing in ι when ι is low and it is decreasing in ι when ι is high; and it jumps down discretely when an increase in ι causes an increase in the optimal cycle length.

As mentioned earlier, an increase in ι directly reduces the expected length of FDI monopoly, thus reducing the profitability of FDI. This is the reason why the minimum cycle length is increasing in ι. To understand the relationship between the optimal p^o and ι when the optimal cycle length remains the same, note that an increase in ι also dampens the incentive of imitation, which indirectly makes FDI more profitable.

126

How big this indirect effect is depends on the intensity of imitation. When p is lower, the imitation intensity is higher, and this effect tends to be stronger. Becuase the optimal p^o is lower when ι is higher, this indrect effect outweighs the direct effect when ι is high and the opposite happens when ι is low.

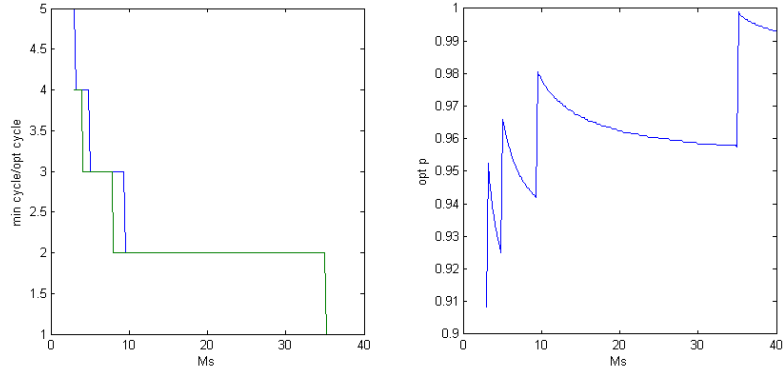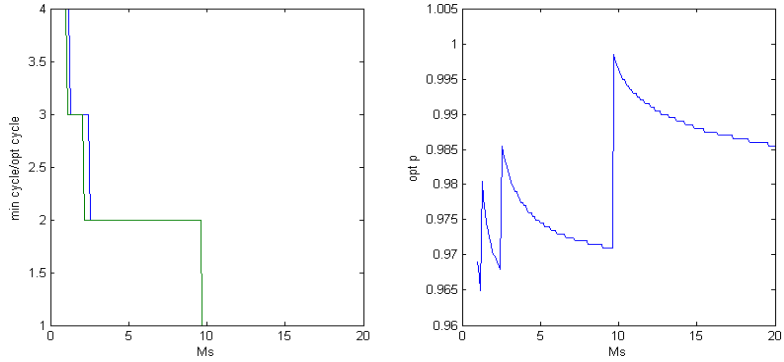Figure 28: Optimal Cycle Length and IPR as M_S Changes-1

Figure 29: Optimal Cycle Length and IPR as M_S Changes-2

Figures 28 and 29 illustrate the impacts of changes in the market size M_S. The parameter values for Figure 28 are: $\lambda = 1.2$, $\rho = 0.05$, $\iota = 0.02$, $\xi = 1.1$, $F = 50$, $a_I = 3$, $n = 1$, and $N = 20$. And those for Figure 29 are: $\lambda = 1.2$, $\rho = 0.05$, $\iota = 0.03$,

127

$\xi = 1.1$, $F = 12$, $a_I = 0.3$, $n = 0$, $N = 10$. As M_S increases, both the minimum cycle length and the optimal cycle length weakly decrease, and they are at most one step apart. As to the optimal IPR strength, it monotonically decreases as M_S increases when the optimal cycle length remains the same; and it jumps up discretely when an increase in M_S causes a decrease in the optimal cycle length. The decreases in the optimal cycle length "lag behind" the decreases in the minimum cycle length, for the same reason as mentioned before.

The above pattern generates two testable empirical implications. First, a developing country with a bigger market size tends to have a smaller FDI cycle length. Second, among developing countries the relationship between the domestic market size and the optimal IPR strength is non-monotonic. For developing countires having the same FDI cycle length, the IPR strength is decreasing in the market size. But a bigger market size implies a smaller FDI cycle length, which tends to increase the IPR strength. The following figure shows the relationship between IPR and the log of GDP between 1995-2005 for 79 developing countries.[14]

---

[14]The IPR data is obtained from Park (2008). The GDP data is based on Penn World Tables: https://pwt.sas.upenn.edu/php_site/pwt71/pwt71_form.php
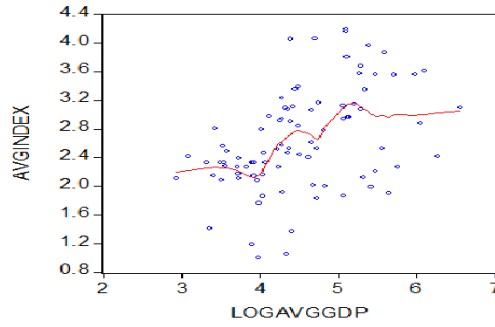
Figure30: The Relationship between IPR and GDP

Note: AVGINDEX is the index of degree of intellectual right protection, and

LOGAVGGDP is the log of average GDP for each country.

In Figure 30, the loss-fit curve clearly indicates that the relationship between IPR and market size is not monotonic. Actually, the shape of the loss-fit curve largely resembles that in the right panels of Figures 28 and 29. Therefore, this shows that the prediction of our model is largely consistent with empirical evidence.[15]
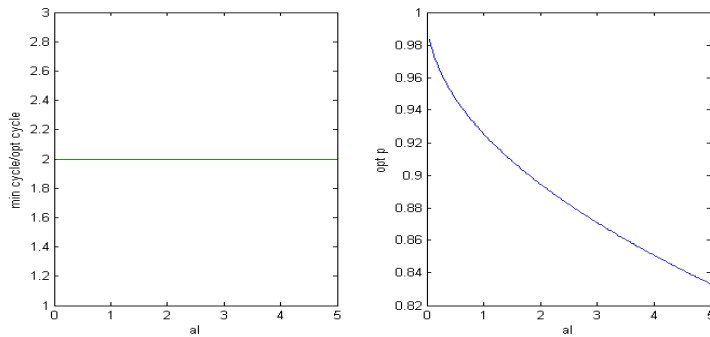
Figure 31: Optimal Cycle Length and IPR as a_I Changes

---

[15]Empirically, Auriol et al. (2012) found that IPR and domestic market size among developing countries have a U-shape relationship.

Figure 31 illustrates the impacts of changes in the imitation cost a_I. The parameter values are $\lambda = 1.26$, $\rho = 0.05$, $\iota = 0.02$, $\xi = 1.25$, $M_S = 10$, $F = 10$, $n = 1$ and $N = 5$. The optimal cycle length does not change as a_I varies, and the optimal IPR strength is monotonically decreasing in a_I. The reason that the optimal cycle length does not change with a_I is that the minimum cycle length is independent pf a_I.

To investigate the impacts of the relationship between imitation cost and North-South technology gap, we consider the following example: $\lambda = 1.2$, $\rho = 0.05$, $\iota = 0.02$, $\xi = 1.1$, $M_S = 10$, $F = 50$. In the first scenario, the imitation cost is independent of the cycle length: $a_I = 6$. In the second scenario, the imitation cost is linear in cycle length $a_I = 3\overline{\Delta}$. Note that the imitation cost is the same when the cycle length is 2, and the imitation cost is higher in the second scenario when $\overline{\Delta} > 2$ and the opposite is true when $\overline{\Delta} = 1$. The three dimensional figures (varying M_S and N) are illustrated below.
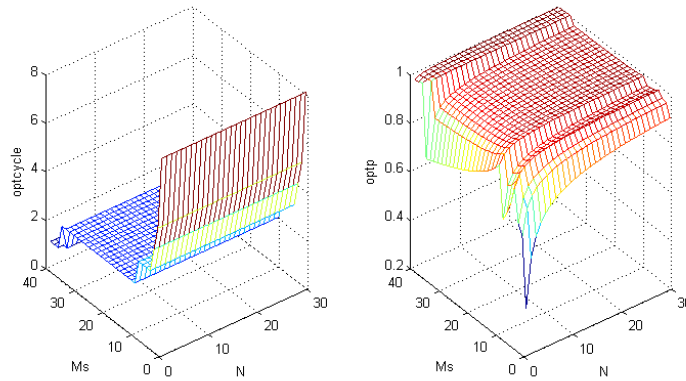
Figure 32: Optimal Cycle Length and Optimal IPR with Cost Structure 1
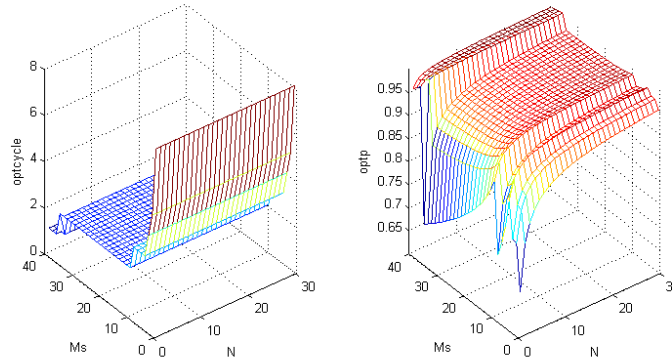
Figure 33: Optimal Cycle Length and Optimal IPR with Cost Structure 2

From Figures 32 and 33, the patterns of the optimal cycle length (the left panels) are almost the same under two different cost structures. This is again due to the following two facts. First, the dominance of the free upgrade effect implies that the optimal cycle length either coinsides with or very close to the minimum cycle length. Second, the minimum cycle length is independent of the imitation cost. From the right panels of Figures 32 and 33, the optimal IPR is lower under the second cost structure when the optimal cycle length is bigger than 2, and it is higher under the second cost structure when the optimal cycle length is 1. This is because the optimal cycle length is almost the same under two cost structures, thus the optimal IPR depends only on the magnitude of the imitation costs.

To summarize, lower imitation costs will lead to higher optimal IPR strength. Since imitation costs are lower in more technologically advanced economies, this implies that IPR strength should be increasing in GDP per capita among developing countries. Empirically, Maskus (2000), Braga et al. (2000), and Chen and Puttitanun

(2005) all found a U-shaped relationship between IPR and per cpita income. However, the negative relationship only holds for countries with very low income levels. For the majority of developing countries, IPR is increasing in income per capita.[16] The following figure shows the relationship between IPR and the log of GDP per capita between 1995-2005 for 79 developing countries.[17]
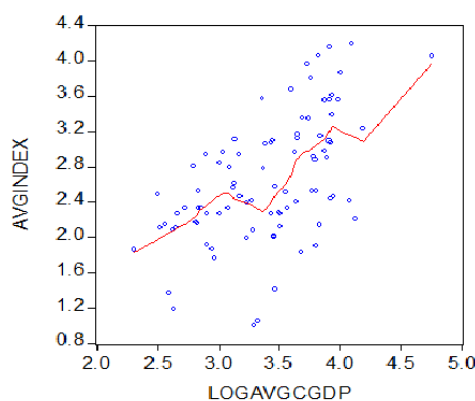
Figure 34: The Relationship between IPR and GDP Per Capita

Note: AVGINDEX is the index of degree of intellectual right protection, and

LOGAVGGDP is the log of average GDP per capita for each country.

In Figure 34, the loss-fit curve clearly indicates that IPR and GDP per capita are positively correlated. Compared to Figure 30, the loss-fit curve in Figure 34 is much steeper. This implies that IPR increases with the level of economic development,

---

[16]In Chen and Puttitanun (2005), IPR reaches its minimum for countries with a GDP per capita of $854 in 1995 prices.

[17]The source is the same as that of Figure 9.

while the relationship between IPR and domestic market size is less clear-cut.[18] This pattern is largely consistent with the prediction our model.

5 Conclusion

This paper develops a quality ladder model in which the technology gap between the North and South is endogenously determined. Equilibrium exhibits FDI cycles: New FDI arrives if and only if the technology gap reaches some threshold. A stronger IPR in the South discourages imitation and reduces the FDI cycle length. A smaller market size and more imitating firms in the South tend to enlarge the FDI cycle length. A weaker IPR in the South brings a short-run benefit: within each FDI cycle it encourages imitation and increases the South welfare. However, it entails a long run cost: it increases the FDI cycle length and makes FDI less frequent, which due to discounting reduces the South welfare. The optimal IPR strength balances these two effects. Our comparative statics results show that the optimal IPR is non-monotonic in the South market size, and increasing in the level of economic development. These two predictions are largely consistent with empirical evidence. Moreover, IPR is positively correlated with the number of firms in the South industry.

---

[18]A big domestic market can be due to a large population, or a high GDP per capita.

References

Aiyagari, R., "Optimal Capital Income Taxation with Incomplete Markets, Borrowing Constraints, and Constant Discounting," *Journal of Political Economy*, 103, 1995: 158-75.

Auriol, E., S. Biancini, and R. Paillacar. "Intellectual Property Rights Protection in Developing Countries," 2012, working paper, Toulouse School of Economics.

Bonham, C., and R. Cohen, "To Aggregate, Pool, or Neither: Testing the Rational Expectations hypothesis Using Survey Data," *Journal of Business and Economic Statistics*, 19,2001, 278-291.

Bonham, C., and D. Dacy, "In Search of a Strictly Rational Forecast," *Review of Economics and Statistics*, 73, 1991,245-253.

Braga, C.E.P., C. Fink, and C.P. Sepulveda. "Intellectual Property Rights and Economic Development," World Bank publications, 2000.

Branstetter, L., and K. Saggi. "Intellectual Property Rights, Foreign Direct Investment, and Industrial Development," *Economic Journal* 121, 2011, 1161-1191.

Brett, C. and J. Pinkse. "The determinants of municipal tax rates in British Colombia," 2000, *Canadian Journal of Economics*, 33, 695-714.

Brueckner, J.K. and L.A. Saavedra. "Do Local Governments Engage in Strategic Property–Tax Competition?" 2001, *National Tax Journal*, 54, 2, 203-230.

Brueckner, J.K. "Strategic Interaction among Governments: An Overview of Empirical Studies," 2003, *International Regional Science Review*, 26, 2, 175-188.

Bucovetsky, S. "Asymmetric Tax Competition," 1991, *Journal of Urban Economics*, 30, 167-181.

Buettner, T. "Tax Base Effects and Fiscal Externalities of Local Capital Taxation: Evidence from A Panel of German Jurisdictions," 2003, *Journal of Urban Economics*, 54, 110-128.

Campbell, John Y. and Shiller, Robert J., "Cointegration and Tests of Present Value Models," *Journal of Political Economy*, 95, 1987: 1062-1088.

Chamley, C., "Optimal Taxation of Capital Income in General Equilibrium with Infinite Lives," *Econometrica*, 54, 1986: 607-622.

Chen, Y., and T. Puttitanun. "Intellectual Property Rights and Innovation in Developing Countries," *Journal of Development Economics* 78, 2205, 474-493.

Chirinko, R.S. and D.J. Wilson. "Tax Competition among U.S. States: Racing to the Bottom or Riding on a Seesaw?" 2013, working paper.

Croushore, D., "The Livingston Survey: Still Useful After All These Years," *Federal Reserve Bank of Philadelphia Business Review*,1997 March/April, 15-26.

Chow, G.C., "Rational versus Adaptive Expectations in Present Value Models," *Review of Economics and Statistics*,71(3), 1989: 376-384.

Chow, G.C.,"Usefulness of Adaptive and Rational Expectations in Economics," *C*EPS Working Paper No. 221, 2011.

Deardoff, A.V. "The growth rate for population: comment." 1976, *Int Econ Rev*, 17:510-515.

Diamond P.A.and Spinnewijn J., "Capital Income Taxes with Heterogeneous Discount Rates," *American Economic Journal: Economic Policy, American Economic Association*, vol. 3(4), 2011, pages 52-76, November.

Eakin, D.H. "Public-Sector Capital and the Productivity Puzzle," 1997, *The Review of Economics and Statistics*, 76 (1), 12-21.

Elhorst, J.P. "Specification and Estimation of Spatial Panel Data Models," 2003, *International Regional Science Review*, 26, 244-268.

Elhorst, J.P. "Spatial Panel Data Models." 2009, *Handbook of Applied Spatial Analysis*, Ch. C.2. Springer: Berlin Heidelberg New York.

Erosa, A. and M. Gervais, "Optimal Fiscal Policy in Life-Cycle Economies," *Journal of Economic Theory*, 105, 2002: 338-369.

Fama, E.F. and French, Kenneth R., "Permanent and Temporary Components of Stock Prices," *Journal of Political Economy*, 96, 1988: 246-273.

Frankel, J., and K. Froot, "Using Survey Data to Test Standard Propositions Regarding Exchange Rate Expectations," *American Economic Review*, 77, 1987b, 133-153.

Frankel, J., and K. Froot, "Chartists, Fundamentalists and the Demand for Dollars," *in* A.S. Courakis and M. P. Taylor, ed.,Private Behaviour and Government Policy in Interdependent Economies, Oxford University Press, Oxford, pp.73-126.

Friedman, B., "Survey Evidence on the 'Rationality' of Interest Rate Expectations," *Journal of Monetary Economics*, 6, 1990, 453-465.

Froot, K.A., "New Hope for the Expectations Hypothesis of the Term Structure of Interest Rates, " *Journal of Finance*, 44, 1989, 283-305.

Galor, O. and D.N. Weil. "The Gender Gap, Fertility, and Growth," 1996, *American Economic Review* 86 (3), 374-387.

Garriga, C., "Optimal Fiscal policy in overlapping generations models," *Working Paper*, Florida State University.

Garofalo, G.A. and Yamarik S., "Regional Convergence: Evidence From A New State-By-State Capital Stock Series," *The Review of Economics and Statistics*, MIT Press, vol. 84(2), 2002, pages 316-323.

Glass, A.J., and K. Saggi. "International Technology Transfer and the Technology Gap," *Journal of Development Economics* 55, 1998, 369-398.

Glass, A.J., and K. Saggi. "Intellectual Property Rights and Foreign Direct Investment," *Journal of International Economics* 56, 2002, 387-410.

Glass, A.J., and X. Wu. "Intellectual Property Rights and Quality Impovement," *Journal of Development Economics* 82, 2007, 393–415.

Grossman, G.M., and E. Helpman. "Quality Ladders and Product Cycles," *Quarterly Journal of Economics* 106, 1991a, 557-586.

Grossman, G.M., and E. Helpman. "Endogenous Product Cycles," *Economic Journal* 101, 1991b, 1214-1229.

Grossman, G.M., and E.L.C. Lai. "International Protection of Intellectual Property," *American Economic Review* 94, 2004, 1635-1653.

Hartman, D.G., "Tax policy and foreign direct investment," 1985, *Journal of Public Economics,* 26, 107-21.

Hayashi, M. and R. Boadway. "An empirical analysis of intergovernmental tax interaction: The case of business income taxes in Canada." 2001, *Canadian Journal of Economics*, 34, 481-503.

Helpman, E. "Innovation, Imitation, and Intellectual Property Rights," *Econometrica* 61, 1993, 1247-1280.

Hoyt, W.H. "Property Taxation, Nash Equilibrium, and Market Power," 1991, *Journal of Urban Economics*, 30, 123-131.

Hubbard, R.G. and Kenneth L. Judd, "Liquidity Constraints, Fiscal Policy, and Consumption," Papers on Economic Activity, Economic Studies Program, The Brookings Institution, vol. 17(1), pages 1-60.

Jeong, J., and G. S. Maddala,"Testing the Rationality of Survey Data Using the Weighted Double-bootstrapped Method of Moments," *Review of Economics and Statistics*, 78, 1996 296-302.

Judd, K.L., "Redistributive Taxation in a Simple Perfect Foresight Model," 1985, *Journal of Public Economics*, 28, 59-83.

Kanbur, R. and M. Keen. "Jeux sans frontieres: tax competition and tax coordination when countries differ in size," *American Economic Review*, 83, 877-92.

Keen, M. J. and C. Kotsogiannis. "Does federalism lead to excessively high taxes?" 2002, *American Economic Review* 92 (1), 363-370.

Keen, M. J. and C. Kotsogiannis. "Tax competition in federations and the welfare consequences of decentralization." 2004, *Journal of Urban Economics* 56 (3), 397-407.

Kelejian, H.H., Prucha, I.R., 1998. "A generalized spatial two-stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbance," *J. Real Estate Finance Econom*, 17(1), 99-121.

Kim, S. "The Effects of Labor Migration on Optimal Taxation: An International Tax Competition Analysis," 2013, working paper.

Knight, B., "State capital taxes and the location of investment: empirical lessons from theoretical models of tax competition," Finance and Economics Discussion Series 2002, Board of Governors of the Federal Reserve System (U.S.).

Koehne, Sebastian and Kuhn, Moritz, "Optimal capital taxation for time-nonseparable preferences," MPRA Paper 45203, University Library of Munich, Germany.

Lai, E.L.C. "International Intellectual Property Rights Protection and the Rate of Product Innovation," *Journal of Development Economics* 55, 1998, 133-153.

Lee, L.F., Yu, J.H., 2014. "Efficient GMM estimation of spatial dynamic panel data models with fixed effects," *Journal of Econometrics*, 180, 174-197.

Mankiw, N.G., D. Romer and D.N. Weil. "A Contribution to the Empirics of Economic Growth," 1992, *The Quarterly Journal of Economics,* 107 (2), 407-437.

Markiewicz, A. and Pick, A., "Adaptive Learning and Survey Data," *C*DMA Working Paper Series 201305, Centre for Dynamic Macroeconomic Analysis.

Markusen, J. "Contracts, Intellectual Property Rights, and Multinational Investment in Developing Countries," *Journal of International Economics* 53, 2001, 189-204.

Maskus, K.E. *Intellectual Property Rights in the Global Economy*, Peterson Institute, 2000.

Moretti, E., "Workers' Education, Spillover and Productivity: Evidence from Plant-Level Production Functions,," *American Economic Review* 94(3),2004, 656-690.

O'Donoghue, T. "A Patentability Requirement for Sequential Innovation," *Rand Journal of Economics* 29, 1998, 654-679.

Park, W. "International Patent Protection: 1960-2005," *Reseach Policy* 37, 2008, 761-766.

Pi, J. and Y. Zhou. "Tax competition in federations revisited," 2014, *Economic Letters*, 123, 104-107.

Pinkse, J., M.E. Slade and C. Brett. "Spatial Price Competition: A Semiparametric Approach," 2002, *Econometrica*, 70 (3), 1111-1153.

Pomerleau K., "The High Burden of State and Federal Capital Gains Tax Rates," *T*he Tax Foundation, 2014.

Poterba, James M. and Summers, Lawrence H., "Mean Reversion in Stock Prices: Evidence and Implications," NBER Working Paper, 1987.

Ramsey, F.P., "A Contribution to the Theory of Taxation," *Economic Journal*, Vol. 37, No 145, 1927, 47-61.

Samuelson, P.A. "The optimum growth rate of a population." 1975 *Int Econ Rev*, 16:531-538.

Swarbrick J., "Optimal social policy in a DSGE model with heterogeneous agents ," Master thesis, 37-62.

Walczak, J., S. Drenkard and J. Henchman. "2017 State Business Tax Climate Index," 2016, Tax Foundation.

Yamarik, S., "State-Level Capital And Investment: Updates And Implications," *Contemporary Economic Policy*, Western Economic Association International, vol. 31(1),2013, pages 62-72, 01.

Young, Kan H., "The effects of taxes and rates of return on foreign direct investment in the United States," *National Tax Journal*, 41 (1),1988: 109-21.

Appendix A: Initial Belief

Initial Belief	States
0.001-0.01	Illinois; South Dakota; West Virginia
0.01-0.02	Florida; Maine; New Jersey; Ohio
0.02-0.03	Delaware; Iowa; Indiana; Kansas; Massachusetts; Michigan; Missouri; North Dakota; New Hampshire; Nevada; Texas
0.03-0.04	Georgia; Hawaii; New Mexico; Oklahoma;
0.04-0.05	Connecticut; Maryland; Utah;
0.05-0.06	Arkansas; Montana; Rhode Island; Tennessee; Vermont; Washington; Alabama; Alaska; California; Colorado; Idaho; Louisiana; Minnesota; Nebraska; New York;
0.06-0.07	Oregon; Pennsylvania; South Carolina; Virginia;
0.07-0.08	Arizona; Kentucky; Mississippi; Wyoming
0.08-0.09	North Carolina; Wisconsin
0.1-	DC

NOTE: Initial beliefs of each state is the tax rate in year 1958.

Table 21: Initial Belief

Appendix B: Sensitivity Pattern

Sensitivity Pattern	States
Decreasing	Alabama; Arizona; Arkansas; Colorado; DC;
	Indiana; Massachusetts; Maryland;
	North Carolina; South Dakota; Utah; Oregon;
Increasing	Iowa; Idaho; Kansas; Kentucky; North Dakota; Nebraska;
	Rhode Island; Wisconsin; Oklahoma;
U-shape	California; Georgia; Maryland; Michigan;
	New Jersey; Nevada; South Carolina; Tennessee;
	Texas; Virginia; West Virginia; Wyoming;
Inverse U-shape	Alaska; Delaware; Florida; Hawaii; Louisiana; Missouri;
	Mississippi; Montana; New Mexico; Pennsylvania;
Cubic or higher	Connecticut; Illinois; Minnesota; New Hampshire;
	New York; Ohio; Vermont; Washington

NOTE: States differ in terms of tax sensitivity pattern,
based on equation $dk_t = c_0 + c_1\theta_t + c_2\theta_t^2 + c_3\theta_t^3 + \epsilon_t$
Table 22: Sensitivity Pattern

Appendix C: Tests for Equality of Means between Series: Gini Coefficient

Method	df	Value	Probability
t-test	19	1.536094	0.1410
Anova F-statistic	(1, 19)	2.359585	0.1410
Category Statistics			
Variable	Count	Mean	Std. Dev.
DECR	12	0.458833	0.027643
INCR	9	0.443000	0.015716

Table 23: Tests for Equality of Means between Series: Gini Coefficient

t-test	18	1.974393	0.0639
Anova F-statistic	(1, 18)	3.898228	0.0639
Category Statistics			
Variable	Count	Mean	Std. Dev.
DECR	11	0.462455	0.025835
INCR	9	0.453700	0.023535

Table 24: Tests for Equality of Means between Series: Gini Coefficient (outlier Utah excluded)

t-test	19	1.400530	0.1775
Anova F-statistic	(1, 19)	1.961485	0.1775
Category Statistics			
Variable	Count	Mean	Std. Dev.
DECR	12	30.26667	8.746151
INCR	9	25.96667	3.155551

Table 25: Tests for Equality of Means between Series: Bachelor Degree Attainment

t-test	18	1.761973	0.0950
Anova F-statistic	(1, 18)	3.104550	0.0950
Category Statistics			
Variable	Count	Mean	Std. Dev.
DECR	11	31.23636	8.469507
INCR	9	25.96667	3.155551

Table 26: Tests for Equality of Means between Series: Bachelor Degree Attainment (outlier Arkansas excluded)

Appendix D: State Abbreviations of the major four areas in the U.S.

Midwest	South	West	Northeast
IL	FL	AZ	CT
IN	GA	CA	RI
IA	MD	CO	NJ
KS	NC	ID	PA
MI	SC	MT	NY
MN	VA	NM	MA
MO	WV	NV	VT
NE	DE	OR	NH
ND	AL	UT	ME
OH	KY	WA	
SD	MS	WY	
WI	TN		
	AR		
	LA		
	OK		
	TX		

NOTE: Alaska and Hawaii are excluded as they are geographic outliers in the U.S.
Table 27: State Abbreviations of the major four areas in the U.S.

Appendix E: Variables

In this section, the data source for and definition of each variable is provided. The panel dataset is for 48 contiguous states from 1958 to 2007.

A. Capital tax rate:

The average capital tax rate for each state s at time t is defined as follows,

$\text{ACTR}_{s,t} = \frac{\text{capital tax revenue}_{s,t}}{\text{taxable capital income}_{s,t}}$

From US Census Bureau, I sum up the two main sources of capital tax revenue: property tax, corporate net income tax.

$\text{capital tax revenue}_{s,t} = \text{property tax revenue}_{s,t} + \text{corporate net income tax revenue}_{s,t}$

Code T01 Property Taxes

Taxes imposed on ownership of property and measured by its value.

Definition: Three types of property taxes, all having in common the use of value as a basis for the tax:

• General property taxes, relating to property as a whole, taxed at a single rate or at classified rates according to the class of property.

Property refers to real property (e.g., land and structures) as well as personal property; personal property can be either tangible (e.g., automobiles and boats) or intangible (e.g., bank accounts and stocks and bonds).

- Special property taxes, levied on selected types of property (e.g., oil and gas properties, house trailers, motor vehicles, and intangibles) and subject to rates not directly related to general property tax rates.

- Taxes based on income produced by property as a measure of its value on the assessment date.

Code T41 Corporation Net Income Taxes

Definitions: Taxes on corporations and unincorporated businesses (when taxed separately from individual income), measured by net income, whether on corporations in general or on specific kinds of corporations, such as financial institutions.

To construct taxable capital income, I use the summation of personal dividend income, personal interest income, and rental income of persons with capital consumption adjustment. This series of taxable capital income is obtained from BEA, where

taxable capital income$_{s,t}$ =

personal dividend income$_{s,t}$+personal interest income$_{s,t}$+rental income$_{s,t}$

Personal dividend income is payments in cash or other assets, excluding the corporations' own stock, that corporations in the United States or abroad make to non-corporate stockholders who are U.S. residents.

Personal interest income is the interest income (monetary and imputed) from all sources that is received by individuals, by private and government employee retire-

ment plans, by nonprofit institutions, and by estates and trusts.

The rental income of persons with capital consumption adjustment is the net current-production income of persons from the rental of real property except for the income of persons primarily engaged in the real estate business; the imputed net rental income received by owner-occupants of dwellings; and the royalties received by persons from patents, copyrights, and rights to natural resources. The estimates include BEA adjustments for uninsured losses to real estate caused by disasters, such as hurricanes and floods.

B. Control Variables

The series of federal effective capital gains tax rate from 1958 to 2007 is obtained from Tax Foundation. It is calculated as follows:

$$\text{ECTR}_t^{\text{fed}} = \frac{\text{taxes paid on capital gains}_t}{\text{realized capital gains}_t}$$

Personal income data are obtained from U.S. CENSUS Bureau.

Data of electoral outcomes are obtained from Council of State Governments-Book of States. For each state from 1958 to 2007, I collect the data "number of members in Lower House that are Democrat" (HD), "number of members in Lower House that are Republican" (HR), "number of members in Upper House that are Democrat" (SD) and "number of members in Upper House that are Republican" (SR).

The political environment variables are caluculated as follows:

For the fraction of state house that is Democrat, $D\_H_{s,t} = \frac{HD_{s,t}}{HD_{s,t}+HR_{s,t}}$;

for the fraction of state senate that is Democrat $D\_S_{s,t} = \frac{SD_{s,t}}{SD_{s,t}+SR_{s,t}}$;

and for the dummy variable, $d_{s,t}$=2 if Democrat is majority in both State Lower House and Senate, $d_{s,t}$=1 if either State Lower house or Senate has Democrat as majority, and $d_{s,t}$=0 Republican is majority in both State Lower House and Senate.

For Nebraska from 1958-2007 and Minnesota from 1958-1973, members were selected in nonpartisan elections. I include missing variables to account for it.

C. Weighting Scheme

Scheme 1: For state i in each of the four areas (South, Midwest, West and Northeast), $w_{ij} = \frac{1}{K}$ if states i and j are located in the same area and share the same border geographically; and $w_{ij} = 0$ otherwise. K is the total number of contiguous states of state i in its area.

Scheme 2: For state i in each of the four areas (South, Midwest, West and Northeast), $w_{ij} = \frac{\text{timeavgpopu}_j}{\sum \text{timeavgpopu}_k}$ if states i and j are located in the same area and share the same border geographically; and $w_{ij} = 0$ otherwise. timeavgpopu_j is the average population size of state j from 1958 to 2007, and $\sum \text{timeavgpopu}_k$ is the sum of average population size from 1958 to 2007 of all the contiguous states of state i in its area.

D. Population

The series of population data for each state from 1958 to 2007 is obtained from U.S. CENSUS Bureau.

Midwest		South		West		Northeast	
IL	0.0056	FL	0.0287	AZ	0.0348	CT	0.0073
IN	0.0068	GA	0.0187	CA	0.0184	RI	0.0043
IA	0.0019	MD	0.0134	CO	0.0221	NJ	0.0079
KS	0.0054	NC	0.0148	ID	0.0174	PA	0.0026
MI	0.0055	SC	0.0135	MT	0.0075	NY	0.0032
MN	0.0092	VA	0.0140	NM	0.0165	MA	0.0053
MO	0.0069	WV	-0.0003	NV	0.0474	VT	0.0101
NE	0.0048	DE	0.0143	OR	0.0160	NH	0.0169
ND	0.0010	AL	0.0081	UT	0.0238	ME	0.0068
OH	0.0040	KY	0.0075	WA	0.0175		
SD	0.0038	MS	0.0068	WY	0.0106		
WI	0.0078	TN	0.0118				
		AR	0.0100				
		LA	0.0071				
		OK	0.0093				
		TX	0.0195				

NOTE: The sample period is 1958-2007.
Table 28: Time-Average Population Growth Rates

Midwest	South	West	Northeast
0.0052	0.0123	0.0211	0.0072
(0.0024)	(0.0066)	0.0112	0.0044

NOTE: The sample period is 1958-2007, with standard deviation in parentheses.
Table 29: Group and Time Averaged Population Growth Rate

E. Capital

Data of capital series at the state level is obtained from Garofalo and Yamarik (2002), and Yamarik (2012).

For year 1958-1990, capital at state level, denoted as K_{st}, is calculated as Net Private Capital Stock created through 1-digit SIC industries, using gross private investment of Net Private Capital Stock created through 1-digit SIC industries and time-varying depreciation rate created through 1-digit SIC industries. The estimates are further revised because many farms declare losses, and thus propreitary income of agriculture was removed.

For year 1991-2008, capital K_{st} is calculated as Net Private Capital Stock created through 1-digit NAIS industries, using gross private investment of K1 using industry-specific time-varying depreciation rate created through 1-digit NAIS industries.

Thus,capital per cap $k_{st} = \frac{K_{st}}{L_{st}}$, where L_{st} is population of state s at time t from Appendix II.D.

Instrumental Variables

$$Y_{nt}^* = \lambda_1 W_{1n} Y_{nt}^* + \lambda_2 W_{2n} Y_{nt}^* + X_{nt}^* \beta + \epsilon_{nt}^*$$

And in this paper, $Y_{nt}^* = (\lambda_1 + \lambda_2 G_n) W_{1n} Y_{nt}^* + X_{nt}^* \beta + \epsilon_{nt}^*$, and it follows that

$Y_{nt}^* = S_n^{-1} X_{nt}^* \beta + S_n^{-1} \epsilon_{nt}^*$, where $S_n = I_n - (\lambda_1 + \lambda_2 G_n) W_{1n}$.

Thus, the optimum IV matrix is $(X_{nt}^*, W_{1n} S_n^{-1} X_{nt}^*)$.

Whenever λ_1 and λ_2 take the values so that S_n is invertible and expandable,

optimum IVs can be chosen as $(W_{1n}(\lambda_1 + \lambda_2 G_n) W_{1n} X_{nt}^*, W_{1n}(\lambda_1 + \lambda_2 G_n) W_{1n}(\lambda_1 +$

$\lambda_2 G_n) W_{1n} X_{nt}^*, ...)$,

and in this paper,

$(W_{1n} X_{nt}^*, W_{1n}^2 X_{nt}^*, W_{1n} G_n W_{1n} X_{nt}^*, W_{1n}^3 X_{nt}^*, W_{1n} G_n W_{1n}^2 X_{nt}^*,$

$W_{1n}^2 G_n W_{1n} X_{nt}^*, W_{1n} G_n W_{1n} G_n W_{1n} X_{nt}^*)$

is chosen as IVs.

Appendix G: Capital allocation regression

Dependent variables: log capital per cap				
Specifications	I	II	III	IV
Explanatory Variables				
OwnTax	-2.766***	-2.764***	-2.775***	-2.772***
	(0.184)	(0.184)	(0.184)	(0.184)
Growth×OwnTax	270.154***	270.280***	268.813***	268.460***
	(14.038)	(14.060)	(14.091)	(14.090)
TaxNeighbor	0.971***	0.977***	0.946***	0.986***
	(0.163)	(0.166)	(0.164)	(0.166)
Growth×TaxNeighbor				
TaxFed	0.149***	0.150***	0.140**	0.140**
	(0.056)	(0.056)	(0.056)	(0.056)
log Personal Income	0.264***	0.264***	0.265***	0.264***
	(0.002)	(0.002)	(0.002)	(0.002)
Democrat_House		-0.003		-0.032
		(0.015)		(0.022)
Democrat_Senate			0.015	0.036*
			(0.014)	(0.020)
Constant	0.404***	0.406***	0.395***	0.402***
	(0.019)	(0.021)	(0.021)	(0.021)

NOTE: These are least squares estimates with fixed-effect of the parameters.
Based on $\log(k_{st}) = \alpha + \beta_1 \cdot OTR_{st} + \beta_2 \cdot (g_{st} \cdot OTR_{st}) + \lambda_1 \cdot TN_{st} + \lambda_2 \cdot (g_{st} \cdot TN_{st}) + \gamma \cdot TF_t + X_{st} \cdot \tau + u_s + \epsilon_{st}$
The influence of population growth rates on the response of capital levels to
own tax rates and neighbors' tax rates are examined, represented by
coefficients in front of GrowthOwnTax and Growth*TaxNeighbor.
Robust standard errors in parentheses. ***p<0.01, **p<0.05, *p<0.1.

Table 30: Capital allocation regression: Response to Own Tax. Scheme 1

Dependent variables: log capital per cap

Specifications	I	II	III	IV
Explanatory Variables				
OwnTax	-0.186*	-0.188*	-0.217**	-0.217**
	(0.107)	(0.108)	(0.109)	(0.109)
Growth×OwnTax				
TaxNeighbor	-2.239***	-2.243***	-2.259***	-2.217***
	(0.237)	(0.239)	(0.238)	(0.239)
Growth×TaxNeighbor	347.450***	347.375***	345.911***	345.470***
	(16.479)	(16.492)	(16.509)	(16.508)
TaxFed	0.093*	0.093*	0.082	0.083
	(0.055)	(0.055)	(0.055)	(0.055)
log Personal Income	0.270***	0.270***	0.271***	0.271***
	(0.002)	(0.002)	(0.002)	(0.002)
Democrat_House		0.002		-0.030
		(0.015)		(0.022)
Democrat_Senate			0.020	0.040**
			(0.013)	(0.020)
Constant	0.339***	0.338***	0.328***	0.334***
	(0.019)	(0.021)	(0.021)	(0.022)

NOTE: These are least squares estimates with fixed-effect of the parameters.

Based on $\log(k_{st}) = \alpha + \beta_1 \cdot OTR_{st} + \beta_2 \cdot (g_{st} \cdot OTR_{st}) + \lambda_1 \cdot TN_{st} + \lambda_2 \cdot (g_{st} \cdot TN_{st}) + \gamma \cdot TF_t + X_{st} \cdot \tau + u_s + \epsilon_{st}$

The influence of population growth rates on the response of capital levels to own tax rates and neighbors' tax rates are examined, represented by coefficients in front of Growth*OwnTax and Growth*TaxNeighbor.

Robust standard errors in parentheses. ***p<0.01, **p<0.05, *p<0.1.

Table 31: Capital allocation regression: Response to Neighbors' Taxes. Scheme 1

Dependent variables: log capital per cap				
Specifications	I	II	III	IV
Explanatory Variables				
OwnTax	-2.759***	-2.764***	-2.772***	-2.771***
	(0.185)	(0.185)	(0.185)	(0.185)
Growth×OwnTax	282.091***	281.352***	279.712***	279.746***
	(13.945)	(13.992)	(14.019)	(14.020)
TaxNeighbor	0.440***	0.422**	0.410**	0.426***
	(0.161)	(0.164)	(0.162)	(0.163)
Growth×TaxNeighbor				
TaxFed	0.127**	0.123**	0.114**	0.114**
	(0.056)	(0.056)	(0.056)	(0.056)
log Personal Income	0.264***	0.264***	0.264***	0.264***
	(0.002)	(0.002)	(0.002)	(0.002)
Democrat_House		0.010		-0.017
		(0.015)		(0.022)
Democrat_Senate			0.022	0.034*
			(0.014)	(0.020)
Constant	0.426***	0.420***	0.413***	0.416***
	(0.020)	(0.022)	(0.022)	(0.023)

NOTE: These are least squares estimates with fixed-effect of the parameters.

Based on $\log(k_{st}) = \alpha + \beta_1 \cdot OTR_{st} + \beta_2 \cdot (g_{st} \cdot OTR_{st}) + \lambda_1 \cdot TN_{st} + \lambda_2 \cdot (g_{st} \cdot TN_{st}) + \gamma \cdot TF_t + X_{st} \cdot \tau + u_s + \epsilon_{st}$

The influence of population growth rates on the response of capital levels to

own tax rates and neighbors' tax rates are examined,

coefficients in front of GrowthOwnTax and Growth*TaxNeighbor.

Robust standard errors in parentheses. ***p<0.01, **p<0.05, *p<0.1.

Table 32: Capital allocation regression: Response to Own Tax. Scheme 2

Dependent variables: log capital per cap				
Specifications	I	II	III	IV
Explanatory Variables				
OwnTax	-0.248**	-0.244**	-0.275**	-0.273**
	(0.106)	(0.107)	(0.108)	(0.108)
Growth×OwnTax				
TaxNeighbor	-2.069***	-2.065***	-2.078***	-2.052***
	(0.194)	(0.195)	(0.195)	(0.195)
Growth×TaxNeighbor	316.290***	316.729***	314.397***	315.535***
	(13.525)	(13.613)	(13.606)	(13.613)
TaxFed	0.139**	0.141**	0.129**	0.129**
	(0.054)	(0.055)	(0.055)	(0.055)
log Personal Income	0.266***	0.266***	0.266***	0.266***
	(0.002)	(0.002)	(0.002)	(0.002)
Democrat_House		-0.004		-0.039*
		(0.015)		(0.022)
Democrat_Senate			0.017	0.043**
			(0.013)	(0.020)
Constant	0.386***	0.389***	0.376***	0.384***
	(0.020)	(0.022)	(0.022)	(0.022)

NOTE: These are least squares estimates with fixed-effect of the parameters.

Based on $\log(k_{st}) = \alpha + \beta_1 \cdot OTR_{st} + \beta_2 \cdot (g_{st} \cdot OTR_{st}) + \lambda_1 \cdot TN_{st} + \lambda_2 \cdot (g_{st} \cdot TN_{st}) + \gamma \cdot TF_t + X_{st} \cdot \tau + u_s + \epsilon_{st}$

The influence of population growth rates on the response of capital levels to own tax rates and neighbors' tax rates are examined, represented by coefficients in front of GrowthOwnTax and Growth*TaxNeighbor.

Robust standard errors in parentheses.   ***p<0.01, **p<0.05, *p<0.1.

Table 33: Capital allocation regression: Response to Neighbors' Taxes. Scheme 2

Proof of Lemma 8.

Proof. Part (i). Define

$$H(\mu_I) \equiv a_I(2N-1)\mu_I^2 + [2a_I\frac{\rho}{1-p} - (N-1)(1-p)V_I]\mu_I - V_I\rho. \qquad (49)$$

By equation (41), $H(\mu_I^*) = 0$. Note that $H(\mu_I)$ is a quadratic function in μ_I, the coefficient of μ_I^2 is strictly bigger than 0, and $H(0) < 0$. This implies that $H(\mu_I) = 0$ has exactly one positive solution, or there is a unique $\mu_I^* > 0$. To show that $\mu_I^* < \frac{(1-p)V_I}{2a_I}$, we compute $H(\frac{(1-p)V_I}{2a_I})$:

$$H(\frac{(1-p)V_I}{2a_I}) = \frac{(1-p)^2V_I^2}{4a_I} > 0.$$

Combining with the fact that $H(0) < 0$, we must have $\mu_I^* < \frac{(1-p)V_I}{2a_I}$. On the other hand,

$$H(\frac{(N-1)(1-p)V_I}{(2N-1)a_I}) = -\frac{V_I\rho}{2N-1} < 0.$$

Combining with the fact that $H(0) < 0$, we must have $\mu_I^* > \frac{(N-1)(1-p)V_I}{(2N-1)a_I}$.

Part (ii). We show the properties hold for μ_I^*. It immediately follows that the properties also hold for μ^*, as $\mu^* = N\mu_I^*$.

Suppose $V_I' > V_I$, and define $H(\mu_I^*, V_I) = 0$ and $H(\mu_I^{*\prime}, V_I') = 0$. Inspecting (49), we can see that, for all $\mu_I \geq 0$, $H(\mu_I, V_I') < H(\mu_I, V_I)$. This implies that $H(\mu_I^*, V_I') < 0$. Therefore, $\mu_I^{*\prime} > \mu_I^*$. This proves that μ_I^* is increasing in V_I.

Suppose $p' > p$, and define $H(\mu_I^*, p) = 0$ and $H(\mu_I^{*\prime}, p') = 0$. Inspecting (49), we can see that, for all $\mu_I \geq 0$, $H(\mu_I, p') > H(\mu_I, p)$. This implies that $H(\mu_I^{*\prime}, p) < 0$. Therefore, $\mu_I^{*\prime} < \mu_I^*$. This proves that μ_I^* is decreasing in p.

Suppose $a_I' > a_I$, and define $H(\mu_I^*, a_I) = 0$ and $H(\mu_I^{*\prime}, a_I') = 0$. Inspecting (49), we can see that, for all $\mu_I \geq 0$, $H(\mu_I, a_I') > H(\mu_I, a_I)$. This implies that $H(\mu_I^{*\prime}, a_I) < 0$. Therefore, $\mu_I^{*\prime} < \mu_I^*$. This proves that μ_I^* is decreasing in a_I.

Suppose $N' = N+1$, and define $H(\mu_I^*, N) = 0$ and $H(\mu_I^{*\prime}, N') = 0$. Computing the difference, $H(\mu_I, N') - H(\mu_I, N) = [2a_I \mu_I - (1-p)V_I]\mu_I$. Thus, for all $\mu_I \in (0, \frac{(1-p)V_I}{2a_I})$, $H(\mu_I, N') - H(\mu_I, N) < 0$. Since by part (i) $\mu_I^* \in (0, \frac{(1-p)V_I}{2a_I})$ and $\mu_I^{*\prime} \in (0, \frac{(1-p)V_I}{2a_I})$, it implies that $H(\mu_I^*, N') < 0$. Therefore, we must have $\mu_I^{*\prime} > \mu_I^*$. This proves that μ_I^* is increasing in N. ∎

Proof of Lemma 11.

Proof. Part (i). Since $V_I(0, \overline{\Delta})$ is increasing in $\overline{\Delta}$ and $V_I(\overline{\Delta} - 1, \overline{\Delta})$ is independent of $\overline{\Delta}$, by Lemma 8, it is sufficent to show that $\lim_{\overline{\Delta} \to \infty} \mu_I^*(0, \overline{\Delta}) \leq \lambda \mu_I^*(\overline{\Delta} - 1, \overline{\Delta})$. The corresponding V_Is are $M_S(\xi - 1)/\rho$ and $M_S(\xi - 1)/(\rho + \iota)$, respectively. By equation

(41),

$$\lambda \mu_I(\overline{\Delta} - 1, \overline{\Delta}) - \lim_{\overline{\Delta} \to \infty} \mu_I(0, \overline{\Delta})$$

$$\propto \quad (N-1)(1-p)M_S(\xi-1)(\frac{\lambda}{\rho+\iota} - \frac{1}{\rho})$$

$$+\lambda\sqrt{[2a_I\frac{\rho}{1-p} - (N-1)(1-p)M_S(\xi-1)\frac{1}{\rho+\iota}]^2 + 4M_S(\xi-1)\frac{\rho}{\rho+\iota}a_I(2N-1)}$$

$$-\sqrt{[2a_I\frac{\rho}{1-p} - (N-1)(1-p)M_S(\xi-1)\frac{1}{\rho}]^2 + 4M_S(\xi-1)a_I(2N-1)}.$$

Since by Condition 10, $\lambda\rho > \rho + \iota$. The above expression being greater than 0 is equivalent to the second term minus the 3rd term being positive, which can be simplified as

$$4M_S(\xi-1)a_I(2N-1)(\frac{\lambda^2\rho}{\rho+\iota} - 1) - 4M_S(\xi-1)a_I\rho(N-1)(\frac{\lambda^2}{\rho+\iota} - \frac{1}{\rho}) > 0.$$

The above inequality holds obviously.

Part (ii). It is enough to show that $V_F(0, \overline{\Delta} + 1) - V_F(0, \overline{\Delta}) > 0$ for all $\overline{\Delta}$. Note that $V_I(i+1, \overline{\Delta}+1) = V_I(i, \overline{\Delta})$ for all $i \leq \overline{\Delta} - 1$. And because $a_I(\overline{\Delta}+1) \geq a_I(\overline{\Delta})$, following part (ii) of Lemma 8 we have $\mu(i+1, \overline{\Delta}+1) \leq \mu(i, \overline{\Delta})$. Let

$$x(i, \overline{\Delta}) \equiv \sum_{j=i}^{\overline{\Delta}-1}[\frac{1}{\rho + (1-p)\mu(j, \overline{\Delta}) + \iota} \prod_{k=i}^{j-1} \frac{\iota}{\rho + (1-p)\mu(k, \overline{\Delta}) + \iota}.$$

By the fact that $\mu(i+1, \overline{\Delta}+1) \leq \mu(i, \overline{\Delta})$ we have $x(i+1, \overline{\Delta}+1) \geq x(i, \overline{\Delta})$. Since $\mu(i, \overline{\Delta}) \geq \mu(\overline{\Delta}-1, \overline{\Delta})$, we have $x(i, \overline{\Delta}) < \frac{1}{\rho+(1-p)\mu(\overline{\Delta}-1, \overline{\Delta})}$.

By (44), $V_F(0, \overline{\Delta}+1) - V_F(0, \overline{\Delta}) > 0$ is equivalent to

$$\frac{\lambda^{\overline{\Delta}+1} - \xi}{\lambda^{\overline{\Delta}} - \xi} > \frac{x(0, \overline{\Delta})}{x(0, \overline{\Delta}+1)}.$$

Since $\frac{\lambda^{\overline{\Delta}+1} - \xi}{\lambda^{\overline{\Delta}} - \xi} > \lambda$, it is enough to show $\lambda x(0, \overline{\Delta}+1) \geq x(0, \overline{\Delta})$, which is equivalent to

$$\frac{\lambda}{\rho+(1-p)\mu(0, \overline{\Delta}+1)+\iota} + \frac{\lambda\iota}{\rho+(1-p)\mu(0, \overline{\Delta}+1)+\iota}x(1, \overline{\Delta}+1) - x(0, \overline{\Delta}) \geq 0.$$

Since $x(i+1, \overline{\Delta}+1) \geq x(i, \overline{\Delta})$, the following inequality is sufficient:

$$\frac{\lambda}{\rho+(1-p)\mu(0, \overline{\Delta}+1)+\iota} + [\frac{\lambda\iota}{\rho+(1-p)\mu(0, \overline{\Delta}+1)+\iota} - 1]x(0, \overline{\Delta}) \geq 0.$$

Since $\lambda\iota \leq \rho + \iota$, the fact that $x(i, \overline{\Delta}) < \frac{1}{\rho+(1-p)\mu(\overline{\Delta}-1, \overline{\Delta})}$ implies that the following is enough

$$\frac{\rho+(1-p)\mu(0, \overline{\Delta}+1)+\iota-\lambda\iota}{\rho+(1-p)\mu(\overline{\Delta}-1, \overline{\Delta})} \leq \lambda$$
$$\Leftrightarrow \quad \rho+\iota+(1-p)\mu(0, \overline{\Delta}+1) \leq \lambda(\rho+\iota)+\lambda(1-p)\mu(\overline{\Delta}-1, \overline{\Delta}).$$

159

Now the following condition is suffiicent: $\mu(0, \overline{\Delta}+1) \leq \lambda \mu(\overline{\Delta}-1, \overline{\Delta})$. Since $\mu(\overline{\Delta}, \overline{\Delta}+1) \leq \mu(\overline{\Delta}-1, \overline{\Delta})$,

$$\mu(0, \overline{\Delta}+1) \leq \lambda \mu(\overline{\Delta}-1, \overline{\Delta}) \Leftarrow \mu(0, \overline{\Delta}+1) \leq \lambda \mu(\overline{\Delta}, \overline{\Delta}+1).$$

But the last inequality holds by part (i). ∎

Proof of Proposition 6.

Proof. Part (i). As either p or a_I increases, by part (ii) of Lemma 8, $\mu(i, \overline{\Delta})$ decreases for all i and $\overline{\Delta}$. And, by (44), $V_F(0, \overline{\Delta})$ increases for all $\overline{\Delta}$. Thus $\overline{\Delta}^*$ is weakly decreasing in p and a_I.

Part (ii). As ξ increases, $V_I(i)$ increases for all i. By part (ii) of Lemma 8, $\mu(i, \overline{\Delta})$ increases for all i and $\overline{\Delta}$. Combining with the fact that $\lambda^{\overline{\Delta}} - \xi$ decreases, by (44) we conclude that $V_F(0, \overline{\Delta})$ decreases for all $\overline{\Delta}$. Thus $\overline{\Delta}^*$ is weakly increasing in ξ.

Part (iii). As λ increases, by (41), $\mu(i, \overline{\Delta})$ remains the same for all i and $\overline{\Delta}$. Since $\lambda^{\overline{\Delta}} - \xi$ increases, by (44), $V_F(0, \overline{\Delta})$ increases for all $\overline{\Delta}$. Thus $\overline{\Delta}^*$ is weakly decreasing in λ.

Part (iv). As N increases, by part (ii) of Lemma 8, $\mu(i, \overline{\Delta})$ increases for all i and $\overline{\Delta}$. By (44) we conclude that $V_F(0, \overline{\Delta})$ decreases for all $\overline{\Delta}$. Thus $\overline{\Delta}^*$ is weakly increasing in N.

Part (v). We first show the following property: $M_S \frac{\partial \mu(i)}{\partial M_S} < \mu(i)$ for all i. In particular,

$$\mu(i) - M_S \frac{\partial \mu(i)}{\partial M_S} \quad \propto \quad \mu_I(i) - V_I(i) \frac{(N-1)(1-p)\mu_I(i) + \rho}{2a_I(2N-1)\mu_I(i) + 2a_I \frac{\rho}{1-p} - (N-1)(1-p)V_I(i)}$$

$$\propto \quad a_I(2N-1)\mu_I(i) - (N-1)(1-p)V_I(i) > 0,$$

where the second equality uses equation (41), and the inequality uses the result that

$\mu_I(i) > \frac{(N-1)(1-p)V_I(i)}{(2N-1)a_I}$ in part (i) of Lemma 8.

Using the property that $M_S \frac{\partial \mu(i)}{\partial M_S} < \mu(i)$ for all i, and collecting terms, we have

$$\frac{\partial V_F(0)}{\partial M_S} \quad > \quad \sum_{z=0}^{\overline{\Delta}-1} \prod_{t=0}^{z} \frac{\iota}{\rho + (1-p)\mu(t) + \iota} \{ \frac{1}{\rho + (1-p)\mu(z) + \iota}$$

$$- \sum_{j=z}^{\overline{\Delta}-1} \frac{(1-p)\mu(z)}{[\rho + (1-p)\mu(z) + \iota]^2} \prod_{k=z+1}^{j} \frac{\iota}{\rho + (1-p)\mu(k) + \iota} \}.$$

Note that the RHS of the above inequality has $\overline{\Delta} - 1$ terms. It is sufficient to show that for every z, the term in the bracket is positive. We will only show it holds for $z = 0$, since the proof for other z is similar. More specifically, we want to show

$$\frac{1}{\rho + (1-p)\mu(0) + \iota} - \sum_{j=0}^{\overline{\Delta}-1} \frac{(1-p)\mu(0)}{[\rho + (1-p)\mu(0) + \iota]^2} \prod_{k=1}^{j} \frac{\iota}{\rho + (1-p)\mu(k) + \iota} > 0. \quad (50)$$

161

To show inequality (50) holds, we proceed recursively. The LHS of (50) is pro-

portinal to

$$[\rho + (1-p)\mu(0) + \iota] - \sum_{j=0}^{\overline{\Delta}-1}(1-p)\mu(0)\prod_{k=1}^{j}\frac{\iota}{\rho + (1-p)\mu(k) + \iota} \tag{51}$$

$$= (\rho + \iota) - (1-p)\mu(0)\sum_{j=1}^{\overline{\Delta}-1}\prod_{k=1}^{j}\frac{\iota}{\rho + (1-p)\mu(k) + \iota}$$

$$\propto (\rho + \iota)[\rho + (1-p)\mu(1) + \iota] - \iota(1-p)\mu(0) - \iota(1-p)\mu(0)\sum_{j=2}^{\overline{\Delta}-1}\prod_{k=2}^{j}\frac{\iota}{\rho + (1-p)\mu(k) + \iota}$$

$$> \iota(\rho + \iota) - \iota(1-p)\mu(0)\sum_{j=2}^{\overline{\Delta}-1}\prod_{k=2}^{j}\frac{\iota}{\rho + (1-p)\mu(k) + \iota}$$

$$\propto (\rho + \iota) - (1-p)\mu(0)\sum_{j=2}^{\overline{\Delta}-1}\prod_{k=2}^{j}\frac{\iota}{\rho + (1-p)\mu(k) + \iota}. \tag{1}$$

In the above derivation, the inequality uses Condition 9, $\rho + \iota > \lambda\iota$, and the property

in part (i) of Lemma 11, $\mu(0) \leq \lambda\mu(i)$ for all $i \leq \overline{\Delta} - 1$. Repeat the same recursive

procedure as in the above derivation, we can show that (51) is bigger than 0. ∎

Proof of Proposition 7.

Proof. We want to show that, for $p \in P(\overline{\Delta}^*)$, $W(p)$ is strictly decreasing in p.

Observing (47), it is sufficent to show that the last term in the bracket of (47) is

strictly increasing in p. Let $K(i) \equiv \frac{M_s}{\rho}[1 - (\frac{\iota}{\rho+\iota})^{\overline{\Delta}^*-i}]$. Since $(1-p)\mu(i)$ is decreasing

162

in p, it is enough to show that, for all j between 0 and $\overline{\Delta}^* - 1$,

$$Z(j) \equiv \frac{M_s \lambda^{\overline{\Delta}^*} + a_I \mu^2(j)/N + (1-p)\mu(j)K(j)}{\rho + (1-p)\mu(j) + \iota} \tag{52}$$

is increasing in p for all j between 0 and $\overline{\Delta}^* - 1$. Now we take partial derivative of (52) with respect to p, which is proportional to:

$$\frac{\partial Z(j)}{\partial p} = [\mu(j) - (1-p)\mu'(j)][M_s \lambda^{\overline{\Delta}^*} - K(j)(\rho + \iota)] + \frac{2a_I}{N}\mu(j)\mu'(j)(\rho + \iota) \tag{53}$$
$$+ \frac{a_I \mu^2(j)}{N}[\mu(j) + (1-p)\mu'(j)].$$

Note that by previous result, $\mu'(j) < 0$. We first show that $\mu(j) + (1-p)\mu'(j) > 0$, which is equivalent to $\mu_I(j) + (1-p)\mu_I'(j) > 0$. For that purpose, we differentiate equation (41) with respect to p:

$$\mu_I'(j) = \frac{-[2a_I \frac{\rho}{(1-p)^2} + (N-1)V_I]\mu_I(j)}{2a_I \frac{\rho}{1-p} - (N-1)(1-p)V_I + 2a_I(2N-1)\mu_I(j)}.$$

Since the numerator is negative, $\mu_I'(j) < 0$ means that the denominator is positive.

163

Using the above equation, we have

$$\mu_I(j) + (1-p)\mu_I'(j) \ \propto \ a_I(2N-1)\mu_I^2(j) - (N-1)(1-p)V_I\mu_I(j)$$

$$= \ \rho[V_I - \frac{2a_I}{1-p}\mu_I(j)] > 0,$$

where the last two steps uses equation (41) and the fact that $\mu_I(j) < \frac{(1-p)V_I}{2a_I}$.

Now to show $\frac{\partial Z(j)}{\partial p} > 0$, it is sufficient to show that

$$[\mu(j) - (1-p)\mu'(j)][M_s\lambda^{\overline{\Delta}^*} - K(j)(\rho+\iota)] + \frac{2a_I}{N}\mu(j)\mu'(j)(\rho+\iota) > 0.$$

Again using the fact that $\mu(j) + (1-p)\mu'(j) > 0$, the following inequality is sufficent:

$$-(1-p)\mu'(j)[M_s\lambda^{\overline{\Delta}^*} - K(j)(\rho+\iota)] + \mu(j)[M_s\lambda^{\overline{\Delta}^*} - K(j)(\rho+\iota) - \frac{2a_I}{N(1-p)}\mu(j)(\rho+\iota)] > 0.$$

Now it is enough to show that $M_s\lambda^{\overline{\Delta}^*} - K(j)(\rho+\iota) - \frac{2a_I}{N(1-p)}\mu(j)(\rho+\iota) \geq 0$. Using the fact that $\mu_I(j) < \frac{(1-p)V_I}{2a_I}$, the following inequality is sufficient:

$$M_s\lambda^{\overline{\Delta}^*} - [K(j) + V_I](\rho+\iota) \quad \geq \quad 0 \tag{54}$$

$$\Leftrightarrow \quad \lambda^{\overline{\Delta}^*} - \frac{\rho+\iota}{\rho}\xi[1 - (\frac{\iota}{\rho+\iota})^{\overline{\Delta}^*-j}] \geq 0$$

$$\Leftarrow \quad \lambda^{\overline{\Delta}^*} - \frac{\rho+\iota}{\rho}\xi[1 - (\frac{\iota}{\rho+\iota})^{\overline{\Delta}^*}] \geq 0. \tag{2}$$

To show that (54) holds, note that when $\overline{\Delta}^* = 1$, it becomes $\lambda - \xi$, which is positive since $\lambda > \xi > 1$. Now for $\overline{\Delta}^* \geq 2$, the LHS of (54) becomes

$$\lambda^{\overline{\Delta}^*} - \frac{\rho + \iota}{\rho}\xi + \frac{\rho + \iota}{\rho}\xi(\frac{\iota}{\rho + \iota})^{\overline{\Delta}^*}] > \lambda^2 - \frac{\rho + \iota}{\rho}\xi > 0,$$

where the last inequality uses $\lambda > \xi$ and Condition 10 that $\lambda > \frac{\rho+\iota}{\rho}$. ∎

Proof of Proposition 8.

Proof. Let $x \equiv \overline{\Delta}^*$ and $a \equiv \frac{\iota}{\rho+\iota}$. Note that x must be integers and $\lambda a < 1$ by Condition 9. It is sufficient to show that $\frac{(\lambda a)^x}{1-(\lambda a)^x}[1 - a^x]$ is strictly decreasing in x. For that purpose, we take the difference

$$\begin{aligned}
\Pi &\equiv \frac{(\lambda a)^{x+1}}{1 - (\lambda a)^{x+1}}[1 - a^{x+1}] - \frac{(\lambda a)^x}{1 - (\lambda a)^x}[1 - a^x] \\
&\propto (\lambda a - 1) + a^x(1 - \lambda a^2) + (\lambda a)^{x+1}a^x(a - 1) \\
&= (\lambda a - 1)[1 - a^x] + a^x(\lambda a)(1 - a)[1 - (\lambda a)^x].
\end{aligned}$$

By the above expression, $\Pi < 0$ is equivalent to

$$a^x(\lambda a)\frac{1 - (\lambda a)^x}{1 - \lambda a} < \frac{1 - a^x}{1 - a} \Leftrightarrow a^x(\lambda a)\sum_{i=0}^{x-1}(\lambda a)^i < \sum_{i=0}^{x-1}a^i.$$

The last inequality obviously holds, since, by $\lambda a < 1$ and $a < 1$, for any $0 \leq i \leq x-1$

165

we have $a^x (\lambda a)^{i+1} < a^i$. ∎

Proof of Proposition 9.

Proof. Part (i). Let $\lambda' > \lambda$, and superscript $'$ denote the endogenous variables under λ'. Note that $p^o = \underline{p}(\overline{\Delta}^o)$. By previous analysis, we have $V_F(0, \lambda, p^o, \overline{\Delta}^o) = F$. Since V_F is increasing in λ by Proposition 6, $V_F(0, \lambda', p^o, \overline{\Delta}^o) > F$. This implies that $p^{o\prime} < p^o$ as $\overline{\Delta}^o$ does not change. It follows that $\mu'(i) > \mu(i)$ for all $i \le \overline{\Delta}^o - 1$ by Lemma 8.

Part (ii). Let $\xi' > \xi$, and superscript $'$ denote the endogenous variables under ξ'. By previous analysis, we have $V_F(0, \xi, p^o, \overline{\Delta}^o) = F$. Since V_F is decreasing in ξ by Proposition 6, $V_F(0, \xi', p^o, \overline{\Delta}^o) < F$. This implies that $p^{o\prime} > p^o$ as $\overline{\Delta}^o$ does not change.

Part (iii). Let $M_S' > M_S$, and superscript $'$ denote the endogenous variables under M_S'. By previous analysis, we have $V_F(0, M_S, p^o, \overline{\Delta}^o) = F$. Since V_F is increasing in M_S by Proposition 6, $V_F(0, M_S', p^o, \overline{\Delta}^o) > F$. This implies that $p^{o\prime} < p^o$ as $\overline{\Delta}^o$ does not change. The fact that $p^{o\prime} < p^o$ and $V_I' > V_I$ (implied by $M_S' > M_S$) means that $\mu'(i) > \mu(i)$ for all $i \le \overline{\Delta}^o - 1$ by Lemma 8.

Part (iv). Let $N' > N$, and superscript $'$ denote the endogenous variables under N'. By previous analysis, we have $V_F(0, N, p^o, \overline{\Delta}^o) = F$. Since V_F is decreasing in N by Proposition 6, $V_F(0, N', p^o, \overline{\Delta}^o) < F$. This implies that $p^{o\prime} > p^o$ as $\overline{\Delta}^o$ does not change. ∎