February 2022 Vol.44, No.1

KDI Journal of Economic Policy

KDI Korea Development Institute

KDI Journal of Economic Policy

Statement of Purpose

The KDI Journal of Economic Policy (KDI JEP) is a professional journal published on a quarterly basis. The Journal publishes papers on the academic and policy issues related to the development of Korea's economy. The KDI Journal of Economic Policy welcomes innovative and insightful academic papers on all areas of economics with an emphasis on empirical analysis that contain solid policy implications. KDI JEP is published in English starting in 2015, volume 37 number 1.

The Journal aims to disseminate research outcomes and policy recommendations not only to experts at academia and research institutes but also to policy-makers and the general public. First published in March 1979, the original objective was to circulate ongoing- and past researches conducted in KDI, a leading economic think-tank of South Korea. Starting in August, 2001, the Journal has accepted manuscripts from outside in order to provide the readers more diverse perspectives on Korea's policy initiatives. The Journal now actively seeks and welcomes submissions by researchers at home and from abroad who have genuine interests in the Korean economy.

The contents of papers published in KDI JEP contain personal views of the authors, and thus do not represent the objectives of the Journal or the mission statements of KDI.

Editorial Board

Editor-in-Chief:	Kim, Taejong	(Professor at KDI School) tjkim@kdischool.ac.kr
Managing Editor:	Oh, Jiyoon	(Fellow at KDI) jiyoon.oh@kdi.re.kr
Editors:	Choi, Yongseok	(Professor at KyungHee University) choiy@khu.ac.kr
	Chun, Youngjun	(Professor at HanYang University) yjchun@hanyang.ac.kr
	Chun, Hyunbae	(Professor at Sogang University) hchun@sogang.ac.kr
	Chung, Wankyo	(Professor at Seoul National University) wankyo@snu.ac.kr
	Hahn, Chinhee	(Professor at Gachon University) chhahn@gachon.ac.kr
	Kim, Jungwook	(Fellow at KDI) awaker2@kdi.re.kr
	Lee, Jongyearn	(Professor at KDI School) jonlee@kdi.re.kr
	Kim, Jongil	(Professor at Dongguk University) jongil@dongguk.edu
	Lee, Chulhee	(Professor at Seoul National University) chullee@snu.ac.kr
	Shin, Kwanho	(Professor at Korea University) khshin@korea.ac.kr
	Shin, Sukha	(Professor at Sookmyung Women's University) shin89kr@sm.ac.kr
	Yoon, Kyungsoo	(Professor at Gachon University) yoonks@gachon.ac.kr
	Kim, Kwang-ho	(Professor at HanYang University) kwanghokim@hanyang.ac.kr
	Lee, Hangyong	(Professor at HanYang University) hl306@hanyang.ac.kr
Administration:	Park, Sejin	(Research Associate at KDI) dolkong0925@kdi.re.kr

KDI Journal of
Economic Policy

February 2022 VOL. 44, NO. 1

Articles

KDI Journal of Economic Policy 2022, 44(1):1–47
http://dx.doi.org/10.23895/kdijep.2022.44.1.1

Political Economy of Immigration and Fiscal Sustainability†

By JINWOOK HUR*

This paper introduces a politico-economic model with a welfare state and immigration. In this model, policies on taxes and immigration are determined through a plurality voting system. While many studies of fiscal implications of immigration argue that relaxing immigration policies can substitute for tax reforms in an aging economy, I show that the democratic voting procedure can dampen the effect of relaxing immigration policies as desired policy reforms are not always implemented by the winner of an election. This political economy results in three types of social welfare losses. First, the skill composition is not balanced at a socially efficient level because workers are motivated to maximize their wages. Second, older retirees implement excessive taxes to maximize the size of the welfare state. Third, the volume of immigration is lower than the optimal level given the incentive by young workers to regain political power in the future.

Key Word: Immigration, Political Economy, Fiscal Sustainability
JEL Code: E60, F22, H20, J61, D72

I. Introduction

In many developed economies, sustainability of the welfare state is an important issue. The aging of the baby-boomer generation and falling fertility rates make it more difficult to maintain welfare programs without reforms of fiscal policies. Tax reform is one option, but increasing the tax burden causes other controversies given the potential negative effects of this strategy on productivity and social welfare.

In this sense, some economists argue that relaxing immigration policies can be an effective alternative to tax reforms. According to these studies, a more liberal immigration policy for young and skilled workers can increase tax revenues and solve fiscal problems in many economies with a pay-as-you-go (PAYG) system of welfare. For example, Storesletten (2000) argues that admitting 1.6 million high-

* Associate Fellow, Korea Development Institute (E-mail: jhur@kdi.re.kr)
* Received: 2021. 11. 28
* Referee Process Started: 2021. 12. 1
* Referee Reports Completed: 2021. 12. 29
† This paper is written based on Chapter 1 of Hur, 2016, *Essays on Fiscal Policy*, University of California, Los Angeles.

skilled immigrants aged 40-44 years old annually can replace tax reforms and resolve the fiscal problems arising in the U.S. economy as the U.S. population ages.

In most studies concentrating on the fiscal implications of immigration policies, the policy variables are assumed to be exogenous. However, during the actual immigration policy decision-making process, there are many factors, especially those that are political, that can keep desired outcomes from being implemented. For instance, young citizens of working age may worry that their wages would decrease due to competition from immigrants, causing them to vote for parties against generous immigration policies. Therefore, the extent to which the desired reforms of immigration policies can be implemented through a democratic decision-making process is important.

In this paper, I analyze political equilibria in a simple overlapping-generation model with two skill levels in labor, proportional taxes on labor income, and a welfare state in the form of simple lump-sum transfers. The native population, or the electorate, is divided into three different political parties depending on their age and skill level. The labor income tax rate, volume of immigration, and skill composition of the immigrants are then determined by the winning candidate after plurality voting. As there are three different political parties, the possibilities of strategic voting and political coalitions are also considered.

Using this model economy, I demonstrate the pattern of different preferences on fiscal and immigration policies by different cohorts. Older retirees want to maximize tax revenue per capita because they do not have labor income; hence, they prefer the Laffer rate of labor income tax and a balanced skill composition between skilled and unskilled workers to maximize production efficiency. On the other hand, the preference of younger workers depends on their skill level. Specifically, skilled workers want immigrants to be unskilled, and unskilled workers want them to be skilled. Moreover, depending on the state of the economy, young workers may not want to allow the maximum number of immigrants in, although their instantaneous utility is maximized by allowing in as many as possible. If young workers allow too many immigrants in, the probability that they lose the election in the next period increases given that the demographic composition of the electorate will be biased toward the younger generation in the next period due to the high fertility of the immigrants, which may remove the current young cohort's opportunity to enjoy the benefit of the welfare state after they retire. Thus, older retirees want the maximum amount of immigration and the strongest welfare state as pure beneficiaries of the welfare program, while skilled young workers want zero taxes with no welfare program in most of the parameter regions. Unskilled young workers are mostly in the middle of the two groups in the sense that they do not mind taxes as much as skilled workers do. This emphasizes the role of unskilled workers as a "swing" cohort in elections.

It is also shown that the skill composition of the economy exhibits "cycling" dynamics. That is, when the skill composition is biased toward unskilled labor at the beginning of a period, unskilled workers are highly likely to sway the election, thereby approving of a high volume of skilled immigrants and balancing the skill composition. Similarly, when the skill composition leans too much toward skilled labor, skilled workers are more likely to sway the election, and the winning candidate will allow a high volume of unskilled immigrants. Since the optimal skill composition of workers is to have more skilled workers than unskilled workers in an economy with the simple Cobb-Douglas technology, the democratic procedure of plurality voting causes welfare losses due to the gap between the optimal skill composition of immigrants and the realized skill composition of immigrants, which tends to be lower than the optimal level.

I also show other factors causing welfare losses. First, the distortionary labor income tax in the model is a source of inefficiency. Because older retirees implement the Laffer rate whenever they win an election, this distortion is most severe when these retirees become the decisive political party. Second, the volume of immigration may be lower than the optimal level. Given that younger workers worry that they may not regain political power when they are older and lose their chance to obtain future social security benefits, young workers have an incentive to adjust the immigration volume to guarantee that they can win not only the election today but also those in the future.

This paper is in line with the literature on the macroeconomic implications of the political economy. Specifically, it builds on several studies on the political economy of immigration, such as Benhabib (1996), Dolmas and Huffman (2004) and Ortega (2005), and on studies of the political economy of social security systems, such as Boldrin and Rustichini (2000) and Galasso and Profeta (2002). Specifically, the main goal of this paper is to examine whether policy implications in the immigration literature[1] are feasible in a positive theory setup.

This study technically follows the analyses of Ortega (2005) and Suwankiri *et al.* (2016). Ortega (2005) introduces a dynamic politico-economic model with immigration and skill upgrading. In his model, the voting decision of the electorate affects the skill composition of the economy, thereby influencing not only the skill complementarity but also the skill composition of the future electorate. Therefore, voters confront the tradeoff between their skill premium and their future political power. On the other hand, the work by Suwankiri *et al.* (2016) abstracts from skill complementarity by assuming that the production function is linear in skilled and unskilled labor. Instead, their work concentrates on a more complicated political situation in which there exists a redistributive welfare state and three political parties. The three parties, formed according to skill levels and ages, make voting decisions in a plurality voting system and, therefore, potentially vote strategically. This paper builds on the techniques introduced in both of these papers. In a state of political equilibrium, each young worker considers not only the effect of the election result on his labor income but also how it will affect his political power when he becomes older and retires.

This paper proceeds as follows. The model economy is described in section II. The political procedure of policy implementation is explained in section III. Section IV describes the political equilibria of the model. In section V, the dynamics of the model and the welfare analysis are discussed. Section VI concludes the paper.

II. Model Economy

The model builds on the dynamic overlapping generation model as in Suwankiri

[1]As explained above, many studies report that relaxing immigration policies can reduce fiscal problems in developed economies. Among others, see Auerbach and Oreopoulos (1999) and Storesletten (2000).

et al. (2016) with the application of the democracy model with strategic voting behaviors in Besley and Coate (1997; 1998). In this model, there are two generations (the young and the old) and two skill levels (skilled and unskilled). Each individual works only in the first period of his life and retires in the second and last period. The government levies proportional labor income taxes from young workers and spends lump-sum transfers via a pay-as-you-go (PAYG) system. Furthermore, the government sets the tax rate, the volume of immigration, and the skill composition of immigrants by democratic voting for each period.

In this model, each period is divided into three sub-periods. In the first sub-period, native individuals - both young and old - hold a plurality election, and the winner of the election decides on the labor income tax rate (τ_t), the volume of immigration (μ_t), and the skill composition of immigrants (σ_t).[2] In the second sub-period, immigrants enter the economy as determined by the winning candidate. In the third and final sub-period, production, taxation, transfers, and consumption take place.

A. *Preference*

The lifetime utility function of a young worker with skill level $i \in \{s,u\}$ is assumed to be quasi-linear in terms of consumption, as in

$$U^y\left(c_t^{y,i}, l_t^i, c_{t+1}^{o,i}\right) = c_t^{y,i} - \frac{(l_t^i)^{1+1/\nu}}{1+1/\nu} + \beta c_{t+1}^{o,i},$$

where l_t^i is the labor supply of skill level $i \in \{s,u\}$, β is the discount factor for the next period, and ν is the Frisch elasticity of the labor supply.

The preference of an old retiree simply consists of the utility from today's consumption because he does not have a future continuation value. Thus, the utility function is given by

$$U^o\left(c_t^{o,i}\right) = c_t^{o,i}.$$

The budget constraints of the young and the old are given by

$$c_t^{y,i} + d_t^i \le (1-\tau_t) w_t^i l_t^i + b_t,$$

$$c_t^{o,i} \le (1+r_t) d_{t-1}^i + b_t,$$

where τ_t is the proportional labor income tax rate at period t, and d_t and b_t are savings and lump-sum transfers from the government, respectively. w_t represents the wage rate.

Given that the utility function is quasi-linear in terms of consumption, the Euler

[2]The process of voting is described in detail in section III.

equation pins down the equilibrium interest rate at a constant, i.e.,

$$r_t = \frac{1}{\beta} - 1 \text{ for all } t.$$

At this r_t, young workers are indifferent with respect to the amount of savings, as the marginal benefit of savings offsets exactly the marginal cost of savings such that any amount of savings can be an equilibrium solution. Throughout this paper, we will only consider the case of $d_t^i = 0$ for all $i \in \{s, u\}$ for simplicity. This makes old retirees simply consume the lump-sum transfer regardless of their skill level, allowing this setup to reduce the two groups of old retirees (skilled & unskilled) into a single group.

The labor supply function of a young worker with skill level $i \in \{s, u\}$ is derived as follows:

$$l_t^i = [(1-\tau_t) w_t^i]^{\nu}$$

Therefore, the indirect lifetime utility of each cohort is expressed as shown below.

$$\begin{aligned} v_t^i &= c_t^{y,i} - \frac{(l_t^i)^{1+\frac{1}{\nu}}}{1+\frac{1}{\nu}} + \beta c_{t+1}^{o,i} \\ &= \frac{[(1-\tau_t) w_t^i]^{1+\nu}}{1+\nu} + b_t + \beta b_{t+1} \end{aligned} \tag{1}$$

$$V_t^o = b_t \tag{2}$$

B. *Production Technology*

This model is abstracted from capital accumulation, meaning that skilled and unskilled labor will become the only factors of production.[3] The output is produced according to the Cobb-Douglas technology,

$$Y_t = (L_t^s)^{\alpha} (L_t^u)^{1-\alpha},$$

where L_t^i is the aggregate labor input of skill level $i \in \{s, u\}$ and α is assumed to be greater than 0.5. The labor demand equations for the two skill groups

[3]An alternative setup for this is to assume that capital stock per capita is fixed and that immigrants bring the same amount of capital stock when they are approved to immigrate.

corresponding to this production function are thus expressed as shown below.

$$w_t^s = \alpha Y_t (L_t^s)^{-1}$$

$$w_t^u = (1-\alpha) Y_t (L_t^u)^{-1}$$

C. *Immigration and Demographics*

Before production takes place in a period, the government sets the volume and the skill composition of immigrants. For simplicity, it is assumed that only young people are allowed to immigrate; thus, all immigrants supply labor immediately upon their entrance into the economy. Then, the aggregate labor supply can be characterized by the following formulae:

$$L_t^s = [s_t + \sigma_t \mu_t] N_t l_t^s,$$

$$L_t^u = [1 - s_t + (1-\sigma_t)\mu_t] N_t l_t^u,$$

in which N_t is the total size of the native young group of individuals and $s_t \in [0,1]$ is the fraction of native-born skilled workers in the labor force in the beginning of period t. $\mu_t \in [0, \overline{\mu}]$ the ratio of approved immigrants to the native-born young population, thereby governing the volume of immigration.[4] Finally, $\sigma_t \in [0,1]$ stands for the fraction of skilled immigrants in the group of new immigrants.

Also, there are certain additional assumptions for convenience. First, the skill levels of offspring follow the parent. Second, the offspring of immigrants are considered native, while the immigrants themselves are not.[5] Then, the demographic dynamics can be described by the following equations:

$$N_{t+1} = [1 + n + (1+m)\mu_t] N_t \quad \text{(level of population)}$$

$$s_{t+1} N_{t+1} = [(1+n)s_t + (1+m)\sigma_t \mu_t] N_t \quad \text{(skill composition)}$$

where n and m represent the population growth rates for natives and migrants, respectively.[6] Combining these two equations, we obtain the following law of motion for the skill composition of the natives:

[4]The volume of immigration has a finite upper bound because otherwise the choice of μt can be infinity in a number of political equilibria, as the production function is a constant-return-to-scale function. In an economy with DRS technology, μt has a finite value even without an exogenous upper bound. In this DRS case, however, the state of the economy depends not only on the relative population size of the voters but also the absolute number of young workers, thereby increasing the dimension of the state vector by one.

[5]In other words, immigrants cannot vote for their lifetimes, while their descendants are legal voters.

[6]In many developed countries, it is consistently observed in census data that the fertility rate is higher for immigrants than for natives, implying $m > n$. One recent example is the report of U.S. Census Bureau by Monte and Ellis (2014).

$$s_{t+1} = \frac{(1+n)s_t + (1+m)\sigma_t \mu_t}{1+n+(1+m)\mu_t} \tag{3}$$

Equation (3) demonstrates the role of immigration policies. For instance, $\sigma_t < s_t$ implies that the fraction of skilled immigrants in the entire group of immigrants exceeds the fraction of skilled workers in the group of native young workers. Then, the fraction of skilled workers in the native young workers will increase in the next period given the assumption that the descendants succeed their parents' skill levels. Also, if $\mu_t = 0$ (no immigrants accepted), then s_t obviously remains the same in the next period. [7] As $\mu_t \to \infty$ (infinitely many immigrants are approved), $s_{t+1} \to \sigma_t$, implying that the skill composition of the natives will converge to the skill composition of accepted immigrants as more immigration is accepted.

D. *Government and the Welfare State*

The government levies proportional labor income taxes at the rate determined by the winner of the vote. The tax revenue is transferred equally regardless of age (young and old), skill level (skilled and unskilled), and/or residence status (natives and immigrants). The government is assumed to balance the budget every period; thus, there is no government debt. The lump-sum transfer is then expressed as

$$b_t = \frac{\tau_t \left[(s_t + \sigma_t \mu_t) w_t^s N_t l_t^s + (1 - s_t + (1-\sigma_t)\mu_t) w_t^u N_t l_t^u \right]}{(1+\mu_t)N_t + (1+\mu_{t-1})N_{t-1}}$$

in which the numerator is the total tax revenue and the denominator stands for the total population, including both natives and immigrants.

E. *Equilibrium-Characterizing Equations*

The general equilibrium of this economy is characterized by the following nine equations.

$$l_t^s = [(1-\tau_t) w_t^s]^\nu \tag{4}$$

$$l_t^u = [(1-\tau_t) w_t^u]^\nu \tag{5}$$

$$Y_t = (L_t^s)^\alpha (L_t^u)^{1-\alpha} \tag{6}$$

[7]In other words, immigration is the only method of changing the aggregate skill composition in this model, and there is no skill acquisition or loss. For studies focusing on the transition of the aggregate skill level or task composition through immigration, readers can refer to Hunt and Gauthier-Loiselle (2010) and Peri (2012), among others.

$$(7) \qquad w_t^s = \alpha Y_t (L_t^s)^{-1}$$

$$(8) \qquad w_t^u = (1-\alpha) Y_t (L_t^u)^{-1}$$

$$(9) \qquad L_t^s = [s_t + \sigma_t \mu_t] N_t l_t^s$$

$$(10) \qquad L_t^u = [1 - s_t + (1-\sigma_t)\mu_t] N_t l_t^u$$

$$(11) \qquad b_t = \frac{\tau_t \left[(s_t + \sigma_t \mu_t) w_t^s N_t l_t^s + (1 - s_t + (1-\sigma_t)\mu_t) w_t^u N_t l_t^u \right]}{(1+\mu_t) N_t + (1+\mu_{t-1}) N_{t-1}}$$

$$(12) \qquad N_t = [1 + n + (1+m)\mu_{t-1}] N_{t-1}$$

In this system of equations, the following variables exist:

- Nine endogenous variables: l_t^s, l_t^u, w_t^s, w_t^u, Y_t, L_t^s, L_t^u, b_t, *and* $\frac{N_t}{N_{t-1}}$
- Two state variables: μ_{t-1} *and* s_t, *and*
- Three policy variables: τ_t, μ_t, *and* σ_t

Note that all nine endogenous variables are uniquely solvable given the state variables and the policy variables for period t. It is obvious that the decisions of output produced and labor input are purely static because there is no inter-temporal component in production, such as capital stock in a neoclassical model. However, the demographic structure of this economy is dynamic as the ratio of the native young to the native old at time $t+1$ (N_{t+1}/N_t) and the skill composition of the natives at time $t+1$ depend on today's decisions with regard to immigration policy (μ_t and σ_t). Therefore, the only dynamic component of this model will be the voting decision, which will be discussed in detail in section III.

The equations are further normalized by dividing by N_t, which represents the population of native young workers. The equations above can then be rewritten as follows:

$$(13) \qquad l_t^s = [(1-\tau_t) w_t^s]^{\nu}$$

$$(14) \qquad l_t^u = [(1-\tau_t) w_t^u]^{\nu}$$

$$(15) \qquad \tilde{Y}_t = (\tilde{L}_t^s)^{\alpha} (\tilde{L}_t^u)^{1-\alpha}$$

$$w_t^s = \alpha \tilde{Y}_t (\tilde{L}_t^s)^{-1} \tag{16}$$

$$w_t^u = (1-\alpha) \tilde{Y}_t (\tilde{L}_t^u)^{-1} \tag{17}$$

$$\tilde{L}_t^s = [s_t + \sigma_t \mu_t] N_t l_t^s \tag{18}$$

$$\tilde{L}_t^u = [1 - s_t + (1-\sigma_t)\mu_t] N_t l_t^u \tag{19}$$

$$b_t = \frac{\tau_t \left[w_t^s \tilde{L}_t^s + w_t^u \tilde{L}_t^u \right]}{(1+\mu_t) + (1+\mu_{t-1}) \dfrac{N_{t-1}}{N_t}} \tag{20}$$

$$\frac{N_t}{N_{t-1}} = 1 + n + (1+m)\mu_{t-1} \tag{21}$$

where the variables with tildes are the original variables divided by N_t. For example, $\tilde{L}_t^s$ is the aggregate input of skilled labor per native worker.

III. Political Decision

As discussed in Section II, the policies pertaining to the income tax rate (τ_t), the volume of immigration (μ_t), and the skill composition of immigrants (σ_t) are decided at the beginning of each period by the winner of plurality voting. In this section, the political decision-making process is explained in detail. This modeling of the voting process is basically identical to the application of Besley and Coate (1997) by Suwankiri *et al.* (2016), while there are slight differences in the electorate composition.[8] In modeling the implementation procedure of the winning candidate, the idea of Markov-perfect equilibria is applied.

A. *Political Decision Process*

The political decision-making process of each period consists of the following three steps:

- Step 1: One individual is selected as a candidate in each of three political

[8]For example, Suwankiri *et al.* (2016) simplifies the political equilibria in two aspects. First, they assume that only the smallest party out of the three can choose to vote strategically, thereby eliminating the possibility that the smallest party wins the election. Second, they assume that immigrants have the right to vote when they become old. This simplifies the equilibrium because the young-to-old ratio in the electorate is consistent with that of the whole pre-immigration population. Both simplifications are relaxed in this paper; the smallest party can win an election if the members of the second-largest party have an incentive to vote strategically and the smallest party's members do not have incentive to vote for the second-largest party. Also, immigrants do not have the right to vote for their lifetimes in this paper.

parties (the "Skilled Young," the "Unskilled Young," and the "Old Retirees").

- Step 2: One and only one of the three candidates is elected by one single plurality voting event.[9]
- Step 3: The winning candidate chooses the set of policy variables (τ_t, μ_t, σ_t).

For convenience, these steps are described in reverse order in the following sub-sections.

1. Policy Implementation of the Winning Candidate

Once one of the three candidates is elected by plurality voting, the third and final step of the political process is decision making by the winning candidate. The winning candidate implements his preferred policies on taxation and immigration. As there is no mechanism to resolve commitment problems, all voters recognize *ex ante* that the winning candidate will pursue his optimum after being elected, and all other pledges are considered as 'cheap talk' under any voting equilibrium. Let the optimal policy of the decisive candidate at time t be denoted by a triplet $(\tau_t^d, \mu_t^d, \sigma_t^d) \equiv \Phi_t^d(\mu_{t-1}, s_t)$ for the winning candidate's party $d \in \{s,u,o\}$. Then, Φ_t^d solves the following maximization problem:

$$\Phi_t^d(\mu_{t-1}, s_t) = \text{argmax}_{\tau_t,\mu_t,\sigma_t} V^d(\tau_t, \mu_t, \sigma_t, \Phi_{t+1}),$$

where Φ_{t+1} is the perception of today's winning candidate of the policy variables in the next period. Obviously there can be multiple different rules of perception Φ_{t+1} that satisfy sub-game perfection; thus, the complexity of this optimization problem depends on the multiplicity of the expectation of the policy rule of the winning candidate in the next period. If there are multiple reasonable perceptions on future policies, this period's winning candidate must then consider multiple possibilities stemming from his decision today.

In order to refine this type of complexity and reduce the number of perceived future policy rules, this paper relies on the Markov-perfect property.[10] That is, the perception function Φ_{t+1} is assumed to be a function of the state variable for the period $t+1$ only without depending on anything else, such as signals to eliminate any types of "sunspot" equilibria. With the introduction of this Markov-perfect property, the winning candidate expects that the next period's winning candidate will make a decision according to a rule identical to those of individuals living today. Formally, the winning candidate's problem is expressed as follows

[9]In other words, there is no second voting such as, for example, the run-off primary in the U.S. presidential election.

[10]The Markov-perfection property in this paper is taken directly from Suwankiri *et al.* (2016), which builds on the Markov-perfection setup used in previous works on short-term government and political economy, such as Krusell and Rios-Rull (1996), Grossman and Helpman (1998), and Hassler *et al.* (2003).

$$\Phi^d(\mu_{t-1},s_t)=\text{argmax}_{\tau_t,\mu_t,\sigma_t}V^d(\tau_t,\mu_t,\sigma_t,\Phi^d(\mu_t,s_{t+1}))$$
$$\text{subject to } s_{t+1}=\frac{(1+n)s_t+(1+m)\sigma_t\mu_t}{1+n+(1+m)\mu_t}$$

if the winning candidate belongs to the cohort $d\in\{s,u\}$. Intuitively, this means that today's winning candidate decides on policy variables with the belief that the winning candidate of the next period will decide on policy variables in exactly the same manner used by this period's candidates, and, in any equilibrium, this belief is precisely realized.

If the winning candidate is an old retiree, his decision rule is simple given that his utility consists only of the consumption of the lump-sum payment this period. Specifically, the decision when winning the election is given by the solution of the following problem:

$$\Phi^o(\mu_{t-1},s_t)=\text{argmax}_{\tau_t,\mu_t,\sigma_t}V^o(\tau_t,\mu_t,\sigma_t)=b_t,$$

where b_t is represented as a function of only the policy variables and the states by solving (4)~(12).

2. Voting Process

One step backward from the policy implementation procedure in section III-A-1, now assume that one candidate is selected from each of the three political parties. The winning candidate will be determined by single plurality voting, meaning that the result of voting mainly depends on the sizes of the political parties. However, when analyzing the voting process, the possibility of strategic voting must be considered, as there are three parties. That is, if a voter believes that the candidate from his cohort has no possibility of winning the election, he may have an incentive to vote for the next-best candidate in order to prevent the worst candidate from winning the election.

Also, in this model all of the voters in the same cohort have identical preferences, implying that voters in the same political party vote identically. This setup means we must concentrate on the behavior of a representative voter instead of analyzing the complicated voting equilibrium within a cohort.

Formally, let $e_t=(p_t^{is},p_t^{iu},p_t^{io})\in\Delta^2$ denote the voting decision of a representative voter in cohort $i\in\{s,u,o\}$, where p_t^{ij} stands for the probability that the voter in party i votes for the candidate from party j, and Δ^2 is a 2-simplex. Then, voting profile $\mathfrak{e}_t=(\mathfrak{e}_t^s,\mathfrak{e}_t^u,\mathfrak{e}_t^o)$ represents voting equilibrium if it solves the following problem: [11]

[11]It is implicitly assumed that there is no additional cost for the electorate to participate in voting, and only equilibria in which none of the voters abstains are considered.

$$\mathfrak{e}_t^i = \operatorname{argmax}_{e_t^j} \left\{ \sum_{j \in \{s,u,o\}} \mathcal{P}^j \left(e_t^i, \mathfrak{e}_t^{-i} \right) V^i \left(\Phi_t^j, \Phi_{t+1}, \mathfrak{e}_{t+1} \right) \right\}$$

for all $i \in \{s,u,o\}$, where $\mathcal{P}^j(e_t)$ is the probability that the candidate from party j wins the election given voting profile e_t. That is, under a state of voting equilibrium, each voter chooses for whom to vote so that the expected value of himself is maximized given the other voter's decision and the perceived future policy of the winning candidate for the next period.

In order to reduce the possibility of multiple equilibria for a given set of state variables, I only consider the equilibria in which none of the voting decisions is a weakly dominated strategy. In addition, I assume the following tie-breaking rules for the election:

i) If there are two parties with the equally the largest number of votes, the candidate whose votes are from the fewest number of cohorts then becomes the winner.

ii) If two groups are willing to form a coalition by voting for either candidate of the two, the coalition follows the larger party of the two. If their sizes are equal, the tie is broken with equal probability.

iii) If the sizes of all three parties are identical, the tie is broken with equal probability.

These tie-breaking assumptions, together with the assumption that equilibria with weakly dominated strategies are excluded, guarantee that there exists a unique pure-strategy voting equilibrium for each state vector of (μ_{t-1}, s_t). In all of these equilibria, the members of the largest (not necessarily majority) cohort always vote for the candidate from their own party regardless of the formation of a coalition; thus, a political coalition is potentially formed only by the two smallest parties.[12]

B. *Sizes of Political Groups*

Because the size of each of the political groups (groups of skilled young workers, unskilled young workers, and old retirees) is a critical component of the political decision,[13] I start by looking at which group becomes the "largest" or the "majority" for all possible combinations of the state variables (μ_{t-1}, s_t).

The group of native skilled young workers becomes the largest group under the following conditions:

(i) $s_t > \dfrac{1}{2}$, and

(ii) $[1 + n + (1+m)\mu_{t-1}]s_t > 1$

[12]See Lemma 4 in Appendix C for proof.

[13]As emphasized before, note that the size of the electorate is not the only component, as whether a coalition is formed or not depends not only on the relative size of the two smallest groups but also on their policy profile preferences with regard to the others parties.

Specifically, the group of skilled young workers becomes the largest if it is larger than the group of unskilled young workers and the group of native old retirees. It is clear that the first condition dominates if $n+(1+m)\mu_{t-1}>1$ while the second condition dominates otherwise.

Similarly, the group of unskilled young workers is the largest group under the following two conditions:

(i) $1-s_t>\frac{1}{2}$, and

(ii) $[1+n+(1+m)\mu_{t-1}](1-s_t)>1$

and the first condition dominates if $n+(1+m)\mu_{t-1}>1$, with the second condition dominating otherwise.

The political party of the native old retirees becomes the largest under the following condition:

$$\max\{[1+n+(1+m)\mu_{t-1}]s_t, [1+n+(1+m)\mu_{t-1}](1-s_t)\}<1$$

Therefore, the relative sizes of the political groups depend on the combination of the state (μ_{t-1}, s_t) and the parameters m (fertility rate of the immigrants) and n (fertility rate of the natives). Specifically, the relative size of the electorate groups depends on the value of $n+(1+m)\mu_{t-1}$, which governs the ratio between the native young and the native old. The following four cases categorize the relative size of each political group at different skill compositions s_t. The important fact here is that the largest group does not always win owing to the possibility that the two smaller parties form a coalition. Accordingly, this section does not describe who the winning candidate of the election is.

Case I. $\frac{N_t}{N_{t-1}}>3$

In this case, the size of the old retirees' party is less than a third of the size of the population of native young workers, meaning that this group can never be the largest regardless of the skill composition of the young generation. Figure 1 displays the relative size of each political group depending on the value of $s_t\in[0,1]$ in this case of $\frac{N_t}{N_{t-1}}>3$. The old retirees' party constitutes the smallest group unless the skill composition is extremely biased toward either skilled or unskilled labor, and either skilled or unskilled workers always become the largest group.

s_t: 0 | $\frac{1}{1+n+(1+m)\mu_{t-1}}$ | $\frac{1}{2}-\frac{1}{2}\frac{1}{1+n+(1+m)\mu_{t-1}}$ | $\frac{1}{2}$ | $\frac{1}{2}+\frac{1}{2}\frac{1}{1+n+(1+m)\mu_{t-1}}$ | $1-\frac{1}{1+n+(1+m))\mu_{t-1}}$ | 1

Largest	Unskilled*	Unskilled*	Unskilled	Skilled	Skilled*	Skilled*
2nd Largest	Old	Skilled	Skilled	Unskilled	Unskilled	Old
Smallest	Skilled	Old	Old	Old	Old	Unskilled

FIGURE 1. RELATIVE SIZES OF POLITICAL GROUPS FOR DIFFERENT SKILL COMPOSITIONS - CASE I

Note: * More than 50%.

Case II. $2<\dfrac{N_t}{N_{t-1}}\leq 3$

This case is similar to Case I in that the old retirees never have a chance to be the largest group given that the population of the native old retirees is still less than half of the population of the native young workers.

s_t: 0 | $\frac{1}{2}-\frac{1}{2}\frac{1}{1+n+(1+m)\mu_{t-1}}$ | $\frac{1}{1+n+(1+m)\mu_{t-1}}$ | $\frac{1}{2}$ | $1-\frac{1}{1+n+(1+m))\mu_{t-1}}$ | $\frac{1}{2}+\frac{1}{2}\frac{1}{1+n+(1+m)\mu_{t-1}}$ | 1

Largest	Unskilled*	Unskilled	Unskilled	Skilled	Skilled	Skilled*
2nd Largest	Old	Skilled	Skilled	Unskilled	Unskilled	Old
Smallest	Skilled	Old	Old	Old	Old	Unskilled

FIGURE 2. RELATIVE SIZES OF POLITICAL GROUPS FOR DIFFERENT SKILL COMPOSITIONS - CASE II

Note: * More than 50%.

Case III. $1<\dfrac{N_t}{N_{t-1}}\leq 2$

This is likely to be the most interesting case. In this case, the native old population is still smaller than the population of native young workers but larger than half of it. Therefore, the group of old retirees has a chance to be the largest group if the number

s_t: 0 | $\frac{1}{2}-\frac{1}{2}\frac{1}{1+n+(1+m)\mu_{t-1}}$ | $1-\frac{1}{1+n+(1+m)\mu_{t-1}}$ | $\frac{1}{2}$ | $\frac{1}{1+n+(1+m))\mu_{t-1}}$ | $\frac{1}{2}+\frac{1}{2}\frac{1}{1+n+(1+m)\mu_{t-1}}$ | 1

Largest	Unskilled*	Unskilled	Old	Old	Skilled	Skilled*
2nd Largest	Old	Old	Unskilled	Skilled	Old	Old
Smallest	Skilled	Skilled	Skilled	Unskilled	Unskilled	Unskilled

FIGURE 3. RELATIVE SIZES OF POLITICAL GROUPS FOR DIFFERENT SKILL COMPOSITIONS - CASE III

Note: * More than 50%.

of skilled young workers is close enough to the number of unskilled workers. However, because the population of old retirees is still less than half of the total native population as a whole, the old retirees' group cannot constitute the majority for any value of s_t.

Case IV. $\dfrac{N_t}{N_{t-1}} \leq 1$

In this case, the native old population exceeds that of the native young, meaning that the old retirees' party always constitutes the majority of the electorate.

s_t	$0 < s_t < \frac{1}{2}$	$\frac{1}{2} < s_t < 1$
Largest	Old*	Old*
2nd Largest	Unskilled	Skilled
Smallest	Skilled	Unskilled

FIGURE 4. RELATIVE SIZES OF POLITICAL GROUPS FOR DIFFERENT SKILL COMPOSITIONS - CASE IV

Note: * More than 50%.

IV. Political Equilibria

As mentioned in section II, while the political decision of the policy-makers is forward-looking, the equilibrium allocation and prices are purely static in this model because there are no inter-temporal components linking different periods, such as the bond market or capital accumulation. Therefore, the equilibrium allocation can be solved analytically with only a few simplifying conditions. For instance, the equilibrium allocation is solved in a closed form if the following hold true:

(i) The marginal rate of substitution between consumption and leisure is multiplicative in labor input.

(ii) The production function is multiplicative in both types of labor (skilled and unskilled).

Therefore, a Cobb-Douglas production function and our functional form of the utility function enable the equilibrium to be solvable in closed form for a given state variable (μ_{t-1}, s_t) and a set of policy variables $(\tau_t, \mu_t, \sigma_t)$.[14] Specifically, the lump-sum transfer per capita and the wage of each type of labor are represented with respect to only (μ_{t-1}, s_t) and $(\tau_t, \mu_t, \sigma_t)$, as follows:

[14]For more details about the procedure of computing the closed-form equilibrium, see Appendix A.

$$(22)\qquad b_t = \frac{\tau_t(1-\tau_t)^{\nu}\alpha^{\alpha\nu}(1-\alpha)^{(1-\alpha)\nu}(s_t+\sigma_t\mu_t)^{a}[1-s_t+(1-\sigma_t)\mu_t]^{1-\alpha}}{1+\mu_t+(1+\mu_{t-1})[1+n+(1+m)\mu_{t-1}]}$$

$$(23)\qquad \omega_t^s = \alpha^{\frac{1+\alpha\nu}{1+\nu}}(1-\alpha)^{\frac{(1-\alpha)\nu}{1+\nu}}(s_t+\sigma_t\mu_t)^{-\frac{1-\alpha}{1+\nu}}[1-s_t+(1-\sigma_t)\mu_t]^{\frac{1-\alpha}{1+\nu}}$$

$$(24)\qquad \omega_t^u = \alpha^{\frac{\alpha\nu}{1+\nu}}(1-\alpha)^{\frac{1+(1-\alpha)\nu}{1+\nu}}(s_t+\sigma_t\mu_t)^{\frac{\alpha}{1+\nu}}[1-s_t+(1-\sigma_t)\mu_t]^{-\frac{\alpha}{1+\nu}}$$

Therefore, the lifetime utility of individuals in each cohort is also represented only with the state variables and the policy variables by (1) and (2).

A. *Equilibria without Forward-looking*

Before analyzing the equilibria of the full model, I start by demonstrating the equilibria without forward-looking behavior by individuals, which is simply equivalent to assuming $\beta = 0$. This simplified version is considered because

i) All equilibria associated with allocation and voting are solvable in a closed form, meaning that we can analyze all details of the equilibria, and

ii) More importantly, this is a good benchmark for analyzing how the forward-looking motive affects the decisions of the winning candidate and thus the voters' choice after all.

By assuming $\beta = 0$, this model becomes abstract from the inter-temporal components, and the winning candidate's policy decisions become a simple optimization problem of maximizing his instantaneous utility. First, we analyze the preference of each cohort with regard to policy variables, after which we examine the political equilibrium.

1. Preferences of Candidates on Policy Variables

After substituting the equilibrium allocation with closed-form solutions, the utility of an individual in each cohort can be represented by the following indirect utility functions with respect to policy variables and states:

$$V_t^i = b_t + \frac{\left[(1-\tau_t)w_t^i\right]^{1+\nu}}{1+\nu},$$

$$V_t^o = b_t,$$

where b_t and w_t^i take the form in (22)~(24) for $i \in \{s,u\}$. Note that the continuation value is zero for young workers as $\beta = 0$ is assumed. The ideal policies of each political party, therefore, can be analyzed with these indirect utility functions. The following proposition displays the old retiree's preferred policies.

Proposition 1. *If an old retiree wins the election, the implemented policy, as a result of his optimum, is characterized as follows:*[15]

$$(\tau_t, \mu_t, \sigma_t) = \begin{cases} \left(\frac{1}{1+\nu}, \bar{\mu}, 1\right) & \text{if } 0 \le s_t < \alpha - (1-\alpha)\bar{\mu} \\ \left(\frac{1}{1+\nu}, \bar{\mu}, \alpha + \frac{\alpha - s_t}{\bar{\mu}}\right) & \text{if } \alpha - (1-\alpha)\bar{\mu} \le s_t < \alpha + \alpha\bar{\mu} \\ \left(\frac{1}{1+\nu}, \bar{\mu}, 0\right) & \text{if } \alpha + \alpha\bar{\mu} \le s_t \le 1 \end{cases}$$

Proof. See Appendix B.1.

The implication of this proposition is clear: as old retirees do not have any labor endowment, they are pure beneficiaries of the welfare state. Therefore, if the candidate from the old retirees wins the election, he will choose policy variables such that the tax revenue of the government per capita is maximized. Figure 5 displays the relationship between the tax rate and the lifetime value of the old retirees. The tax rate is decided at $1/(1+\nu)$, which is the Laffer rate of the economy. Regarding the skill composition of immigrants (σ_t), the old retiree chooses $\sigma_t = 1$ if s_t is small enough and $\sigma_t = 0$ if s_t large enough, , as production and thus tax revenue is maximized when the proportion of skilled workers in the labor force equals α, the exponent in the Cobb-Douglas production function. To gain a better understanding of this implication, we assume that $\bar{\mu} \to \infty$. Then, the decision of σ_t converges to α, which means that the old retirees want the skill composition of the economy's workers to be consistent with the Cobb-Douglas exponent and the wage rates of skilled and unskilled labor groups to be equalized. Figure 5 shows that

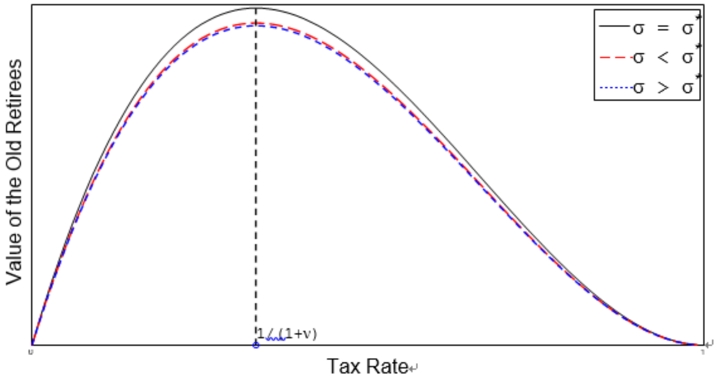


FIGURE 5. TAX RATES AND THE INDIRECT UTILITY OF THE OLD RETIREES

[15]Obviously, some of the ranges of s_t can be an empty set if $\bar{\mu}$ is large enough. For instance, if $\bar{\mu}$ is larger than $\alpha/(1-\alpha)$, the optimal σ_t always equals $\alpha + \frac{\alpha - s_t}{\bar{\mu}}$ because the first and the third intervals become empty sets.

the value of the old retirees is maximized when σ_t is optimal, as in Proposition 1. Finally, the immigration volume is decided at $\mu_t = \bar{\mu}$. This result relies on the assumption of constant-returns-to-scale technology. In other words, as long as the skill composition satisfies the optimality condition, the tax revenue is monotone increasing in the volume of immigration because the production efficiency is robust with regard to the scale of factor inputs.

The next two propositions show the optimality of the young workers with each skill level:

Proposition 2. *If a skilled young worker wins the election, the implemented policy, as a result of his optimum, is characterized as follows:*

$$(\tau_t, \mu_t, \sigma_t) = \begin{cases} (0, \bar{\mu}, 0) \;\; if \;\; 0 \le s_t < \alpha\left[1 + \bar{\mu} + (1+\mu_{t-1})\dfrac{N_{t-1}}{N_t}\right] \\ (\tau_t^s, \bar{\mu}, 0) \;\; if \;\; \alpha\left[1 + \bar{\mu} + (1+\mu_{t-1})\dfrac{N_{t-1}}{N_t}\right] < s_t \le 1 \end{cases}$$

where

$$\tau_t^s = \frac{s_t - \alpha\left[1 + \bar{\mu} + (1+\mu_{t-1})\dfrac{N_{t-1}}{N_t}\right]}{s_t(1+\nu) - \alpha\left[1 + \bar{\mu} + (1+\mu_{t-1})\dfrac{N_{t-1}}{N_t}\right]} \in \left(0, \frac{1}{1+\nu}\right)$$

Proof. See Appendix B.2.

Proposition 3. *If an unskilled young worker wins the election, the implemented policy, as a result of his optimum, is characterized as follows:*

$$(\tau_t, \mu_t, \sigma_t) = \begin{cases} (\tau_t^u, \bar{\mu}, 1) \;\; if \;\; 0 \le s_t < 1 - (1-\alpha)\left[1 + \bar{\mu} + (1+\mu_{t-1})\dfrac{N_{t-1}}{N_t}\right] \\ (0, \bar{\mu}, 1) \;\; if \;\; 1 - (1-\alpha)\left[1 + \bar{\mu} + (1+\mu_{t-1})\dfrac{N_{t-1}}{N_t}\right] < s_t \le 1 \end{cases}$$

where

$$\tau_t^u = \frac{1-s_t-(1-\alpha)\left[1+\bar{\mu}+(1+\mu_{t-1})\dfrac{N_{t-1}}{N_t}\right]}{(1-s_t)(1+\nu)-(1-\alpha)\left[1+\bar{\mu}+(1+\mu_{t-1})\dfrac{N_{t-1}}{N_t}\right]} \in \left(0, \frac{1}{1+\nu}\right)$$

Proof. The proof is omitted as it is symmetric to the proof of Proposition 2. Proof is provided simply by substituting s_t with $1-s_t$, σ_t with $1-\sigma_t$, and α with $1-\alpha$ in proposition 2.

The main implication of Propositions 2 and 3 is that young workers consider not only the lump-sum transfer but also their labor income. For example, when s_t is sufficiently small such that there are only a small number of skilled young workers in the economy, the wage rate of the skilled labor is so high that skilled young workers become net contributors to the welfare state, meaning that they prefer a zero tax rate. The upper panel in Figure 6 shows the utility of the skilled young workers for different tax rates and σ_t, the skill composition of the immigrants, when their optimal tax rate equals zero. Given a zero tax rate, the transfers also equal zero; thus, the utility of the skilled young workers depends only on their labor income. Because the wage rate for skilled labor is increasing in the input of unskilled labor due to skill complementarity, skilled workers prefer as many unskilled immigrants as possible.

The lower panel in Figure 6 displays the opposite case, i.e., where s_t is large enough. In this case, skilled workers now become net beneficiaries of the welfare state because there are already excessively many skilled workers such that the equilibrium wage for skilled labor is too small, meaning that skilled workers prefer a strictly positive tax rate τ_t^s. Nevertheless, their ideal tax rate is still smaller than the Laffer rate because they are in any case taxpayers, while the old retirees are not. In this case, how the skilled young workers select σ_t depends on two factors. First, their wage rate increases as more unskilled labor enters the economy, as in the previous case. Second, the direction of the changes of the lump-sum transfer depends

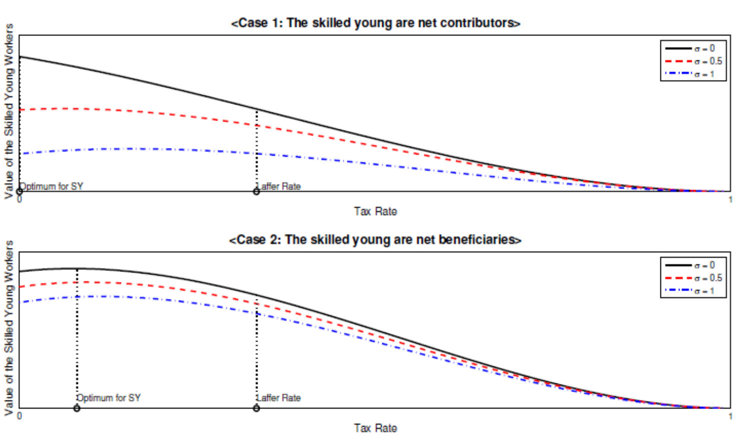


FIGURE 6. TAX RATES AND THE INDIRECT UTILITY OF THE SKILLED YOUNG WORKERS

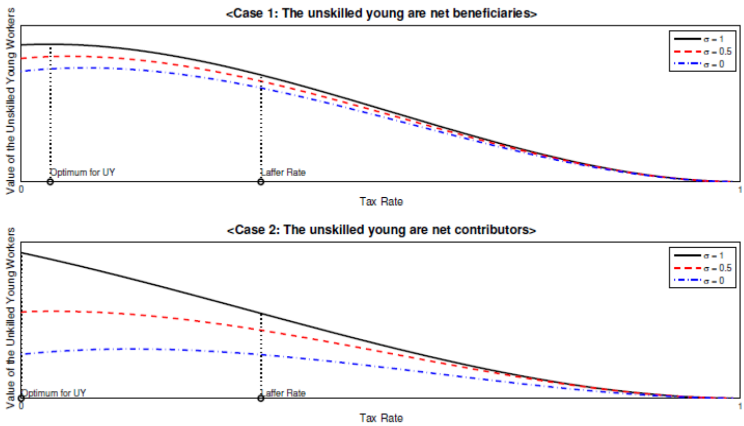


FIGURE 7. TAX RATES AND THE INDIRECT UTILITY OF THE UNSKILLED YOUNG WORKERS

on the pre-immigration skill composition of the economy. Specifically, the transfer increases in σ_t if s_t is relatively small and decreases if s_t is large. Regardless of the size of s_t, however, it can be shown that the wage effect dominates the transfer effect. Thus, skilled workers choose to approve of unskilled immigrants only.

The optimum level of unskilled young workers is symmetric to that of skilled workers, as shown in Figure 7. They set a zero tax rate when s_t is so large that the unskilled workers are net contributors to the welfare state, and they prefer a positive tax rate when s_t is small and they are net beneficiaries of the welfare state. In either case, unskilled workers prefer the maximum number of skilled immigrants.

In summary, old retirees want to maximize the government's tax revenue by levying a high tax rate (at the Laffer rate) and balancing the skill composition of workers for efficient production. On the other hand, young workers prefer a lower (even zero) tax rate and want all immigrants to have opposite skills to maximize their disposable labor income.

2. Political Equilibria

Essentially, the result of the election depends on the relative sizes of the three political parties, as described in Section III-B, while there is no guarantee that the largest party always wins the election given that the two smallest parties potentially form a coalition when the largest party's preferred policy is believed to be the worst for either of the smaller groups. Therefore, not only the relative size of the electorate but also the preference of each party with regard to the preferred policies of one another will determine the result of the election.

The pattern of political equilibria depends on the relative values of the parameters. For instance, a simple case is when $\frac{N_t}{N_{t-1}} > 1$ (the size of the native old population exceeds that of the native young population), in which the candidate of the old retirees' party always wins the election by a majority. While equilibria can be represented in a closed form in this non-forward-looking case, I will describe only one form of equilibrium with interesting implications rather than considering every possible case of parameter values and states. Specifically, the analysis concentrates on the case in which

i) $1 < \frac{N_t}{N_{t-1}} \leq 2$

ii) α is sufficiently large and $\bar{\mu}$ is sufficiently small such that $\frac{1}{2} < \alpha - (1-\alpha)\bar{\mu} < \alpha(1+\bar{\mu})$

iii) μ_{t-1} is sufficiently large such that $\alpha - (1-\alpha)\left[\bar{\mu} + (1+\mu_{t-1})\frac{N_{t-1}}{N_t}\right] < \frac{1}{2}$

In this case, the patterns of political equilibrium can be described by the range of s_t, as follows:

Case I. $0 \leq s_t \leq \frac{1}{2} - \frac{1}{2} \cdot \frac{N_{t-1}}{N_t}$

In this case, the population of native unskilled young workers constitutes the majority, meaning that the winning candidate's policy decision maximizes the lifetime value for the unskilled young. Because s_t is so small in this case that the unskilled young workers become net beneficiaries of the welfare state, the choice of the winner's policy variables is $\Phi^u(\mu_{t-1}, s_t) = (\tau_t^u, \bar{\mu}, 1)$, where τ_t^u is defined as in Proposition 3. Whether a political coalition arises is immaterial as it cannot take the majority.

Case II. $\frac{1}{2} - \frac{1}{2} \cdot \frac{N_{t-1}}{N_t} \leq s_t \leq 1 - \frac{N_{t-1}}{N_t}$

In this case, the party of the unskilled young workers is still the largest but does not secure a majority. However, the smallest parties (the skilled young workers and the old retirees) do not have any incentive to form a coalition because they perceive that the policies of the other party are the worst. For the old retirees, the preferred policy of the skilled young includes a zero tax rate, which is the worst for the old retirees as it leads to a zero welfare state. For the skilled young workers, the policies of σ_t for both the unskilled young workers and the old retirees are equally the worst because both prefer $\sigma_t = 1$, while the old retirees even prefer a higher tax rate. Therefore, for both the old retirees and the skilled young workers, it is the next-best choice to let the unskilled workers win the election, thereby forgoing the possibility of a political coalition. As a result, the unskilled young workers' party wins the election and the winner implements $(\tau_t, \mu_t, \sigma_t) = (\tau_t^u, \bar{\mu}, 1)$ according to Proposition 3.

Case III. $1-\frac{N_{t-1}}{N_t} < s_t \leq \alpha-(1-\alpha)\left[\bar{\mu}+(1+\mu_{t-1})\frac{N_{t-1}}{N_t}\right]$

In this case, the population of old retirees is the largest, the unskilled workers are the second, and the skilled workers form the smallest group. Therefore, it is important as to whether the skilled and unskilled workers have an incentive to form a coalition with each other. It is clear that the skilled workers have an incentive to vote strategically for the unskilled candidate. The preferred tax rate of the unskilled workers is $\tau^u \in (0, 1/(1+\nu))$, which is advantageous for the skilled workers as it levies fewer taxes than the old retirees would levy. Both the old retirees and the unskilled workers prefer the maximum number of skilled immigrants. Therefore, the skilled workers are willing to vote for the political party of the unskilled workers. Because the population of unskilled workers exceeds that of skilled workers, the coalition is formed such that the unskilled candidate represents the overall coalition. Therefore, the unskilled workers win the election.[16] The policy implementation becomes $(\tau_t, \mu_t, \sigma_t) = (\tau^u, \bar{\mu}, 1)$ according to Proposition 3.

Case IV. $\alpha-(1-\alpha)\left[\bar{\mu}+(1+\mu_{t-1})\frac{N_{t-1}}{N_t}\right] < s_t < \frac{1}{2}$

In this case, the population size is still the order of the old, the unskilled, and the skilled. The only difference from the previous case is that now the unskilled workers are net contributors and thus prefer zero taxes, as also wanted by the skilled workers. Therefore, the coalition is formed so that the unskilled party wins the election, and the policy of $(\tau_t, \mu_t, \sigma_t) = (0, \bar{\mu}, 1)$ is implemented according to Proposition 3.

Case V. $\frac{1}{2} \leq s_t < \frac{N_{t-1}}{N_t}$

In this case, the old retirees are the largest group, the skilled are the second-largest, and the unskilled are the smallest in size. Also, because both skilled and unskilled workers are net contributors to the welfare state, both groups prefer a zero tax rate; thus, the old retirees equally dislike both young candidates and want neither to win. The skilled workers obviously prefer the unskilled workers' policy profile to that of the old retirees. Therefore, the important problem is whether the unskilled prefer the skilled workers' ideal policy to the old retirees'. If the unskilled workers prefer skilled workers, the skilled candidate would then represent the coalition. Otherwise, the unskilled workers would have no reason to vote for the skilled workers, meaning

[16]I do not consider whether the unskilled workers have incentive to vote strategically for the skilled candidate because the unskilled workers' party is a larger group than that of the skilled workers and the skilled group does have an incentive to vote for the unskilled candidate. In this case, regardless whether the unskilled party's second-best choice is the skilled workers' candidate or not, the coalition is formed such that the unskilled workers' candidate represents the overall coalition according to the equilibrium refinement procedure introduced in Section III-A-2. Obviously, the case in which the skilled party wins the election by forming a coalition with the unskilled workers can be a form of Nash equilibrium if the unskilled workers have incentive to do so.

that the skilled workers strategically vote for the unskilled workers' candidate, and the unskilled candidate will represent the coalition.

For the unskilled workers, the preferred policies of the other two parties have the following trade-offs: Though the skilled workers' preferred policy, denoted by $(\tau_t, \mu_t, \sigma_t) = (0, \bar{\mu}, 0)$, is preferable because it levies zero taxes, it will allow many unskilled immigrants in and make the competition in the unskilled labor market tighter. Though the old retirees' preferred policy, denoted by $(\tau_t, \mu_t, \sigma_t) = (1/(1+\nu), \bar{\mu}, 0)$, is beneficial because it will allow in a maximum number of skilled immigrants, it levies high taxes on labor income. Therefore, the problem is whether the unskilled workers can bear the high tax rate (at the Laffer rate) preferred by the old retirees in order to allow as many skilled immigrants to enter as possible. Formally, this trade-off is characterized by the threshold tax rate, denoted by $\tilde{\tau}(\mu_{t-1}, s_t)$ and determined implicitly by solving for the following equation:

$V^u(\mu_{t-1}, s_t; \tau_t = \tilde{\tau},\ \mu_t = \bar{\mu},\ \sigma_t = 1) = V^u(\mu_{t-1}, s_t; \tau_t = 0,\ \mu_t = \bar{\mu},\ \sigma_t = 0)$, where the left-hand side is the utility of the unskilled workers when the tax rate is at $\tilde{\tau}$ and the maximum number of skilled immigrants are approved, and the right-hand side is the utility of the unskilled workers when the skilled young workers' candidate wins the election.[17] In other words, $\tilde{\tau}$ is the maximum tax rate that the unskilled young workers can bear in return for approving the maximum number of skilled immigrants. If $\tilde{\tau} < 1/(1+\nu)$, then the tax rate to be levied by the old retirees is excessively burdensome to the unskilled, meaning that the unskilled will strategically vote for the skilled workers' party. Otherwise, if $\tilde{\tau} \geq 1/(1+\nu)$, the unskilled workers then prefer the old retirees' policy at least as much as the skilled workers'; thus, a coalition is formed such that the unskilled candidate represents the overall coalition and wins the election.

As a result, the skilled young workers' candidate wins the election with the support of the unskilled party if $\tilde{\tau} < 1/(1+\nu)$, which implements $(\tau_t, \mu_t, \sigma_t) = (0, \bar{\mu}, 0)$ according to Proposition 2. If $\tilde{\tau} \geq 1/(1+\nu)$, then the unskilled workers win the election with the support of the skilled workers and the policy implementation is $(\tau_t, \mu_t, \sigma_t) = (0, \bar{\mu}, 1)$ via Proposition 3.

Case VI. $\dfrac{N_{t-1}}{N_t} \leq s_t < \alpha - (1-\alpha)\bar{\mu}$

In this case, the order of electorate size is as follows: skilled, the old retirees, and then the unskilled. As in the previous case, the old retirees do not have an incentive for strategic voting because all of the young workers prefer a zero tax rate. Therefore, the problem is whether the unskilled are willing to vote strategically for the old candidate or not, and this decision depends on the level of the unskilled workers' threshold tax rate $\tilde{\tau}$. If $\tilde{\tau} \leq 1/(1+\nu)$, he unskilled do not participate in the coalition with the old retirees and the skilled party wins; thus , the policy implementation will be $(\tau_t, \mu_t, \sigma_t) = (0, \bar{\mu}, 0)$ according to Proposition 2. If $\tilde{\tau} > 1/(1+\nu)$, then the old

[17]The existence of this form of $\tilde{\tau}$ is guaranteed by the Intermediate Value Theorem.

retirees win the election with the support of the unskilled workers, and the resulting policy implementation is given by $(\tau_t, \mu_t, \sigma_t) = (1/(1+\nu), \bar{\mu}, 1)$ according to Proposition 1.

Case VII. $\alpha - (1-\alpha)\bar{\mu} < s_t \leq \alpha + \alpha\bar{\mu}$

In this case, the order of the electorate size is identical to that in the previous case (the skilled, the old retirees, and then the unskilled). The difference is that now s_t is so large that the old retirees prefer a "balanced" skill composition of immigrants at $\sigma_t = \alpha + \frac{\alpha - s_t}{\bar{\mu}}$ instead of allowing skilled immigrants only. This makes the unskilled workers less favorable to the old retirees, thereby decreasing the threshold tax rate for strategic voting. Formally, the new threshold tax, denoted by $\hat{\tau}(\mu_{t-1}, s_t)$, is determined implicitly by the following equation:

$$V^u\left(\mu_{t-1}, s_t; \hat{\tau}, \bar{\mu}, \alpha + \frac{\alpha - s_t}{\bar{\mu}}\right) = V^u(\mu_{t-1}, s_t; 0, \bar{\mu}, 0)$$

If $\hat{\tau} \leq 1/(1+\nu)$, the unskilled do not participate in the coalition with the old retirees and the skilled party wins; thus, the policy implementation will be $(\tau_t, \mu_t, \sigma_t) = (0, \bar{\mu}, 0)$ according to Proposition 2. If $\hat{\tau} > 1/(1+\nu)$, then the old retirees win the election with the support of the unskilled workers; thus, $(\tau_t, \mu_t, \sigma_t) = \left(1/(1+\nu), \bar{\mu}, \alpha + \frac{\alpha - s_t}{\bar{\mu}}\right)$ is implemented via Proposition 1.

Case VIII. $\alpha + \alpha\bar{\mu} < s_t < \frac{1}{2} + \frac{1}{2} \cdot \frac{N_{t-1}}{N_t}$

In this case, the electorate size is identical to that in the previous case (the skilled, the old retirees, and then the unskilled). However, now s_t is so high that the old retirees want to implement $\sigma_t = 0$. Therefore, neither the old retirees nor the unskilled workers have an incentive to support each other. As a result, no coalition is formed and the skilled candidate wins the election as the largest party. The policy of $(\tau_t, \mu_t, \sigma_t) = (0, \bar{\mu}, 0)$ is implemented according to Proposition 2.

Case IX. $\frac{1}{2} + \frac{1}{2} \cdot \frac{N_{t-1}}{N_t} < s_t \leq 1$

In this case, the skilled young workers' party constitutes the majority of the electorate. The skilled young candidate wins the election and implements $(\tau_t, \mu_t, \sigma_t) = (0, \bar{\mu}, 0)$ via Proposition 2.

The nine cases above describe the patterns of political equilibria depending on the skill composition and the age composition of the economy. In short, a highly unbalanced skill composition makes the larger skill group more likely win the election, and the old retirees are likely to win when the young-to-old ratio is small enough and the skill composition of the young voters is relatively balanced. The forward-looking case will be shown to exhibit a similar pattern of equilibria in the next section.

B. *Equilibria with Forward-Looking Characteristics*

In this section, the young workers now care not only the instantaneous utility but also for their own future value (alternatively, $\beta > 0$). This results in a trade-off for the young workers. For example, for an unskilled young worker, it is still advantageous to approve of many skilled immigrants because the skill complementarity will increase his wage during this period, as discussed in the previous section, without forward-looking behavior. However, forward-looking behavior now imposes a drawback on this strategy. If too many skilled young immigrants enter the economy during this period, the descendants of those immigrants will be native skilled young workers in the next period (meaning that they have the right to vote in the next period as members of the skilled young party). Therefore, it is highly likely that the skilled young workers' party, which is most hostile to the welfare state, will win the election during the next period. This will deprive the present unskilled workers of the future welfare state after they retire. The unskilled young workers of this period can predict this and thus may choose to approve of only a small number of skilled workers so that the skilled young workers will not constitute a majority in the next period.

Formally, if a young candidate with skill level d wins the election, his policy rule is determined by the following optimization problem:

$$\Phi^d(\mu_{t-1}, s_t) = \text{argmax}_{\tau_t, \mu_t, \sigma_t} \frac{(1-\tau_t)\left[w^j(\tau_t, \mu_t, \sigma_t; \mu_{t-1}, s_t)\right]^{1+\nu}}{1+\nu} + b\left(\tau_t, \mu_t, \sigma_t; \mu_{t-1}, s_t\right) + \beta \cdot b\left(\Phi^d(\mu_t, s_{t+1})\right) \tag{25}$$

subject to

$$s_{t+1} = \frac{(1+n)s_t + (1+m)\sigma_t \mu_t}{1+n+(1+m)\mu_t}.$$

An old retiree's decision when he wins is identical to that in the previous section when $\beta = 0$ because his utility does not depend on the future.

Due to the complexity of this model, I rely on a numerical solution for this forward-looking case. Essentially, the solution numerically searches for the

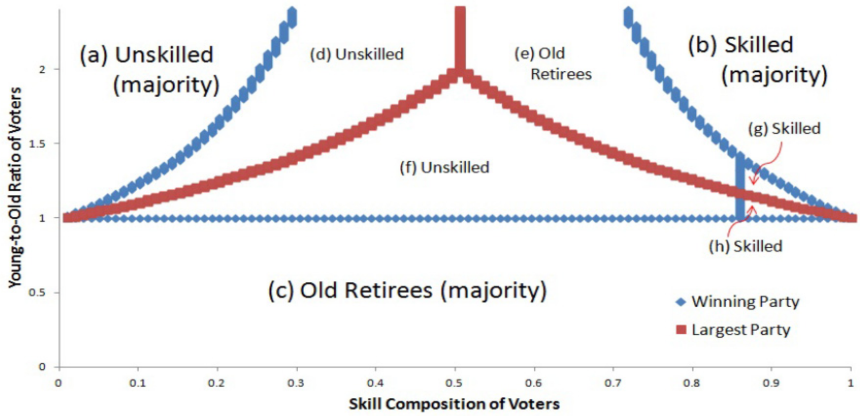


FIGURE 8. WINNING CANDIDATE FOR EACH STATE ($\alpha = 0.8$)

functional form of $\Phi^d(\mu_{t-1}, s_t)$ which solves the functional equation (25).[18] In other words, the solution describes what the winning candidate's optimal choice of the policy profile is given the state vector.

Figure 8 displays which cohort actually wins the election given the state vector (skill composition of the electorate on the horizontal axis and the young-to-old ratio of the electorate on the vertical axis). First, the red line divides the state space into three regions according to in which the largest party is. For instance, regions (a) and (d) are where the unskilled workers' party forms the largest group, regions (b), (e), and (g) are where the skilled workers' party is the largest, and the old retirees' party is the largest in (c), (f), and (h). The blue line further divides the regions according to which group becomes the winner and how this group wins. The following details can explain this:

- In regions (a), (b), and (c), either one of the groups constitutes the majority (more than 50%) of the electorate. Therefore, the political party with the majority wins the election.
- In region (d), the unskilled party is the largest. It wins the election because the two smaller parties (the skilled and the old retirees) are expected to implement a policy profile which is the worst to the other; thus, no coalition is organized.
- In region (e), the old retirees' party wins while the skilled workers group is the largest because the unskilled young workers prefer the old retirees' policy profile to that of the skilled. In other words, the unskilled workers can bear the high tax rates that the old retirees are expected to levy as the old retirees will approve of many skilled immigrants and increase the wages of the unskilled labor.

[18]For the solution to (25), a grid for the state vector (μ, s') is constructed and fixed points are found by iteration. Though the solution is checked for several initial points for robustness, uniqueness is not mathematically proved in this study. It is hoped that this issue will be solved in future studies.

- In region (f), the unskilled workers' party wins the election though the old retirees are the largest because the skilled workers strategically vote for the unskilled candidate to keep the old retirees from winning the election.
- In region (g), the skilled workers' party wins the election as the largest party. The other two parties do not form a coalition.
- In region (h), the skilled workers' party wins the election by forming a coalition with the unskilled. The old retirees' party is the largest but fails to win because their preferred tax rate (the Laffer rate) is excessively burdensome to the unskilled workers.

Therefore, a political coalition is formed in (e), (f), and (h), while the smaller parties let the largest to win without the formation of a coalition in (d) and (g).

Figure 9 shows how the winning party and the equilibrium volume of immigration differ according to the skill composition of the voters for a fixed young-to-old ratio of the electorate. While the old retirees always approve of the maximum number of immigrants, young workers can choose a value smaller than the maximum for certain values of s_t. Note that this occurs solely due to forward-looking behavior, as the desired immigration volume is always the maximum level without the forward-looking component, as shown in Propositions 1~3.

Figures 10~12 describe why unskilled workers prefer $\mu_t < \bar{\mu}$ in some states and $\mu_t = \bar{\mu}$ in others. In Figure 10, the current state is given by point A. Path (i) shows how the economy will move to a new state in the next period if the winning candidate decides to approve of the maximum number of immigrants. As the diagram demonstrates, allowing the maximum number of skilled immigrants will make skilled workers the majority group in the next period, thereby removing the possibility of receiving transfers after they retire. To prevent this, an alternative option for the unskilled workers is to approve of a slightly smaller number of immigrants instead (path ii), which will guarantee that today's unskilled young workers will win the election again in the next period (and thus receive the maximum lump-sum transfer), though it will reduce today's labor income of the unskilled workers relative to path (i). Figure 11 shows an example in which it is too costly to reduce the immigration volume so as to win in the next period if the economy is currently at point B. In this case, the unskilled workers will abandon the possibility of winning in the next period and attempt to maximize today's utility by allowing the maximum number of immigrants, which explains why the immigration volume increases to its maximum at around $s = 0.6$ in Figure 9.

Finally, Figure 12 depicts a case in which it is not feasible for unskilled workers to regain political power in the next period by adjusting the immigration volume. If the skill composition is sufficiently biased toward unskilled labor, unskilled workers cannot make the economy move toward a state in which old retirees will win the election in the next period even if the maximum number of skilled immigrants is accepted. In this case, the unskilled workers' optimal choice is simply to approve of the maximum number of skilled immigrants to maximize the instantaneous piece of their utility.

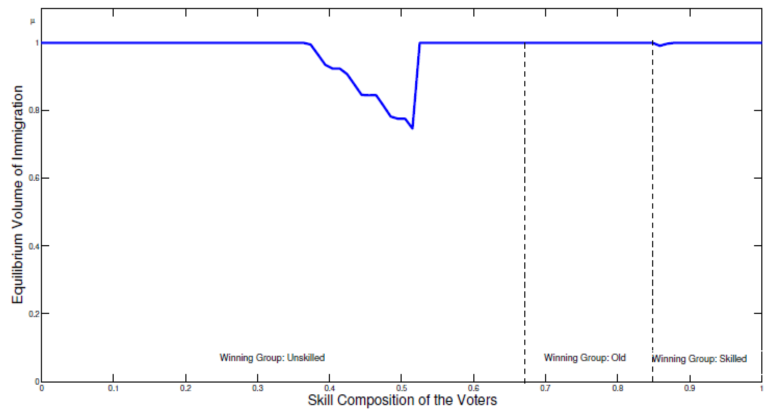


FIGURE 9. DECISION OF IMMIGRATION VOLUME BY SKILL COMPOSITION

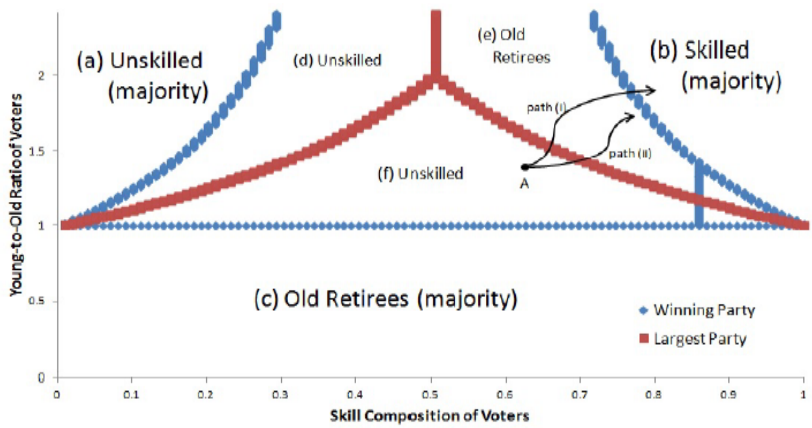


FIGURE 10. CASE I: LESS THAN MAXIMUM VOLUME OF IMMIGRATION IS OPTIMAL

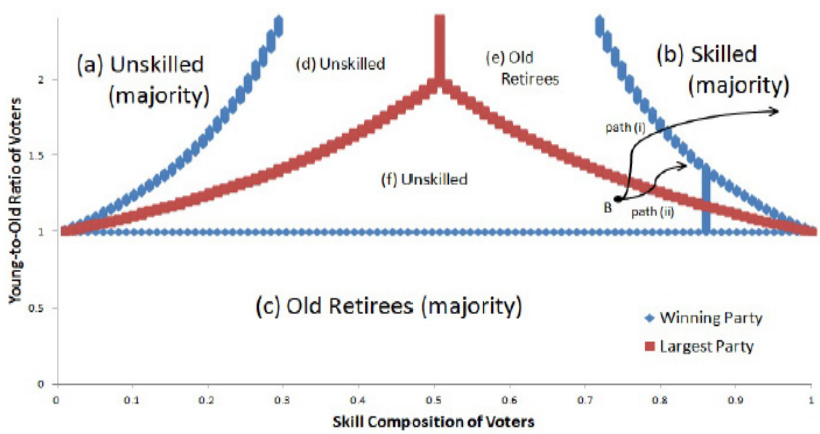


FIGURE 11. CASE II: LESS THAN MAXIMUM VOLUME OF IMMIGRATION IS TOO COSTLY

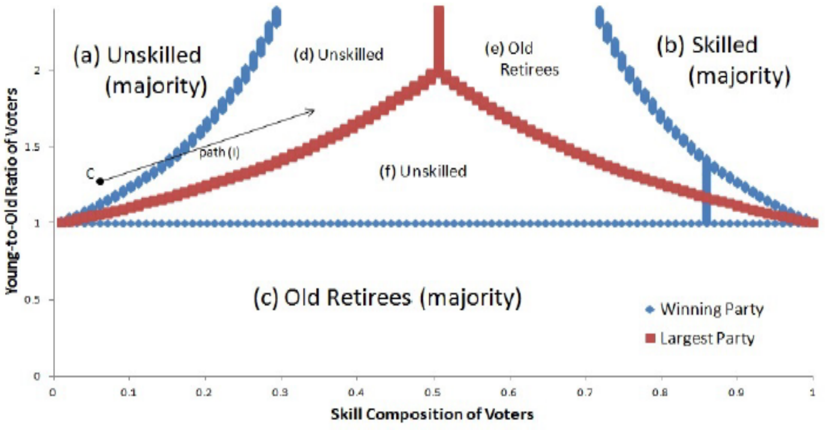


FIGURE 12. CASE III: LESS THAN MAXIMUM VOLUME OF IMMIGRATION OFFERS NO GAIN

V. Dynamics and Social Welfare

In this section, the dynamics of the key macroeconomic variables and their implications are analyzed. According to the analysis of political equilibria shown in the previous section, the skill composition of the economy is the key component which determines who wins the election and how the policies are implemented. Moreover, based on the results, it is expected that the skill composition exhibits cyclical movement; if the skill level of the young natives (i.e., young voters) is biased toward unskilled labor, the skill composition will move toward skilled labor in the next period because the unskilled workers' party is highly likely to win the election and approve of the maximum number of skilled immigrants, whose descendants will be skilled workers in the next period. Similarly, if the proportion of skilled workers is very large, the skill composition will move toward unskilled labor in the next period by the same logic.

This cyclical behavior of the skill composition leads to welfare loss in the economy. Because the production function of this model is the Cobb-Douglas aggregation of skilled and unskilled labor with exponent α for skilled labor, the socially optimal skill composition is obviously such that α is the proportion of skilled labor and $1-\alpha$ is that of unskilled labor. In this sense, the skill composition selected by the old retirees is socially optimal, while the young workers' ideal decision may be suboptimal.

In this model, immigration per se does not cause any inefficiency due to the constant-returns-to-scale technology. In other words, it is socially optimal to approve of as many immigrants as possible as long as the post-immigration skill composition becomes more balanced toward α. Therefore, the welfare loss in this economy is three-fold. First, the skill composition is not always balanced because young workers prefer maximal skill complementarity to maximize their own wages. Second, the proportional labor income tax is distortionary. Third, young workers may not choose to allow the maximum number of immigrants in to regain their political power in the following period.

A. *Welfare Loss due to a Skill Imbalance*

In order to check whether the skill composition exhibits a suboptimal path, the dynamics of the skill composition is examined in this section. Figure 13 displays the dynamics of the post-immigration skill composition of workers when $\alpha = 0.8$, and the initial skill composition equals 0.5. As surmised in the previous paragraphs, the skill composition fluctuates over time. For instance, skilled workers win the election at $t = 2$, decreasing the proportion of skilled workers by allowing the maximum number of unskilled immigrants. This causes, at $t = 3$, the unskilled workers to win the election and pushes the skill composition upward. The old retirees win the election at $t = 4$ because the unskilled workers adjust the immigration volume at $t = 3$ so that they can be the decisive party again when they retire. Given that the old retirees want the skill composition to be balanced at α, the skill composition of immigrants is adjusted such that the skill composition of workers is precisely at α after the immigrants enter the economy. This causes the skilled workers to win the election at $t = 5$, and this pattern is repeated later on.

It is notable that the value of s_t always stays below α. Because α is assumed to be larger than 0.5 for the skill premium, the skilled workers' party is likely to win the election when s_t is close to α. Therefore, the welfare loss from the skill imbalance is mainly caused by the skilled workers in this model. This skill imbalance not only causes a welfare loss but also dampens the effect of immigration in sustaining the welfare state. Given that the welfare state is most efficiently financed when $s_t = \alpha$, $\forall t$, the gap between α and s_t caused by the skilled workers' ideal policy deteriorates the inefficiency of the welfare state.

Furthermore, the fluctuation of the skill composition converges to the same pattern of cycles regardless of the initial skill composition. Figure 14 shows how different initial skill compositions in the end exhibit the same cyclical behavior. The skill composition converges to identical cycles because the value of s_t is pulled into α whenever the old retirees win an election. In other words, whenever the old retirees' party wins an election, the economy is absorbed into the state of $s_t = \alpha$ in the next period.

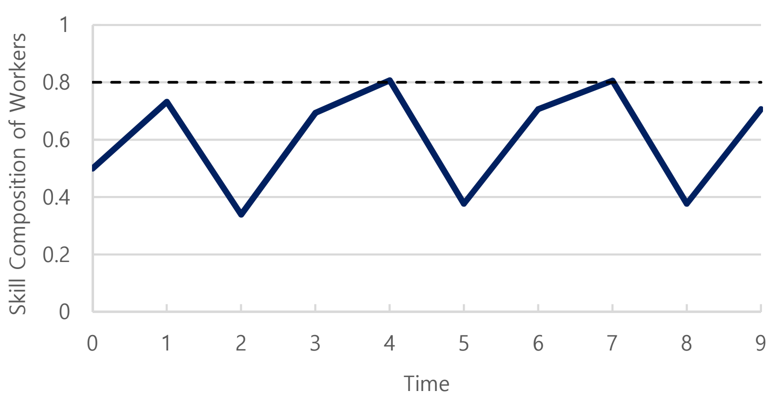


FIGURE 13. DYNAMICS OF THE SKILL COMPOSITION OF WORKERS ($\alpha = 0.8$)

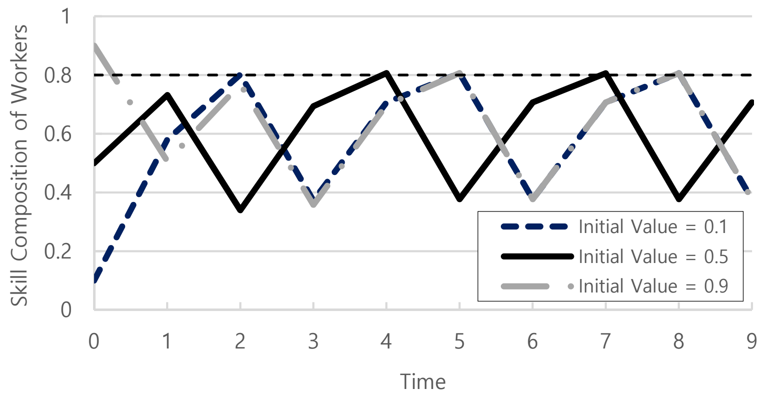


FIGURE 14. DYNAMICS OF THE SKILL COMPOSITION OF WORKERS FOR DIFFERENT INITIAL VALUES

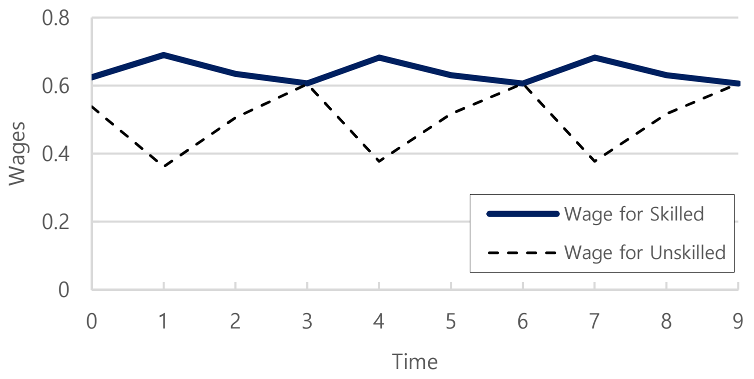


FIGURE 15. DYNAMICS OF PRE-TAX WAGES

Figure 15 displays the dynamics of wage rates for skilled and unskilled labor, respectively. It is clear that the skilled workers implement a policy profile to maximize their own wage (e.g., $t=1$) and that unskilled workers maximize their own wage (e.g., $t=2$). The wage rates are equalized at $t=3, t=6,$ and $t=9$, denoting the period when the old retirees win an election. The implication is simple; as discussed above, the Cobb-Douglas exponent and the actual skill composition will be equalized by the old retirees. At the same time, the income share of the skilled workers is always α according to the property of the Cobb-Douglas technology. Therefore, the wage rates for both skills are equalized when the old retirees win an election and implement the “skill-balancing” immigration policy.

B. *Welfare Loss by Distortionary Taxation*

It is obvious that the labor income tax in this model distorts the labor supply of both skills. This type of welfare loss is mainly caused by the old retirees because they prefer the Laffer rate to maximize the welfare state. As proved in Proposition 2 and Proposition 3, the tax distortion is either zero or is smaller than that caused by

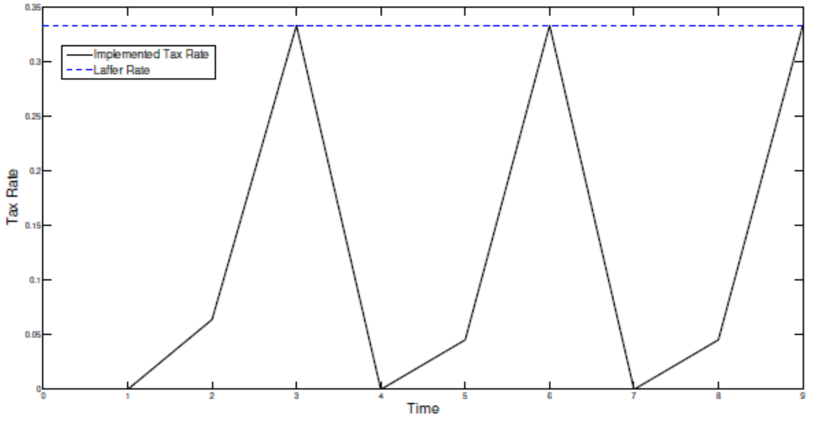


FIGURE 16. DYNAMICS OF TAX RATES

the old retirees if any cohort of the young workers becomes the decisive party. Therefore, the principal contributors of the welfare loss by taxes are the opposite of those in the case of the skill imbalance: the old retirees are the main contributor, and the skilled workers' preferred tax policy is the least distortionary in terms of taxation.[19]

Figure 16 shows the implementation of the tax rate over time. It is straightforward to observe that the old retirees levy $\tau_t = \frac{1}{1+\nu}$ at the Laffer rate, while the skilled workers levy zero taxes. The unskilled young workers implement a small non-zero tax rate when they win, as proved in Proposition 3.

C. *Welfare Loss by an Immigration Volume Adjustment*

In this model, the only reason for the immigration volume to be smaller than $\bar{\mu}$ is the young workers' motive to regain political power after they retire, as discussed in section VI-B. Figure 17 shows the implemented volume of immigration over time. Note that the volume does not reach its maximum at $t = 5$ and $t = 7$, both of which are the periods when the unskilled workers win an election. However, in this numerical example, the gap between $\bar{\mu}$ and the implemented μ_t is very small; moreover, such a value of μ_t that is smaller than $\bar{\mu}$ is still very large (nearly 1, which means that an identical number of immigrants equal to that of the existing native young workers enters the economy).

[19]Although skilled workers potentially prefer a non-zero tax rate, as proved in Proposition 2, it is mostly not realized because these workers prefer a non-zero rate only if $\alpha\left[1+\bar{\mu}+(1+\mu_{t-1})\frac{N_{t-1}}{N_t}\right] < s_t \leq 1$, which is an empty set for most reasonable values of $\frac{N_{t-1}}{N_t}$, $\bar{\mu}$, and α. For instance, the interval becomes empty for the following conservative choice of parameters: $\alpha = 0.8$, $\bar{\mu} = 0$, $\frac{N_{t-1}}{N_t} = 0.25$.

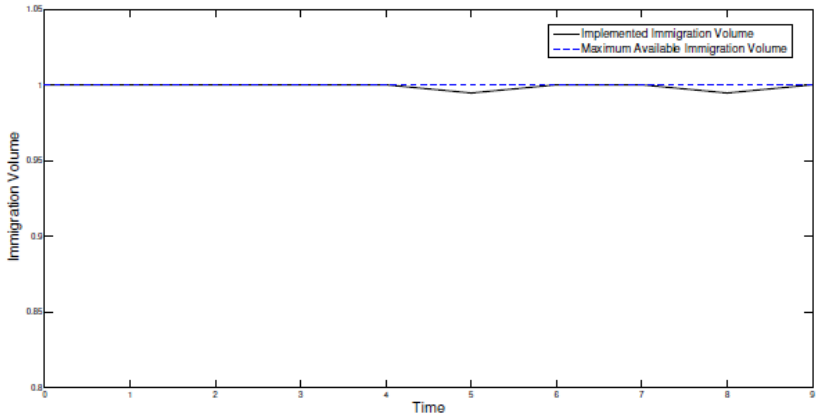


FIGURE 17. DYNAMICS OF THE IMMIGRATION VOLUME

VI. Concluding Remarks

This paper analyzes a politico-economic model of immigration and the welfare state to demonstrate the patterns of the political equilibria and the theoretical implications of sustainability of the welfare state and efficiency of the economy. In this two-period overlapping- generation model, the size of the welfare state, the volume of immigration, and the skill composition of immigrants are determined by the winner of a plurality voting system for each period. Because there are more than two political parties, the possibilities of strategic voting and consequent political coalitions among small political parties are also considered.

In the political processes, there are two key conflicts among the cohorts: the old vs. the young (maximizing the welfare state vs. low taxes) and skilled workers vs. unskilled workers (more unskilled immigrants vs. more skilled immigrants). In equilibrium, the skill composition of immigrants is biased toward unskilled labor relative to a socially efficient level because, at the socially efficient level of skill composition, the skilled workers' party is more likely to win the election. Due to this inefficiency, the effect of the immigration policy on the sustainability of the welfare state is weaker than expected when considering the literature.

Different sources of welfare loss are also analyzed. Not only does the imbalance of the skill composition contribute to the welfare loss but so also do distortionary taxation and the suboptimal volume of immigration. Specifically, the ruling of the old retirees causes maximal tax distortion, as these retirees want to maximize tax revenues by the government. If a young cohort becomes the ruling party, the volume of immigration can be determined at a level lower than the optimal level due to the young workers' incentive to regain political power in the future.

Finally, this study does not consider the possibility that the winning candidate's implementation of immigration policy is not feasible due to, for example, a lack of foreign-born workers who want to migrate to the host country. Also, another interesting case not studied here is one in which there are masses of refugees potentially entering several countries, and the refugees' country choice depends on the policy of each country as well as the economic state, such as the skill composition, aggregate productivity, and demographic structure.

APPENDIX

A. *Full Equilibrium Solution of Allocation and Prices*

If the production function is a typical Cobb-Douglas function in skilled and unskilled labor, the equilibrium can be represented by closed-form expressions.[20] In other words, all instances of equilibrium allocation and prices are represented in closed-form functions of the policy variables (the labor-income tax rate, the volume of immigration, and the skill composition of immigrants), state variables (μ_{t-1} and s_t), and the parameters. For a Cobb-Douglas production function, the equilibrium is characterized by the following nine equations:

$$l_t^s = [(1-\tau_t)w_t^s]^\nu \tag{26}$$

$$l_t^u = [(1-\tau_t)w_t^u]^\nu \tag{27}$$

$$\tilde{Y}_t = (\tilde{L}_t^s)^\alpha (\tilde{L}_t^u)^{1-\alpha} \tag{28}$$

$$\omega_t^s = \alpha \tilde{Y}_t (\tilde{L}_t^s)^{-1} \tag{29}$$

$$\omega_t^u = (1-\alpha)\tilde{Y}_t (\tilde{L}_t^u)^{-1} \tag{30}$$

$$\tilde{L}_t^s = [s_t + \sigma_t \mu_t] l_t^s \tag{31}$$

$$\tilde{L}_t^u = [1 - s_t + (1-\sigma_t)\mu_t] l_t^u \tag{32}$$

$$b_t = \frac{\tau_t \left[\omega_t^s \tilde{L}_t^s + \omega_t^u \tilde{L}_t^u \right]}{(1+\mu_t) + (1+\mu_{t-1})\dfrac{N_{t-1}}{N_t}} \tag{33}$$

$$\frac{N_t}{N_{t-1}} = 1 + n + (1+m)\mu_{t-1} \tag{34}$$

with nine unknowns $\left(l_t^s, l_t^u, \tilde{Y}_t, \tilde{L}_t^s, \tilde{L}_t^u, b_t, \dfrac{N_t}{N_{t-1}}, \omega_t^s, \omega_t^w \right)$, each of which is defined as in section 2.

[20]More rigorously, closed-form solutions are available when the marginal rate of substitution between the labor input and consumption is represented as a multiplicative function of the labor input, and the production function is multiplicative in both skills.

Plugging (29) and (30) into (26) and (27), respectively, leads to

$$(l_t^s)^{\frac{1}{\nu}} = (1-\tau_t)\alpha(\tilde{L}_t^s)^{\alpha-1}(\tilde{L}_t^u)^{1-\alpha}$$
$$(l_t^u)^{\frac{1}{\nu}} = (1-\tau_t)(1-\alpha)(\tilde{L}_t^s)^{\alpha}(\tilde{L}_t^u)^{-\alpha}.$$

Because at this point the equations are expressed in terms of variables for labor input only, this becomes a system of two equations and two unknowns (l_t^s and l_t^u) when substituting $\tilde{L}_t^s$ and $\tilde{L}_t^u$ with the (31) and (32) solutions. After some algebra, labor input per skilled and unskilled worker is represented only by the following policy variables and parameters:

$$l_t^s = (1-\tau_t)^{\nu}\alpha^{\frac{(1+\alpha\nu)\nu}{1+\nu}}(1-\alpha)^{\frac{(1-\alpha)\nu^2}{1+\nu}}(s_t+\sigma_t\mu_t)^{\frac{(\alpha-1)\nu}{1+\nu}}[1-s_t+(1-\sigma_t)\mu_t]^{\frac{(1-\alpha)\nu}{1+\nu}}$$
$$l_t^u = (1-\tau_t)^{\nu}\alpha^{\frac{\alpha\nu^2}{1+\nu}}(1-\alpha)^{\frac{[1+(1-\alpha)\nu]\nu}{1+\nu}}(s_t+\sigma_t\mu_t)^{\frac{\alpha\nu}{1+\nu}}[1-s_t+(1-\sigma_t)\mu_t]^{\frac{-\alpha\nu}{1+\nu}}$$

By plugging these back into (31) and (32), we obtain the following closed-form expressions of total skilled and unskilled labor divided by the native young population ($\tilde{L}_t^s$ and $\tilde{L}_t^u$, respectively):

$$\tilde{l}_t^s = (1-\tau_t)^{\nu}\alpha^{\frac{(1+\alpha\nu)\nu}{1+\nu}}(1-\alpha)^{\frac{(1-\alpha)\nu^2}{1+\nu}}(s_t+\sigma_t\mu_t)^{\frac{1+\alpha\nu}{1+\nu}}[1-s_t+(1-\sigma_t)\mu_t]^{\frac{(1-\alpha)\nu}{1+\nu}}$$
$$\tilde{l}_t^u = (1-\tau_t)^{\nu}\alpha^{\frac{\alpha\nu^2}{1+\nu}}(1-\alpha)^{\frac{[1+(1-\alpha)\nu]\nu}{1+\nu}}(s_t+\sigma_t\mu_t)^{\frac{\alpha\nu}{1+\nu}}[1-s_t+(1-\sigma_t)\mu_t]^{\frac{1+(1-\alpha)\nu}{1+\nu}}$$

By plugging these further back into (28), the output per native young worker ($\tilde{Y}_t$) is also represented in a closed form, as follows:

$$\tilde{Y}_t = (1-\tau_t)^{\nu}\alpha^{\alpha\nu}(1-\alpha)^{(1-\alpha)\nu}(s_t+\sigma_t\mu_t)^{\alpha}[1-s_t+(1-\sigma_t)\mu_t]^{1-\alpha}$$

Consequently the amount of the lump-sum transfer (b_t) is represented in the function of policy variables and states according to the results above.

$$b_t = \frac{\tau_t(1-\tau_t)^{\nu}\alpha^{\alpha\nu}(1-\alpha)^{(1-\alpha)\nu}(s_t+\sigma_t\mu_t)^{\alpha}[1-s_t+(1-\sigma_t)\mu_t]^{1-\alpha}}{(1+\mu_t)+(1+\mu_{t-1})\dfrac{N_{t-1}}{N_t}} \tag{35}$$

Finally, the wage rates of skilled and unskilled labor are calculated as shown below.

(36)
$$\omega_t^s = \alpha^{\frac{1+\alpha\nu}{1+\nu}}(1-\alpha)^{\frac{(1-\alpha)\nu}{1+\nu}}(s_t+\sigma_t\mu_t)^{-\frac{1-\alpha}{1+\nu}}[1-s_t+(1-\sigma_t)\mu_t]^{\frac{1-\alpha}{1+\nu}}$$
$$\omega_t^u = \alpha^{\frac{\alpha\nu}{1+\nu}}(1-\alpha)^{\frac{1+(1-\alpha)\nu}{1+\nu}}(s_t+\sigma_t\mu_t)^{\frac{\alpha}{1+\nu}}[1-s_t+(1-\sigma_t)\mu_t]^{-\frac{\alpha}{1+\nu}}$$

B. *Proofs*

B.1. Proof of Proposition 1

The old retirees' objective function at time t is simply the lump-sum transfer per capita:

$$V_t^o = b_t$$

Thus, we can differentiate it by each of τ_t, μ_t, and σ_t using (35).

First, the derivative with respect to the tax rate is

(37)
$$\frac{\partial V_t^o}{\partial \tau_t} = \frac{\partial b_t}{\partial \tau_t},$$

(38)
$$= B_1 \cdot [(1-\tau_t)^{\nu} - \tau_t \nu (1-\tau_t)^{\nu-1}]$$

where

$$B_1 = \frac{\alpha^{\alpha\nu}(1-\alpha)^{(1-\alpha)\nu}(s_t+\sigma_t\mu_t)^{\alpha}[1-s_t+(1-\sigma_t)\mu_t]^{1-\alpha}}{1+\mu_t+(1+\mu_{t-1})\dfrac{N_{t-1}}{N_t}}$$
$$> 0$$

Obviously $\frac{\partial V_t^o}{\partial \tau_t} = 0$ at either $\tau_t = 1/(1+\nu)$ or $\tau_t = 1$. However, $\tau_t = 1$ implies no lump-sum payment by (35) and thus zero utility for the old retirees, meaning that $\tau_t = 1/(1+\nu)$ is the optimal tax rate when the old retiree group wins the election.

Second, the derivative of V_t^o with respect to the skill composition of immigrants is as follows:

(39)
$$\frac{\partial V_t^o}{\partial \sigma_t} = \frac{\partial b_t}{\partial \sigma_t},$$
$$= B_2 \cdot (\alpha + \alpha\mu_t - s_t - \sigma_t\mu_t)$$

where

$$B_2 = \frac{\tau_t(1-\tau_t)^{\nu}\alpha^{\alpha\nu}(1-\alpha)^{(1-\alpha)\nu}(s_t+\sigma_t\mu_t)^{\alpha-1}[1-s_t+(1-\sigma_t)\mu_t]^{-\alpha}\mu_t}{1+\mu_t+(1+\mu_{t-1})\dfrac{N_{t-1}}{N_t}}$$

$$>0 \quad (\because 0<\tau_t^* = 1/(1+\nu)<1 \text{ as shown above})$$

Note that $\dfrac{\partial V_t^o}{\partial \sigma_t}<0$ for any $\sigma_t \in (0,1)$ if $s_t > \alpha+\alpha\mu_t$, and $\dfrac{\partial V_t^o}{\partial \sigma_t}>0$ for any $\sigma_t \in (0,1)$ if $s_t < \alpha+\alpha\mu_t-\mu_t$. Therefore, the optimal choice of the skill composition of immigrants when the old retirees' party wins an election is expressed as shown below.

$$(40) \qquad \sigma_t = \begin{cases} 1 & \text{if } 0 \le s_t < \alpha+\alpha\mu_t-\mu_t \\ \alpha+\dfrac{\alpha-s_t}{\mu_t} & \text{if } \alpha+\alpha\mu_t-\mu_t \le s_t \le \alpha+\alpha\mu_t \\ 0 & \text{if } \alpha+\alpha\mu_t < s_t \le 1 \end{cases}$$

Finally, the derivative of V_t^o with respect to the volume of immigration is

$$(41) \qquad \begin{aligned} \frac{\partial V_t^o}{\partial \mu_t} &= \frac{\partial b_t}{\partial \mu_t} \\ &= B_3 \left[\begin{aligned} &\left\{1+\mu_t+(1+\mu_{t-1})\frac{N_{t-1}}{N_t}\right\}\{\alpha\sigma_t[1-s_t+(1-\sigma_t)\mu_t] \\ &+(s_t+\sigma_t\mu_t)(1-\alpha)(1-\sigma_t)\}-(s_t+\sigma_t\mu_t)[1-s_t+(1-\sigma_t)\mu_t] \end{aligned} \right], \end{aligned}$$

where

$$B_3 = \frac{\tau_t(1-\tau_t)^{\nu}\alpha^{\alpha\nu}(1-\alpha)^{(1-\alpha)\nu}(s_t+\sigma_t\mu_t)^{\alpha-1}[1-s_t+(1-\sigma_t)\mu_t]^{-\alpha}}{\left[1+\mu_t+(1+\mu_{t-1})\dfrac{N_{t-1}}{N_t}\right]^2}$$

$$>0 \quad (\because 0<\tau_t = 1/(1+\nu)<1 \text{ as shown above}).$$

There can be three cases according to the optimal decision for σ_t.

Case I. $s_t \in (\alpha + \alpha\mu_t, 1]$

In this case, $\sigma_t = 0$ as shown in (40). After substitution, (41) is rewritten as follows:

$$\begin{aligned}
\frac{\partial V_t^o}{\partial \mu_t} &= B_3 \cdot \left\{ \left[1 + \mu_t + (1+\mu_{t-1})\frac{N_{t-1}}{N_t} \right](1-\alpha)s_t - (1+\mu_t - s_t)s_t \right\} \\
&= B_3 \cdot s_t \cdot \left\{ \left[1 + \mu_t + (1+\mu_{t-1})\frac{N_{t-1}}{N_t} \right](1-\alpha) - (1+\mu_t - s_t) \right\} \\
&> B_3 \cdot s_t \cdot \left\{ \left[1 + \mu_t + (1+\mu_{t-1})\frac{N_{t-1}}{N_t} \right](1-\alpha) - (1+\mu_t)(1-\alpha) \right\} \\
&\qquad (\because s_t > \alpha + \alpha\mu_t) \\
&> 0
\end{aligned}$$

Therefore, because $\frac{\partial V_t^o}{\partial \mu_t} > 0$ for any value of μ_t, it is optimal to choose $\mu_t = \bar{\mu}$.

Case II. $s_t \in [0, \alpha + \alpha\mu_t - \mu_t)$

In this case, $\sigma_t = 1$ as shown in (40). After substitution, (41) is rewritten as follows:

$$\begin{aligned}
\frac{\partial V_t^o}{\partial \mu_t} &= B_3 \cdot \left\{ \left[1 + \mu_t + (1+\mu_{t-1})\frac{N_{t-1}}{N_t} \right]\alpha(1-s_t) - (s_t + \mu_t)(1-s_t) \right\} \\
&= B_3 \cdot (1-s_t) \cdot \left\{ \left[1 + \mu_t + (1+\mu_{t-1})\frac{N_{t-1}}{N_t} \right]\alpha - s_t - \mu_t \right\} \\
&> B_3 \cdot (1-s_t) \cdot \left\{ \left[1 + \mu_t + (1+\mu_{t-1})\frac{N_{t-1}}{N_t} \right]\alpha - (1-\mu_t)\alpha \right\} \\
&\qquad (\because s_t < \alpha + \alpha\mu_t - \mu_t) \\
&> 0
\end{aligned}$$

Therefore, because $\frac{\partial V_t^o}{\partial \mu_t} > 0$ for any value of μ_t, it is optimal to choose $\mu_t = \bar{\mu}$.

Case III. $s_t \in [\alpha + \alpha\mu_t - \mu_t, \alpha + \alpha\mu_t]$

In this case, $\sigma_t = \alpha + \frac{\alpha - s_t}{\mu_t}$ as shown in (40). After reorganization, (41) is rewritten as follows:

$$\frac{\partial V_t^o}{\partial \mu_t} = B_3 \cdot \left\{ \begin{array}{l} \left[1 + \mu_t + (1+\mu_{t-1}) \frac{N_{t-1}}{N_t} \right] \sigma_t \{ \alpha [1 - s_t + (1-\sigma_t)\mu_t] - (1-\alpha)(s_t + \sigma_t \mu_t) \} \\ + \left[1 + \mu_t + (1+\mu_{t-1}) \frac{N_{t-1}}{N_t} \right] (s_t + \sigma_t \mu_t)(1-\alpha) - (s_t + \sigma_t \mu_t)[1 - s_t + (1-\sigma_t)\mu_t] \end{array} \right\}$$

After substituting $\sigma_t = \alpha + \frac{\alpha - s_t}{\mu_t}$, the first term equals zero, and the entire expression is simplified into the following form:

$$\begin{aligned} \frac{\partial V_t^o}{\partial \mu_t} &= B_3 \cdot \left\{ \left[1 + \mu_t + (1+\mu_{t-1}) \frac{N_{t-1}}{N_t} \right] (1-\alpha) - (1+\mu_t)(1-\alpha) \right\} \\ &> 0 \end{aligned} \tag{42}$$

Therefore the optimal choice is $\mu_t = \bar{\mu}$.

In conclusion, given that $\frac{\partial V_t^o}{\partial \mu_t}$ is strictly positive, the optimal choice of the old retiree is to choose $\mu_t = \bar{\mu}$ regardless of the state of the economy, which ends the proof.

B.2. Proof of Proposition 2

The skilled young worker's objective function at time t consists of two terms: the utility from the lump-sum transfer and the utility from the labor income net of the disutility of working:

$$V_t^s = b_t + \frac{[(1-\tau_t)\omega_t^s]^{1+\nu}}{1+\nu}$$

First, the derivative of V_t^s with respect to the tax rate is expressed as follows:

$$\frac{\partial V_t^s}{\partial \tau_t} = \frac{\partial b_t}{\partial \tau_t} + [(1-\tau_t)\omega_t^s]^{\nu}\left[(1-\tau_t)\frac{\partial \omega_t^s}{\partial \tau_t} - \omega_t^s\right]$$
$$= \frac{\partial b_t}{\partial \tau_t} - (1-\tau_t)^{\nu}(\omega_t^s)^{1+\nu}$$

Because $\frac{\partial \omega_t^s}{\partial \tau_t} = 0$ by (36). Plugging (36) and (38) into this expression yields

$$\frac{\partial V_t^s}{\partial \tau_t} = B_4 \cdot \left[\frac{s_t + \sigma_t \mu_t}{1+\mu_t + (1+\mu_{t-1})\frac{N_{t-1}}{N_t}}[1-(1+\nu)\tau_t] - (1-\tau_t)\alpha\right]$$
$$= B_4 \cdot \left\{\frac{s_t + \sigma_t \mu_t}{1+\mu_t + (1+\mu_{t-1})\frac{N_{t-1}}{N_t}} - \alpha - \left[\frac{s_t + \sigma_t \mu_t}{1+\mu_t + (1+\mu_{t-1})\frac{N_{t-1}}{N_t}}(1+\nu) - \alpha\right]\cdot \tau_t\right\},$$

where

$$B_4 = (1-\tau_t)^{\nu-1}\alpha^{\alpha\nu}(1-\alpha)^{(1-\alpha)\nu}(s_t + \sigma_t\mu_t)^{\alpha-1}[1-s_t+(1-\sigma_t)\mu_t]^{1-\alpha}$$
$$> 0.$$

Therefore, the optimal tax rate for a skilled young worker depends on the state of the economy (s_t, μ_{t-1}), and the following two cases are possible:

Case I. $s_t < \alpha\left[1+\mu_t+(1+\mu_{t-1})\frac{N_{t-1}}{N_t}\right] - \sigma_t\mu_t$

In this case we can observe the following two consequences:

i) $\frac{s_t + \sigma_t \mu_t}{1+\mu_t + (1+\mu_{t-1})\frac{N_{t-1}}{N_t}} - \alpha < 0$, and

ii) $\frac{s_t + \sigma_t \mu_t}{1+\mu_t + (1+\mu_{t-1})\frac{N_{t-1}}{N_t}}(1+\nu) - \alpha$ is either is either a positive number or a negative number smaller in terms of the absolute value compared to $\frac{s_t + \sigma_t \mu_t}{1+\mu_t + (1+\mu_{t-1})\frac{N_{t-1}}{N_t}} - \alpha < 0$.

These two observations guarantee that $\frac{\partial V_t^s}{\partial \tau_t} < 0$. Therefore, the optimal tax rate for a skilled young worker in this case is given by $\tau_t = 0$.

Case II. $s_t \geq \alpha\left[1+\mu_t+(1+\mu_{t-1})\frac{N_{t-1}}{N_t}\right]-\sigma_t\mu_t$

In this case, we have an interior solution at

$$\tau_t = \frac{\dfrac{s_t+\sigma_t\mu_t}{1+\mu_t+(1+\mu_{t-1})\frac{N_{t-1}}{N_t}}-\alpha}{\dfrac{s_t+\sigma_t\mu_t}{1+\mu_t+(1+\mu_{t-1})\frac{N_{t-1}}{N_t}}(1+\nu)-\alpha}$$

$$= \frac{s_t+\sigma_t\mu_t-\alpha\left[1+\mu_t+(1+\mu_{t-1})\frac{N_{t-1}}{N_t}\right]}{(s_t+\sigma_t\mu_t)(1+\nu)-\alpha\left[1+\mu_t+(1+\mu_{t-1})\frac{N_{t-1}}{N_t}\right]}$$

$$\in (0,\ 1)$$

Therefore, these two cases show that the optimal decision of τ_t for the skilled young worker is given by

$$\tau_t = \begin{cases} 0 & \text{if } s_t < \alpha\left[1+\mu_t+(1+\mu_{t-1})\frac{N_{t-1}}{N_t}\right]-\sigma_t\mu_t \\ \dfrac{s_t+\sigma_t\mu_t-\alpha\left[1+\mu_t+(1+\mu_{t-1})\frac{N_{t-1}}{N_t}\right]}{(s_t+\sigma_t\mu_t)(1+\nu)-\alpha\left[1+\mu_t+(1+\mu_{t-1})\frac{N_{t-1}}{N_t}\right]} & \text{if } s_t \geq \alpha\left[1+\mu_t+(1+\mu_{t-1})\frac{N_{t-1}}{N_t}\right]-\sigma_t\mu_t \end{cases} \quad (43)$$

The derivative of V_t^s with respect to the skill composition of immigrants is

$$\frac{\partial V_t^s}{\partial \sigma_t} = \frac{\partial b_t}{\partial \sigma_t} + (1-\tau_t)^{1+\nu}(\omega_t^s)^{\nu}\frac{\partial \omega_t^s}{\partial \sigma_t}.$$

Plugging (36) and (39) leads to

$$(44)\qquad \frac{\partial V_t^s}{\partial \sigma_t} = B_5 \cdot \left\{ \begin{array}{l} \dfrac{\tau_t(s_t+\sigma_t\mu_t)}{1+\mu_t+(1+\mu_{t-1})\dfrac{N_{t-1}}{N_t}}(\alpha - s_t + \alpha\mu_t - \sigma_t\mu_t) \\ -\dfrac{(1-\tau_t)(1+\mu_t)\alpha(1-\alpha)}{1+\nu} \end{array} \right\}$$

where

$$B_5 = (1-\tau_t)^{\nu}\mu_t\alpha^{\alpha\nu}(1-\alpha)^{(1-\alpha)\nu}(s_t+\sigma_t\mu_t)^{\alpha-2}[1-s_t+(1-\sigma_t)\mu_t]^{-\alpha}$$
$$>0.$$

Again, the optimal decision of σ_t can be separated into two cases, as in (43).

Case I. $s_t < \alpha\left[1+\mu_t+(1+\mu_{t-1})\dfrac{N_{t-1}}{N_t}\right] - \sigma_t\mu_t$

In this case, the optimal τ_t is zero according to (43). After plugging $\tau_t = 0$ into (44), obviously $\dfrac{\partial V_t^s}{\partial \sigma_t} < 0$, implying that the optimal σ_t equals zero.

Case II. $s_t \geq \alpha\left[1+\mu_t+(1+\mu_{t-1})\dfrac{N_{t-1}}{N_t}\right] - \sigma_t\mu_t$

This case is equivalent to $-s_t - \sigma_t\mu_t \leq -\alpha\left[1+\mu_t+(1+\mu_{t-1})\dfrac{N_{t-1}}{N_t}\right]$. Plugging this inequality into (44),

$$\frac{\partial V_t^s}{\partial \sigma_t} \leq B_5 \cdot \left\{ \begin{array}{l} \dfrac{\tau_t(s_t+\sigma_t\mu_t)}{1+\mu_t+(1+\mu_{t-1})\dfrac{N_{t-1}}{N_t}}\left[\alpha(1+\mu_t) - \alpha\left[1+\mu_t+(1+\mu_{t-1})\dfrac{N_{t-1}}{N_t}\right]\right] \\ -\dfrac{(1-\tau_t)(1+\mu_t)\alpha(1-\alpha)}{1+\nu} \end{array} \right\}$$
$$<0.$$

Because $\dfrac{\partial V_t^s}{\partial \sigma_t} < 0$ for both cases, the optimal decision of σ_t is always zero for the skilled young worker.

Finally, the derivative of the skilled young worker's objective function with respect to the volume of immigration is expressed as follows:

$$\frac{\partial V_t^s}{\partial \mu_t} = \frac{\partial b_t}{\partial \mu_t} + (1-\tau_t)^{1+\nu}(\omega_t^s)^{\nu}\frac{\partial \omega_t^s}{\partial \mu_t}$$

As it was already shown that $\sigma_t = 0$ in the optimum state, we can use (42) to derive the following:

$$\left.\frac{\partial b_t}{\partial \mu_t}\right|_{\sigma_t=0} = B_3 \cdot s_t \cdot \left\{(1-\alpha)\left[1+\mu_t+(1+\mu_{t-1})\frac{N_{t-1}}{N_t}\right]-(1-s_t+\mu_t)\right\}$$

and there can be two cases depending on the value of s_t:

Case I. $s_t < \alpha\left[1+\mu_t+(1+\mu_{t-1})\frac{N_{t-1}}{N_t}\right]$

In this case, $\tau_t = 0$ as shown above; therefore, $b_t = 0$ and thus $\frac{\partial b_t}{\partial \mu_t} = 0$.

Case II. $s_t \geq \alpha\left[1+\mu_t+(1+\mu_{t-1})\frac{N_{t-1}}{N_t}\right]$

In this case:

$$\begin{aligned}\left.\frac{\partial b_t}{\partial \mu_t}\right|_{\sigma_t=0} &= B_3 \cdot s_t \cdot \left\{(1-\alpha)\left[1+\mu_t+(1+\mu_{t-1})\frac{N_{t-1}}{N_t}\right]-(1-s_t+\mu_t)\right\} \\ &\geq B_3 \cdot s_t(1+\mu_{t-1})\frac{N_{t-1}}{N_t} \\ &\geq 0\end{aligned}$$

Therefore, $\frac{\partial b_t}{\partial \mu_t}$ is non-negative in either case. Now, the derivative of the wage rate is written in detail as follows:

$$\frac{\partial \omega_t^s}{\partial \mu_t} = \alpha^{\frac{1+\alpha\nu}{1+\nu}}(1-\alpha)^{\frac{(1-\alpha)\nu}{1+\nu}}\left[\begin{array}{l} -\frac{1-\alpha}{1+\nu}\sigma_t(s_t+\sigma_t\mu_t)^{-\frac{1-\alpha}{1+\nu}-1}\left[1-s_t+(1-\sigma_t)\mu_t\right]^{\frac{1-\alpha}{1+\nu}} \\ +\frac{1-\alpha}{1+\nu}(1-\sigma_t)(s_t+\sigma_t\mu_t)^{-\frac{1-\alpha}{1+\nu}}[1-s_t+(1-\sigma_t)\mu_t]^{\frac{1-\alpha}{1+\nu}-1}\end{array}\right]$$

Substituting $\sigma_t = 0$, it is readily apparent that $\frac{\partial \omega_t^s}{\partial \mu_t} > 0$, thereby implying that $\frac{\partial V_t^s}{\partial \mu_t} > 0$. Therefore, the optimal choice of the volume of immigration for the skilled young worker is $\bar{\mu}$.

At this point, we return to the intervals of s_t in (43). Because now we know that $\mu_t = \bar{\mu}$ and $\sigma_t = 0$ in optimality, we can substitute μ_t and σ_t with $\bar{\mu}$ and 0, respectively, ending the proof.

C. *Political Coalition and Equilibria*

In this section, I discuss the details of the assumption of the determination process of political equilibria as discussed in section III-A-2. As in the model, there are three political parties, and all of the individuals in one party have identical preferences. Throughout this section, let $\mathcal{A}$, $\mathcal{B}$, and $\mathcal{C}$ denote the three political parties in the order of their population sizes and $\#(\mathcal{A})$, $\#(\mathcal{B})$, and $\#(\mathcal{C})$ denote the relative sizes of the three parties, respectively. The sizes are normalized such that $\#(\mathcal{A}) + \#(\mathcal{B}) + \#(\mathcal{C}) = 1$ such that the measure of the size of the overall electorate equals 1.[21] We assume that all parties have at least one member; formally, this is $\#(\mathcal{X}) > 0$ for all $\mathcal{X} \in \{\mathcal{A}, \mathcal{B}, \mathcal{C}\}$.

First, it will be proved that the largest party's members always vote for their own party regardless of whether they constitute the majority or not. It is obvious that $\#(\mathcal{A}) > 0.5$, which is a trivial case. Hence, we assume that it does not exceed 50%. Then, Lemma 4 follows:

Lemma 4. A member of $\mathcal{A}$ votes only for $\mathcal{A}$ in any political equilibrium.

This implies that the members of the largest political party do not vote strategically for a candidate from another party in any of the equilibria after repeatedly eliminating weakly dominating strategies. The proof is shown below.

Proof. It is obvious that the members of $\mathcal{A}$ prefer $\mathcal{A}$'s policy profile most. Accordingly, we can assume, without a loss of generality, that the members of $\mathcal{A}$ prefer $\mathcal{B}$ to $\mathcal{C}$. Then, it is obvious that $\mathcal{C}$ is weakly dominated by the other two strategies. Therefore, we assume that there is a state of equilibrium in which $\mathcal{A}$'s members vote for $\mathcal{B}$. Then, the following possible cases exist:

Case I. $\mathcal{B}$'s members vote for $\mathcal{B}$, and $\mathcal{C}$'s members vote for $\mathcal{C}$.

In this case, for $\mathcal{C}$'s members, voting for $\mathcal{B}$ gives a weakly smaller payoff than voting for $\mathcal{A}$ because if $\mathcal{A}$'s members deviate to voting for $\mathcal{A}$, then $\mathcal{A}$ wins the election.

Case II. $\mathcal{B}$'s members vote for $\mathcal{C}$, and $\mathcal{C}$'s members vote for $\mathcal{B}$.

Similar to Case I.

[21]Therefore, $\#(\mathcal{A}) \geq \#(\mathcal{B}) \geq \#(\mathcal{C})$ holds because A, B, and C are labeled in the order of size.

Case III. $\mathcal{B}$'s members vote for $\mathcal{B}$, and $\mathcal{C}$'s members vote for $\mathcal{B}$.

Then, for the members of $\mathcal{A}$, voting for $\mathcal{A}$ and $\mathcal{B}$ results in an identical payoff because $\mathcal{C}$ wins after all.

Case IV. $\mathcal{B}$'s members vote for $\mathcal{C}$, and $\mathcal{C}$'s members vote for $\mathcal{C}$.

Similar to Case III.

Therefore, in any of the cases above, voting for $\mathcal{B}$ is at least weakly dominated by voting for $\mathcal{A}$, which contradicts the assumption that voting for $\mathcal{B}$ occurs in equilibrium and ends the proof.

Therefore, it is clear that the members of the largest cohort always vote for the candidate of their own cohort. Together with this lemma, the following assumptions refine the political equilibrium so that there is a unique form of pure-strategy equilibrium per state.

(i) Weakly dominated strategies are repeatedly eliminated.

(ii) If there are two parties with the equally the largest number of votes, the candidate whose votes are from the least number of cohorts then becomes the winner.

(iii) Voters strategically vote for the candidate of another cohort only if it is strictly preferred to not doing so.

(iv) Suppose there are two parties. If all members of both parties prefer to form a coalition with each other, the candidate of the larger party out of the two becomes the candidate of the overall coalition. If the sizes of the two parties are identical, the candidate is determined by equal probability.

(v) Any remaining possibility of ties is broken with equal probability.

These assumptions, together with Lemma 4, imply the following results of the election.

Case I. $\#(\mathcal{A}) \geq 0.5$

This is the simplest case to solve, where $\mathcal{A}$'s candidate wins the election by a majority. Note that $\#(\mathcal{A}) = 0.5$ also guarantees $\mathcal{A}$'s victory due to assumption (ii) above.

Case II. $0.5 > \#(\mathcal{A}) > \#(\mathcal{B}) \geq \#(\mathcal{C})$

This is the most interesting case. Because $\mathcal{A}$ does not take the majority, it is important as to whether $\mathcal{B}$ and $\mathcal{C}$ form a political coalition or not, which depends on the preference of $\mathcal{B}$'s and $\mathcal{C}$'s members. The following exclusive and exhaustive cases are possible:

- **Case II-1.** $\mathcal{B}$'s members at least weakly prefer $\mathcal{A}$ to $\mathcal{C}$, and $\mathcal{C}$'s members at least weakly prefer $\mathcal{A}$ to $\mathcal{B}$.

  In this case, neither $\mathcal{B}$ nor $\mathcal{C}$ wants a coalition because they perceive the other party's ideal policy to be the worst. Therefore, $\mathcal{A}$ wins the election by receiving the largest number of votes.

- **Case II-2.** $\mathcal{B}$'s members at least weakly prefer $\mathcal{A}$ to $\mathcal{C}$, and $\mathcal{C}$'s members strongly prefer $\mathcal{B}$ to $\mathcal{A}$.

  In this case, while $\mathcal{B}$'s members do not have any incentive to vote for $\mathcal{C}$ strategically, $\mathcal{C}$'s members have an incentive to vote for $\mathcal{B}$'s candidate. Therefore, $\mathcal{B}$'s candidate represents the political coalition by $\mathcal{B}$ and $\mathcal{C}$ and wins the election in the equilibrium.

- **Case II-3.** $\mathcal{B}$'s members strongly prefer $\mathcal{C}$ to $\mathcal{A}$, and $\mathcal{C}$'s members at least weakly prefer $\mathcal{A}$ to $\mathcal{B}$.

  In this case, $\mathcal{C}$'s candidate represents the coalition by $\mathcal{B}$ and $\mathcal{C}$ wins the election via the same logic presented in Case II-2. Note that $\mathcal{C}$ wins the election despite the fact that it is the smallest cohort in the electorate.

- **Case II-4.** $\mathcal{B}$'s members strongly prefer $\mathcal{C}$ to $\mathcal{A}$, and $\mathcal{C}$'s members strongly prefer $\mathcal{B}$ to $\mathcal{A}$.

  In this case, both of the two smaller parties want to form a coalition. If $\#(\mathcal{B}) > \#(\mathcal{C})$, then the candidate of $\mathcal{B}$ represents the entire coalition according to assumption (iv) and wins the election. If $\#(\mathcal{B}) = \#(\mathcal{C})$, then the candidate of the coalition is determined by a coin toss, and the resulting candidate wins the election.

Case III. $\#(\mathcal{A}) = \#(\mathcal{B}) > \#(\mathcal{C})$

In this case, $\mathcal{C}$ serves as the swing cohort. If $\mathcal{C}$'s members strongly prefer $\mathcal{A}$ to $\mathcal{B}$, then $\mathcal{A}$ wins the election by forming a coalition with $\mathcal{C}$. If' $\mathcal{C}$'s members strongly prefer $\mathcal{B}$ to $\mathcal{A}$, $\mathcal{B}$ wins the election similarly. If $\mathcal{C}$'s members are indifferent, then they do not vote strategically, and either $\mathcal{A}$ or $\mathcal{B}$ then wins the election with equal probability (a coin toss).

Case IV. $\#(\mathcal{A}) = \#(\mathcal{B}) = \#(\mathcal{C})$

In this case, each party wins the election with a probability of 1/3.

References

Auerbach, Alan J. and Philip Oreopoulos. 1999. "Analyzing the Fiscal Impact of US Immigration," *The American Economic Review*, 89(2): 176-180.

Benhabib, Jess. 1996. "On the political economy of immigration." *European Economic Review*, 40(9): 1737-1743.

Besley, Timothy and Stephen Coate. 1997. "An economic model of representative democracy," *The Quarterly Journal of Economics*, 112(1): 85-114.

Besley, Timothy and Stephen Coate. 1998. "Sources of inefficiency in a representative democracy: a dynamic analysis," *American Economic Review*, 88(1): 139-156.

Boldrin, Michele and Aldo Rustichini. 2000. "Political equilibria with social security," *Review of Economic Dynamics,* 3(1): 41-78.

Dolmas, Jim and Gregory W. Huffman. 2004. "On the political economy of immigration and income redistribution," *International Economic Review,* 45(4): 1129-1168.

Galasso, Vincenzo and Paola Profeta. 2002. "The political economy of social security: a survey," *European Journal of Political Economy,* 18(1): 1-29.

Grossman, Gene and Elhanan Helpman. 1998. "Intergenerational redistribution with short-lived governments," *The Economic Journal*, 108(450): 1299-1329.

Hassler, John, *et al*. 2003. "The survival of the welfare state," *American Economic Review*, 93(1): 87-112.

Hunt, Jennifer and Marjolaine Gauthier-Loiselle. 2010. "How much does immigration boost innovation?" *American Economic Journal: Macroeconomics*, 2(2): 31-56.

Krusell, Per and Jose-Victor Rios-Rull. 1996. "Vested interests in a positive theory of stagnation and growth," *The Review of Economic Studies,* 63(2): 301-329.

Monte, Lindsay M. and Renee Ellis. 2014. *Fertility of women in the United States: 2012*, US Department of Commerce, Economics and Statistics Administration, US Census Bureau.

Ortega, Francesc. 2005. "Immigration quotas and skill upgrading," *Journal of Public Economics,* 89(9-10): 1841-1863.

Peri, Giovanni. 2012. "The effect of immigration on productivity: Evidence from US states," *Review of Economics and Statistics*, 94(1): 348-358.

Storesletten, Kjetil. 2000. "Sustaining Fiscal Policy through Immigration," *Journal of Political Economy*, 108(2): 300-323.

Suwankiri, Benjarong, Assaf Razin, and Efraim Sadka. 2016. "The welfare State and migration: a dynamic analysis of political coalitions," *Research in Economics*, 70(1): 122-142.

KDI Journal of Economic Policy 2022, 44(1):49–75
http://dx.doi.org/10.23895/kdijep.2022.44.1.49

Franchise Market, Contract Conditions, and Welfare Implications: Evidence from Korea†

By JINKOOK LEE AND MYOUNGSHIK SEO*

This paper analyzes how franchise contract conditions are influenced by business structures as well as how contract conditions affect producer surplus by utilizing Korean franchise Information Disclosure Documents for the years 2014-2016. We find that franchise fees tend to increase in line with increases in the numbers of direct stores or the business period. Accordingly, it would be reasonable to check whether the franchise fee is excessive compared to the amount of reputation capital rather than to criticize the absolute level of the franchise fee. Regarding royalty contracts, the larger the discount in the raw materials purchase is, the higher the initial royalty is. Although this appears to be a royalty discount, it can be a means of inducing a raw materials purchase contract by initially setting a high royalty rate and then lowering it after the purchase contract is signed. Concerning the effect on producer surplus, the results show that an increase in franchise fees and royalties negatively affects the franchisee's operating profits but positively affects those of the franchisor's, leading to conflicts over the distribution of economic value added. Based on the findings here, we propose various policy recommendations, specifically reinforcing the contents in the Information Disclosure Document, further activating fixed-rate royalties, and strengthening the qualifications of franchisors when recruiting franchisees.

Key Word: Korean Franchise Market, Franchise Fee, Royalty Rate, Franchise Contract Conditions
JEL Code: D22, L13, L16, L22, L50

* Lee (First Author): Fellow, Korea Development Institute (E-mail: ljk@kdi.re.kr); Seo: Senior Research Associate, Korea Development Institute (E-mail: msseo@kdi.re.kr).
* Received: 2021. 9. 13
* Referee Process Started: 2021. 9. 23
* Referee Reports Completed: 2021. 11. 15

† This paper is written based on Lee, 2018, *Study on Franchise Contract and Franchise Market System*, KDI Policy Study 2018-04 and: Lee, 2019, "Conflicts in the Franchise Industry: How to Forge a Win-Win Partnership," KDI Focus No.96.

I. Introduction

Korea's franchise market is growing rapidly. From 2014 to 2016, the number of franchisors increased from 3,923 to 4,664, and that of franchisees rose from 208,099 to 230,054. Accordingly, the overall market size of the franchise business grew to 114 trillion won in 2016, an increase of 24.3% compared to 2014.

Such rapid growth was made possible by several business advantages arising from the franchise contractual relationship. First, franchisors can spread their own brands and business models through the recruitment of franchisees, which eventually raise brand value as well as profits. Also, they can acquire various types of market information through local franchisees so as to improve their business strategies. From the franchisees' perspective, they can easily and quickly start a business while receiving marketable products, learning sales techniques, and gaining management know-how from their franchisors. Furthermore, the tendency of consumers to prefer standardized products in terms of price and quality over relatively non-standardized goods may support the expansion of this market.

However, this rapid growth has been accompanied by various internal conflicts between franchisors and their franchisees. There were 779 applications for dispute mediation in 2017, more than double the 357 filed in 2008. Looking at the types of conflicts, disputes related to information asymmetry account for 30% of the total, and cases caused by the imbalance of bargaining power are also on the rise (Korea Fair Trade Mediation Agency, 2013; 2018).[1]

Based on this background, this study examines the growth trend and market structure of the domestic franchise market, utilizing brand-level microdata from Information Disclosure Documents.[2] Subsequently, we quantify the effect of the market structure on franchise contract conditions and analyze in depth the effects of contract conditions on producer welfare. The empirical findings lead to policy suggestions for improving the franchise market system.

II. Related Literature

This section reviews existing literature, focusing on franchise contract conditions, and identifies the various contributions of this paper.

First, Gallini and Lutz (1992) argue that franchisors use royalty rates as a signal to convey the profitability of their products. A franchisor after developing a highly profitable product sets a high initial royalty rate to signal the quality of the product. As time passes, her brand becomes more recognized and a reputation for product profitability is formed, the incentive for signaling is weakened and the royalty rate

[1]Disputes related to information asymmetry include cases in which Information Disclosure Documents were not submitted or those in which false (or exaggerated) information was provided. The imbalance of bargaining power can be related to issues such as unfair liability for damages, an unjustified termination of a contract, and territorial encroachment.

[2]An Information Disclosure Document refers to a document that specifies general information such as the business status of the franchisor, the burden on the franchisees, and various contract conditions. In accordance with the "Act on the Fairness of Franchise Transactions," all franchisors must file this document before recruiting franchisees.

therefore gradually decreases, though the franchise fee increases instead.

Mathewson and Winter (1985) finds that franchisors use franchise fees as a means of recovering reputation capital. Their findings through a reputation model imply that a new franchisor sets the franchise fee low due to the low reputation level, later increasing it when the business period is extended and the reputation for brand profitability is established.

These theoretical results pertaining to the volatility of franchise fees and royalties were subsequently verified empirically. Lafontaine (1992) analyzes data from 548 franchisors in the service sector in the US and finds that royalty rates tended to decrease as the number of years in business increased. The franchise fee, however, was not significantly affected by the business period, from which she suggests that the franchise fee and royalty rate do not necessarily have a negative relationship.

Lafontaine and Shaw (1999) examine contract data of one thousand franchisors who were active during 1980-92. In the OLS model, a longer business period was likely to decrease the royalty rate and increase the franchise fee. However, in the fixed-effect model, these tendencies were mostly insignificant.

As such, various opinions exist depending on the data and methodologies, confirming a gap between theory and empirical results. In particular, for the Korean franchise market, even basic data on contract conditions are insufficient, and a relationship between business characteristics and contract conditions has not yet been proven. These conditions serve as the motivation for this study.

Park (2013) focuses on how franchisors adjusted the number of franchisees and directly managed stores, especially when the conditions of the business area change. Analyzing Information Disclosure Documents (as of the second half of 2010) with an ANOVA test, he finds that the rate of increase in the number of franchisees and directly managed stores, whose business areas become unprotected, was higher than that of stores facing the opposite situation. Accordingly, the sales of franchisees in the former situation were lower than those of rate stores in the latter.

When the profit structure of franchisors is excessively dependent on the franchise fee and store opening cost, as in Korea, the market is likely to have increasingly more franchisees, possibly driving the sales of individual stores down. Park (2013) demonstrates that such concerns are emerging in the domestic franchise market. However, the effect of business area protection on the eventual profits of franchisees is debatable. When business area is protected, some part of the franchisee's income from the exclusive territory may be transferred to the franchisors (in the form of franchise fees or royalty rates), ultimately not contributing to the profits of the franchisee. Our study differs in that it quantifies the business area based on distance, population, and administrative district and uses this new definition in a regression analysis to find how business area protection affects both contract conditions and producer welfare, respectively.

Meanwhile, Klein and Saft (1985) qualitatively discuss essential commodity contracts (tie-in contracts).[3] Without a tie-in contract, the free-rider problem may arise in which opportunistic franchisees use cheap and low-quality intermediate

[3]An essential commodity refers to raw and subsidiary materials that franchisees must purchase from their headquarters or from a supplier designated by the headquarters for the purpose of standardizing products and services produced by franchisees.

goods to secure high profits. If consumer utility is reduced due to quality degradation, it may place a negative externality on the demands of other franchisees using standardized intermediate goods. Thus, they argue that commodity contracts that lower monitoring cost required to ensure product standardization and to prevent the free-ride problem. If so, the distribution margin included in the price of the essential commodity is understood as a reward for maintaining the demand for individual franchisees due to product standardization and the lowering of monitoring costs that must be paid by franchisees in part.

However, there has been strong criticism in the domestic franchise market, pointing out that the distribution margin for raw and subsidiary materials is excessive and related to the inclusion of items irrelevant to quality in the contract. Accordingly, the appropriate level of margin and the range of commodities to be included in the tie-in contract are becoming important issues. However, it is very difficult to obtain objective data necessary to examine these issues. Therefore, this study intends to examine the possibility of and the extent to which the tie-in contract is linked to the royalty contract and the effect it has on the welfare of producers.

III. Structure of the Korean Franchise Market

In this section, we examine the growth trend of the domestic franchise market (in terms of the numbers of franchisors, brands, franchisees, and directly managed stores) and then identify the different characteristics between industries. The data in this section are mainly constructed based on Information Disclosure Documents (2014-2016).

A. *Number of Franchisors and Their Brands*

According to Figure 1, the number of franchisors increased from 3,923 in 2014 to 4,664 in 2016, as has the number of brands, from 4,843 to 5,741. During this period, both figures increased by 19%, indicating that new franchisors have actively entered the market.

Table 1 shows the rise and fall of franchise brands. In 2015 alone, domestic

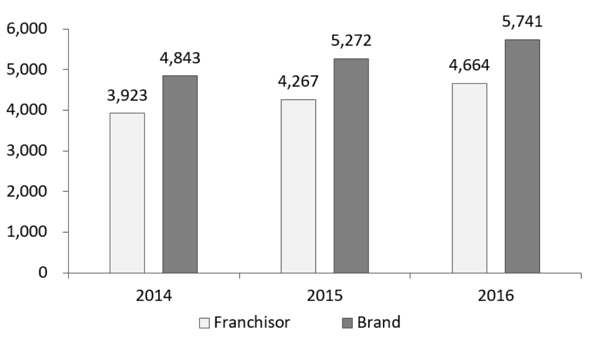


FIGURE 1. NUMBER OF FRANCHISORS AND BRANDS

Source: Calculated by the author using microdata from Information Disclosure Documents (2014-2016).

TABLE 1—RISE AND FALL OF FRANCHISE BRANDS

	No. of brands	2014	2015	2016	Life-span
①	1,046		←→		1 year
②	1,178		←—	—→	2 years
③	1,497			←→	1 year
④	17	—→		←→	-
⑤	3,049	←—	——	—→	3+ years
⑥	1,778	—→			-

Source: Calculated by the author using microdata from Information Disclosure Documents (2014-2016).

franchisors newly launched 2,224 brands (①+②), and a staggering 47% (1,046) failed to last for more than a year, possibly due to low profitability and failures to recruit new franchisees.

Of the 5,741 active brands in 2016 (②~⑤), only half were older than three years, while those who remained were fairly new (maximum of two years). Nevertheless, the market supply of brands was growing in net terms as new entrants outnumbered exiters.

Regarding the distribution of brands by industry (Figure 2), there were 4,341 (75.6%) in food services, 1,076 (18.7%) in the service sector, and 324 (5.6%) in wholesale and retail businesses out of 5,741 brands in 2016. With food services taking up an overwhelming proportion, about 70% of brand growth from 2015 to 2016 also occurred in this sector.

A closer examination reveals that fried chicken (22%), Korean food (Hansik, 18%), and coffee franchises (11%) account for more than 50% of all food services. The concentration on food services is largely dependent on overconcentration in some specific food sectors.

In Korea, the proportion of self-employment is higher than those in other countries with similar economies (Lee, 2020). In addition, establishments are concentrated in a few industries, such as restaurants and accommodation, causing fierce competition among these small business owners (Lee, 2017). These types of industrial concentration and overcrowding are also occurring in the franchise market.

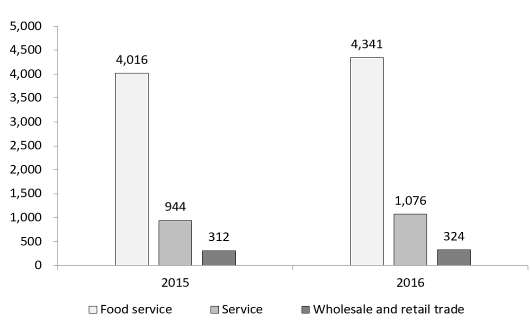


FIGURE 2. NUMBER OF BRANDS BY LARGE SECTORS

Source: Calculated by author using microdata from Information Disclosure Documents (2015-2016).

B. *Number of Franchisees and Directly Managed Stores*

The growth of the franchise market is also evident when examining the number of stores. As shown in Figure 3, franchisees increased from 208,099 in 2014 to 230,054 in 2016, and directly managed stores increased from 15,459 to 17,033 during the same period.

A closer look at the number of franchisees in Figure 4 reveals that food services account for 112,533 (48.9%) of the total, followed by the service industry at 68,090 (29.6%) and wholesale and retail trade at 49,431 (21.5%).

However, in terms of the number of franchisees under each brand, (the scale of the franchise system), the order is wholesale and retail, services, and food services. This occurs because small-scale franchises are concentrated in food services, whereas wholesale and retail businesses consist of many large-scale and corporate types of franchise systems.[4]

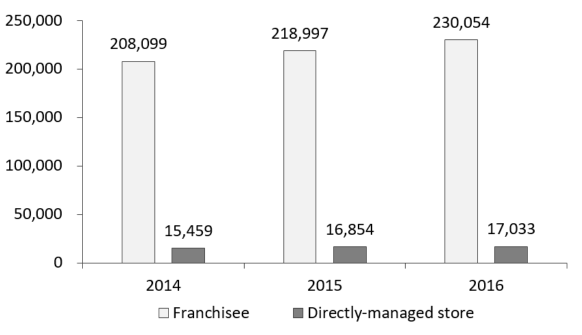


FIGURE 3. NUMBER OF FRANCHISE STORES AND DIRECTLY MANAGED STORES

Source: Calculated by the author using microdata from Information Disclosure Documents (2014-2016).

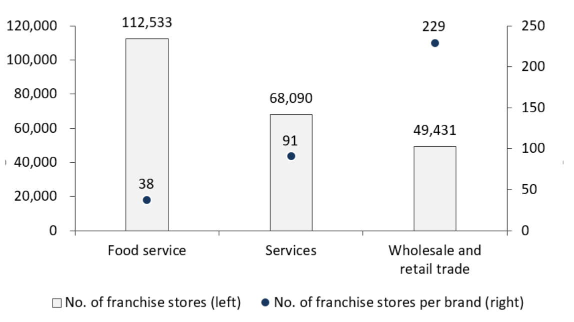


FIGURE 4. NUMBER OF FRANCHISEES AND NUMBER OF FRANCHISE STORES PER BRAND

Source: Calculated by the author using microdata from Information Disclosure Documents (2016).

[4]In fact, in the wholesale and retail industry, convenience stores have the largest number of franchisees per brand. Looking at the status of the three major convenience stores, CU has 10,746, GS25 has 10,604, and 7-Eleven has 8,206 stores. This contrasts with the situation in the food service industry, such as Korean food (24), chicken (80), and coffee (50).

C. *Sales of Franchisees*

Total franchisee sales increased from 50.1 trillion won in 2014 to 64.4 trillion won in 2016, and sales per franchise also increased from 286 million won to 325 million won during the same period. Along with more franchisors, brands, and franchisees, the growth in franchisee sales also underpins the quantitative expansion of the market.

At each industry level, total franchisee sales in the food services sector were largest at 30.6 trillion won in 2016, while the wholesale and retail sector earned 27.9 trillion won. Given that the number of franchisees in food services is more than double that in the wholesale and retail sector, the difference between the two appears to be surprisingly small.

Moreover, the growth rate of sales compared to 2015 is highest at 3.9 trillion won in the wholesale and retail sector; hence it is likely that the market size will reverse in the future.

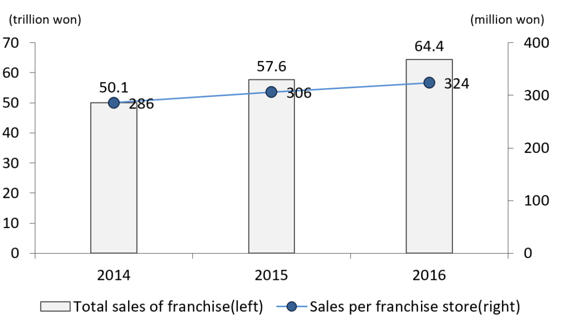


FIGURE 5. TOTAL SALES OF FRANCHISES AND SALES PER FRANCHISE STORE

Source: Calculated by the author using microdata from Information Disclosure Documents (2014-2016).

TABLE 2—RISE AND FALL OF SALES BY SECTOR

	Total sales of franchise (trillion won)		Sales per franchise store (100 million won)	
	Year 2016	YoY	Year 2016	YoY
Total	**64.4**	**+6.7**	**3.2**	**+0.2**
Food services	30.6	+2.5	3.0	+0.1
Services	5.9	+0.3	1.2	+0.1
Wholesale and retail	27.9	+3.9	5.8	+0.3

Source: Calculated by the author using microdata from Information Disclosure Documents (2015-2016).

D. *Brands without Directly Managed Stores*

Table 3 indicates that nearly 60% of brands do not have directly managed stores and rely solely on franchisees. More specifically, this applies to 60.8% in food services, 58.3% in services, and 45.7% in wholesale and retail.

Such a practice raises significant concerns because by not running directly managed stores, franchisors may not have the opportunity to gain insight into their

TABLE 3—PERCENTAGES OF BRANDS WITHOUT DIRECTLY MANAGED STORES

	Brands without directly managed stores (%)
Total	**59.5**
Food services	60.8
Services	58.3
Wholesale and retail trade	45.7

Source: Calculated by the author using microdata from Information Disclosure Documents (2016).

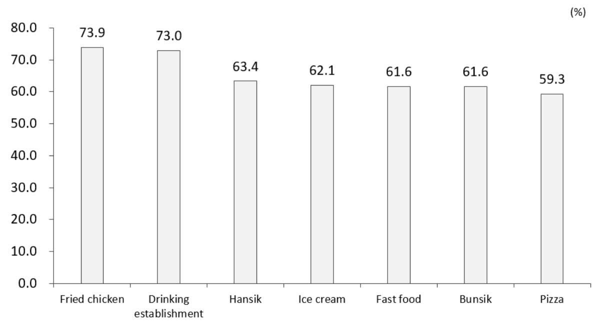


FIGURE 6. PERCENTAGE OF BRANDS WITHOUT DIRECTLY MANAGED STORES (FOOD SERVICES)

Source: Calculated by the author using microdata from Information Disclosure Documents (2014-2016).

products' competitiveness and learn from trial and error. This, in turn, can negatively impact their franchisee sales through transferred business risks. Indeed, Figure 6 shows that brands without directly managed stores are particularly prevalent in businesses with low sales performance, such as fried chicken, Hansik, and drinking establishments.

Also, according to the correlation analysis, sales are negatively correlated with the number of franchise stores (-0.104) but positively correlated with the number of directly managed stores (+0.345). In other words, the more franchisors are exposed to the market by operating directly managed stores, the more likely franchise sales will increase.

IV. Conditions of the Franchise Agreement

The franchise agreement involves various costs the franchisee has to bear, including the initial costs before the start of the business, the costs incurred when running the business, and several conditions imposed on business activities.

A. *Setup Costs (Prior to Opening)*

In order to start a franchise store, prospective franchisees pay an average of 117.6 million won to franchisors. The setup costs include the franchise fee (10.1 million won, 8.6%), the training fee and deposit (7.6 million won, 6.5%), the interior fit-out (53.4 million won, 45.4%), and other expenses (46.5 million won, 39.5%), as shown

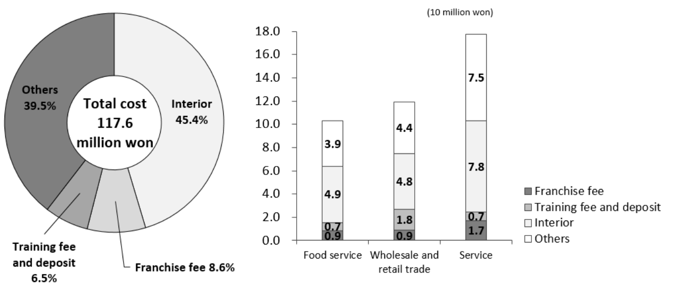


FIGURE 7. FRANCHISE SETUP COST OVERALL (LEFT) AND BY SECTOR (RIGHT)

Source: Calculated by the author using microdata from Information Disclosure Documents (2016).

in the panel on the left in Figure 7.

Franchisors reap rewards for transferring their accrued brand reputation and management know-how. The franchise fee is a price of using a brand with reputational capital. From the franchisees' perspectives, they are willing to pay more for reputable brands. Accordingly, rather than criticizing the absolute amount of the franchise fee, it would be more logical to examine whether the high price is justified by the franchisor's level of experience and expertise.

The interior fit-out of a franchise store accounts for the greatest share of the initial outlay. With regard to this fee, franchisees are often forced into contracts with certain interior design firms designated by the franchisors. This has recently led to some controversy. Of course, franchisors should be able to provide interior blueprints or guidelines in order to maintain consistency of their stores. However, binding contracts to a specific firm may violate Article 12 (Prohibition on Unfair Trade Practices) of the Fair Transactions in Franchise Business Act. Recognizing this, most franchisors accept the franchisee's choice of contractor. This, however, entails an exorbitant supervision fee.

When the interior cost becomes a common source of revenue for franchisors, this will further incentivize them to open more stores, ultimately damaging the win-win relationship between franchisors and franchisees.

B. *Types of Royalties*

A royalty payment is a monthly payment made by the franchisee to the franchisor after the opening of the store, and it is another form of remuneration for using the franchisor's reputational capital and business know-how.

According to the panel at the top of Table 4, the average adoption rate in royalty contracts is 68.3%, which appears to be relatively high figure. However, a fixed-rate royalty (a certain percentage of the franchisee's sales), which is most suitable due to the nature of royalties and which induces mutual cooperation by linking the profits of franchisors and franchisees, is only used by 25.4% of brands.

A fixed-sum royalty is more commonly adopted with 41.8% of brands. Within such a contract, the royalty revenue is fixed regardless of the franchisee's sales; hence, franchisors are less inclined to support their "progenies" in their efforts to improve their performance.

TABLE 4—ADOPTION RATE OF ROYALTIES

		Total	Fixed rate	Fixed sum	Other
Adoption Rate of Royalties (%)	**Total**	**68.3**	**25.4**	**41.8**	**2.8**
	Food services	67.5	27.5	40.6	0.8
	Services	76.0	20.3	48.6	10.6
	Wholesale and retail trade	52.0	14.6	36.2	3.5

Note: Because certain brands accept multiple royalty types, the sum of the types does not equal the total.

Source: Calculated by the author using microdata from Information Disclosure Documents (2016).

TABLE 5—ROYALTY RATES

		Fixed-rate (%)	Fixed-sum (1,000 won)
Royalty Rate	Food services	2.5	285
	- Bunsik (Snack food)	2.1	220
	- Hansik (Korean food)	2.1	258
	- Fried chicken	2.4	233
	Services	6.1	739
	Wholesale and retail trade	5.4	449

Source: Calculated by the author using microdata from Information Disclosure Documents (2016).

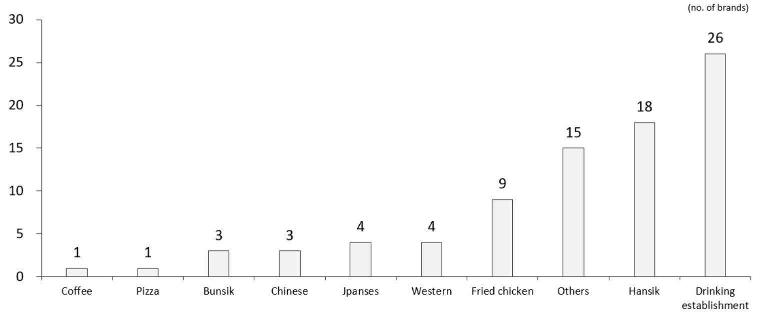


FIGURE 8. NUMBER OF BRANDS PROVIDING ROYALTY DISCOUNTS CONDITIONAL ON RAW MATERIALS PURCHASES

Source: Calculated by the author using microdata from Information Disclosure Documents (2016).

There are other types of royalty schemes as well. In 2.8% of brands, royalties are set in proportion to the price of raw materials purchased by franchisees or in proportion to the number of facilities or even the furniture in the store.

According to Table 5, royalty rates tend to be quite low in food services, such as the bunsik, Hansik, and fried chicken types, where franchisees are concentrated and sales per store are relatively low. In such overcrowded markets, the survival rate of stores is quite low. Therefore, franchisors may try to secure a one-time initial payment (franchise fee) by recruiting franchisees or may attempt to include the franchisors' margin in the price of the raw materials.

In particular, we found many cases in which the royalty rate was discounted when franchisees purchased raw materials from their franchisors or designated firms. Such cases frequently arose in the food services sector, such as drinking establishments (26 cases), Hansik stores (18 cases), and fried chicken stores (nine cases), where the royalty discount could be in the range of 70-100%. This implies that royalties and raw materials margins are linked and that a royalty discount strategy is used to induce raw materials purchase contracts.

C. *Business Area*

The business area refers to the range of the commercial area in which franchisees can exclusively operate under the brand. In accordance with the Franchise Business Act (Article 5), all franchisors in Korea must set up a business area for each franchisee and must not open directly managed stores or other franchisee stores within the area.[5]

As of 2016, approximately 80% of brands had established areas based on distance (radius or walking). The distance scheme is used the most in food services (87%), with the narrowest range (0.69km).

TABLE 6—TYPES OF BUSINESS AREAS

(Unit: %)

	Distance (radius/walking)		Population/ households	Administrative area (eup/myeon/dong)	Other
		Territorial range			
Total	**79.7**	**0.73km**	**8.6**	**16.0**	**3.5**
Food services	87.0	0.69 km	7.1	11.2	2.3
Services	52.2	1.04 km	15.0	33.8	7.8
Wholesale and retail trade	72.0	0.75 km	7.5	20.9	4.7

Source: Calculated by the author using microdata from Information Disclosure Documents (2016).

V. Effects of Franchise Market Structures on Contract Conditions

This section analyzes the relationship between the market structure and contract conditions of domestic franchise businesses. First, we explain the main variables and models, after which we estimate models of franchise fees and royalties.

A. *Variables and Models*

We established a 'basic database' with the cooperation of the Korea Fair Trade Mediation Agency. The basic database contains various summary information on all brands in the Information Disclosure Documents (2015-2017). Given that the summary information for each year reflects the situation of the previous year, our

[5]Despite concerns that this practice hinders market competition, it has been maintained as an obligatory clause for all franchisors according to the legislative purpose of the business stability of franchisees.

basic database contains information from 2014 to 2016.

However, contract terms such as royalties and business areas, which are essential for this study, were missing from the basic database. In order to create these variables, we searched all brands in the basic database on the franchise business website and downloaded the original Information Disclosure Documents.[6] Through this work, royalties, business areas, opening dates of directly managed stores, the existence of the regional headquarters, and other factors were extracted and quantified for use in the ensuing quantitative analysis.

In particular, the average franchise fee in Table 7 is 9.8 million won, which is lower than the average of 10.1 million won mentioned in Section 5; in Section 5, the

TABLE 7—SUMMARY STATISTICS OF VARIABLES

Variables		No. of Obs.	Mean	Std. Dev	Min	Max
Franchisor Level						
	Franchise fee (thousand won)	10,917	9831.47	14194.23	0.00	330,000.00
Royalty	Fixed-sum (thousand won)	3,407	394.54	578.97	10.00	11,000.00
	Fixed-rate (%)	2,043	3.13	2.95	0.30	38.50
	No. of stores	10,917	44.06	269.33	0.00	10,857.00
	No. of directly managed stores	10,917	3.07	25.66	0.00	777.00
	Ratio of directly managed stores (%)	8,679	22.41	34.91	0.00	100.00
	No. of regions with directly managed stores	10,917	0.40	1.36	0.00	17.00
	No. of franchisee stores	10,917	40.99	264.28	0.00	10,746.00
	Annual operating months	10,917	44.51	56.55	0.00	474.00
	Store area (m^2)	10,917	28.88	54.10	0.00	2,200.00
Territorial range	Radius (km)	4,692	0.53	3.87	0.00	250.00
	Walking distance (km)	4,692	0.05	0.19	0.00	3.00
	Vehicle distance (km)	4,692	0.02	0.78	0.00	50.00
	Population (thousand)	4,692	3.06	22.93	0.00	550.00
	Households (thousand)	4,692	0.22	1.74	0.00	50.00
	Administrative area (eup/myeon/dong)	4,692	0.16	0.37	0.00	1.00
	Others	4,692	0.03	0.17	0.00	1.00
Brand Level						
	No. of franchisor's brands	10,917	1.75	1.83	1.00	20.00
	No. of employees (person)	10,872	75.96	843.50	0.00	28389.00
	Advertising expenses (thousand won)	10,917	51.40	1,426.43	0.00	123,246.70
	Promotional expenses (thousand won)	10,917	28.64	1,097.72	0.00	88,600.00
	No. of three-year law-violations	10,917	0.05	0.35	0.00	27.00
Others						
	No. of contract terminations	10,917	1.54	10.06	0.00	278.00
	No. of contract cancellations	10,917	3.21	19.62	0.00	1076.00
	Population (ten thousand)	10,917	780.50	415.17	20.41	1,267.20
Sector	Food services	10,917	0.06	0.23	0.00	1.00
	Services	10,917	0.18	0.39	0.00	1.00
	Wholesale and retail trade	10,917	0.76	0.43	0.00	1.00
	Year 2015	10,917	0.48	0.50	0.00	1.00
	Year 2016	10,917	0.52	0.50	0.00	1.00

[6]The web address is 'https://franchise.ftc.go.kr/index.do'.

average value was calculated for brands with a positive franchise fee, whereas in the regression analysis in this section, all brands even with a zero franchisee fee were included. In 2016, the total number of brands was 5,741, of which 245 (4.3%) had a franchise fees of zero. This is done to reflect franchisors' recent strategy of not charging franchise fees in the fierce competition to recruit franchisees.

In the regression analysis, only data from 2015-2016 were used, with data from 2014 excluded. This was done because the key variables in this study (industry, business start date, number of franchisees, number of directly managed stores, franchise fees, royalties, business area, etc.) do not exist in the 2014 data. The final dataset used consists of 48% of 2015 brands and 52% of 2016 brands.

The estimation model is as follows:

$$y_{ijst} = X_{ijst}\alpha + \sum_j \beta_j I_j + \sum_s \gamma_s D_s + \sum_t \delta_t T_t + u_{ijst}$$

The dependent variable y_{ijst} is the contract terms (franchise fee, fixed-sum or fixed-rate royalty) of brand i with its franchisor located in region s, in industry j and in year t. For example, it can be the franchise fee for the restaurant brand J&M Chicken located in Seoul in 2016.

The independent variable X_{ijst} includes various franchise characteristics according to the model. At the brand level, it includes the number of stores, the number of directly managed stores, the number of regions where directly managed stores exist, the number of franchisees, and the duration of the franchise business. These characteristics are closely related to the size of the franchise system as well as to brand awareness and can thus affect the reputation capital of the brand. In addition, various types of business areas are controlled as operating conditions faced by franchisees.

Meanwhile, at the level of the establishment (trade name), the number of brands owned is included in X_{ijst}. When a single franchisor operates multiple brands, not only the brand reputation but also the franchisor's reputation can be formed. Specifically, the franchisor's reputation may have a common influence on each brand or may induce interactions between brands. For this reason, it must be controlled in the model. In addition, the number of employees, advertising expenses, promotional expenses, and the number of violations of the law are controlled.

I_j is an industry dummy controlling the fixed effect by industry at the level of a middle or large sector. D_s is a regional dummy, set at the metropolitan city level. T_t is a year dummy, representing year-by-year economic fluctuations, and finally u_{ijst} is an error term that satisfies the IID condition. If u_{ijst} satisfies the IID assumption, it means that the unobserved characteristics of different brands are independently extracted from an identical distribution, which is a rather strong assumption. Other estimation methods can be explored if more abundant data exist, but for now, we have controlled sector, region, and year as measures that can contribute to satisfying the IID assumption.

As an estimation method, we basically use the Ordinary Least Square approach considering the structure of pooled data, but also use instrumental variable estimations under the potential endogeneity of a few variables.

B. *Estimation Results: Franchise Fee Model*

Table 8 reports the influence of various market and business characteristics on the level of the franchise fee. According to model (1), when the number of directly managed stores and the period of the franchise business increase by one unit, the franchise fee significantly increases by 13.8 and 33.4 thousand won, respectively.

TABLE 8—ESTIMATION RESULTS OF THE FRANCHISE FEE MODEL (1)

Dep. Var.: Franchise fee (thousand won)	Model (1)	Model (2)	Model (3)	Model (4)
No. of directly managed stores	13.82*** (5.19)	13.20** (5.19)	12.83** (6.12)	11.62* (6.08)
No. of months of franchise business	33.42*** (5.15)	37.09*** (5.21)	36.08*** (5.24)	29.48*** (5.17)
No. of months of franchise business^2	-0.06*** (0.02)	-0.08*** (0.02)	-0.07*** (0.02)	-0.05** (0.02)
No. of franchisor's brands	211.03*** (71.65)	221.09*** (71.64)	179.93** (73.29)	199.77*** (72.66)
No. of franchisee stores	-1.59*** (0.54)	0.49 (0.71)	0.64 (0.73)	0.23 (0.72)
No. of contract terminations		-45.61*** (16.09)	-48.95*** (16.29)	-58.72*** (16.04)
No. of contract cancellations		-29.56*** (7.78)	-30.32*** (7.80)	-37.31*** (7.69)
Store area (m^2)	64.37*** (2.41)	64.12*** (2.41)	64.19*** (2.41)	58.91*** (2.39)
No. of brand per total sales	-2.53e+09*** (4.96e+08)	-2.41e+09*** (4.98e+08)	-2.40e+09*** (5.00e+08)	-1.27e+09** (5.00e+08)
No. of brand within a sector	-1.39*** (0.34)	-1.47*** (0.34)	-1.25*** (0.34)	0.59 (0.38)
Total sales of a sector	-8.37e-07*** (1.10e-07)	-8.23e-07*** (1.10e-07)	-8.40e-07*** (1.11e-07)	-1.82e-07 (1.26e-07)
Advertising expenses			-0.14 (0.35)	-0.27 (0.35)
Promotional expenses			0.17 (1.09)	0.56 (1.07)
No. of three-year law-violations			-435.33 (372.19)	-474.26 (368.89)
No. of employees			0.09 (0.16)	0.22 (0.17)
Population			1.78*** (0.32)	9.96 (23.16)
Year 2015				406.00 (259.33)
Wholesale and retail trade				-7,180.97*** (746.86)
Food services				-7,149.40*** (393.87)
Constant	9,570.48*** (371.11)	9,509.73*** (371.25)	8,162.69*** (451.64)	332.80 (29,150.38)
Region dummy	X	X	X	O
Observations	10,917	10,917	10,872	10,872
R^2	0.08	0.09	0.09	0.12

Note: Standard errors are in parentheses. *, **, and *** denote statistical significance at the 10%, 5%, and 1% level, respectively.

This implies that franchisors transfer their business know-how acquired from operating directly managed stores and franchise businesses to the franchisees and receive a franchise fee in return. Meanwhile, a negative coefficient appears in the squared term of the franchise business period, indicating that the rate of increase is gradually slowing down.

The number of brands owned by the franchisor also has a significant impact, with the franchise fee increasing by 211 thousand won as one more brand is added. When operating multiple brands, awareness of the franchisor can be increased and various trials and errors of the franchisor can have a positive impact on the reputation capital of each brand. These results are consistent with Mathewson and Winter (1985), who suggest that the franchise fee can be set high as the brand reputation capital accumulates.

On the other hand, as the number of franchisee stores increases, the franchise fee tends to decrease. However, this appears less convincing because an increase in the number of franchisees can also be positive, increasing brand awareness and accumulating reputation capital.

Models (2) to (4) control for factors related fluctuations in the number of franchisees. Interestingly, as the number of contract terminations or cancellations of franchisees increases, the franchise fee significantly decreases. Contract terminations and cancellations can negatively affect the brand reputation capital by lowering the reliability and degrading the cooperative relationship between the two parties, possibly becoming a direct reason for the suspension of new contracts. At the same time, the number of franchisees, which previously showed a negative sign, no longer shows a significant effect in models (2) to (4). Considering a highly positive and significant correlation between the number of franchisees, the number of contract terminations and cancellations,[7] the negative effect previously shown appears to have been induced by the effect of contract terminations and cancellations.

In addition, models (1) to (3) show that the franchise fee is likely to decrease as the number of brands per total sales and the number of brands within a sector increase. This outcome supports the franchisor's strategy of lowering the franchise fee amid the competition to recruit franchisees.[8] Model (4) controls for year dummies and industry dummies, measured at the large classification level. Recalling that the food services sector has significantly more brands than others, it is possible for the negative effect of the number of brands within a sector to be included in the fixed effect of the industry.

In order to verify the robustness of the above results, we examine various models, as shown in Table 9. Model (5) includes 44 middle-class industry dummies to control for fixed effect by sector. As a result, key results pertaining to the number of directly managed stores, the period of the franchise business, the number of franchisor’s

[7]Usually, as the number of contract terminations and cancellations increases, the number of franchisees is expected to decrease. However, the number of franchisees showed high correlations with the number of contract terminations and cancellations, at 0.63 and 0.57, respectively. This can be understood considering that when the franchise market grows, the number of new openings increases to an extent that overwhelms the number of cancellations. In short, the number of franchisees reflects the effects of both new contracts and contract terminations.

[8]The number of brands per total sales and the number of brands within a sector are measured at the middle class of the industry sector.

TABLE 9—ESTIMATION RESULTS OF THE FRANCHISE FEE MODEL (1)

Dep. Var.: Franchise fee (thousand won)	Model (5)	Model (6)	Model (7)	Model (8)
No. of directly managed stores	12.15** (6.11)			24.43*** (8.06)
No. of directly managed stores (*t*-1)			11.80** (6.23)	
Ratio of directly managed stores		25.91*** (4.60)		
No. of regions with directly managed stores		416.95*** (114.99)		
No. of months of franchise business	27.59*** (5.04)	31.88*** (6.04)	27.17*** (5.04)	29.30*** (5.28)
No. of months of franchise business^2	-0.04* (0.02)	-0.06** (0.02)	-0.04** (0.02)	-0.07*** (0.02)
No. of franchisor's brands	158.45** (70.90)	133.63* (77.07)	151.50** (70.89)	164.57** (77.67)
No. of franchisee stores	0.71 (0.71)			2.64*** (0.99)
No. of franchise stores (*t*-1)			0.30 (0.83)	
No. of stores		0.57 (0.72)		
No. of contract terminations	-95.59*** (15.63)	-84.83*** (15.71)		-103.31*** (26.41)
No. of contract terminations (*t*-1)			-109.80*** (17.09)	
No. of contract cancellations	-46.34*** (7.86)	-42.98*** (8.21)		-73.65*** (15.02)
No. of contract cancellations (*t*-1)			-33.92*** (8.72)	
Territorial range (Radius/ Walking)				71.45** (28.25)
Store area (m^2)	55.41*** (2.34)	110.64*** (3.51)	55.56*** (2.34)	60.18*** (2.56)
No. of brand per total sales	5.30e+08 (1.58e+09)	1.81e+09 (1.81e+09)	1.20e+09 (1.57e+09)	-5.88e+07 (1.73e+09)
No. of brand within a sector	3.00 (3.68)	2.46 (4.15)	2.51 (3.69)	2.66 (4.02)
Total sales of a sector	-7.08e-07 (7.32e-07)	-5.83e-07 (8.26e-07)	-4.77e-07 (7.34e-07)	-9.93e-07 (8.28e-07)
Advertising expenses	-0.33 (0.34)	-0.26 (0.34)	-0.39 (0.34)	0.98 (1.16)
Promotional expenses	0.59 (1.05)	0.47 (1.05)	0.76 (1.05)	0.25 (1.28)
No. of three-year law-violations	-411.92 (356.29)	-383.90 (367.67)	-378.60 (356.50)	-553.62* (325.31)
No. of employees	0.15 (0.17)	0.05 (0.17)	0.16 (0.17)	0.10 (0.16)
Population	8.57 (22.38)	1.85 (25.10)	6.78 (22.40)	0.70 (23.81)
Year 2015	440.16* (265.16)	879.40*** (315.54)	475.90* (265.20)	388.51 (282.28)
Constant	3,719.63 (29,632.81)	8,744.39 (33,300.05)	3,759.60 (29,657.00)	16,224.62 (31,861.90)
Region dummy	O	O	O	O
Middle-sector dummy	O	O	O	O
Observations	10,872	8,636	10,872	6,422
R^2	0.19	0.25	0.19	0.28

Note: Standard errors are in parentheses. *, **, and *** denote statistical significance at the 10%, 5%, and 1% level, respectively.

brands, and the number of franchisees are maintained.

In model (6), the number of stores, the ratio of directly managed stores, and the number of regions in which directly managed stores exist are controlled for, instead of the number of directly managed stores and franchisees. While the number of stores is insignificant, the ratio of directly managed stores and the number of regions with directly managed stores have significantly positive effects. This reaffirms that the experience of running directly managed stores is an important factor with regard to the accumulation of brand reputation capital.

Meanwhile, the possibility that the franchise fee can affect the number of franchisees needs to be checked. If prospective franchise founders prefer brands with low franchise fees, an endogeneity problem arises in relation to the number of franchisees. Also, if a certain brand has a reputation for being highly profitable, this increases the franchise fee and makes it easier to recruit franchisees; accordingly, the number of franchisee stores may be vulnerable to the endogeneity problem due to simultaneity.[9,10] Therefore, model (7) considers previous values of the variables, but it shows no significant differences from the previous results.[11]

In Model (8), we estimate the equation using only the observations with the business area set. These results show that as the distance range of the business area increases by 1km, the franchise fee tends to increase by 71.5 thousand won. As the scope of the exclusive commercial district widens, the survival probability of a franchisee may increase. When uncertainties regarding business stability are reduced, it will be easier for franchisors to recover their reputation capital at the time of the contract.

C. *Estimation Results: Royalty Model*

According to the literature, royalties tend to decrease as the period of the franchise business becomes longer. Therefore, it is necessary to examine whether such a relationship exists in the domestic franchise industry as well.

Looking at models (1) and (2), where the fixed-sum type of royalty is a dependent variable, the royalty tends to increase as the number of directly managed stores increases. This suggests that the fixed-sum royalty is being utilized as a tool similar to a franchise fee. A franchisor would like to receive all of the compensation for reputational capital when signing a franchise contract, but it is impossible to recover

[9]It is unclear as to whether a reputation for profitability affects the number of directly managed stores or brands, as in a situation where market risks exist, business expansion strategies through franchisees rather than directly managed stores can be considered first due to the nature of franchise businesses. In addition, there is a possibility that the strategic judgment and expectations of entrepreneurs play a greater role than the reputation in the expansion of directly managed stores.

[10]The number of months of a franchise business cannot be completely free from the problem of endogeneity. This is true because if a product has high marketability and profitability, a positive reputation for the brand will be accumulated, which will increase the franchise fee and, at the same time, increase the survival rate of the business and increase the number of months of the franchise business. In order to remove the endogeneity caused by this simultaneity issue, it is necessary to control for the variables of product quality or marketability. Improvements in this area are left for future research.

[11]An alternative method that can be considered is to estimate a fixed-effect model by constructing panel data for each franchisor (or brand). In future follow-up studies, it will be desirable to apply an effective methodology to eliminate the fixed effect based on franchise market data compiled over a longer term.

TABLE 10—ESTIMATION RESULTS OF THE ROYALTY MODEL

Dep. Var.	Fixed-sum royalty (thousand won)		Fixed rate royalty (%)	
	Model (1) (2015~16)	Model (2) (2016)	Model (3) (2015~16)	Model (4) (2016)
Degree of royalty reduction_ Fixed-sum	0.98*** (0.03)	0.97*** (0.04)		
Degree of royalty reduction_ Fixed-rate			0.32*** (0.11)	0.35** (0.15)
No. of directly managed stores	6.86*** (1.97)	7.14*** (2.76)	-0.00 (0.00)	-0.00 (0.00)
No. of months of franchise business	-0.33 (0.28)	-0.04 (0.38)	0.00 (0.00)	0.00 (0.00)
No. of months of franchise business^2	0.00* (0.00)	0.00 (0.00)	-0.00 (0.00)	-0.00 (0.00)
No. of franchisor's brands	3.01 (3.80)	3.17 (5.02)	0.08** (0.03)	0.09** (0.04)
No. of franchisee stores	-0.11 (0.12)	-0.19 (0.17)	0.002*** (0.00)	0.002*** (0.00)
No. of contract terminations	-1.21 (1.25)	-0.76 (1.92)	0.03 (0.02)	0.02 (0.03)
No. of contract cancellations	-0.13 (1.15)	-0.92 (1.64)	-0.01* (0.01)	-0.02** (0.01)
Territorial range(Radius/ Walking)	-5.18 (9.81)	-6.69 (12.94)	0.09 (0.07)	0.07 (0.09)
Store area (m^2)	2.94*** (0.24)	2.18*** (0.28)	0.00 (0.00)	0.00 (0.00)
No. of brand per total sales	5.08e+07 (8.42e+07)		-1.18e+06* (631,101)	
No. of brand within a sector	-0.03 (0.18)	0.03 (0.23)	-0.01*** (0.00)	-0.00*** (0.00)
Total sales of a sector	1.06e-08 (3.76e-08)	8.77e-09 (2.71e-08)	1.02e-09*** (3.67e-10)	7.26e-10*** (1.92e-10)
Advertising expenses	0.05 (0.05)	0.40 (0.40)	0.00*** (0.00)	0.00*** (0.00)
Promotional expenses	0.61** (0.26)	0.86*** (0.31)	-0.00 (0.00)	0.00 (0.00)
No. of three-year law-violations	1.15 (10.98)	121.81 (77.14)	-0.22 (0.20)	-0.36 (0.67)
No. of employees	-0.11 (0.16)	-0.22 (0.25)	-0.00*** (0.00)	-0.00*** (0.00)
Population	0.00 (1.12)	0.20 (0.28)	0.00 (0.01)	0.00 (0.00)
Year 2015	5.82 (13.37)		0.09 (0.11)	
Constant	-274.91 (1,571.79)	-127.13 (485.29)	7.51** (3.02)	4.07 (2.91)
Region dummy	O	O	O	O
Meddle-sector dummy	O	O	O	O
Observations	2,727	1,587	1,713	1,013
R^2	0.49	0.46	0.41	0.43

Note: Standard errors are in parentheses. *, **, and *** denote statistical significance at the 10%, 5%, and 1% level, respectively.

all in such cases due to uncertainty about the future. Accordingly, they are likely to receive a fixed-sum royalty during the business period as a complement.

In models (3) and (4), in which the fixed-rate royalty is set as the dependent variable, a significant positive effect is found in the number of brands, while the number of directly managed stores and the franchise business period became insignificant.

Another interesting point is that as the number of franchisees increases, the fixed-rate royalty tends to increase. If there are more franchisees in the market, competition intensifies, which may reduce the sales of individual franchisees as well as the royalty income of franchisors. Accordingly, the franchisor has an incentive to cover the reduced income by increasing the royalty rate. This interpretation is consistent with Rubin (1978)'s argument that, when the market is saturated, franchisees earn less in sales and thus franchisors increase royalties.

We knew cases in which royalties were discounted conditional on the purchasing of raw materials from franchisors or designated firms. Considering this, models (1) to (4) include the degree of the royalty reduction. These results show that both fixed-sum and fixed-rate royalties tend to be set significantly higher as the royalty reduction is larger. This suggests that franchisors adopt a strategy of setting royalties high initially and then lowering them after the signing of a purchase contract. They can secure profits by including a margin in the price of the raw materials, and a royalty discount strategy can be introduced to induce purchase contracts.

VI. Effects of Contract Conditions on Producer Welfare

At this stage, we proceed to discuss to effect of franchise contract terms on business performance outcomes and provide implications on how these effects are related to conflicts between two parties over the distribution of value added.

A. *Merging Contract Condition Variables into MDIS*

Tackling this issue requires additional data about the business performance outcomes of franchisors and franchisees. Accordingly, we construct the data from the Microdata Integrated Service (MDIS) of wholesale and retail surveys as well as from service industry surveys of Statistics Korea. The MDIS dataset contains a wealth of information on sales and operating profits and various characteristics of individual firms. Using MDIS data, we can distinguish franchise and non-franchise businesses and identify franchisors and franchisees.

Specifically, the contract condition variables from the Information Disclosure Documents are merged into MDIS at the five-digit level of the standard industrial classification (KSIC). Table 11 shows summary statistics of the wholesale and retail surveys and the service industry surveys used for the analysis.

TABLE 11—SUMMARY STATISTICS OF RETAIL DATA

Variables	No. of Obs.	Mean	Std. Dev	Min	Max
Franchisee					
Sales (million won)	21,075	336.18	380.00	11.00	2,576.00
Operating profit (million won)	21,075	35.67	37.50	-37.00	228.00
Regular holidays of businesses	21,075	1.34	1.50	0.00	5.00
No. of months of operation per year	21,075	11.32	2.11	1.00	12.00
Year 2016	21,075	0.52	0.50	0.00	1.00
Year 2014	21,075	0.48	0.50	0.00	1.00
Franchisor					
Sales (million won)	3,031	2,699.90	4,395.00	30.00	30,598.00
Operating profit (million won)	3,031	93.90	167.44	-195.00	1,077.00
Individual businesses	3,031	0.16	0.36	0.00	1.00
Regular holidays of business	3,031	1.00	1.49	0.00	5.00
No. of months of operation per year	3,031	11.10	2.42	1.00	12.00
Year 2016	3,031	0.99	0.08	0.00	1.00
Year 2014	3,031	0.01	0.08	0.00	1.00

Note: Regular business holidays indicate one day of the month, two represent two to three days of the month, three indicates four to five days of the month, four indicates six to seven days of the month, and five indicates eight or more days of the month.

Source: Microdata on franchise disclosure documents (2014, 2016), Wholesale and Retail Trade Survey MDIS (2014, 2016), Service Industry Survey MDIS (2014, 2016).

B. *Effects of Contract Conditions on the Sales of Franchisees*

Examining model (1) in Table 12, the sales of franchisees tend to increase as the ratio of directly managed stores and the period of the franchise business increase. Specifically, a 1%p increase in the ratio of directly managed stores and one more month of the franchise business are likely to raise sales by 1.47 and 0.78 million won, respectively. It is possible that reputation capital contributes to attracting consumers and thus increases the demand for individual franchisees. Also, rich business know-how passed down from franchisors may have raised franchisees' sales.

In addition, as the number of stores increases, franchisee sales tend to decrease. An increase in the number of stores can help increase sales by improving brand awareness, whereas fierce competition among stores may lead to a decrease in the sales of an individual franchisee. The results suggest that the latter effect is greater.

Regarding the effect of protecting the business area, the sales of franchisees are significantly higher when the established business area is the walking, vehicle, or population type. When the scope of the business area increases by 1km for the walking or vehicle distance, sales tend to increase by 114.4 million won and 57.9 million won, respectively; if the population within the business area increases by 1,000, sales are likely to rise by 1.2 million won.

In models (2) to (4), the franchise fee and various types of royalty variables are set as independent variables. Regarding these models, it should be noted that if reputation capital-related variables are set as independent variables at the same level

TABLE 12—ESTIMATION RESULTS OF FRANCHISEES' SALES

Dep. Var.: Sales (million won)	OLS Model (1)	IV Model (2)	IV Model (3)	IV Model (4)
Ratio of directly managed stores	1.472*** (0.349)			
No. of months of franchise business	0.783*** (0.044)			
No. of franchisee stores	-0.013*** (0.003)			
Franchise fee		11.665*** (2.996)	3.739 (3.059)	18.958*** (3.204)
Royalty: Fixed-sum		-21.617 (23.202)		
Royalty: Fixed-rate			4.401*** (0.554)	
Royalty: % of raw materials price				-34.978*** (4.318)
Territorial range: Radius	0.917 (2.512)	-3.346 (2.639)	-0.375 (2.649)	-5.633** (2.757)
Territorial range: Walking	114.434*** (14.464)	110.962*** (16.577)	108.169*** (15.383)	67.786*** (16.761)
Territorial range: Vehicle	57.947*** (2.373)	54.087*** (2.380)	54.082*** (2.364)	68.689*** (3.057)
Territorial range: Population	1.226*** (0.082)	0.947*** (0.082)	0.737*** (0.086)	0.947*** (0.085)
Territorial range: Administrative area	6.805 (6.846)	-26.014*** (8.025)	2.779 (8.769)	-4.913 (8.753)
Store area	0.341*** (0.121)	0.014 (0.127)	0.161 (0.128)	-0.457*** (0.144)
Regular holidays of business	-43.703*** (1.799)	-43.593*** (2.234)	-40.228*** (1.907)	-43.016*** (1.975)
No. of months of operation per year	31.039*** (1.163)	31.641*** (1.177)	31.545*** (1.169)	31.573*** (1.224)
Year 2014	-91.721*** (6.337)	-4.846 (24.715)	-63.210** (25.070)	56.991** (26.553)
Constant	35.300 (92.977)	-56.449 (102.449)	20.708 (102.277)	-187.927* (107.844)
Region dummy	O	O	O	O
Meddle-sector dummy	O	O	O	O
Observations	21,075	21,075	21,075	21,075
R^2	0.159	0.143	0.154	0.072

Note: Standard errors are in parentheses. *, **, and *** denote statistical significance at the 10%, 5%, and 1% level, respectively.

as franchise fees and royalties, endogeneity problems may arise in these contract condition variables. In addition, the effect of reputation capital is dispersed into direct effects through the ratio of directly managed stores and the period of the franchise business and indirect effects through franchise fees and royalties, causing the difficulties in interpretation. Therefore, in Models (2) to (4), the reputation capital variables are set as instrumental variables of franchise fees and royalties, respectively.

The estimates in Models (2) and (4) suggest a tendency for franchisees' sales to increase significantly when the franchise fee increases. A positive coefficient is also estimated in Model (3), but it is not significant.

On the other hand, royalty contracts show different results by type. The fixed-sum royalty in Model (2) does not have a significant effect. In such a contract, a fixed amount of money is collected every month regardless of the level of franchisees' sales, meaning that a franchisor's incentive for raising their stores' sales may be low. Conversely, the fixed-rate royalty contract of Model (3) shows a significant positive effect. When the royalty increases by 1%p, the franchisees' sales increase by approximately 4.4 million won on average, supporting the contention that franchisors can have a strong incentive to increase their stores' sales.

Model (4) considers another type of royalty, corresponding to a certain percentage of the raw materials price. Having negative coefficients, this type of royalty is commonly observed in wholesale and retail and in food services. In a situation where new franchisors who have not accumulated reputation capital cannot set high franchise fees and royalties, they may try to secure profits by including margins in the raw materials price. However, as brand awareness and the reputation are insignificant, it appears to be difficult to generate the effect of increasing demand for franchisees.

C. *Effects of Contract Conditions on Franchisees' Operating Profits*

The effect of increasing sales from franchise fees and fixed-rate royalty contracts appears to be good for franchisees. Nevertheless, setting them high may not be necessarily beneficial to their welfare because franchise fees and royalties are what they eventually have to pay to franchisors. If the effect of increasing costs is greater than the effect of increasing sales, the operating profit that franchisees can secure will actually decrease. In particular, considering the industry criticism that the costs paid by franchisees to the franchisor are excessive and cause management instability, verification of this aspect is very important.

Indeed, it is found that increases in franchise fees and all types of royalties have a negative effect on operating profits (Table 13). This implies that the cost-increasing effect is stronger than the sales-increasing effect of reputation capital. Even if the franchisor sets franchise fees and royalty levels that reflect the reputation capital accumulated in the brand, the sales of franchisees may not improve, as can be expected due to the fierce competition in the market. Notably, the fixed-rate royalty type had the smallest negative effect in terms of the extent of the decrease in the operating profit.

On the other hand, all types of business area protection tend to raise operating profit significantly. The business area guarantees exclusive commercial territories. Because it increases demand but does not increase costs, operating profit is likely to improve.

TABLE 13—ESTIMATION RESULTS OF FRANCHISEES' OPERATING PROFIT

Dep. Var.: Operating profit (million won)	IV Model (1)	IV Model (2)	IV Model (3)
Franchise fee	-0.658** (0.307)	-0.321 (0.314)	0.015 (0.325)
Royalty: Fixed-sum	-4.108* (2.380)		
Royalty: Fixed-rate		-0.266*** (0.057)	
Royalty: % of raw materials price			-3.471*** (0.438)
Territorial range: Radius	1.189*** (0.271)	0.985*** (0.272)	0.953*** (0.280)
Territorial range: Walking	9.201*** (1.700)	7.983*** (1.579)	4.414*** (1.699)
Territorial range: Vehicle	4.181*** (0.244)	4.199*** (0.243)	5.637*** (0.310)
Territorial range: Population	0.075*** (0.008)	0.091*** (0.009)	0.076*** (0.009)
Territorial range: Administrative area	4.514*** (0.823)	2.701*** (0.900)	6.581*** (0.887)
Store area	0.030** (0.013)	0.020 (0.013)	-0.017 (0.015)
Regular holidays of business	-1.951*** (0.229)	-1.879*** (0.196)	-1.794*** (0.200)
No. of months of operation per year	2.381*** (0.121)	2.377*** (0.120)	2.370*** (0.124)
Year 2014	-17.189*** (2.535)	-14.743*** (2.574)	-11.444*** (2.692)
Constant	28.463*** (10.508)	23.801** (10.502)	15.416 (10.933)
Region dummy	O	O	O
Meddle-sector dummy	O	O	O
Observations	21,075	21,075	21,075
R^2	0.075	0.084	0.021

Note: Standard errors are in parentheses. *, **, and *** denote statistical significance at the 10%, 5%, and 1% level, respectively.

D. *Effects of Contract Conditions on Franchisors' Sales and Profits*

The sales structure of franchisors is quite different from those of franchisees. The latter depends entirely on the purchases of end-consumers, but the former can be generated through various channels according to the formula below.

$$
\begin{aligned}
\textit{Franchisor's sales}_i = & \sum_i (\textit{Franchise fee}_{ij} + \textit{Education expenses}_{ij} + \textit{Other expenses}_{ij}) \\
& \cdot \textit{Number of franchise stores}_{ij} \\
& + \sum_j (\textit{Fixes sum royalty}_{ij}) \cdot \textit{Number of franchise stores}_{ij} \\
& + \sum_j \sum_s \textit{Fixed percentage royalty}_{ij} \cdot \textit{Sales of franchise}_{ijs} \\
& + \sum_j \sum_s (1 + \textit{Raw material royalty}_{ij}) \cdot \textit{Supply price of raw materials} \\
& + \sum_j \sum_t \textit{Sales of directly managed store}_{ijt}
\end{aligned}
$$

where i : Franchisor, j :Brand, s : Franchisee, t : Directly managed store.

According to model (1) in Table 14, a franchisor’s sales tend to rise as the ratio of directly managed stores and the period of the franchise business increase, similar to the trend found for franchisees. The reputation capital and business know-how accumulated in the brand act as positive factors not only for the sales of franchisees but also for the growth of the franchisors.

The number of stores also has a positive effect, as opposed to a negative effect on franchisees. As the number of stores increases, franchisors come to have more channels for securing franchising fees, education fees, and royalties, meaning that their sales can increase. This confirms that franchisors have economic incentives to recruit more franchisees.

Interestingly but naturally, models (2) to (4) clearly show that franchise fees and all types of royalties are factors that increase the sales of franchisors. If the franchise fee increases by 1 million won, the franchise sales tend to increase by 440 million won, and when the fixed-sum, fixed-rate, and raw materials royalty levels increase by one unit, the franchise sales tend to increase by 4.35, 0.09, and 0.59 billion won, respectively. As shown in the formula above, these contract terms correspond to the franchisor's revenue source, making it likely that positive effects will result.

Models (5) to (7) estimate the effects of contract terms on the operating profit and reveal a distinct positive effect. This result is contrary to the situation of franchisees, where the operating profit decreases as the franchise fees and royalties increase. From this, we can guess the cause of the ongoing conflicts between the two parties over the distribution of value added.

The effects of business area protection also reveal contradictory results. While it helps franchisees have more sales and gain more operating profits, it diminishes those of franchisors by limiting the opening of additional stores. This is how the economic interests collide, from which we may forecast that disputes over infringements of business areas are likely to continue in the future.

TABLE 14—ESTIMATION RESULTS OF FRANCHISOR'S SALES AND OPERATING PROFIT

Dep. Var.:	Sales (million won)				Operating profit (million won)		
	OLS Model (1)	IV Model (2)	IV Model (3)	IV Model (4)	IV Model (5)	IV Model (6)	IV Model (7)
Ratio of directly managed stores	68.101*** (4.670)						
No. of months of franchise business	2.846*** (1.062)						
No. of total franchise store	0.493*** (0.071)						
Franchise fee		442.556*** (24.126)	395.150*** (21.718)	351.247*** (23.643)	7.753*** (0.884)	7.230*** (0.844)	6.602*** (0.884)
Royalty: Fixed-sum		4,354.754*** (508.886)			47.673** (18.642)		
Royalty: Fixed-rate			90.399*** (10.645)			0.959** (0.414)	
Royalty: % of raw materials price				586.346*** (125.183)			9.264** (4.679)
Territorial range: Radius	-117.407 (92.717)	-619.907*** (106.288)	-508.312*** (99.597)	-534.894*** (104.195)	-4.520 (3.894)	-3.312 (3.870)	-3.516 (3.895)
Territorial range: Walking	305.435 (366.073)	-1,442.919*** (429.909)	-1,040.907*** (391.952)	-282.786 (396.058)	-55.027*** (15.748)	-50.315*** (15.228)	-43.059*** (14.805)
Territorial range: Vehicle	164.420** (71.287)	373.236*** (80.652)	337.336*** (75.552)	108.124 (91.977)	16.521*** (2.954)	16.125*** (2.935)	12.552*** (3.438)
Territorial range: Population	4.810* (2.672)	11.821*** (2.999)	7.025** (2.809)	9.966*** (2.937)	0.154 (0.110)	0.102 (0.109)	0.138 (0.110)
Territorial range: Administrative area	-2,301.321*** (196.767)	-2,976.654*** (220.187)	-2,350.277*** (214.176)	-3,058.390*** (220.683)	-36.001*** (8.066)	-29.312*** (8.321)	-37.940*** (8.249)
Individual businesses	-1,346.589*** (208.548)	-2,096.117*** (237.405)	-1,618.997*** (216.365)	-1,653.400*** (226.711)	-34.021*** (8.697)	-28.794*** (8.406)	-29.387*** (8.475)
Constant	O	O	O	O	O	O	O
Control variables, Meddle-sector dummy, Region_dummy	-3,010.994 (2,773.208)	657.574 (3,075.487)	587.409 (2,887.477)	1,641.021 (3,020.655)	110.961 (112.661)	110.526 (112.184)	122.108 (112.915)
Observations	3,031	3,031	3,031	3,031	3,031	3,031	3,031
R^2	0.285	0.103	0.209	0.133	0.170	0.178	0.166

Note: Standard errors are in parentheses. *, **, and *** denote statistical significance at the 10%, 5%, and 1% level, respectively.

VII. Policy Recommendations

Based on the empirical results, we can provide the following policy recommendations. First, the experience of running directly managed stores of franchisors should be clearly revealed in the Information Disclosure Documents. It is necessary to set items such as the 'start date of directly managed stores' and the 'period of the operation of directly managed stores before the franchise business' as mandatory provision items. Information on direct management experience can be used as an important criterion for franchisees to judge the degree of reputation capital of the brand and the appropriateness of franchise fees and royalties.

Next, it is necessary to pay more attention to the positive relationship between the business performance of franchisees and fixed-rate royalties, which are adopted by only a quarter of brands. Fixed-rate royalties significantly increase franchisee sales and have the smallest negative effect on operating profit compared to other types of royalties. This type also provides a strong incentive for franchisors to increase franchisees’ sales, which can enhance the win-win growth of the franchise market. In order to revitalize fixed-rate royalties, franchisors must be able to observe their franchisees’ sales transparently. As part of the small business supporting policy, the central government can consider supporting POS systems for franchisees of brands meeting certain criteria.

Also, it is necessary to strengthen the qualifications of franchising business so that the franchisor does not recruit franchisees with untested brands in the market. As in the overseas market, where the franchise industry has developed, the government needs to consider an institutional device that allows only franchisors with a certain degree of experience in operating a directly managed store to recruit franchisees.

Policy efforts are also needed to alleviate the reckless entry of franchisees into the market. In the domestic market, unprepared start-ups and overcrowding are aggravating factors, mainly in the food services and wholesale and retail businesses. This is highly likely to lower the profitability of individual franchisees by limiting the demand-increasing effect of reputation capital. Therefore, in some overcrowded industries, it is necessary to induce contemplated start-ups by restricting governmental or franchisee loans to a certain extent.

References

Fair Trade Commission Franchise. Franchise Disclosure Document (http://franchise.ftc.go.kr, last access: Nov. 11, 2018).

Gallini, Nancy and Nancy Lutz. 1992. "Dual Distribution and Royalty Fees in Franchising," *Journal of Law Economics and Organization*, Vol. 8, No. 3.

Klein, Benjamin and Lester Saft. 1985. "The Law and Economics of Franchise Tying Contracts," *The Journal of Law & Economics*, Vol. 28, No. 2.

Korea Fair Trade Mediation Agency. 2013. "2012 Dispute Mediation Results," press release, Mar. 2013 (in Korean).

Korea Fair Trade Mediation Agency. 2018. "2017 Dispute Mediation Results," press release, Jan. 2018 (in Korean).

Lafontaine, Francine. 1992. "Agency Theory and Franchising: Some Empirical Result," *The RAND Journal of Economics*, Vol. 23, No. 2.

Lafontaine, Francine and Kathryn Shaw. 1999. "The Dynamics of Franchise Contracting: Evidence from Panel Data," *Journal of Political Economy*, Vol. 107, No. 5.

Lee, Jinkook. 2017. "Performance evaluation for small business support policies," in Jaejoon Lee, (ed.) Imp*roving the Performance Management System of Budgetary Programs: Focusing on In-depth Evaluation*, Research Monograph, KDI (in Korean).

Lee, Jinkook. 2018. *Study on Franchise Contract and Franchise Market System*, Policy Study 2018-04, Korea Development Institute (in Korean).

Lee, Jinkook. 2019. "Conflicts in the Franchise Industry: How to Forge a Win-Win Partnership," KDI Focus No. 96, Korea Development Institute (in Korean).

Lee, Jinkook. 2020. "Determinants of the Share of Self-Employed and Excess Level: Focusing on Comparisons Between OECD Countries," in Jinkook Lee, (ed.) *Study on Self-Employment and Policy Suggestions*, Research Monograph, KDI (in Korean).

Mathewson, Frank and Ralph Winter. 1985. "The Economics of Franchise Contracts," *The Journal of Law and Economics*, Vol. 28, No. 3.

Park, Jooyoung. 2013. "A Study on Improving Territorial Encroachment of Franchisors," Legal Research Group Research Report, Korea Fair Trade Mediation Agency (in Korean).

Statistics Korea. 2014; 2016. "Service Industry Survey MDIS" (in Korean).

Statistics Korea. 2014; 2016. "Wholesale and Retail Trade Survey MDIS" (in Korean).

Statistics Korea. 2015. "Wholesale and Retail Trade Survey MDIS" (in Korean).

Statistics Korea. "Korea Standard Industrial Classification, Rev 9" (in Korean).

LITERATURE IN KOREAN

공정거래조정원. 2013. 「공정거래조정원 2012년도 분쟁조정 실적」, 보도자료, 2013. 3. 11.

공정거래조정원. 2018. 「공정거래조정원 2017년도 분쟁조정 실적」, 보도자료, 2018. 1. 23.

박주영. 2013. 『가맹본부의 배타적 영업지역 보호에 대한 개선방안 연구』, 법경제분석그룹(LEG) 연구보고서, 한국공정거래조정원.

이진국. 2017. 「소상공인 지원정책의 성과평가와 성과지표 개선방안」, 이재준 편, 『재정사업 성과관리체계 개선방향: 심층평가를 중심으로』, 제1장, KDI 연구보고서, 한국개발연구원.

이진국. 2018. 『가맹계약과 가맹사업 시장제도 연구』, 정책연구시리즈 2018-04, 한국개발연구원.

이진국. 2019. 「가맹업계의 갈등, 상생협력을 위한 방안은?」, KDI Focus No.96, 한국개발연구원.

이진국. 2020. 「자영업자 비중의 결정요인과 과잉 수준 분석: OECD 국가 간 비교를 중심으로」, 이진국 편, 『자영업에 대한 종합적 분석과 정책제언』, 제1장 KDI 연구보고서, 한국개발연구원.

통계청. 2014; 2016. 「도소매업조사 마이크로데이터」.

통계청. 2014; 2016. 「서비스업조사 마이크로데이터」.

통계청. 2015. 「경제총조사 마이크로데이터」.

통계청. 「제9차 표준산업분류 개정」.

KDI Journal of Economic Policy 2022, 44(1):77–98
http://dx.doi.org/10.23895/kdijep.2022.44.1.77

Investment and Business Cycles: Focusing on Firms' Capital Adjustment Costs†

By CHANGWOO NAM*

This paper empirically verifies that the types of capital adjustment costs serve as an important mechanism in relation to investment decision-making after confirming that the investment dispersion of Korean firms is pro-cyclical and can affect business cycles. Specifically, it is found through empirical methods using corporate financial data that capital adjustment costs generally assumed to take a quadratic form in macroeconomics are asymmetric and irreversible in the Korean economy. In particular, capital adjustment costs are empirically proven to cause investment dispersion to expand given that the substitution effect of the marginal value to the marginal cost for one unit of investment in the inter-temporal investment decision is affected by that cost with regard to the resale of owned equipment assets, as opposed to new investments in equipment assets. We ultimately show, albeit indirectly, that investment dispersion can affect business cycles as capital adjustment costs influences investment decisions. What is implied is that the capital adjustment cost is not merely an exogenously deep parameter that fits the dynamics of business cycles in a macroeconomic model but could instead be a policy variable that can be endogenized through government policies.

Key Word: Investment, Business Cycles, Capital Adjustment Costs, GMM
JEL Code: C58, D24, D25

I. Introduction

After the government experienced an industrial crisis due to insolvencies in the shipbuilding and shipping industry in the 2010s, it became necessary to consider various types of support for marginal firms so that firms could preemptively

* Senior Fellow, Korea Development Institute (E-mail: cnam@kdi.re.kr)
* Received: 2021. 10. 19
* Referee Process Started: 2021. 10. 25
* Referee Reports Completed: 2021. 12. 13

† This paper is written based on Nam, 2020, *Investment and Business Cycles: focusing on Investment Dispersion*, KDI Policy Study 2020-07 (in Korean).

restructure or reorganize their businesses to prevent insolvency. In general, alleviating the adjustment costs or frictional costs incurred when firms purchase or resell equipment assets as a business reorganization strategy can be considered as a policy measure. Accordingly, it is necessary to examine whether capital adjustment costs (CAC, hereafter), assumed to arise when a firm makes an investment, for instance, can be considered as a policy measure and how the CAC can affect a firm's investment decisions and business cycles. Therefore, here we attempt to verify whether a firm’s investment decisions can cause business cycles. This is also a topic that has been actively discussed recently. In particular, the parameters of CAC are estimated by separating the adjustment cost incurred when purchasing equipment assets and the adjustment cost when reselling them. In addition, this paper empirically demonstrates that the lower the estimated CAC is, the greater the business cycles become. For example, if the adjustment costs incurred when purchasing or reselling equipment assets decrease, there will be incentives for firms to expand these purchases or resales further. This means that the investment can expand even more during economic upturns and that the redistribution of equipment assets can be more efficient with lower frictional costs during economic downturns. Overall, this means that when exogenous shocks are identical, the reduction in the adjustment cost can affect the amplitude of the business cycles (Hamermesh and Pfann, 1996).

Although in the literature on macroeconomics this is usually estimated by defining the CAC of the investment as a quadratic function that is downwardly convex, our paper assumes that the convexity is different when buying and reselling the equipment assets. This is based on the assumption that if a firm sells its owned equipment assets for various reasons, such as becoming insolvent or when undergoing a transfer to another area, that cost may differ from the adjustment cost of new equipment investment. This asymmetry of CAC may also include differences in various institutional aspects between factor markets. We verify the asymmetry of CAC through an empirical analysis using corporate financial data rather than a macroeconomic model based on the equilibrium model. Moreover, we intend to derive policy implications for equipment investments in the future by analyzing changes in the adjustment cost function of equipment investments given that the existence of asymmetry in CACs can affect business cycles. In particular, our paper examines the possibility of investment irreversibility, referring to how the resale of equipment assets is not carried out smoothly due to the higher adjustment cost incurred when reselling the equipment. In other words, investment irreversibility makes it difficult to resell equipment assets when the adjustment cost incurred when purchasing equipment assets is identical during an economic upturn, thereby limiting active investments and thus limiting the accumulation of capital.

This paper is organized as follows. Section II explains the literature on the corporate investments, the CAC, and the relationship between corporate investments and business cycles, and then explains the research purpose. Section III describes the structural model and empirical methodology used to estimate the CAC. Section IV presents the empirical results of the estimates of the CAC and shows the correlations between the CACs, investment dispersion, and business cycles through various methods. Section V presents the conclusion.

II. Relationship between Corporate Investment and Business Cycles

A. *Firm's Capital Adjustment Cost*

Essentially, the CAC refers to the comprehensive cost borne by firm in addition to prices formed in markets during the process of purchasing and reselling (or hiring/dismissing) production factors (capital such as equipment and labor, etc.). However, these adjustment costs do not appear in ordinary financial data. In particular, these costs can reflect the economic environment the firm faces in factor markets, the characteristics of the technology used by the firm, and the direct or indirect effects of government policies (Hamermesh and Pfann, 1996).

> *"Moreover, understanding of the nature of adjustment costs is vital for the evaluation of policies, such as tax credits, that attempt to influence investment and thus aggregate activity. Despite the obvious importance of investment to macroeconomics, it remains an enigma. Costs of adjusting the stock of capital reflect a variety of interrelated factors that are difficult to measure directly or precisely so that the study of capital adjustment costs has been largely indirect through studying the dynamics of investment itself."* (in Motivation by Cooper and Haltiwanger, 2005)

As Cooper and Haltiwanger (2005) point out, understanding the CAC as it pertains to investments is very important when evaluating corporate-related policies that affect investments, such as tax deductions and subsidies. However, despite the importance of investments, the CAC function for investments in macroeconomics remains a difficult problem. Moreover, because the CAC function of investments reflects various economic characteristics but is difficult to measure directly or indirectly, research on the CAC function is only conducted indirectly by studying corporate investment dynamics.

Early research on corporate investments mainly assumes the CAC to be a symmetric convex function. For example, Holt (1960) shows that the quadratic functional form of the CAC in the manufacturing industry feasibly explains firms' hiring or layoff costs, as well as the cost of installing machinery. Cooper and Haltiwanger (2005) find that these factors as well as other external factors are reflected in the CAC. The CAC in the form of such a symmetrical quadratic function can be differentiated at all investment rates (generally net investment size/total assets), and the marginal CAC increases as the investment rate increases positively or negatively based on zero. Subsequently, in a macro model that does not take into account the heterogeneity of firms, a symmetrical convex function is mainly assumed when explaining the investment dynamics of the corporate sector. In this case, the investment level, investment dispersion, and business cycles are mainly determined by the parameters of the symmetric CAC function.

Later, in studies of firms' heterogeneity using plant-level micro-data and corporate financial statements, the asymmetry of the CACs began to be considered in earnest. In fact, there is no reason for the marginal CAC of purchasing one unit of investment to be identical to the marginal CAC of reselling one unit. For example, if there is no secondary market for capital goods market, firms will hesitate to purchase new equipment due to uncertainty about future shocks, meaning that additional costs for capital adjustment will be incurred. The frictional cost of reselling these equipment assets is defined as irreversibility (Hamermesh and Pfann, 1996). In other words, if the CAC for reselling the equipment assets is higher than that for purchasing them due to various environmental and policy factors in the aforementioned production factors markets, it becomes difficult to sell equipment assets such as the machinery, and this irreversibility hinders firms' active investments. Typical studies assuming an asymmetric convex function as a CAC function of investment include those by Abel and Eberly (1994) and Zhang (2005). Abel and Eberly (1994) consider the gap and irreversibility between the purchase and sale prices of capital goods as an asymmetric convex CAC, showing that the size of the investment is a non-decreasing function. In addition, Zhang (2005) shows that the irreversibility of corporate investments can generate a value premium in the stock market. In particular, he analyzes and explains the dynamics of firms in which the risk premium caused by the irreversibility of installed capital expands firms' financial friction in the event of an economic slowdown.

On the other hand, in the manufacturing industry in the United States, several studies focus on the occurrence of lumpy investments rather than continuous investments (or the investment rate) in plant-level micro-data. In order to explain the characteristics of a period of inaction in which corporate investments do not occur, especially at the plant level, these studies report that investments do not occur continuously at all times. Instead, determination is made as to whether to invest at each time. Firms' investment behaviors are analyzed through a discrete choice model in which the CAC function is non-convex and where irreversibility does not incur a negative investment, although there can be periods of non-investment. In relation to this, Cooper *et al.* (1999) argue that with regard to manufacturers, the investment is not made until equipment assets (machinery) are aged (depreciated) to a certain level, with productivity then falling below a certain level. In this model, the CAC function shows non-convexity because only fixed costs exist for a certain size of investment. In particular, this type of replacement of equipment assets often occurs when there is a positive impact on productivity, which feasibly explains how corporate investments are pro-cyclical. In the end, Cooper *et al.* (1999) show that this discrete investment dynamics can suitably explain the interactions among the replacement cycle of equipment assets, the cross-sectional distribution of equipment assets, and the business cycle.

B. *Corporate Investments and Business Cycles*

> *"Most economists agree that the primary source of cyclical instability is to be found in the determinants of investment behavior."* (at the beginning of Gordon, 1955)

Since the beginning of modern economics, corporate investments have been among the main factors inducing business cycles because the total factor productivity shock, a major factor inducing business cycles, occurs in the production sector. If only an exogenous productivity shock occurs, ceteris paribus, a negative shock will induce a decrease in planned investments and a positive shock will induce a further expansion of planned investments. Moreover, because corporate investments are a mechanism by which to generate future cash flows, investment fluctuations can eventually widen the amplitude of business cycles. Conversely, if there is no reduction or expansion relative to a planned investment despite an exogenous productivity shock, the total production level will be determined according to the exogenous shock and the firm's planned production capacity.

Assuming that this is the case, how is the CAC function of investment related to the business cycle? As explained earlier, CACs basically refer to not only the additional costs incurred when a firm installs and employs production factors (e.g., capital, labor) but also the market environment (including the secondary market) of production factors, as well as the direct or indirect effects of government policies. Naturally, this CAC function affects a firm’s investment decisions and plays a role in expanding or reducing the amplitude of corporate investments (Hamermesh and Pfann, 1996). For example, if the CAC increases suddenly for any reason, when exogenously positive shocks occur and thus firms plan to expand their investments, they will not be able to invest as much as originally planned and the amplitude of the economic expansion will be reduced as compared to when there is no CAC. Conversely, if the CAC is lower than before due to policies that support firms, the amplitude of economic expansion may be greater. However, considering the opportunity cost for the financial resources of government policies that support firms, it cannot be inferred that the expansion of the amplitude of the business cycle necessarily maximizes social welfare.

Recent studies have shown that rational investment decision models with various CAC functions of the types described above explain various economic phenomena such as business cycles using sophisticated models with corporate financial statements or plant-level micro-data. In particular, Bloom (2009) and Bachmann and Bayer (2014) accurately show how corporate investment behaviors interact with uncertainty and affect business cycles. First, Bloom (2009) shows that a macro-uncertainty shock expands the fluctuations of gross output and employment given the temporary suspension of corporate investments and employment due to uncertainty shocks. In particular, Bloom (2009)’s model uses a mixture of convex and non-convex functions as a CAC function to assume transaction costs in the capital goods market and partial irreversibility due to resale losses in the secondary market. The parameters of the CAC function determine the period of inaction of the investment and the period in which the investment proceeds continuously.

In addition, Bachmann and Bayer (2014, expressed as BB) focus on investment dispersion of firms, showing that when the dispersion of productivity shocks for individual firms is pro-cyclical, the investment dispersion by firms is pro-cyclical. In particular, BB’s corporate model uses a discrete choice model for investment, and the CAC function is assumed to be a fixed cost. The important point here is that the investment dispersion and pro-cyclicality are strengthened as the average level of the CAC increases regardless of the distribution of individual firms’ productivity shocks and the presence of counter-cyclicality (Table 9 of BB, 2014). If the CAC rises, the investment inaction period is extended but the absolute size of the investment becomes comparatively large, meaning that the dispersion of the investment rate and the pro-cyclicality are strengthened. In this way, changes in CAC the function can affect the amplitude of business cycles.

We assume that CACs are market costs resulting from direct and indirect effects of government policies and estimate them using accounting data from financial statements. In addition, our study assesses the possibility that the CAC can significantly affect business cycles, finding correlations among changes in the CAC function, investment dispersion, and business cycles. For example, ceteris paribus, if the CAC for reselling equipment assets increases during an economic boom, the firm will make an investment smaller than the planned amount in consideration of the CAC that may be higher in the future. At the same time, firms that have suffered a negative productivity shock will reduce the scale of their restructuring of equipment assets. This means that investment dispersion is reduced compared to when the CAC for reselling assets is low, and as the efficient allocation of equipment assets is restricted, the amplitude of economic growth could be limited. Even during an economic downturn, an increase in the CAC for reselling assets causes both planned purchases and resales of equipment assets to decrease, in turn causing the investment dispersion to decrease. However, the expansion of frictional costs (CACs) for the allocation of equipment assets can cause inefficiency in the economy as a whole, which can widen the extent of an economic downturn. Changes in the CAC in relation to purchasing equipment are much clearer than those in the CAC with regard to reselling equipment assets. If the CAC for purchasing equipment fall, the economic recovery will accelerate and the extent of the economic downturn will decrease.

In fact, this interpretation is the logic of the general equilibrium model. We analyze whether the CAC function when estimated using corporate data is significantly correlated with realized investment variance and business cycles. When analyzing the relationship between the CAC function and business cycles, accurately identifying the actual exogenous shock and decomposing the effect of the CAC shock on the economy become necessary. However, because this empirical analysis model has limitations related to this type of identification, we leave this issue as a future research topic.

III. Corporate Investment Model and Empirical Methodology

In macroeconomics, the CAC for a firm's investment is an index referring to the efficiency or fractional cost of the allocation of equipment assets not shown in the firm's financial statements. Investment irreversibility is a measure for considering the efficiency of restructuring tactics, such as equipment resales, when a firm faces bankruptcy due to a negative productivity shock or corporate insolvency. In other words, the CAC can be interpreted as representing additional costs due to changes in the prices of the equipment to be sold, the supply and demand environment in the equipment market, and the restructuring of employment following the resale of equipment. This means that the parameter of the CAC function is not necessarily positive because tax benefits for equipment investments, tax benefits for the resale of used equipment, or various forms of support for restructuring can be reflected in the CAC. In particular, subsidies and tax benefits for corporate relocation after the 2000s show that there may be benefits in the form of management costs that are not observed in financial statements. In such a case, additional adjustment costs in addition to the purchase and sale prices for equipment assets may in fact be negative (Whited, 1998). Zhang (2015) also shows that such asymmetric irreversibility of investment plays an important role in determining the value premium. The basic model of this paper is based on Zhang (2015), and the empirical method is based on Whited (1992; 1998).

A. *Corporate Investment and Business Cycles in Korea*

This part analyzes the relationship between corporate investment behavior and business cycles through corporate data before establishing a corporate investment model. As explained earlier, to determine the shape of the CAC function, we initially check the distribution of corporate investments for every decade from 1990 to 2019. Also, in this paper, it is important to determine whether we should assume corporate investment as a continuous type of behavior or as a discrete choice in which investment inaction can also exist (Cooper *et al*., 1999). Given that the scope of our paper encompasses not only the manufacturing industry but also all industries, it is necessary to review in advance the possibility of the discrete choice model to analyze manufacturing businesses.

Table 1 shows the distribution of the equipment investment rates in all industries and in the manufacturing industry. First, the proportion of investment rates whose absolute value is 1% or less, indicating inactivity of investments, is 3% or less in all industries and in the manufacturing industry in all periods. These rates are much lower than the 8.1% in Cooper *et al*. (1999). In addition, the proportion of negative investment rates in all industries and in manufacturing exceeds 20% in all periods. This is twice as high as the rate of 10.4% in Cooper *et al*. (1999). Additionally, the proportions of the absolute values of positive and negative investment rates exceeding 20% are around 45% and 10%, respectively, much higher than the rates of 20% and 2% in Cooper *et al*. (1999). Thus, it is appropriate to use a continuous investment decision model rather than a discrete choice model. In particular, because negative investments occur frequently in the data used in our paper, we decided to use an asymmetric convex function rather than a discrete choice model, in which negative investments rarely occur.

TABLE 1—DISTRIBUTION OF REAL INVESTMENT RATES

(Unit: %)

	Investment Inaction ($\lvert i/k \rvert < 1\%$)	$i/k < -20\%$	$i/k > 20\%$	$\lvert i/k \rvert > 20\%$	$i/k < 0\%$
			All Industries		
1990s	2.5	13.2	45.2	58.4	29.3
2000s	2.2	12.5	50.9	63.4	25.5
2010s	3.0	9.9	47.4	57.3	24.4
			Manufacturing		
1990s	2.6	11.1	46.6	57.7	26.3
2000s	2.1	10.8	52.4	63.2	22.6
2010s	2.7	8.3	48.7	56.9	21.9

Note: Investment rates for each year are calculated by the author using the BB model.

Source: KISData (KDI DB, 20 Jul 2020).

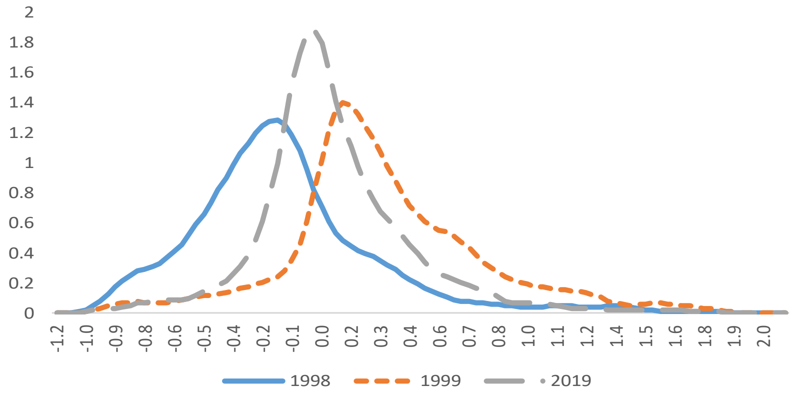


FIGURE 1. DISTRIBUTION OF REAL INVESTMENT RATES

Note: Investment rates for each year are calculated by the author using the BB model.

Source: KISData (KDI DB, 20 Jul 2020).

TABLE 2—CORRELATION BETWEEN REAL GDP AND VARIABLES RELATED TO REAL INVESTMENT RATES (IRS)

	S.D. (IRs)	Props of $\lvert i/k \rvert > 20\%$	S.D. ($\lvert i/k \rvert < 20\%$)	Mean (IRs)
From 1990	0.58***	0.70***	-0.12	0.10
From 2000	0.42*	0.67***	0.19	0.24
From 2010	-0.27	0.01	0.49	0.67**

Note: 1) ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively, 2) Data are detrended by HP-filter.

Source: Quantiwise7; KISData.

In addition, the correlation between investment behavior and business cycles was examined using BB's analysis method. BB shows that investment dispersion is pro-cyclical, and Korea has a similar tendency. Figure 1 shows the distributions of investment rates in 1998, 1999, and 2019. During 1998, a severe economic slump occurred due to the Asian Financial Crisis, and 1999 was a period of recovery from this crisis. As shown in the figure, the distribution is thicker at the bottom during a sluggish economy and thicker at the top during an upturn. This occurs because the number of firms making large-scale investments (investment rates greater than 20%) increases when the economy recovers, and large-scale corporate restructuring (investment rates less than -20%) occurs when the economy is sluggish (Bachmann and Bayer, 2014).

Table 2 shows the correlation between investment-rate-related variables and business cycles (GDP growth rate). It is found that the standard deviation of investment rates and GDP has a significantly positive correlation at the 1% level. Also, the correlation between the proportion of firms whose absolute value of investment rate exceeds 20% and the GDP, which affects investment dispersion, is also shown to be significantly positive.

However, recently, the correlation can be seen to weaken. This has occurred because the period of data is short (especially from 2010) and does not include events such as the Asian Financial Crisis or the Global Financial Crisis. At the same time, it is found that the positive correlation between the average investment rate and GDP has increased.

In fact, as explained above, an important investment mechanism is at work here (Bachmann and Bayer, 2014). If macroeconomic uncertainty decreases as the economy recovers, the uncertainty of firms' productivity shocks also decreases (the pro-cyclicality of productivity shocks). Then, in anticipation of the expansion of future cash flows, firms expand their investment more than originally planned (the pro-cyclicality of investment dispersion). Thus, the additional CAC incurred by the investment will be one of the important factors determining the size of the investment. In addition, during persistent macroeconomic shocks, the CAC for positive investment is important during an economic upturn, and the CAC for negative investment is an important factor in determining the size of the investment when the economy is sluggish. However, if the CACs for production factors are large enough, the investment must be lower than the optimal investment without CACs, which will be identical even in the case of a negative shock. According to this inference, it may be reasonable to assume asymmetry of the CAC function for the investment. We will not presuppose investment irreversibility in the model. Investment irreversibility defines not only the asymmetry of the CAC but also indicates that the CAC when reselling equipment is greater than the CAC when purchasing equipment. If it is shown that the asymmetry of the CAC reflecting the direct and indirect effects of production factor markets and government policies, as explained in the previous part, has an effect on investment dispersion, the CAC will then be considered as a policy variable.

B. *Corporate Investment Decision Model*

In this paper, the production function of a firm is composed of individual productivity shocks (A), equipment assets (capital, k), and labor (n) according to a general model. Recent macroeconomic models set production functions by classifying intangible capital to reflect the impacts of technological innovations as well as equipment capital. However, we do not consider intangible assets separately because we focus on decision-making with regard to equipment investments, not on intangible capital. Moreover, in our paper, for simplicity of the model, it is assumed that there are no macroeconomic shocks, with only productivity shocks of individual firms following AR (1). Accordingly, a firm's production function is expressed as

$$y_t = A_t k_t^{\alpha} n_t^{\gamma},\ \alpha + \gamma = 1, \tag{1}$$

in which it is assumed that the production function follows the CRS principle (constant return to scale). Through this production function, the operating profit function is defined as follows:

$$\pi(k_t, A_t) = y_t p_t - w_t n_t, \tag{2}$$

where p_t denotes the product price and w_t denotes the wage. Maximization of the firm value is expressed as follows when the corporate tax rate (τ), equipment price ($p_{i,t}$), and future discount rate (β_t) are given.

(3) $$V_t = \max_{i_t} (1-\tau)[\pi(k_t, A_t) - \Phi(i_t, k_t)] - p_{i,t} i_t + \beta_t E_t[V_{t+1}],$$

in which i_t denotes the investment executed by the firm in period t, V_t denotes the firm value, and $E_t[\bullet]$ denotes the expectation operator. Also, the CAC function (Φ) is basically defined as a convex function with the form of a quadratic function; it is differentiable in the domain except at 0 given its asymmetry (Whited, 1998; Zhang, 2005).

(4) $$\Phi(i_t, k_t) = \left[\frac{(\phi^+ 1_{i\geq 0} + \phi^- 1_{i<0})}{2} \left(\frac{i_t}{k_t} \right)^2 + \alpha \right] k_t.$$

Here, ϕ^+ and ϕ^- are defined as parameters of the CAC function for positive (purchasing) and negative (reselling) investments, respectively. In addition, here it is not assumed that the two parameters are identical according to the assumption of asymmetry.

In the firm value maximization problem (3), the Euler equation for i_t is established as follows:

(5) $$\Phi(i_t, k_t) + \frac{p_{i,t}}{1-\tau} = \beta_t E_t \left[\begin{array}{l} \pi_k(k_{t+1}, A_{t+1}) - \Phi_k(i_{t+1}, k_{t+1}) \\ +(1-\delta)\left(\Phi_i(i_{t+1}, k_{t+1}) + \dfrac{p_{i,t+1}}{1-\tau} \right) \end{array} \right],$$

Here, the FOC of the CAC function with respect to i_t is

(6) $$\Phi_k = \alpha - \frac{(\phi^+ 1_{i\geq 0} + \phi^- 1_{i<0})}{2} \left(\frac{i_t}{k_t} \right)^2,$$
$$\Phi_i = (\phi^+ 1_{i\geq 0} + \phi^- 1_{i<0}) \frac{i_t}{k_t}.$$

Equation (5) shows the effect of the CAC function on the Euler equation. First, the left side of Equation (5) shows the marginal cost for one unit of investment. The CAC parameters ϕ^+ and ϕ^- as well as the investment rate affect the marginal cost of one unit of positive or negative investment. The right side of Equation (5) shows the future marginal return for one unit of investment, and the CAC parameters affect it in two directions; one direction is the effect of lowering future adjustment costs by investing in the present, and the other is the effect of raising future marginal costs that arise when the investment is postponed into the future. Finally, it is assumed that equipment investment and equipment assets have the following relationship:

$$i_t = k_{t+1} - (1-\delta)k_t = \Delta k_{t+1} + \delta k_t . \tag{7}$$

To ensure a better understanding of the concept of the CAC function based on Equations (4) and (6), various shapes of the CAC function can be confirmed by explaining Figure 2, where (A) shows that the CAC function is both convex and symmetric. As described in the literature, this is a form that is generally assumed in a typical macroeconomic model. (B) is when the parameter of the CAC function for a negative investment is larger than the CAC for a positive investment, indicative of the assumption of investment irreversibility. Contrary to (B), (C) is a case in which the CAC function for a negative investment is lower than the CAC function for a positive investment.

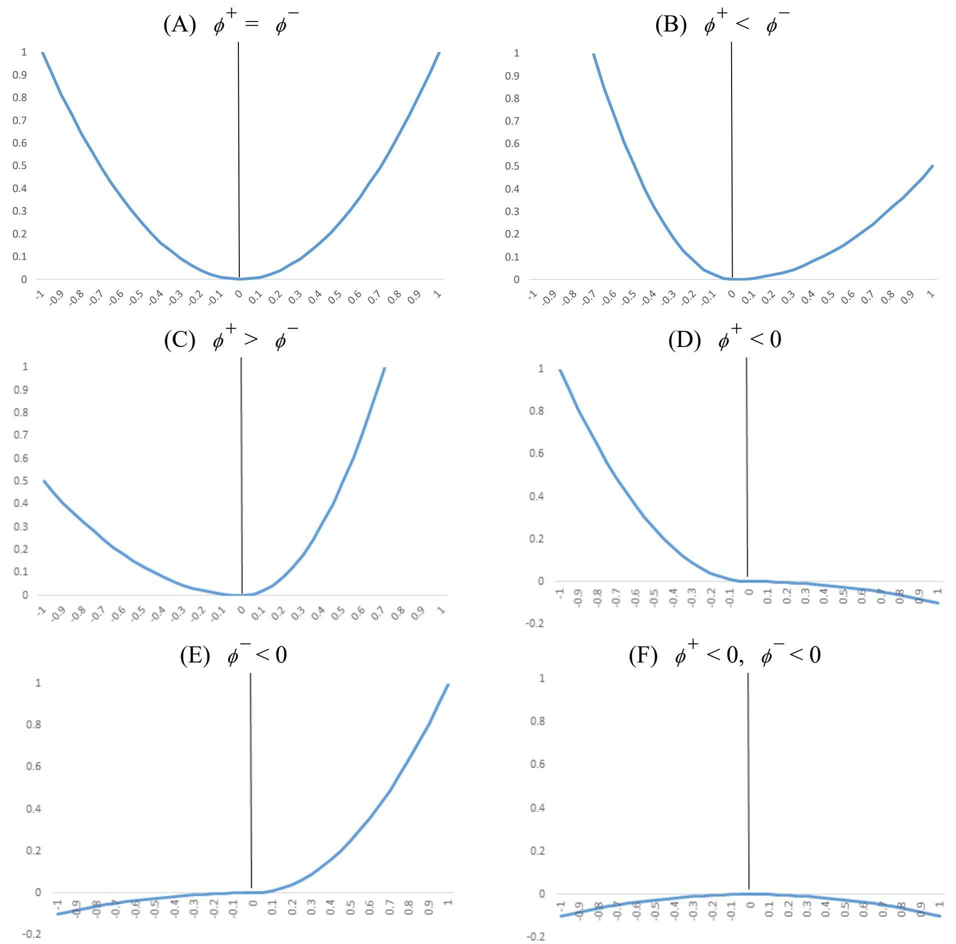


FIGURE 2. CASES OF THE CAPITAL ADJUSTMENT COST FUNCTION

As described above, this may occur when support for corporate restructuring processes such as personnel restructuring, support for the resale of used equipment assets, and deregulation are reflected. (D) refers to cases where the CAC function for a positive investment is negative. In other words, it represents a case where the total adjustment cost is lower than the theoretical equipment price in the market due to deregulation and support for equipment investment. In (E), as in (D), the CAC function for the resale of equipment is negative. Here, therefore, it is considered that used equipment assets are allocated at a price lower than the market's theoretical price due to deregulation and efforts to support restructuring, such as the reorganization of the business. Finally, in (F), the CACs of purchasing or reselling equipment assets are all negative, meaning that they are lower than the theoretical prices in the market overall. This situation can be seen as a case where regulations on corporate investments have improved overall.

C. *Empirical Data*

We use the accounting data of externally audited firms (49,644 firms) for the period of 1975 to 2019 obtained from KISData and use the methods of Whited (1998), Bachmann and Bayer (2014) to convert the basic data into real variables. First, regarding the equipment asset (k_t) used to calculate the equipment investment rate, the sum of machinery (excluding facilities) and transportation equipment, excluding structures, is used. For the equipment investment (i_t), we use the sum of changes in equipment assets and accumulated depreciation according to Equation (7). Each variable is converted into a real variable using a GDP deflator, an investment deflator, and the PPI by industry (data from the BOK). In particular, the price of equipment investment by industry is converted into a real variable as follows using an equipment investment deflator and the PPI by industry (Whited, 1992):

$$p_t = \frac{P_{k,t}}{PPI_{i,t}}$$

In addition, the real investment price reflecting the tax rate for each firm is calculated as follows:

$$\frac{p_{i,t}}{1-\tau} = \frac{1 - tax\ reduction - tax\ credit}{1-\tau} p_t \simeq (1+\tau_{i,t}) p_t,$$

in which $\tau_{i,t}$ is the average effective tax rate, calculated as (Corporate Tax Expenses / Continuing Business Profit before corporate tax deduction), considering that it is impossible to use tax data held by the National Tax Service.

Also, except for the CAC function in Equation (5), a variable that cannot be inferred from the data is π_k , the marginal operating profit function. Whited (1998) estimates the first derivative for investment by defining it as follows:

$$\pi_k(k_t, A_t) = \frac{y_t - \mu m_t}{k_t}, \tag{8}$$

in which μ denotes the mark-up for the variable cost, m_t. In other words, under the CRS assumption, the marginal operating profit is considered as the ratio of equipment assets to the margin of total sales (y_t). Such an estimation formula makes it easy to interpret the estimated value in the revenue function and has the advantage of estimating the markup variable at the same time. In particular, Whited (1992) estimates the parameters of CRS in the estimation equation at the same time. In this paper, μ can be interpreted as a parameter that reflects both the markup and CRS parameters (Whited, 1998). Also, y_t is the real variable of total sales, k_t is the real variable of equipment assets, and m_t is the real variable of the raw material cost.

Table 3 presents the basic statistics of 73,763 observations excluding outliers. Figure 3 shows the trends of the GDP growth rate and investment rate. As shown in Section II, the mean, standard deviation of investment rates, and the proportion of firms whose absolute values of investment rates exceed 20% appear to be pro-cyclical.

TABLE 3—STATISTICS OF BASIC DATA

	Average	Median	S.D.
β	0.97	0.99	0.04
Investment Rate (%)	25.01	18.62	41.87
δ	0.11	0.12	0.02
$(1+\tau)p$	1.39	1.29	0.38
y / k	24.85	16.86	22.43
m / y	0.46	0.46	0.22
Debt / Assets	0.58	0.61	0.24
Operating profit / Sales	0.08	0.07	0.07

Source: Quantiwise7; KISData.

(A) GDP, mean (IRs)

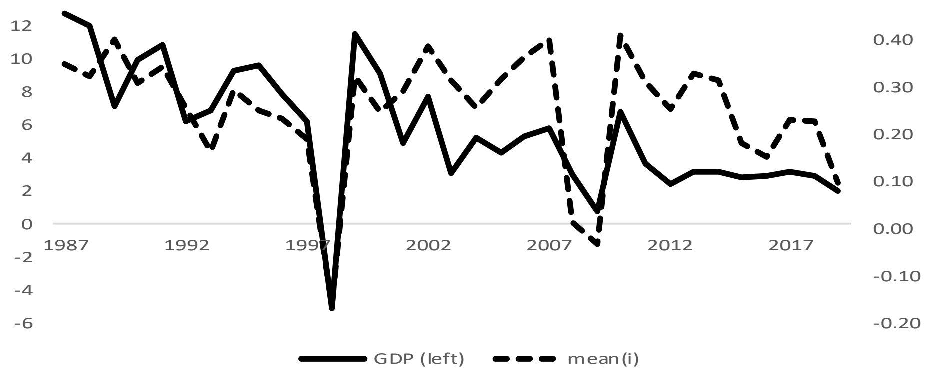
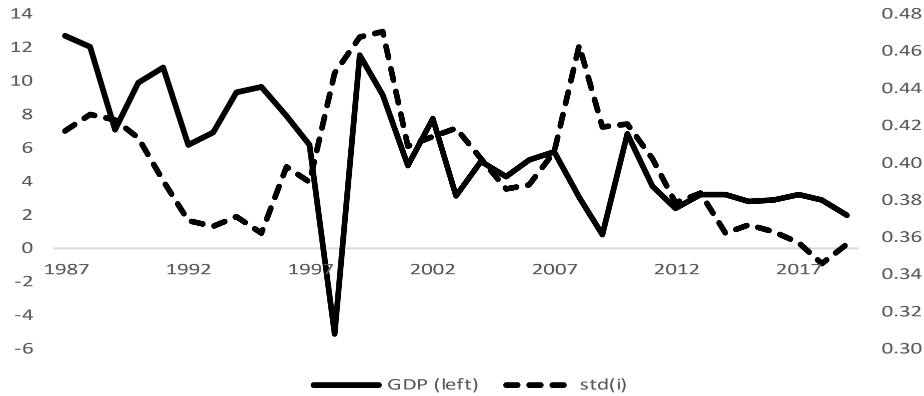


FIGURE 3. TRENDS OF GDP AND VARIABLES RELATED TO INVESTMENT RATES (IRS)

(B) GDP, s.d. (IRs)



(C) GDP, proportion of $|i/k| > 20\%$

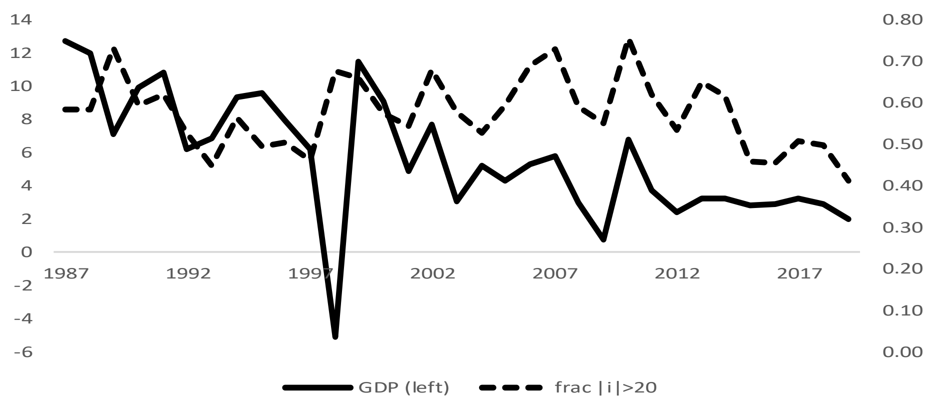


FIGURE 3. TRENDS OF GDP AND VARIABLES RELATED TO INVESTMENT RATES (IRS) (CONT'D)

Source: Quantiwise7; KISData.

D. *Empirical Methodology: GMM*

The empirical methodology used in this paper is defined as a dynamic panel GMM because the firm and year data used hear are based on the generalized method of moments (henceforth GMM) method. The specific estimation equation used as the dynamic panel GMM is derived from Equation (5) as follows:

$$\Phi_i(i_t,k_t)+(1+\tau_t)p_t = \beta_t E_t\left[\begin{array}{l}\dfrac{y_{t+1}-\mu m_{t+1}}{k_{t+1}}-\Phi_k(i_{t+1},k_{t+1}) \\ +(1-\delta_{t+1})(\Phi_i(i_{t+1},k_{t+1})+(1+\tau_t+1)p_{t+1})\end{array}\right]+\varepsilon_{t+1}, \tag{9}$$

Here, β_t is defined as the real discount rate obtained by subtracting the CPI growth rate from the one-year KTB interest rate, and δ_t is the economic depreciation rate, calculated using the formula below with the BOK's real capital stock (K_t) and the real fixed capital formation (I_t):

$$\delta_t = \frac{\Delta K_t + I_t}{K_{t-1}}.$$

The parameter vector to be estimated in Equation (9) is $\theta = (\mu, \phi^+, \phi^-, \alpha)'$, correspondingly referring to the operating profit function, the markup versus variable cost, the marginal CAC for a positive equipment investment, the marginal CAC for a negative equipment investment, and the constant term of the CAC. In addition, because GMM mostly uses corporate financial statements, it is difficult completely to rule out endogeneity in the data. Therefore, we define the following instrument variable group (z_t) with reference to Whited (1992; 1998):

$$\left(\frac{i_t}{k_t}, \frac{y_t}{k_t}, \frac{m_t}{y_t}, \frac{b_t}{k_t}, \frac{\pi_t}{y_t}, \iota_t\right) = z_t,$$

in which represents equipment investments to equipment assets, sales to equipment assets, raw material costs to sales, liabilities to equipment assets, operating profit to sales, and a year dummy variable, correspondingly. Given that each instrument variable must be orthogonal to the residual term (ε_t) defined in Equation (9), the following moment is finally derived using an instrument variable with a lag of two periods:

$$\Delta\varepsilon_t \otimes z_{t-2} = M_t(\theta). \tag{10}$$

Here, in order to remove individual and industrial effects, an estimation formula is established by placing the difference in the residual term in a cross-relationship with the instrument variable. In addition, our GMM uses a symmetric weighting matrix for efficiency of the test statistic for the estimation according to Newey and West (1987). This is a HAC (heteroscedasticity and autocorrelation consistent covariance) matrix (W). Therefore, the final GMM estimation formula is expressed as follows:

$$\hat{\theta} = \arg\min E[M(\theta)'WM(\theta)]. \tag{11}$$

In this paper, in particular according to Hansen *et al.* (1996) considering the finite-sample property of the GMM estimate, the weighting matrix (W) is modified every time from the initial value until it finally converges to find the estimate.

IV. Empirical Results: Investments and Business Cycles

A. *Estimation of the Capital Adjustment Cost Function*

We estimate the CAC function for the period from 1978 to 2019. In particular, the total number of observations is 1,563, with only firms having more than 30 consecutive years of observations used.[1]

Table 4 distinguishes the CAC for a negative investment and the CAC for a positive investment. That is, after the asymmetry of the CAC function is assumed (there is no sign constraint for ϕ^+, ϕ^-), the CAC function is re-estimated with the constraint that the CAC is symmetric ($\phi^+ = \phi^-$). It is important to note here that asymmetry does not necessarily imply irreversibility ($\phi^+ < \phi^-$).

In the estimation when assuming asymmetry, ϕ^+ and ϕ^- are estimated to be 0.13 and 1.05, respectively, and ϕ^- is estimated to be statistically significant at the 1% level. This means that the additional costs incurred during the capital adjustment process appear higher when the equipment asset is sold. This confirms that investment irreversibility as discussed in the literature may exist in Korea as well. In particular, the parameter is estimated to be 0.36 assuming symmetry of the CAC function, though the statistical significance is reduced to the 5% level. In particular, when the CAC function is constrained to be symmetric, the CAC parameter ϕ^+ increases while ϕ^- decreases. Therefore, the symmetry constraint is likely to underestimate the CAC (ϕ^-) resulting from the resale of equipment assets. However, the LM test (Lagrange multiplier test) using the constraint of $\phi^+ = \phi^-$ as the null hypothesis cannot be rejected, indicating that symmetry cannot be completely excluded (Breusch and Pagan, 1980). Also, when compared to the estimates of Whited (1992; 1998), it is judged that the CAC estimated in the Korean economy is not excessive.[2] In particular, the ratio of the estimates $\phi^- / \phi^+ = 8.1$ is not an excessive gap compared to the benchmark model of Zhang (2005), $\phi^- / \phi^+ = 10$.[3] However, the fixed cost (α) of the CAC function appears to be relatively high, which means that the fixed cost of maintaining equipment assets can be relatively high in the Korean economy.

In addition, in Table 4, μ, which is the markup of total sales for the variable cost, is estimated to be 1.25 at the 1%, with or without the restriction of symmetry. This value appears to be reasonable even compared to the range of parameter, [0.99, 2.26], estimated by Whited (1992; 1998). Finally, because the GMM uses more instrumental variables than the number of parameters to be estimated, over-identification tests are conducted. It appears that all null hypotheses (H0: over-identification does not exist) are rejected; accordingly, the problem of over-identification does not seem to be solved in our model, as shown in Whited (1998).

[1]The number of observations in this paper is not small compared to 1,024, the number of observations in Whited (1992).

[2]In Whited (1998), which uses a non-parametric method, the squared term of the CAC function (α_2) was estimated to be negative. This means, as explained in Figure 1, that the firm's actual marginal investment cost was lower than the market price.

[3]Zhang (2005) set $\phi^- / \phi^+ = 10$ as a benchmark model in a simulation considering economic data.

TABLE 4—ESTIMATION OF THE CAPITAL ADJUSTMENT COST

(A) No Constraint

	μ	ϕ^+	ϕ^-	α
value	1.25***	0.13	1.05***	11.3***
t-value	24.82	0.59	3.03	4.61
OT	0.027			

(B) Constraint: $\phi^+ = \phi^-$

	μ	ϕ^+	ϕ^-	α
value	1.25***	0.36**	NA	11.6***
t-value	25.17	2.52	NA	4.81
OT	0.025			
LM	0.553			

Note: 1) ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively, 2) OT is the p-value for the over-identification test, 3) LM is the p-value for $\phi^+ = \phi^-$.

Source: Quantiwise7; KISData.

B. *Capital Adjustment Cost and Business Cycles*

The CAC function of the Korean economy is estimated with corporate micro-data rather than running an macroeconomic simulation as in the previous section under the assumption that the CAC function does not change from the 1970s to 2019. However, as the Korean economy has developed considerably over the past 40 years, corporate dynamics have arisen, new industries have been established, and insolvent industries have disappeared such that it is too strict to assume that the CAC function does not change such a time frame.

Moreover, as explained above, the CAC function is determined by various forms of support and by regulations affecting the equipment investment market, or the corporate restructuring support system. Therefore, it is necessary to examine changes in the CAC due to changes in government policies and labor market regulations over the past 40 years and examine how closely these are related to business cycles. To this end, if we recall Equation (5), the left side of the equation refers to the marginal cost for one unit of investment, and the right side is the expected marginal value (or expected marginal return) generated by one unit of investment. Under Equation (6), if Φ_i is higher than Φ_k due to an increase in the marginal cost, the investment will decrease. If the opposite is true, the investment will be activated. If there is asymmetry between ϕ^+ and ϕ^- here, there are different policy implications. For example, if irreversibility ($\phi^+ < \phi^-$) arises under identical conditions, the current marginal cost per unit of investment will be the same, but the future expected marginal value per unit of investment may be reduced. This eventually becomes a factor that lowers the current investment scale and can act as a mechanism affecting business cycles by reducing investment dispersion in situations where macroeconomic or individual productivity shocks are identical.[4] For another example, if the CAC of equipment resales increases during an economic upturn under identical economic conditions, the firm will invest less than the planned amount of investment in consideration of uncertainty in the future. In addition, a firm that resells equipment assets in response to a current negative shock will reduce the size of the equipment to be sold if the CAC for the resale of the equipment suddenly increases. In the end, an increase in the CAC for resale of the equipment can reduce the investment

[4]A typical macroeconomic model considers the CAC function as a deep parameter that reflects macroeconomic dynamics and business cycles and does not consider it as a policy variable that can be endogenous.

dispersion and restrict the efficient allocation of equipment assets, thereby limiting the amplitude of economic growth. Looking at this in another way, if the equipment resale CAC rises even during an economic downturn, the planned purchase and resale amounts of equipment assets will both decrease. However, the expansion of frictional costs for equipment asset allocation that occurs during an economic downturn can cause inefficiency in the economy as a whole, which can widen the depth of the economic downturn. In other words, the effect of the CAC on business cycles does not simply mean the expansion or contraction of the vertical amplitude; instead, it can restrain the efficient allocation of equipment assets and negatively affect both the rising and falling periods of business cycles.

Therefore, in order to assess the changes of the CAC in the Korean economy, we estimate the CACs of firms with more than five consecutive years of data in ten-year moving windows from 1987 to 2019.[5] Figure 4 shows the trends in GDP, investment-related variables, and estimates of the CAC function. For example, the estimates for 1987 are CAC functions using GMM for firms with more than five consecutive years of data from 1978 to 1987. The figure shows that the volatility of ϕ^{-} is greater than that of ϕ^{+}. This occurs because corporate restructuring costs such as support for corporate reorganization or regulations on manpower restructuring following the downsizing of equipment assets are more dependent on government policies. In (A) of Figure 4, the overall CAC of reselling equipment assets before the Asian Financial Crisis rises more steeply than the CAC of equipment investment, and GDP tends to decrease. During the Asian Financial Crisis, the CAC of reselling equipment assets drops sharply, as corporate restructuring was carried out smoothly as layoffs became flexible due to the pressure to undertake overall corporate restructuring and enact a non-regular-worker system in the labor market due to the crisis. However, the CAC parameters for positive equipment investment move in the opposite direction. In particular, from the early 2000s to 2016 after the financial crisis, ϕ^{+} is estimated to be larger than ϕ^{-}, indicating that the cost for economic growth through new equipment investments was higher than the cost of resource allocation between firms. However, because the absolute value of ϕ^{-} during this period was higher than ϕ^{+}, it can be inferred that even if the marginal cost of one unit of investment is high, the future marginal return is higher and the overall investment incentive increases. This can be seen in (B) and (C) of Figure 4.

(B) shows the standard deviation of the investment rate, and (C) shows the proportion of firms whose absolute value of the investment rate exceeds 20% ($|i/k| > 20\%$). We can confirm that the investment dispersion is large when ϕ^{-} is smaller than ϕ^{+}. In addition, (D) shows the median value of the investment rates, which is also confirmed to increase when irreversibility decreases, similar to the distribution of investment rates. However, during the Global Financial Crisis, the median value of the investment rates decreases sharply. Finally, we find that the estimate of ϕ^{-} increases rapidly after 2017. It should also be noted that, as before the Asian Financial Crisis, the CACs for the reselling of equipment assets and supply/demand conditions worsened, and corporate restructuring costs including

[5]We also estimate the CACs of firms with more than ten consecutive years of data in 15-year moving windows for robustness in the appendix.

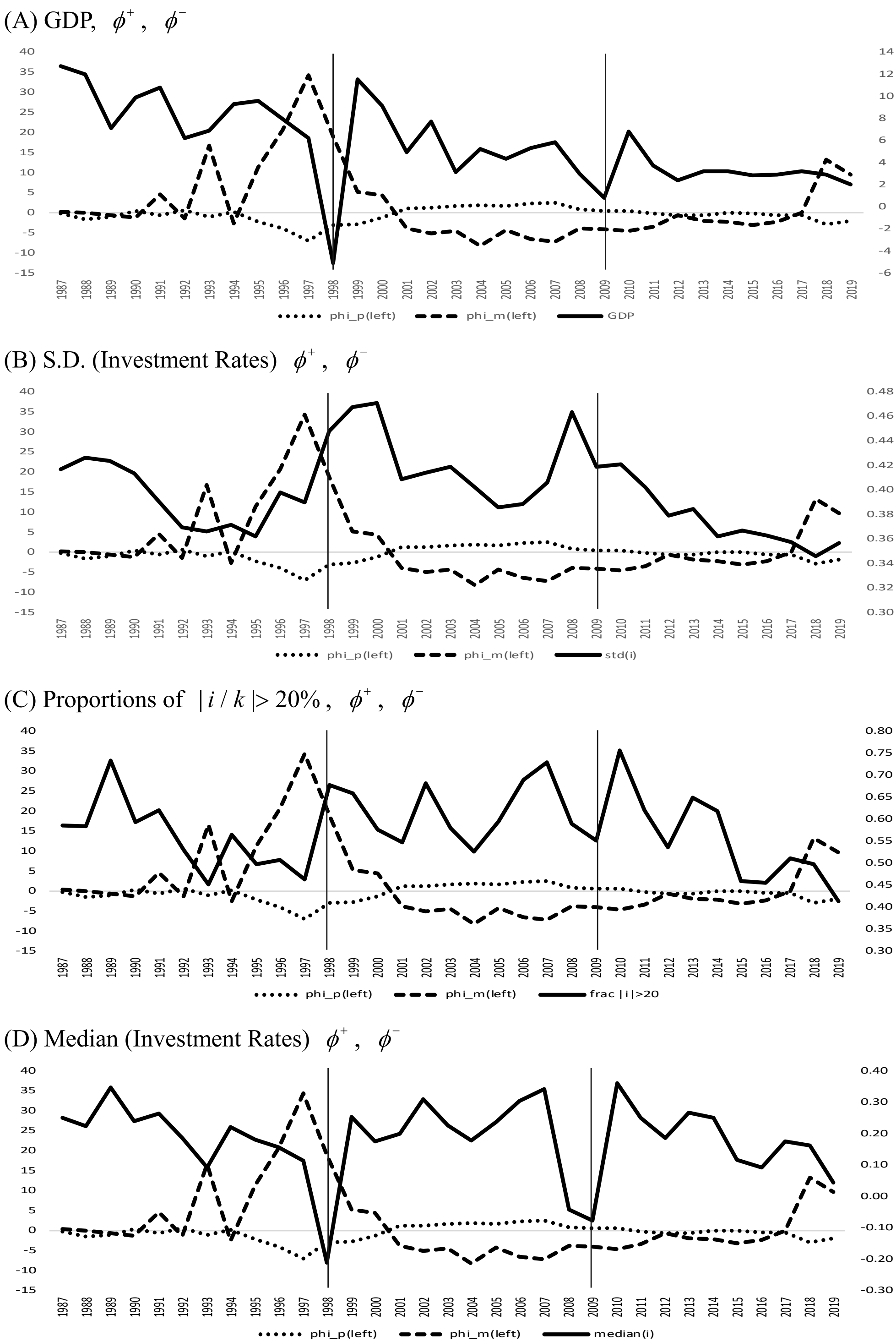


FIGURE 4. CAPITAL ADJUSTMENT COSTS AND BUSINESS CYCLES

Note: The vertical lines represent economic crises: 1998 and 2009.

Source: Quantiwise7; KISData.

TABLE 5—CORRELATION BETWEEN CAPITAL ADJUSTMENT COSTS AND BUSINESS CYCLES

	S.D. (i/k)	GDP	$\mid i/k \mid > 20\%$	$i/k > 20\%$	$i/k < -20\%$	Mean (i/k)
μ	0.38	-0.08	0.18	-0.05	0.19	-0.11
ϕ^+	0.25	**-0.39****	**0.37****	-0.02	**0.29***	-0.13
ϕ^-	**-0.33***	0.15	**-0.39****	-0.03	-0.23	0.06
α	-0.16	0.04	-0.12	-0.02	-0.05	0.00

Note: 1) ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively, 2) Data are detrended by HP-filter.

Source: Quantiwise7; KISData.

personnel restructuring costs increased. It is important to note that on the other hand an increase in the cost of business reorganization may represent a burden on firms during the COVID-19 crisis.

Table 5 statistically shows the correlation between the business cycles and estimators of the CACs described above. Some may point out the possibility of endogeneity in the correlation between the estimators and GDP and investment-related variables, but the parameters are estimated using corporate micro-data for ten years prior to the year of the macro variable. Therefore, it is important to note that endogeneity may not exist due to the averaging effect. It can be seen that μ, the markup of the raw material cost, and a, the constant term of the CAC function, do not have a high statistical correlation with economic variables. This stems from the fact that μ and a in the inter-temporal optimization condition (5) of the firm value maximization problem do not affect the changes in the marginal cost of investment and the expected marginal return. The important concepts here are the correlations among ϕ^+, ϕ^-, and the variables related to the business cycles. ϕ^+ has a significantly positive correlation with the proportion of firms whose absolute value of the investment rate exceeds 20% and has a positive correlation, although not statistically significant, with the standard deviation of the investment rates of firms. In particular, ϕ^+ has a significantly positive correlation with the proportion of firms with investment rates lower than -20% as compared to the proportion of firms with investment rates exceeding +20%. At the same time, ϕ^+ appears to have a significantly negative correlation with the GDP growth rate, which means that an increase in ϕ^+ can have a negative effect on GDP by restricting investments. However, the fact that ϕ^+ shows a positive relationship with the proportion of firms whose absolute value of the investment rates exceeds 20% is in conflict with the existing theory, but this is due to a statistically significant relationship with large-scale negative investment. In other words, when ϕ^+ rises, a faster decrease in ϕ^- appears to expand the resale of equipment assets. Consistent with this is that ϕ^- has significantly negative correlations with the proportion of firms whose absolute value of the investment rates exceeds 20% and the standard deviation of the investment rates. However, GDP and ϕ^- do not appear to be statistically significant. Although the degree of the reduction in the CAC of reselling equipment assets directly affects the investment decision process and thus affects investment dispersion, this is partially offset by the efficiency of the used equipment market, which can actually weaken the upward pressure on cycles. Moreover, given that output growth is not accurately decomposed with exogenous macroeconomic shocks, the impact of changes in the CACs on the economy may be mitigated.

V. Conclusions

We confirm that Korea's investment dispersion is pro-cyclical. In addition, it is empirically verified that the shape of the CAC function is an important factor with regard to corporate investment decisions. Specifically, it is shown for the first time that the CAC function, which is generally assumed in macroeconomics, is asymmetric and irreversible in the Korean economy through a dynamic panel GMM using corporate data rather than a macroeconomic model. In particular, we show that the reduction in the CAC for reselling equipment assets rather than an investment in equipment can affect the substitution effect of the marginal value versus the marginal cost during the inter-temporal decision-making process. Moreover, it has been empirically shown that a reduction in the CAC for a negative investment can expand the degree of investment dispersion. In the end, it is shown indirectly that the cost incurred when reselling equipment among CACs can affect investment decisions and facilitate the reallocation of equipment assets and that the CAC for purchasing equipment can have a certain impact on business cycles. Specifically, immediately after the AFC, 168.7 trillion won as public funds were used to bail out the industrial and financial sectors through KAMCO and KDIC. In addition, a workout program in the form of a British-style restructuring program was introduced to support large-scale corporate restructuring. According to the analysis result, it is judged that these policies functioned as factors that decreased ϕ^{-}. Recently, as policies affecting manpower restructuring have been strengthened overall since 2017, it is believed that the cost of manpower restructuring following the resale of equipment assets exerted pressure to increase ϕ^{-}. Therefore, it can be inferred that changes in labor policies and financial support may have some impact on the cost of reselling equipment assets according to mid- to long-term trends. This implies that the parameters of the CAC are not simply deep parameters for simulating business cycles in macroeconomic models but are rather policy variables that can be endogenized by government policies.

APPENDIX

TABLE A1—CORRELATION BETWEEN CAPITAL ADJUSTMENT COSTS AND BUSINESS CYCLES: 15-YEAR MOVING WINDOW (MORE THAN TEN CONSECUTIVE YEARS OF DATA)

	S.D. (i/k)	GDP	$\|i/k\|>20\%$	$i/k>20\%$	$i/k<-20\%$	Mean (i/k)
μ	0.42	-0.29	0.11	-0.19	0.33	-0.24
ϕ^{+}	0.22	-0.25	**0.21***	0.02	0.21	-0.06
ϕ^{-}	-0.22	0.21	**-0.38****	-0.10	-0.16	-0.01
α	-0.17	0.13	0.15	**0.31***	**-0.28***	**0.31***

Note: 1) ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively, 2) Data are detrended by HP-filter.

Source: Quantiwise7; KISData.

REFERENCES

Abel, A. B. and J. C. Eberly. 1994. "A Unified Model of Investment under Uncertainty," *American Economic Review*, 84(1): 1369-1384.

Bachmann, R. and C. Bayer. 2014. "Investment dispersion and the business cycle," *American Economic Review*, 104(4): 1392-1416.

Bloom, N. 2009. "The impact of uncertainty shocks," *Econometrica*, 77(3): 623-685.

Breusch, T. S. and A. R. Pagan. 1980. "The Lagrange multiplier test and its applications to model specification in econometrics," *The Review of Economic Studies*, 47(1): 239-253.

Cooper, R. and J. Haltiwanger. 2005. *On the nature of capital adjustment costs*, NBER Working Paper.

Cooper, R., J. Haltiwanger, and L. Power. 1999. "Machine replacement and the business cycle: lumps and bumps," *American Economic Review*, 89(4): 921-946.

Gordon, R. A. 1955. "Investment behavior and business cycles," *The Review of Economics and Statistics*, 37(1): 23-34.

Hamermesh, D. S. and G. A. Pfann. 1996. "Adjustment costs in factor demand," *Journal of Economic Literature*, 34(3): 1264-1292.

Hansen, L. P., J. Heaton, and A. Yaron. 1996. "Finite-sample properties of some alternative GMM estimators," *Journal of Business & Economic Statistics*, 14(3): 262-280.

Holt, C. C. 1960. *Planning Production, Inventories, and Work Force*.

Nam, Changwoo. 2020. *Investment and Business Cycles: focusing on Investment Dispersion*, KDI Policy Study 2020-07 (in Korean).

Newey, W. and K. West. 1987. "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica,* 55(3): 703-708.

Whited, T. M. 1992. "Debt, liquidity constraints, and corporate investment: Evidence from panel data," *The Journal of Finance,* 47(4): 1425-1460.

Whited, T. M. 1998. "Why do investment Euler equations fail?" *Journal of Business & Economic Statistics,* 16(4): 479-488.

Zhang, L. 2005. "The value premium," *The Journal of Finance,* 60(1): 67-103.

LITERATURE IN KOREAN

남창우. 2020. 『투자와 경기변동: 기업의 투자의사결정 모형을 중심으로』, 정책연구시리즈 2020-07, 한국개발연구원.

KDI Book Membership Information

Exclusive Offer (Members-Only)

• All KDI publications, with the exception of those classified as confidential or limited, are to be mailed to members
• Preferential invitations to special events hosted by KDI including seminars, policy discussion forums, public hearings, etc., are to be mailed.
• A 10% discount on the online purchases of additional copies of the published research monographs (printed-only) from the KDI homepage.

KDI Publications

• Publications include books, research monographs, policy studies, KDI policy forums, KDI FOCUS, research papers and policy-information materials.
• Three types of periodicals are available:
  - Monthly: KDI Monthly Economic Trends, KDI Review of the North Korean Economy, Economic Bulletin, Narakyungje
  - Quarterly: KDI Journal of Economic Policy, KDI Analysis on Real Estate Market Trends
  - Biannual: KDI Economic Outlook

Annual Fees

• Individual Purchase: 100,000 KRW
• Institutional Purchase: 300,000 KRW

Sign-Up

• You may sign up for membership via KDI homepage. Please register on the homepage by completing and submitting the registration form. Possible payment methods are as follows:
  - Bank-Transfer: Woori Bank, 254-012362-13-145 (account holder name: Korea Development Institute)
  - GIRO System: No. 6961017 (Credit Card and Mobile Payments available)
  - Direct Payment to the Research Outcome Dissemination Unit of KDI Division of External Affairs.

Contact

• Publication personnel in charge: Research Outcome Dissemination Unit, Division of External Affairs, KDI.
  - Tel: 82-44-550-4346 / Fax: 82-44-550-4950 / E-mail: book@kdi.re.kr
• Sales Distributors
  - Kyobo Bookstore (Gwanghwamun Branch: Government Publications Section)
    Tel: 82-2-397-3628
  - Yongpoong Bookstore (Jongno Branch: Policy & Economy Section)
    Tel: 82-2-399-5632

KDI Journal of Economic Policy
韓國開發研究

Registration Number 세종 바00002호
Printed on February, 25, 2022
Published on February, 28, 2022
Published by Jang-Pyo Hong, President of KDI
Printed by Good Idea Good Peoples

Price : 3,000 KRW
@Korea Development Institute 2021
Registered on March, 13, 1979

Korea Development Institute
263, Namsejongro, Sejong-Si, Korea I **TEL** 82-44-550-4114
FAX 82-44-550-4310 I www.kdi.re.kr

ISSN 2586-2995