FinanzArchiv
Public Finance Analysis

2
Volume 73
June
2017

Mohr Siebeck

Call for Papers
Insignificant Results in Public Finance

There is considerable concern among scholars that empirical papers are facing a drastically smaller chance of being published if the results looking to confirm an established theory turn out to be statistically insignificant. If true, such a publication bias can provide a wrong picture of economic magnitudes and mechanisms.

Against this background, *FinanzArchiv / Public Finance Analysis* is posting a call for papers for a special issue on "Insignificant Results in Public Finance". The editors are inviting the submission of carefully executed empirical papers that – despite using state of the art empirical methods – fail to find significant effects for important economic effects that have widespread acceptance.

The editors would like to note that if any insignificant results transform into statistically significant results as an outcome of the refereeing process, this will not be held as an argument against publication. In this case, the paper may be shifted into a different issue of the Journal.

While this call is specifically connected to a special issue, the editors of *FinanzArchiv / Public Finance Analysis* would like to confirm their general willingness to accept papers with insignificant results provided the indeterminacy is not a result of using inappropriate data or methods, or from a lack of care in execution.

The link for submissions using Editorial Express can be found at *www.mohr.de/fa* and submissions to the special issue should reach us before 15 September 2017. Submitting authors should indicate that their paper is being submitted to the special issue "Insignificant Results in Public Finance".

Katherine Cuff (Editor)
Ronnie Schöb (Editor)
Alfons Weichenrieder (Managing Editor)

FinanzArchiv 73 (2017), 133
ISSN 0015-2218
doi: 10.1628/001522117X14950986463925

Endogenous Financial Literacy, Saving, and Stock Market Participation

Luca Spataro and Lorenzo Corsini*

Received 11 June 2015; in revised form 8 November 2016; accepted 14 November 2016

Recent empirical literature provides evidence that financial literacy, human capital, education, saving, and stock market participation are interconnected decisions. However, a consolidated theoretical explanation of such connections is missing. We contribute to this topic by building a framework that includes all these decisions in an encompassing model. We build a two-period model in which individuals acquire education, work, and save for retirement; financial literacy reduces the costs of managing risky assets available on the stock market. Our results, besides providing a theoretical foundation for the role and the determinants of the above-mentioned decisions, can also explain several stylized facts on literacy, human capital, and stock market participation.

Keywords: financial literacy, portfolio decisions, human capital, saving for retirement, stock market participation

JEL classification: D 14, D 91, G 11, J 24

1. Introduction

Recent reforms of pension systems in several countries entail a higher degree of responsibility on the workers' side. This fact, together with the increasing complexity of individuals' financial decisions, has induced both scholars and policymakers to focus on the determinants and the role of financial decisions, such as saving, stock market participation, and financial education.

While quite a large number of empirical studies have shed light on the relevance of financial literacy as a key determinant of crucial lifetime decisions such as retirement planning and stock market participation (we review these contributions in section 2; see also Mitchell and Lusardi, 2011, for a full coverage on this subject), the theoretical literature on these issues is rather thin.

* Spataro: Università di Pisa, Via Ridolfi 10, 56124 Pisa, Italy (luca.spataro@unipi.it); Corsini: Università di Pisa, Via Ridolfi 10, 56124 Pisa, Italy (lorenzo.corsini@unipi.it). We are grateful for their useful comments to two anonymous referees, and to the participants at the seminars held at the Dipartimento di Economia e Management in Pisa, at the PET 13 Conference held in Lisbon, and at the 11th Workshop on Pensions, Insurance, and Saving held in Paris. The usual disclaimer applies.

FinanzArchiv 73 (2017), 135–162 doi: 10.1628/001522117X14877521353555
ISSN 0015-2218

Among the few studies that cover theoretical aspects, Jappelli and Padula (2013) develop a model where financial literacy affects the returns of non-stochastic savings. Lusardi et al. (2011) develop a numerically simulated life-cycle model where endogenous financial literacy affects the stochastic returns from saving and where exogenous education determines income. Finally, Corsini and Spataro (2015) develop a theoretical model where individuals' decisions on pension plans are affected by complexity costs embedded in more sophisticated plans, and they investigate how the acquisition of financial literacy influences pension-saving decisions.

On similar lines, even if not directly tackling financial literacy, some works analyze the influence that the acquisition of information and the time spent on asset management can exert on investment behavior. In particular, Ehrlich et al. (2008) analyze how time spent in acquiring information reduces the volatility of risky investments and connect this activity to the level of (nonendogenous) human capital. Ehrlich and Shin (2010) and Ehrlich et al. (2011) further develop this aspect by distinguishing between the role of general human capital and the role of "specific" human capital, that is, the component of human capital that is particularly useful for managing risky assets of a given country or context.

In this work we aim at taking a step further in the theoretical literature on individual financial decisions by providing a unified framework that encompasses human capital formation, financial education, savings, and capital market participation. To the best of our knowledge, this has not been done so far.

In particular, differently from previous literature, we focus (i) on the fact that the acquisition of financial literacy is beneficial to individuals for managing risky assets on the stock market and thus for diversifying their savings, and (ii) on the interaction between financial literacy and education, with the latter being endogenously determined and affecting both lifetime income and the cost of acquiring financial literacy. Consequently, within our model, savings, financial literacy, education, and investment decisions strongly interact and are endogenously determined.

By doing this, we obtain a relationship between financial literacy, education, income, and wealth that, although complex, is in line with several observed facts and also raises some further empirical questions that are worth addressing.

The work is organized as follows. In section 2 we sketch some stylized facts on financial literacy acquisition and stock market participation unveiled by recent empirical studies. In section 3 we lay out the model, and sections 4 and 5 are devoted to the determinants of stock market participation, financial-literacy acquisition, and wealth accumulation. The empirical implications of our analysis are discussed in section 6. Concluding remarks end the work.

2. Some Stylized Facts on Financial Education, Saving, Human Capital, and Stock Market Participation

A number of empirical studies have analyzed the subject of financial literacy, its determinants, and its consequences for saving decisions and investment behavior. Usually, in these works financial literacy is measured through surveys that contain questions on the concepts of interest compounding, inflation, purchasing power, and risk diversification; from the number of correct answers a literacy score is computed for each individual.

2.1. Determinants of Financial Literacy

Several works analyze the determinants of literacy; some of them are single-country studies[1], while others, like Jappelli (2010) and Lusardi and Mitchell (2011a), contain cross-country analyses. These studies highlight some common patterns in both the determinants and the effects of financial literacy, and they indicate that financial literacy:

- is lower among younger and older individuals;
- is lower among female individuals;
- is usually higher for self-employed workers and for individuals working in business related sectors;
- is higher for higher-income individuals;
- is higher for individuals with higher mathematical and numeracy skills;
- increases with education (schooling), but the actual relationship can be complex.

Among these stylized facts, the evidence on the relation between education and financial literacy requires some further specification. First, education and financial literacy appear to be correlated, but there is evidence that the field of study is also relevant in explaining how the level of schooling actu-

1 Analyses from Alessie et al. (2011), Almenberg and Säve-Söderbergh (2011), Crossan et al. (2011), Bucher-Koenen and Lusardi (2011), Fornero and Monticone (2011), Klapper and Panos (2011), Lusardi and Mitchell (2011b), and Sekita (2011) cover, respectively, the Netherlands, Sweden, New Zealand, Germany, Italy, Russia, the United States, and Japan. Evidence from the United States on the effect of financial literacy on planning and wealth accumulation is also contained in Lusardi and Mitchell (2011b). Finally, evidence from Italy on the effect of literacy on portfolio diversification is contained in Guiso and Jappelli (2009), and evidence from Sweden on how less-sophisticated investors tend to achieve under diversification is contained in Anderson (2013).

ally affects financial literacy (see Almenberg and Save-Soderbergh, 2011). Moreover, when both education and literacy are used as explicative variables in econometric regressions, they are usually both significant, so that literacy appears to have an effect above and beyond education (see Lusardi and Mitchell, 2011a).

Finally, some works have also revealed that age has a relevant effect on literacy and the relationship has an inverted U-shape. Empirical analyses found that people in the 35–65 age group have higher literacy than those outside it. The location of the peak in literacy depends on the country under investigation; in some it falls in the 35–50 age group, and in others in the 51–65 age group (see Alessie et al., 2011; Bucher-Koenen and Lusardi, 2011; Fornero and Monticone, 2011; Lusardi and Mitchell, 2011b).

2.2. The Effects and Consequences of Financial Literacy

From the empirical works we have mentioned above, some clear-cut empirical results on the effects of financial literacy emerge. In particular, financial literacy:

- affects positively the probability of planning for retirement;
- affects positively, other things being equal, wealth accumulation;
- is positively correlated with the degree of portfolio diversification;
- is positively correlated with stock market participation.

Evidence that financial literacy affects positively, other things being equal, wealth accumulation is also given in Lusardi et al. (2011), Behrman et al. (2012), and van Rooij et al. (2012); in particular, this causal effect persists even after controlling for factors like income, age and education.

2.3. Participation in Stock Markets

The question of participation in stock markets has also been the object of empirical analyses. As already mentioned in the previous subsection, several works confirm that participation is higher for individuals endowed with higher financial literacy. Moreover, Bertaut and Haliassos (1995), Christiansen et al. (2008), and van Rooij et al. (2011) have highlighted several common features that, even after controlling for the positive effect of financial literacy, explain participation in the stock market. The main evidence on participation is that it:

- increases with income level;
- increases with education level and depends on the field of study;
- increases when stock markets are more attractive;

– decreases with increasing risk aversion;
– is lower for females.[2]

The effect of the field of study is addressed by Christiansen et al. (2008), who find that individuals with higher education in economics-related disciplines are, all things considered, more likely to invest in the stock market. As for the role of gender, Powell and Ansic (1997) find that females are less likely to take risks in financial investments. The evidence on the role of the attractiveness of stock markets is contained in the works by Malmendier and Nagel (2011) and Thomas and Spataro (2015), who argue that households observing higher stock returns, even in early adult life, are more likely to participate in stock markets.

2.4. Stylized Facts and Our Theoretical Contribution

All the above stylized facts highlight how financial literacy and education, on one hand, and savings and participation in stock markets, on the other hand, are strongly related, not only because the former variables strongly affect the latter, but also because several factors appear to concurrently play a role in determining all these lifetime investment decisions.

In what follows we present a model that aims at encompassing all these issues in a unified framework, thus providing both a possible explanation of the stylized facts listed above and some new testable predictions on the role and determinants of saving behavior, stock market participation, and financial literacy acquisition.

3. A Theoretical Model of Human Capital and Saving

We imagine an economy populated by individuals that live for two periods: in the first one they make a choice between education and labor supply and between consumption and saving; in the second, they consume what they have saved.

Within this basic structure, we add three characterizing features: (i) savings can be invested either in a safe or in a risky asset (or both), and the latter can only be bought on the stock market; (ii) the management of complex financial assets is time-consuming, and to reduce the time cost, individuals may find it convenient to acquire or increase their level of financial literacy; (iii) literacy acquisition is also time-consuming, but the share of time needed for this investment decreases with the level of human capital. Points (ii) and (iii) reflect in particular the fact that investing in risky assets

2 Among other factors not addressed here, Renneboog and Spaenjers (2012) show that also religious beliefs affect saving attitude and stock market participation.

requires some knowledge of the mechanism behind stocks (something that is also suggested in Bertaut and Haliassos, 1995) as well as some effort to track asset performance. It follows that individuals may find it convenient to invest some time in obtaining financial education; our assumption is that the knowledge obtained from generic education effectively reduces the amount of time needed to acquire financial literacy, although to an extent dependent on the field of study.

Basically we are assuming that individuals' actions reflect the following timeline: in period 1 they (i) choose how much human capital to acquire, (ii) work and obtain a given income, (iii) choose whether to sustain financial management costs, and (iv) make consumption and investment decisions. In period 2, they simply consume what they have saved.

For the investment decision, we assume that individuals not only choose how much to save, but also decide whether and how much to invest in financial education. In fact, the latter investment may help them diversify their savings between two assets (by participating in the stock market). We assume that risky assets provide higher returns although they entail higher volatility and complexity. Consequently, since the management cost of risky assets is higher, in this case the investment in financial literacy can be particularly rewarding; on the contrary, we imagine that the cost to manage safe assets is negligible.[3]

3.1. Basic Definitions

In this subsection we characterize the structure of the model and define its components.

3.1.1. Lifetime Utility

Individuals live for two periods (working life and retirement), and their lifetime utility is given by

$$U = -e^{-ac_1} - \beta e^{-ac_2}\,, \tag{1}$$

where c_1 and c_2 are the consumption in periods one and two respectively, a is the absolute risk aversion coefficient, and β measures time preference. The

3 We should mention that financial literacy could also have a direct effect on returns to or volatility of assets. In our contribution, although not treating this case explicitly, we allow for similar effects in that we assume that literacy facilitates the managing of assets and thus can increase the net returns (i.e., net of management costs) to financial investments. For an example in which financial literacy directly affects the returns to (safe) assets see Jappelli and Padula (2013). For works where the acquisition of information makes an impact on the volatility of investments see Ehrlich et al. (2008), Ehrlich and Shin (2010), and Ehrlich et al. (2011).

above utility function displays the CARA property and is chosen to obtain closed-form solutions; however, as we describe in detail in section 3.1.4, we also allow for an inverse relationship between the risk coefficient a and total resources to be invested, which is in line with observed facts.

3.1.2. The Time Constraint

According to our setup, in the first period of life individuals split their time endowment (normalized to 1) between time devoted to general education (θ), labor supply l, and time spent for financial management (c). The time constraint is thus

$$\theta + l + c = 1 . \tag{2}$$

3.1.3. Human Capital

For the investment in general education[4], we assume that the human capital (h) production function is given by

$$h = \theta^{1/x} , \tag{3}$$

where x measures the degree of effectiveness of the education process in producing human capital. We assume that $x > 1$, so that human capital production has decreasing returns, in line with the empirical evidence (see Blundell et al., 1999; Blundell et al., 2000; Dearden,1999).

3.1.4. Income

The income per time unit, w, depends on the amount of human capital acquired through education and on a scale factor $k > 1$. The amount of income per time unit worked is

$$w = kh . \tag{4}$$

The working-life income W is the product of income per time unit (w) and labor supply l:

$$W = khl . \tag{5}$$

Note that W is an increasing function of k.

4 For the sake of simplicity we do not allow human capital acquisition before the working life. Although this assumption does not affect our results qualitatively, it probably increases the optimal time spent in education and consequently reduces the optimal time spent on the job. As a consequence, if we added pre-working-life human capital acquisition, the determinants of stock market participation and financial literacy would not change, although we would probably obtain, other things being equal, more participation in the stock markets and more individuals investing in financial literacy. We thank an anonymous referee for pointing us to the consequences of this assumption.

3.1.5. Risk Aversion

To make our model more realistic, we also allow for an inverse relationship between risk aversion and an individual's lifetime resources, and therefore we assume that

$$a = k^{-\alpha}, \tag{6}$$

where $\alpha > 0$ is the elasticity of risk aversion with respect to income.[5] In choosing the above function, we are assuming that only the exogenous determinant of lifetime resources (that is, the parameter k) influences risk aversion. We make this assumption to guarantee that human capital does not directly influence risk aversion and to avoid assuming that individuals, when determining their optimal level of human capital and wealth, are also choosing their own degree of risk aversion.

Given equation (6), an increase in α implies, other things being equal, a decrease in risk aversion. Moreover, relative risk aversion is increasing for $0 < \alpha < 1$, constant for $\alpha = 1$, and decreasing for $\alpha > 1$.

3.1.6. Savings

Individuals can save through two possible assets: a safe asset yielding a certain rate of return s, and a risky asset, available on the stock market, providing a rate of return that is normally distributed with mean r and variance σ^2. We denote by S the amount invested in the safe asset, and by R the amount invested in the risky asset.

3.1.7. Costs for Financial-Asset Management

Managing financial assets is a complex task, and some time is needed to handle it and to track stock performance. Consequently, there are some opportunity costs associated with the possibility to invest in assets, which cover the time spent to manage them and to acquire the needed financial knowledge. Given the complexity of stock markets, we can assume that these costs are particularly relevant for individuals that enter the stock market, while they are much smaller or even negligible for individuals that only invest in safe assets.

We assume that the total time cost of managing financial investments, c (measured in time units), depends on two components: q, related to the time spent managing financial assets, and f, related to the time spent investing in

5 This methodology of dealing with lifetime resources and decreasing risk aversion was first proposed by Makarov and Schornick (2010) and can help explain the empirical finding that richer individuals have a preference for riskier assets (see, for example, Vissing-Jorgensen, 2002).

financial knowledge, L. Moreover, it is likely that the former component is affected by the latter: namely, acquiring financial knowledge strictly reduces the time needed to manage financial assets. The total cost in units of time thus reads as

$$c = d(q + f), \tag{7}$$

where $d \geq 0$ is a parameter that measures the degree of complexity of this activity; clearly, the corresponding opportunity cost is $c \cdot w$. We develop our analysis in such a way that c will depend inversely on the stock of human capital of the individual. In particular, it can be shown that when individuals are allowed to optimally choose q and f (and thus acquire financial education L), equation (7) reads as (see Appendix 8.1 for details)

$$c(h) = d(1 - h^{1-z}), \tag{8}$$

where the parameter $0 < z < 1$ is a measure of the effectiveness of human capital (i.e., education) in facilitating the acquisition of financial knowledge. We note that, by this specification, different fields of education have different degrees of complementarity with financial skills, and the degree of closeness of human capital to financial education is in fact captured by the parameter z. In other words, the latter parameter measures to what extent human capital is specifically useful in facilitating asset management and in acquiring financial literacy. From this point of view there could be some similarities with the concept of "specific" human capital introduced in Ehrlich and Shin (2010) and Ehrlich et al. (2011). However, while in those contributions specific human capital represents the knowledge about a specific country or region, in the present work z captures those skills and knowledge that enhance the acquisition of financial literacy and increase efficiency in managing any asset.[6]

3.2. The Optimization Problem

For financial investments, individuals must choose whether to enter the stock market and buy risky assets (thus paying also higher costs) or to simply invest in safe assets. In the first case they choose the optimal values of h, S, and R, whereas in the second case they simply choose h and S. Depending on the complexity of assets, agents might sustain a time cost: for the sake of simplicity here we assume that the parameter d in equation (8) is equal to 0 for safe assets and 1 for risky assets.

6 We recognize that z, interpreted as the field of study, could, at least in part, be endogenous. However, for the sake of tractability we leave the analysis of this case for future research. For the present paper, we could also interpret this parameter as the intensity with which the education system of a country or of a region provides economics-related topics in compulsory education, which is exogenous to individuals.

The optimal amounts chosen under the two alternatives (safe assets only or a combination of risky and safe assets) determine two different indirect expected utilities, one for each investment strategy. By convention, we assume that individuals do not enter the stock market when the two indirect utilities are equal. At the end of this subsection we will be able to show the optimal choice of human capital, consumption, and saving in the two cases and the resulting indirect expected utilities.

3.2.1. Safe Assets Only

If an individual does not join the stock market, all savings will be invested in the safe asset only. In this case, asset management is rather simple, and, given our assumption, no time cost need be paid for it or to acquire incremental financial education. The individual time constraint implies then

$$l = 1 - \theta = 1 - h^x \,, \tag{9}$$

and the following budget constraints hold:

$$c_1 = W - S = kh(1 - h^x) - S \,, \tag{10}$$

$$c_2 = S(1 + s) \,. \tag{10b}$$

Substituting (10) and (10a) into (1), the individual maximization problem is found to be

$$\max_{h,S} -e^{-a[kh(1-h^x)-S]} - \beta e^{-a[S(1+s)]} \,, \tag{11}$$

and the solution implies

$$h^* = \left(\frac{1}{1+x}\right)^{\frac{1}{x}} , \tag{12}$$

$$S^* = \left[kh^*(1 - (h^*)^x) + \frac{\log \beta(1+s)}{a}\right] \frac{1}{2+s} , \tag{13}$$

where h^* and S^* are the optimal levels of human capital and saving, respectively (the formal derivation of (12) and (13) is provided in Appendix 8.2).

The above solutions also provide the following optimal lifetime income W^*:

$$W^* = kh^*(1 - h^*) = k\left(\frac{x}{1+x}\right)\left(\frac{1}{1+x}\right)^{\frac{1}{x}} . \tag{14}$$

Finally, the associated indirect utility when investing in the safe asset only is

$$EU^* = -\left(\frac{2+s}{1+s}\right) e^{-a\left(\frac{1}{2+s}\right)\left((1+s)W^* - \frac{\log\beta(1+s)}{a}\right)} . \tag{15}$$

3.2.2. Safe and Risky Assets

An individual choosing to enter the stock market can invest their savings in either the safe or the risky asset (or both), but in any case has to pay the time cost c for asset management and to invest in extra financial education. Given our assumptions, the time constraint is

$$l = 1 - \theta - c = h^{1-z} - h^{x}\,, \tag{16}$$

and the associated budget constraints are

$$c_1 = W - c(h)w - S - R = kh(-h^{x} + h^{1-z}) - S - R\,, \tag{17}$$

$$c_2 \sim N(S(1+s) + R(1+r), R^2\sigma^2)\,, \tag{17a}$$

where c_2 is now a stochastic variable (in fact, returns from risky assets are stochastic). Given the above equations for consumption, the maximization problem becomes[7]

$$\max_{h,S,R} - e^{-a[kh(-h^{x}+h^{1-z})-S-R]} - \beta e^{-a\left[S(1+s)+R(1+r)-\frac{a\sigma^2 R^2}{2}\right]}\,. \tag{18}$$

The solution of (18) implies

$$h_F^* = \left(\frac{2-z}{1+x}\right)^{\frac{1}{x-1+z}}\,, \tag{19}$$

where h_F^* is the optimal amount of human capital when the individual decides to invest also in risky assets. Interestingly, by comparing (12) and (19) one finds that $h_F^* > h^*$: in fact, given that human capital reduces the cost of asset management, individuals joining the stock market also find it convenient to invest more in education. The above equation also yields the following optimal lifetime income (W_F^*) and savings (R_F^*, S_F^*):

$$W_F^* = k\left(\frac{x-1+z}{2-z}\right)\left(\frac{2-z}{1+x}\right)^{\frac{1+x}{x-1+z}}\,, \tag{20}$$

$$R_F^* = \frac{r-s}{a\sigma^2}\,, \tag{21}$$

$$S_F^* = \left[W_F^* + \frac{\log\beta(1+s)}{a} - \left(2+\frac{r+s}{2}\right)\frac{r-s}{a\sigma^2}\right]\left(\frac{1}{2+s}\right)\,, \tag{22}$$

where S_F^* and R_F^* are, respectively, the optimal amounts of safe and risky investments (the formal derivation of (19), (21), and (22) is provided in Appendix 8.2). Finally, the indirect utility in this case is

$$EU_F^* = -\frac{2+s}{1+s}e^{-a\left[(1+s)W_F^* - \frac{\log\beta(1+s)}{a} + \frac{(r-s)^2}{2a\sigma^2}\right]\left(\frac{1}{2+s}\right)}\,. \tag{23}$$

7 We exploit here a well-known property that, for any given stochastic variable n_j distributed normally with mean n and variance σ_n^2, we have that $E(e^{-a\,n_j}) = e^{-a(n-a\,\sigma_n^2/2)}$: see Varian (1992).

3.3. Participation in the Stock Market and Acquisition of Financial Literacy

As already mentioned, individuals choose to enter the stock market when the resulting indirect utility is higher than the one obtaining in the alternative strategy. In terms of our model this happens if $EU_F^* > EU^*$, that is, given (15) and (23), if

$$-\left(\frac{2+s}{1+s}\right)e^{-a\left[(1+s)W_F^*-\frac{\log\beta(1+s)}{a}+\frac{(r-s)^2}{2a\sigma^2}\right]\left(\frac{1}{2+s}\right)} > -\left(\frac{2+s}{1+s}\right)e^{-a\left(\frac{1}{2+s}\right)\left[(1+s)W^*-\frac{\log\beta(1+s)}{a}\right]}, \tag{24}$$

which implies

$$W^* - W_F^* < \frac{1}{1+s}\frac{(r-s)^2}{2a\sigma^2}. \tag{24a}$$

Exploiting (6), (14), and (20), the inequality (24a) reads

$$\left(\frac{x}{1+x}\right)\left(\frac{1}{1+x}\right)^{\frac{1}{x}} - \left(\frac{x-1+z}{2-z}\right)\left(\frac{2-z}{1+x}\right)^{\frac{1+x}{x-1+z}} < \frac{1}{k^{1-\alpha}}\left[\frac{1}{1+s}\frac{(r-s)^2}{2\sigma^2}\right]. \tag{24b}$$

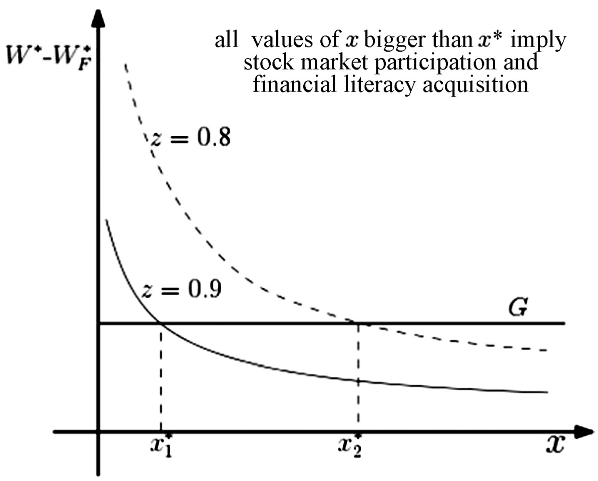
The above inequality determines univocally whether individuals invest in the stock market and acquire financial literacy or not.

The left-hand side of the inequality (24b) is a decreasing function of x and z (the proof is contained in Appendix 8.3). Thus, individuals whose education investment is more effective in producing human capital (i.e., higher x) or in abating portfolio management costs (i.e., higher z) are more likely to participate in the stock market and to acquire financial education.

The right-hand side of the inequality (24b) is increasing in the risk premium $(r - s)$ and decreasing in the degree of risk aversion (in fact, the RHS is increasing in α), the variance of risky-asset returns, and the return from the safe asset, which implies that individuals are more likely to invest in risky assets and acquire financial literacy when the excess return of the risky asset is high, when its volatility is low, and when their risk aversion is low. The RHS also depends on lifetime income, as the latter influences both risk aversion and the opportunity costs of risky-asset management. However, these effects are only due to the component k of lifetime resources, and the sign of the total effect depends on the value of α: for $\alpha < 1$ the component k has a negative effect on the RHS, for $\alpha > 1$ it has a positive effect, and for $\alpha = 1$ it has no effect. Accordingly, for a given level of human capital, richer individuals are more (less) likely to enter the stock market and acquire financial literacy when the value of the elasticity of risk aversion with respect to lifetime income is larger (smaller) than 1.

For illustrative purposes, in figure 1 we present a graphic representation of the inequality (24b), where $G \equiv \frac{1}{k^{1-\alpha}}\left[\frac{1}{1+s}\frac{(r-s)^2}{2\sigma^2}\right]$ is the RHS and the two decreasing curves represent the LHS (that is, the difference between $W^* - W_F^*$) as a function of x, for two values of z ($z = 0.8$ and $z = 0.9$ for the higher and lower curve, respectively). For points on the curve below the G line individuals enter the stock market and invest in financial education. As is clear from the decreasing pattern of the curves, only individuals with a value of x higher than a certain threshold x^* actually invest in risky-asset management. Moreover, an increase in the value of z shifts the curve downward so that the threshold value of x becomes smaller.

Figure 1
Threshold Levels of x Determining the Incentives to Invest in Stock Market and Financial Literacy, for Different Levels of z.



Finally, our model also determines the optimal amount of literacy that individuals acquire to abate the management costs of risky assets. This optimal amount is obtained through the optimization process on time allocation that delivered equation (8) and is equal to (see appendix 8.1)

$$L^* = \left(\frac{2-z}{1+x}\right)^{\frac{1-z}{x-1+z}}. \qquad (25)$$

4. The Determinants of Financial Literacy and of Stock Market Participation

In the light of the analysis carried out so far, we can summarize our findings in several propositions. The following two propositions stem from the inequality (24b).

Proposition 1 The share of individuals that acquire financial literacy and participate in the stock market is an increasing function of the excess return to risky assets $(r - s)$ and a decreasing function of the rate of return of the safe asset s and of the variance of risky-asset returns (σ^2).

Proof. The proof follows by inspection of the RHS of the inequality (24b).■

Quite obviously, the share of individuals participating in the stock market increases when the risky asset becomes more attractive in terms of mean returns and volatility. In addition, an increase of the return to the safe asset (keeping the risk premium constant) reduces participation. This is due to the fact that when s increases, the overall wealth increases as well, and this magnifies the difference between W^* and W_F^* (as the former is strictly larger than the latter).

By exploiting the LHS of the inequality (24), we can write the following:

Proposition 2 Individuals endowed with larger values of x and z are more likely to acquire financial literacy and to participate in the stock market.

Proof. See Appendix 8.3.

We can provide an economic interpretation of the content of proposition 2. In fact, the parameter x measures how effective individuals are in acquiring education and human capital (or how effective the education system is in providing individuals with human capital). Hence, individuals with larger values of x, *ceteris paribus*, acquire more human capital and, as a side effect of human capital and education, sustain lower costs for asset management and financial-literacy acquisition. Consequently, these individuals are more likely to enter the stock market and acquire literacy. As for the parameter z, it measures how effective human capital and education are in abating the financial management costs: individuals with higher values of z have engaged in education activities that are closer to finance, and their costs of managing risky assets and of acquiring financial literacy are smaller.

Finally, we can also provide the following proposition on the optimal amount of financial literacy:

Proposition 3 The amount of financial literacy acquired is larger for individuals endowed with larger x and larger z.

Proof. See Appendix 8.3.

This result reflects two relevant economic mechanisms: first, when the acquisition of human capital is more effective (higher x), individuals end up having more human capital, and this makes it easier to acquire literacy; second, human capital that is closer to the financial sector (higher z) further reduces the cost of acquiring literacy, and this provides incentives to invest in it. This result also implies that the level of financial literacy, for individuals of a given degree of education, depends on z; that is, individuals with the same number of years of schooling could have acquired different levels of financial literacy, depending on their field of study.

5. Savings and Wealth Accumulation

In this section we explore the role that individuals' financial investment strategy has in determining wealth accumulation.

For the sake of simplicity, we define wealth as the total amount that has been accumulated during the working life, net of capitalization of interest: at the end of the section we will discuss how the results are even more robust if we also include the capitalization of interest.

We compute first the total amounts saved by individuals who invested in safe assets and did not invest in the stock market. In particular, if an individual does not invest in risky assets, the resulting total lifetime savings are equal to the amount invested in the safe asset:

$$S^* = \left[W^* + \frac{\log\beta(1+s)}{a}\right]\frac{1}{2+s}. \tag{26}$$

On the contrary, if an individual enters the stock market, the total savings are given by the sum of the amounts invested in the safe and in the risky asset:

$$S_F^* + R_F^* = \left[W_F^* + \frac{\log\beta(1+s)}{a}\right]\frac{1}{2+s} + \left(1 - \frac{r-s}{2}\frac{1}{2+s}\right)\frac{r-s}{2a\sigma^2}. \tag{27}$$

Given the above results, we can evaluate whether joining the stock market and acquiring financial literacy implies higher wealth accumulation. Suppose then that we observe two different individuals: agent 1 has entered the stock market while agent 2 has not, either because they have different preferences (identified by different values of β and a) or because they have different values of any of the triple (x, z, k). The difference in their total savings is given by the following (the subscript identifies the agent):

$$S^*_{F,1} + R^*_{F,1} - S_2 = \left[W^*_{F,1} - W^*_2 + \frac{\log \beta_1(1+s)}{a_1} - \frac{\log \beta_2(1+s)}{a_2} \right] \frac{1}{2+s} + \left(1 - \frac{r-s}{2}\frac{1}{2+s}\right)\frac{r-s}{2a_1\sigma^2}. \tag{28}$$

We can exploit the above equation to formulate the following proposition:

Proposition 4 Suppose that two individuals are identical in lifetime income (W), risk aversion (a), and time preference (β), but they differ in that the former has entered the stock market and acquired financial literacy and the latter has not. Then, if $r < 4 + 3s$, the former individual accumulates a larger amount of wealth.

Proof. If individuals are identical in their lifetime income, risk aversion, and time preferences, we have, by assumption, $W^*_{F,1} - W^*_2 = 0$ (by (14) and (20)), and, given a triple (x_1, z_1, k_1), we can always find a combination of x_2, z_2, and k_2 providing the equalities $a_1 = a_2$ and $\beta_1 = \beta_2$. Consequently, from (28) we have $S^*_{F,1} + R^*_{F,1} - S_2 = \left(1 - \frac{r-s}{2}\frac{1}{2+s}\right)\frac{r-s}{2a_1\sigma^2}$. The latter is strictly positive for $r < 4 + 3s$. ■

The condition $r < 4 + 3s$ is satisfied for realistic values of the parameters, so that our results imply that individuals that have invested in the stock market and acquired financial literacy accumulate a higher amount of wealth[8] than individuals who have the same preferences and lifetime income but did not invest in stock markets or in financial literacy.

In our definition of wealth we did not include the capitalization of returns. If we had also included this factor, given that $E(r) > s$, it is easy to see that the wealth of individuals investing in the stock market and financial literacy would be, on average, larger than the amount accumulated by the others, for an even less stringent condition than $r < 4 + 3s$. That is, engaging in risky-asset management would imply, on average, an even larger accumulation of wealth.

8 Assuming that a period is 30 years long, a sufficient (but not necessary) condition for the inequality $r < 4 + 3s$ to be satisfied is that the excess return does not exceed 4%. Therefore, only for extremely high excess return could the conclusion could be reversed. This theoretical occurrence is due to the income effect that, when r is extremely high relative to s, induces individuals that are investing in the stock market to consume more in both periods, thus reducing the amount they are saving. A numerical exercise can clarify the above inequality: in the last 50 years, in the U.S. the geometric averages of the real returns from the stock market and from Treasury bills were 5.3% and 0.08%, respectively (see Board of Governors of the Federal Reserve System US, 2016a, 2016b and Crawford et al., 2016). If we use these data to compute r and s over a horizon of 30 years, we obtain that r is 3.7 and s is 0.27; these values satisfy the inequality of proposition 4. Given that Treasury-bill returns should be considered a lower bound for safe-asset returns, the inequality of proposition 4 appears to be satisfied for historical U.S. data of the last 50 years. We thank an anonymous referee for pointing out this empirical check.

6. Main Findings and Reconciling Theory with Evidence

The model we have presented delivers some clear-cut results on the determinants of financial literacy acquisition and stock market participation. In this section we compare our theoretical findings with the stylized facts presented in section 2 and we also provide some further empirical insights.

6.1. Education and Numeracy

According to our results, there is a positive relationship between education and financial literacy. However, the degrees of education and financial literacy are not perfectly correlated: in fact, the field of study (possibly represented by the parameter z in our model) is also a key determinant. This is in line with the observed facts (see section 2.1) that, on the one hand, schooling exerts a positive effect on financial knowledge (although with a wide variability that depends, among other things, on the field of study) and, on the other hand, literacy has an effect that goes beyond mere education level. Our model also indicates that education systems that are more effective in providing human capital (i.e., higher x) not only improve human capital accumulation, but also capital market participation and financial-literacy acquirement. As an empirically testable result, we predict that for a given level of schooling, financial literacy should vary across individuals, depending on the field of study and on the efficiency of the education system (at cohort, regional, or country level) in producing human capital.

Moreover, individuals that have developed good abilities in math and numeracy can be thought as having, according to our model, a higher level of z and thus should be more likely to acquire financial literacy and invest in stock markets, which is line with the observed evidence (see section 2.1).

6.2. Income

Our results provide a complex relationship between income and literacy, which can be broken down into three components. First, both income and literacy are positively influenced by education; this component implies a positive correlation but not a causal relationship between income and literacy. Second, higher income implies a higher amount of lifetime resources invested and also lower risk aversion. This in turn implies that individuals with higher income are keen to buy a higher amount of stocks and, consequently, of financial literacy in order to enter the stock market; therefore, this component entails a causal positive relationship between income and literacy. Finally, higher income also implies higher opportunity costs, and this produces a negative effect of income on the incentives to acquire financial literacy. Our model allows us to determine the overall effect of the latter two components, and in particular we have that for $\alpha \geq 1$ (that is, when relative risk aversion is nonincreasing in income) the overall effect is nonnegative, whereas it is negative otherwise. As a consequence, if we consider all the three components, we can be certain that for $\alpha \geq 1$ income and literacy (and stock market participation) are positively correlated, whereas for $\alpha < 1$ the sign of the relationship is undetermined. The assumption that relative risk aversion is nonincreasing in income is usually the most realistic one (see, for example, MasColell et al., 1995, chapter 6), and thus our model fully explains the observed positive correlation between income and financial literacy (discussed in section 2.1) and between income and stock market participation (discussed in section 2.3). Further empirical analysis aimed at disentangling the role of income from that of education would however be helpful to shed further light on the ultimate nature of the relationship between these two factors.

6.3. Employment Condition

Individuals employed in sectors close to business are likely to have acquired a type of human capital that displays higher complementarities with financial knowledge. In the analytical terms of our model, this implies higher z, and thus these individuals should be more likely to acquire financial literacy and join the stock market. A similar reasoning can be used for self-employed workers: in fact, it is likely that, on the one hand, the education that the self-employed acquired to perform their work and, on the other hand, the experience they have gained on the job entail a certain degree of complementarity with financial education. In terms of our model, this translates into higher z. Moreover, given the entrepreneurial nature of self-employment, they should probably have lower risk aversion (higher α in our model). According to these lines of reasoning, our model predicts that self-employed persons have higher probability of acquiring literacy and investing in risky assets. These theoretical predictions are in line with the observed evidence described in sections 2.1 and 2.3.

6.4. Gender

Our model does not include any gender differences and thus cannot account for gender gaps. However, we find that higher risk aversion reduces the incentive to acquire financial literacy and invest in stocks. Given that there exists some evidence that females have higher risk aversion (see Borghans et al., 2009; Croson and Gneezy, 2009), this could explain, at least to some extent, the observed gender differences in financial literacy and stock market participation that we discussed in sections 2.1 and 2.3. However, Barber and Odean (2001) show that risk aversion alone is not always enough to explain gender differences in investment behavior; therefore, further empirical analyses are in any case needed to discern whether risk aversion, *ceteris paribus*, fully explains the gender gap in financial literacy or not.

6.5. Wealth

Our analysis shows that, for realistic values of the parameters, individuals who have acquired financial literacy and have invested in stocks tend to accumulate larger amounts of wealth. This result, which is in line with the empirical evidence discussed in section 2.2, is due to the fact that these individuals save larger amounts and is even truer if the stock market yields returns that are, on average, higher. A testable implication of our model is that individuals with larger financial literacy have, *ceteris paribus*, larger propensity to save.

6.6. Stock Market Participation

Our model strongly relates participation in stock markets to the acquisition of financial literacy. As a result, education and income, influencing literacy, also increase the likelihood of acquiring stocks, something that has been observed in reality. Moreover, we highlight how schooling attainment alone is not enough to explain participation, given that the field of study plays a relevant role too. Finally, and somewhat obviously, higher risk aversion reduces the likelihood of acquiring stocks. All these theoretical results are in line with the empirical evidence on the relationship between financial literacy and stock market participation (see section 2.1) and on the other determinants of participation (see section 2.3).

6.7. The Attractiveness of Risky Assets

We find that the more attractive (in terms of Sharpe ratios) the risky assets, the larger the incentives to acquire financial literacy and to invest in stocks. This result is in line with evidence described in section 2.3. The similar prediction on financial literacy acquisition has not been tested so far, and it would be interesting to test it with respect to differences at the cohort and/or the cross-country level.

7. Conclusions

In the present work we have built a model that encompasses in a unified framework human-capital investment, financial literacy acquirement, saving behavior, and participation in the stock market. Our model provides a clear account of the effects that education, field of study, income, and employment conditions have on financial literacy and stock market participation. In addition, even if we do not address it directly, the observed effect of gender on financial decisions can find a partial rationale within our framework in terms of differences in risk aversion.

Some of our theoretical results stimulate further empirical analysis. A first empirical implication is that individuals or countries with education systems that perform better in producing human capital should also have higher levels of financial literacy and higher participation in stock markets. A second implication is that individuals living in countries with education systems that provide basic financial education should display higher participation rates in the stock market, higher financial literacy, and higher propensity to save. Finally, our findings put an emphasis on the role of other institutional factors, such as the gender wage gap or more attractive capital markets, as possible determinants of differences in financial literacy attainment or stock market participation, both at cohort and at country level. All these empirical implications are left for future empirical research. Another interesting avenue for future research is the possible extension of the present analysis to a multi-period framework, in such a way that the role of age and family composition, both documented by the empirical literature and disregarded here, could be assessed.

Finally, the results of our model offer some policy implications. Given the increasing importance of private savings in determining individuals' economic well-being in current and future societies, the adequacy of households' saving is an issue of major concern. Our model predicts that saving behavior (and the propensity to save) depends on financial literacy, which in turn depends, among other things, on education. Therefore, our findings suggest that policies aimed at fostering general education and economics-related knowledge are highly beneficial both for increasing financial literacy and for promoting economic well-being of individuals through higher saving and portfolio diversification.

Indeed, the interaction between education and financial literacy is at the center of the current political debate of many countries: the Programme for International Student Assessment (PISA), for example, has recently added specific modules to measure and evaluate the financial literacy skills of young students (see OECD, 2014), and several OECD countries are discussing how to promote financial education (see OECD, 2016). However, only a few of them (the UK is one of the rare examples; see OECD, 2016) have actually implemented compulsory financial education curricula within their general education systems. Our analysis provides solid arguments in support of these policies and calls for their enhancement at the international level.

8. Appendix

8.1. Micro Foundation of Asset Management Costs and Direct Acquisition of Financial Literacy

We provide here the micro foundation of asset management costs (equation (8) in the text). The basic idea we develop here is that investing in risky assets is a complex activity and requires time to evaluate, choose, and keep track of asset performance. Given this complexity, it is sensible to expect that the time spent in this process depends inversely on the level of financial literacy of individuals. Therefore, individuals choosing to invest in risky assets may find it convenient to invest also part of their time not only in general education but also in financial education: in fact, this investment will allow them to acquire financial literacy and thus to lessen the time spent on participating in financial markets.

Hence, we assume that the time cost of financial management is made of two components: q and f (see (7)).The former is associated with the time spent to evaluate, choose, and track risky-asset performance, which is assumed to depend inversely on the level of financial literacy L, that is,

$$q = 1 - 2L\,. \tag{29}$$

The latter component f is related to the time spent acquiring financial education L according to the following production function:

$$L = (fh^{1-z})^{1/2}\,. \tag{30}$$

The output L is increasing in f, and its marginal productivity depends positively on the level of general education of individuals (h) and on how useful the specific education is in the acquisition of literacy (e.g., on the field of study, measured by z). Finally, we assume that the marginal productivity of f is decreasing. By exploiting (7), (29), and (30), we get

$$c(L, h) = d(q + f) = d(1 - 2L + L^2h^{z-1})\,. \tag{31}$$

By plugging (31) and the equations for the budget constraint, (17) and (17b), into (1), the maximization problem is expressed as

$$\max_{h,L,S,R} -e^{-a[kh(1-h^x-d2dL-dL^2h^{z-1})-S-R]} - \beta e^{-a\left[S(1+s)+R(1+r)-\frac{a\sigma^2R^2}{2}\right]}\,. \tag{32}$$

The solution of (32) with respect to L yields

$$\frac{\partial EU_R}{\partial L} = 0 \Rightarrow -akhd(2 - 2Lh^{z-1})e^{-a[kh(1-h^x-d+2dL-dL^2h^{z-1})-S-R]} = 0\,, \tag{33}$$

so that

$$L^* = h^{1-z}\,, \tag{34}$$

which, together with (31), implies

$$c(L^*, h) = d(1 - h^{1-z}), \tag{35}$$

which is equation (8) in the text. Moreover, equations (34) and (19) determine the optimal level of financial literacy acquired by individuals:

$$L^* = \left(\frac{2-z}{1+x}\right)^{\frac{1-z}{x-1+z}}. \tag{25}$$

8.2. Optimal Levels of Human Capital and Investment in Safe and Risky Assets

We derive here the optimal levels h^* and S^* as given, respectively, by (12) and (13), and the optimal levels h_F^*, R_F^*, and S_F^* as given, respectively, by (19), (20), and (21).

8.2.1. Derivation of h^* and S^*

Consider the optimization problem $\max_{h,S} -e^{-a[kh(1-h^x)-S]} - \beta e^{-a[S(1+s)]}$. The first-order conditions imply

$$\frac{\partial\left\{-e^{-a[kh(1-h^x)-S]} - \beta e^{-a[S(1+s)]}\right\}}{\partial h} = 0 \Rightarrow akh[1-(1+x)h^x]e^{-a[kh(1-h^x)-S]} = 0, \tag{36}$$

$$\frac{\partial\left\{-e^{-a[kh(1-h^x)-S]} - \beta e^{-a[S(1+s)]}\right\}}{\partial S} = 0 \Rightarrow ae^{-a[kh(1-h^x)-S]} - a\beta(1+s)e^{-a[S(1+s)]} = 0. \tag{37}$$

From (36) we directly obtain the optimal level of human capital, h^*:

$$h^* = \left(\frac{1}{1+x}\right)^{\frac{1}{x}}, \tag{38}$$

and from (37), inserting h^*, we obtain the optimal level of saving, S^*:

$$S^* = \left[kh^*(1-h^{*x}) + \frac{\log\beta(1+s)}{a}\right]\frac{1}{2+s}. \tag{39}$$

8.2.2. Derivation of h_{F^*}, R_{F^*}, and S_{F^*}

Consider the optimization problem

$$\max_{h,S} -e^{-at[kh(-h^x+h^{1-z})-S-R]} - \beta e^{-a\left[S(1+s)+R(1+r)-\frac{a\sigma^2R^2}{2}\right]}.$$

The first-order conditions imply

$$\frac{\partial\left\{-e^{-a[kh(h^{1-z}-h^{x})-S-R]}-\beta e^{-a\left[S(1+s)+R(1+r)-\frac{a\sigma^{2}R^{2}}{2}\right]}\right\}}{\partial h}=0\Rightarrow \tag{40}$$

$$ak[(2-z)h^{1-z}-(1+x)h^{x}]e^{-a[kh(h^{1-z}-h^{x})-S-R]}=0\,,$$

$$\frac{\partial\left\{-e^{-a[kh(-h^{x}+h^{1-z})-S-R]}-\beta e^{-a\left[S(1+s)+R(1+r)-\frac{a\sigma^{2}R^{2}}{2}\right]}\right\}}{\partial S}=0\Rightarrow \tag{41}$$

$$ae^{-a[kh(-h^{x}+h^{1-z})-S-R]}=a\beta(1+s)e^{-a\left[S(1+s)+R(1+r)-\frac{a\sigma^{2}R^{2}}{2}\right]}\,,$$

$$\frac{\partial\left\{-e^{-a[kh(-h^{x}+h^{1-z})-S-R]}-\beta e^{-a\left[S(1+s)+R(1+r)-\frac{a\sigma^{2}R^{2}}{2}\right]}\right\}}{\partial R}=0\Rightarrow \tag{42}$$

$$ae^{-a[kh(-h^{x}+h^{1-z})-S-R]}=a\beta[(1+r)-a\sigma^{2}R]e^{-a\left[S(1+s)+R(1+r)-\frac{a\sigma^{2}R^{2}}{2}\right]}\,.$$

From (40) we directly obtain the optimal level of human capital, h_F^*:

$$h_F^*=\left(\frac{2-z}{1+x}\right)^{\frac{1}{x-1+z}}\,. \tag{43}$$

Moreover, combining (41)–(43), we obtain the optimal level of investment in risky assets, R_F^*, and the optimal level of investment in safe assets, S_F^*:

$$R_F^*=\frac{r-s}{a\sigma^2}\,, \tag{44}$$

$$S_F^*=\left[W_F^*+\frac{\log\beta(1+s)}{a}-\left(2+\frac{r+s}{2}\right)\frac{r-s}{a\sigma^2}\right]\left(\frac{1}{2+s}\right)\,. \tag{45}$$

8.3. Proof of Proposition 2 and Proposition 3

8.3.1. Proof of Proposition 2

We prove here proposition 2. From the inequality (24a) we know that individuals acquire financial literacy and enter the stock market if $W^*-W_F^*$ is lower than $\frac{1}{1+s}\frac{(r-s)^2}{2a\sigma^2}$. Here we prove that $W^*-W_F^*$ is a decreasing function of x and z, and therefore individuals with higher x or z are more likely to acquire financial literacy and enter the stock market.

Formally, we need to prove that

$$\frac{\partial\left(W^*-W_F^*\right)}{\partial x}<0\,, \tag{46}$$

$$\frac{\partial\left(W^*-W_F^*\right)}{\partial z}<0\,. \tag{47}$$

Preliminarily, we define the following variables and functions:

$$g \equiv \frac{1}{1+x}\,, \tag{48}$$

$$p \equiv \frac{2-z}{1+x}\,, \tag{49}$$

$$b(y) = \ln y + 1 - y\,, \tag{50}$$

$$f(y) = y^{\frac{y}{1-y}}\,, \tag{51}$$

$$h(y) = f(y)^{f(y)}\,. \tag{52}$$

Note that by definition, we have that $0 < g < p < 1$ and that $f(y) > 0$ and $h(y) > 0$ for $y > 0$.

Starting from (14) and given the definition (48), we have

$$W^* = k\left(1 - \frac{1}{1+x}\right)\left(\frac{1}{1+x}\right)^{\frac{1}{x}} = k(1-g)(g)^{\frac{g}{1-g}}\,. \tag{53}$$

Similarly, starting from (20) and given the definitions (48) and (49), we have

$$W_F^* = k\left(1 - \frac{2-z}{1+x}\right)\left(\frac{2-z}{1+x}\right)^{\frac{2-z}{x-1+z}} = k(1-p)(p)^{\frac{p}{1-p}}\,. \tag{54}$$

We now present four lemmas and we use them to prove (46) and (47).

Lemma 1 The following properties hold true: $y > 0 \Rightarrow b(y) \leq 0$ and $y > 0 \cap y \neq 1 \Rightarrow b(y) < 0$.

Proof. Starting from the definition (50), we have that $\frac{\partial b(y)}{\partial y} = \frac{\partial(\ln y + 1 - y)}{\partial y} = \frac{1}{y} - 1$. Therefore $\frac{\partial[b(y)]}{\partial y}$ is positive for $y < 1$, zero for $y = 1$, and negative for $y > 1$. It follows that $b(y)$ has a global maximum for $y = 1$. We have then that $b(y) = b(1) = 0$, which implies that $b(y) \leq 0$. Moreover, given that $b(y)$ is zero only for $y = 1$, we also have that $b(y) < 0 \forall y > 0 \cap y \neq 1$. ■

Lemma 2 The following property holds true: $y > 0 \cap y \neq 1 \Rightarrow \frac{\partial f(y)}{\partial y} < 0$.

Proof. From the definitions (50) and (51) we compute $\frac{\partial f(y)}{\partial y} = \frac{y^{\frac{y}{1-y}}(\ln y + 1 - y)}{(y-1)^2} = \frac{y^{\frac{y}{1-y}} b(y)}{(y-1)^2}$. From lemma 1 we know that $b(y) < 0 \forall y > 0 \cap y \neq 1$, thus $\frac{\partial f(y)}{\partial y} < 0 \forall y > 0 \cap y \neq 1$. ■

Lemma 3 The following property holds true: $1/e < f(p) < f(g)$.

Proof. Given that $0 < g < p < 1$, from lemma 2 we necessarily have $f(p) < f(g)$. Moreover, given lemma 2, the infimum of $f(p)$ is obtained for the supremum of p, which, given the definition of p, is obtained for $z = 0$ and $x \to 1$. Therefore, $\inf f(p) = \lim_{x \to 1} \left(\frac{2}{1+x}\right)^{\frac{1+x}{x-1}}$. We prove now that $\lim_{x \to 1} \left(\frac{2}{1+x}\right)^{\frac{1+x}{x-1}} = 1/e$. We define $v \equiv \frac{1-x}{1+x}$, and rearranging $\lim_{x\to 1}\left(\frac{2}{1+x}\right)^{\frac{1+x}{x-1}}$, we obtain $\lim_{v\to 0}\left[\left(1+\frac{1}{v}\right)^{\frac{1}{v}}\right]^{-1}$, which, by definition, is equal to $1/e$. We then have $\frac{1}{e} < f(p) < f(g)$. ■

Lemma 4 The following property holds true: $0 < y \le p \Rightarrow \frac{\partial h(y)}{\partial y} < 0$.

Proof. From the definitions (50), (51), and (52) we compute $\frac{\partial h(y)}{\partial y} = \frac{h(y)\frac{\partial f(y)}{\partial y}[1+\ln f(y)]}{y}$. For $0 < y \le p$ we know: (i) by definition, that $y > 0$ and $h(y) > 0$; (ii) from lemma 2, that $\frac{\partial f(y)}{\partial y} < 0$; and (iii) from lemma 3, that $[1+\ln f(y)] > 0$. Therefore we necessarily have that, for $0 < y \le p$, $\frac{\partial h(y)}{\partial y} < 0$. ■

We use now the four lemmas to prove (46) and (47).

Proof of (46). We can rearrange $\frac{\partial(W^*-W_F^*)}{\partial x}$ as $\frac{\partial(W^*-W_F^*)}{\partial x} = \left(\frac{\partial W^*}{\partial g}\frac{\partial g}{\partial x} - \frac{\partial W_F^*}{\partial p}\frac{\partial p}{\partial x}\right)$. From the latter, using the definitions (48) and (49) and equations (53) and (54), we obtain $\frac{\partial(W^*-W_F^*)}{\partial x} = -kg\left(g^{\frac{g}{1-g}}\ln g^{\frac{g}{1-g}} - p^{\frac{p}{1-p}}\ln p^{\frac{p}{1-p}}\right)$, and from the definitions (51) and (52), we obtain $\frac{\partial(W^*-W_F^*)}{\partial x} = -kg[\ln f(g)^{f(g)} - \ln f(p)^{f(p)}] = -kg[\ln h(g) - \ln h(p)]$. Consider now $[\ln h(g) - \ln h(p)]$: we know from lemma 4 that $0 < y \le p \Rightarrow \frac{\partial h(y)}{\partial y} < 0$, and, given that $g < p$, it follows that $h(g) > h(p)$ and $[\ln h(g) - \ln h(p)] > 0$. Then $\frac{\partial(W^*-W_F^*)}{\partial x} = -kg[\ln h(g) - \ln h(p)] < 0$. ■

Proof of (47). We can rearrange $\frac{\partial(W^*-W_F^*)}{\partial z}$ as $\frac{\partial(W^*-W_F^*)}{\partial z} = \left(\frac{\partial W^*}{\partial g}\frac{\partial g}{\partial z} - \frac{\partial W_F^*}{\partial p}\frac{\partial p}{\partial z}\right)$. From the latter, given that $\frac{\partial g}{\partial z} = 0$ and given (54), we obtain $\frac{\partial(W^*-W_F^*)}{\partial z} = -k\frac{p^{\frac{p}{1-p}}}{1-p}\frac{\partial p}{\partial z}\ln p$. From the definition (49) we have that $\frac{\partial p}{\partial z} = -g < 0$, and, since $0 < p < 1$, we have that $-k\frac{p^{\frac{p}{1-p}}}{1-p}\frac{\partial p}{\partial z}\ln p < 0$. ■

8.3.2. Proof of Proposition 3

We prove here proposition 3. Formally, we need to prove that

$$\frac{\partial L^*}{\partial x} > 0\,, \tag{55}$$

$$\frac{\partial L^*}{\partial z} > 0\,, \tag{56}$$

where $L^* = \left(\frac{2-z}{1+x}\right)^{\frac{1-z}{x-1+z}}$. In what follows we define $q \equiv \left(\frac{2-z}{1+x}\right)^{\frac{1}{x-1+z}}$ and we use the previous definitions $p \equiv \frac{2-z}{1+x}$ and $b(y) = \ln y + 1 - y$. Note that, by definition, $0 < q < 1$ and moreover $q = p^{\frac{1}{(x+1)(1-p)}}$, $\frac{\partial p}{\partial x} = -\frac{2-z}{(1+x)^2} < 0$, and $\frac{\partial p}{\partial z} = -\frac{1}{1+x} < 0$.

Proof of (55). Given the equation for L^*, we simply have to prove that $\frac{\partial q}{\partial x} > 0$. Consider now $\frac{\partial q}{\partial x}$: its sign is the same as that of $\frac{\partial \log q}{\partial x}$, with $\log q = \frac{1}{1+x}\frac{1}{1-p}\log p$. It follows that $\frac{\partial \log q}{\partial x} = \frac{\partial \log q}{\partial q}\frac{\partial q}{\partial x} = \frac{1}{q}\frac{\partial q}{\partial x} = -\frac{1}{(1+x)^2}\frac{1}{1-p}\log p + \frac{1}{1+x}\frac{1}{(1-p)^2}\frac{\partial p}{\partial x}\log p + \frac{1}{1+x}\frac{1}{1-p}\frac{1}{p}\frac{\partial p}{\partial x}$, and through computation we obtain $\frac{\partial \log q}{\partial x} = -\frac{1}{(1+x)^2}\frac{1}{1-p}\log p + \frac{1}{1+x}\frac{1}{(1-p)^2}\frac{\partial p}{\partial x}\log p + \frac{1}{1+x}\frac{1}{1-p}\frac{1}{p}\frac{\partial p}{\partial x} = -\frac{qb(p)}{(1+x)^2(1-p)^2}$, which, given the definitions of q, p, and $b(p)$, is strictly positive, so that $\frac{\partial q}{\partial x} > 0$ and $\frac{\partial L^*}{\partial x} > 0$. ■

Proof of (56). Given the equation for L^* and given that $\frac{2-z}{1+x} < 1$, we simply have to prove that $\frac{\partial q}{\partial z} > 0$. Consider now $\frac{\partial q}{\partial z}$: its sign is the same as that of $\frac{\partial \log q}{\partial z}$, with $\log q = \frac{1}{1+x}\frac{1}{1-p}\log p$. We then have

$$\frac{\partial \log q}{\partial z} = \frac{\partial \log q}{\partial q}\frac{\partial q}{\partial z} = \frac{1}{q}\frac{\partial q}{\partial z} = -\frac{1}{1+x}\frac{\log p}{(1-p)^2}\frac{\partial p}{\partial z} + \frac{1}{1+x}\frac{1}{1-p}\frac{\partial p}{\partial z}$$
$$= \frac{\partial p}{\partial z}\frac{1}{1+x}\frac{1}{(1-p)^2}\left(\log p + \frac{1-p}{p}\right).$$

Given that $p < 1$ strictly implies $\left(\log p + \frac{1-p}{p}\right) < 0$ and that $\frac{\partial p}{\partial z} < 0$, it follows that $\frac{\partial \log q}{\partial z} > 0$ and thus $\frac{\partial q}{\partial z} > 0$. ■

References

Alessie, R., Van Rooij, M., and Lusardi, A. (2011), Financial Literacy and Retirement Preparation in the Netherlands, Journal of Pension Economics and Finance 10, 527–545.

Almenberg, J., and Säve-Söderbergh, J. (2011), Financial Literacy and Retirement Planning in Sweden, Journal of Pension Economics and Finance 10, 585–598.

Anderson, A. (2013), Trading and Under-Diversification, Review of Finance 17, 1699–1741.

Behrman, J. R., Mitchell, O. S., Soo, C. K., and Bravo, D. (2012), How Financial Literacy Affects Household Wealth Accumulation, American Economic Review 102, 300–304.

Bertaut, C. C., and Haliassos, M. (1995), Why Do So Few Hold Stocks?, Economic Journal 105, 1110–1129.

Blundell, R., Dearden, L.,Meghir, C., and Sianesi, B. (1999), Human Capital Investment: The Returns from Education and Training to the Individual, the Firm and the Economy, Fiscal Studies 20, 1–23.

Blundell, R., Dearden, L., Goodman, A., and Reed, H. (2000), The Returns to Higher Education in Britain: Evidence from a British Cohort, Economic Journal 110, 82–99.

Board of Governors of the Federal Reserve System (US) (2016a), 3-Month Treasury Bill: Secondary Market Rate [TB3MS], retrieved from FRED, Federal Reserve Bank of St. Louis, https://fred.stlouisfed.org/series/TB3MS (Access Date: 2016-07-26).

Board of Governors of the Federal Reserve System (US) (2016b), 10-Year Treasury Constant Maturity Rate [DGS10], retrieved from FRED, Federal Reserve Bank of St. Louis, https://fred.stlouisfed.org/series/DGS10 (Access Date: 2016-07-26).

Borghans, L., Golsteyn, B., Heckman, J., and Meijers, H. (2009), Gender Differences in Risk Aversion and Ambiguity Aversion, Journal of the European Economic Association 7, 649–658.

Barber, B. M., and Odean, T. (2001), Boys Will Be Boys: Gender, Overconfidence, and Common Stock Investment, Quarterly Journal of Economics 116, 261–292.

Bucher-Koenen, T., and Lusardi, A. (2011), Financial Literacy and Retirement Planning in Germany, Journal of Pension Economics and Finance 10, 565–558.

Christiansen, C., Joensen, J. S., and Rangvid, J. (2008), Are Economists More Likely to Hold Stocks?, Review of Finance 12, 465–496.

Corsini, L., and Spataro, L. (2015), Optimal Decisions on Pension Plans in the Presence of Information Costs and Financial Literacy, Journal of Public Economic Theory 17, 383–414.

Croson, R. and Gneezy, U. (2009), Gender Differences in Preferences, Journal of Economic Literature 47, 448–474.

Crossan, D., Feslier, D., and Hurnard, R. (2011), Financial Literacy and Retirement Planning in New Zealand, Journal of Pension Economics and Finance 10, 619–635.

Crawford, M., Church, J., and Akin, B. (2016), CPI Detailed Report, Data for June 2016, Bureau of Labor Statistics.

Dearden, L. (1999), The Effects of Families and Ability on Men's Education and Earnings in Britain, Labour Economics 6, 551–567.

Ehrlich, I., Hamlen W., and Yin, Y. (2008), Asset Management, Human Capital, and the Market for Risky Assets, Journal of Human Capital 2, 217–262.

Ehrlich, I., and Shin, J. (2010), Human Capital and Imperfectly Informed Financial Markets, American Economic Review 100, 244–249.

Ehrlich, I., Shin, J., and Yin, Y. (2011), Private Information, Human Capital, and Optimal "Home Bias" in Financial Markets, Journal of Human Capital 5, 255–301.

Fornero, E., and Monticone, C. (2011), Financial Literacy and Pension Plan Participation in Italy, Journal of Pension Economics and Finance 10, 547–564.

Guiso, L., and Jappelli, T. (2009), Financial Literacy and Portfolio Diversification, CSEF Working Papers 212.

Jappelli, T. (2010), Economic Literacy: An International Comparison, Economic Journal 120, 429–451.

Jappelli, T., and Padula, M. (2013), Investment in Financial Literacy and Saving Decisions, Journal of Banking and Finance 37, 2779–2792.

Klapper, L., and Panos, G. A. (2011), Financial Literacy and Retirement Planning: The Russian Case, Journal of Pension Economics and Finance 10, 599–618.

Lusardi, A., and Mitchell, O. S. (2011a), Financial Literacy Around the World: An Overview, Journal of Pension Economics and Finance 10, 497–508.

Lusardi, A., and Mitchell, O. S. (2011b), Financial Literacy and Retirement Planning in the United States, Journal of Pension Economics and Finance 10, 509–525.

Lusardi A., Michaud, P., and Mitchell, O. S. (2011), Optimal Financial Literacy and Saving for Retirement, Working Papers 905, RAND Corporation Publications Department.

Malmendier, U., and Nagel, S. (2011), Depression babies: Do Macroeconomic Experiences Affect Risk-Taking?, Quarterly Journal of Economics 126, 373–416.

Makarov, D., and Schornick, A. V. (2010), A Note on Wealth Effect under CARA Utility, Finance Research Letters 7, 170–177.

Mas-Colell, A., Whinsthon, M. D., and Green, J. R. (1995), Microeconomic Theory, Oxford University Press, Oxford.

Mitchell, O. S., and Lusardi, A. (2011), Financial Literacy: Implications for Retirement Security and the Financial Marketplace, Oxford University Press, Oxford.

OECD (2014), PISA 2012, Results: Students and Money (Volume VI): Financial Literacy Skills for the 21st Century, OECD Publishing, Paris.

OECD (2016), Financial Education in Europe: Trends and Recent Developments, OECD Publishing, Paris.

Powell, M., and Ansic, D. (1997), Gender Differences in Risk Behaviour in Financial Decision-Making: An Experimental Analysis, Journal of Economic Psychology 18, 605–628.

Renneboog, L., and Spaenjers, C. (2012), Religion, Economic Attitudes, and Household Finance, Oxford Economic Papers 64, 103–127.

Sekita, S. (2011), Financial Literacy and Retirement Planning in Japan, Journal of Pension Economics and Finance 10, 637–656.

Thomas, A., and Spataro, L. (2015), Financial Literacy, Human Capital and Stock Market Participation in Europe: An Empirical Exercise under Endogenous Framework, Department of Economics and Management, University of Pisa, Discussion Paper No. 2015/194.

Van Rooij, M., Lusardi, A., and Alessie, R. (2011), Financial Literacy and Stock Market Participation, Journal of Financial Economics 101, 449–472.

Van Rooij, M., Lusardi, A., and Alessie, R. (2012), Financial Literacy, Retirement Planning and Household Wealth, Economic Journal 122, 449–478.

Varian, H. R. (1992), Microeconomic Analysis, Norton & Company, third edition, New York.

Vissing-Jorgensen, A. (2002), Limited Asset Market Participation and the Elasticity of Intertemporal Substitution, Journal of Political Economy 110, 825–853.

Aging, Pensions, and Growth

Tetsuo Ono*

Received 27 September 2015; in revised form 27 September 2016; accepted 10 October 2016

This study presents an endogenous-growth overlapping-generations model featuring probabilistic voting over public pensions. The analysis shows that (i) the pension–GDP ratio increases as life expectancy increases in the presence of an annuity market, while it may show a hump-shaped pattern in its absence; (ii) the growth rate is higher in the presence of an annuity market than in its absence, but its presence implies an intergenerational trade-off in terms of utility.

Keywords: economic growth, population aging, probabilistic voting, public pensions, annuity market

JEL classification: D 70, E 24, H 55

1. Introduction

Many Organisation for Economic Co-operation and Development (OECD) countries have experienced declining population growth rates and increasing life expectancy over the past decades (OECD, 2011). This demographic change raises the share of the elderly in the population, which is expected to strengthen their political power in voting. Therefore, government spending for the elderly, such as on public pensions and long-term care, is likely to increase. One of the expected side effects of this trend is an increase in the tax burden on the young, which may result in a declining growth rate over time.

The prediction about the effect of population decline on pensions is in line with observations in OECD countries. Panel (a) of figure 1 suggests that the public-pension spending-GDP ratio is positively correlated with the declining population growth rate. On the other hand, the aforementioned prediction on the effect of higher life expectancy is not likely to fit the observation. In panel (b) of figure 1, the pension-GDP ratio shows a weak positive

* Osaka University, 1–7, Machikaneyama, Toyonaka, Osaka 560-0043, Japan (tono@econ.osaka-u.ac.jp). This paper is a substantially revised version of an earlier paper entitled "Economic Growth and the Politics of Intergenerational Redistribution." This work was supported in part by a Grant-in-Aid for Scientific Research (C) from the Ministry of Education, Science and Culture of Japan No. 15K03509. The author would like to thank Yuki Uchida for his research assistance.

doi: 10.1628/001522117X14877521353519
ISSN 0015-2218

correlation with higher life expectancy. However, France and Italy show more than five times higher ratios than Iceland, Israel, and Mexico, although they all share similar life expectancy. Therefore, the empirical evidence seems to be mixed. The first aim of this study is to present a political-economy theory that explains the diverging evidence observed in figure 1.

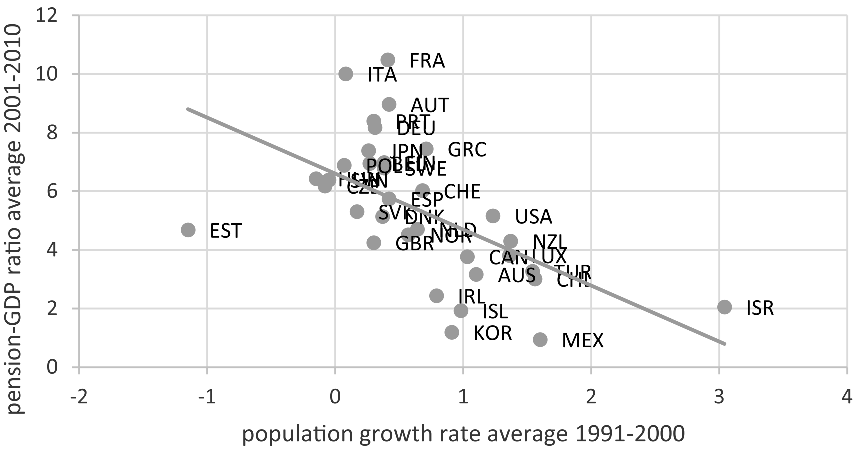
The second aim is to provide a theory that fits the evidence on aging and economic growth. The aforementioned argument suggests that lower population growth has a negative effect on economic growth because of the increasing pension burden. However, the evidence in panel (a) of figure 2 shows that lower population growth is associated with higher per capita GDP growth. In addition, higher life expectancy is associated with lower per capita GDP growth, as depicted in panel (b) of figure 2. The two aging factors have opposite implications for economic growth. The present study demonstrates a model that explains these contradictory results.

To fulfill the aims of this study, we develop an overlapping-generations model with individuals who live for a maximum of two periods, youth and old age, and competitive firms endowed with AK technology, as in Romer (1986). Government spending financed by a tax on the young includes public pensions that benefit the elderly. Within this framework, we consider and compare the following two cases. The first is a perfect annuity case in which an agent's wealth is annuitized and transferred to the other agents, who live throughout old age, if he or she dies young (Sheshinski and Weiss, 1981). The second is a case of no annuity, in which agents' unannuitized wealth is bequeathed to children as an unintentional bequest (Abel, 1985).

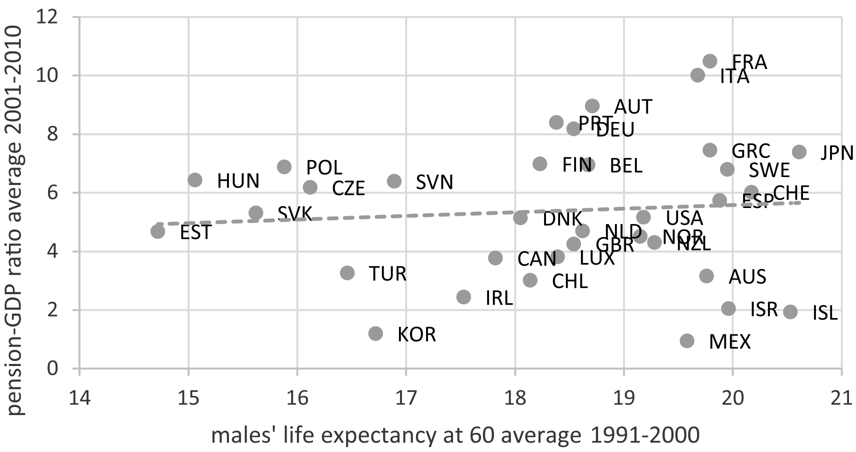
Within the abovementioned framework, we demonstrate the conflict of interest between generations by assuming probabilistic voting à la Lindbeck and Weibull (1987), where the government's objective is to maximize the weighted sum of the utility of the young and elderly. We employ a Markov strategy in which the policy variable of pensions is conditioned on a payoff-relevant state variable, namely, the beginning-of-period capital in the present framework. This implies that the expected level of public pensions in the next period depends on the next-period stock of capital, which is affected by policy decisions in the current period. Forward-looking individuals consider this intertemporal effect when they vote.

We characterize a political equilibrium of probabilistic voting, and obtain the following three results. First, diverging evidence on pensions in figure 1 could be explained by focusing on the presence or absence of an annuity market. The model analysis shows that the pension-GDP ratio increases as life expectancy increases in the presence of an annuity market. In the absence of an annuity market, however, an additional effect appears. Agents increase savings as life expectancy increases. Given that savings are perfect substitutes

Figure 1
Relationship between Pension-GDP Ratio and Population Growth and Life Expectancy



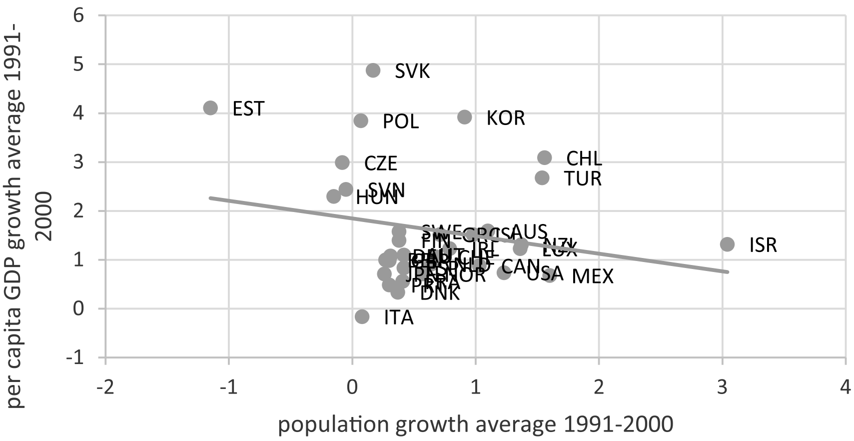
(a) Population growth and pension-GDP ratio



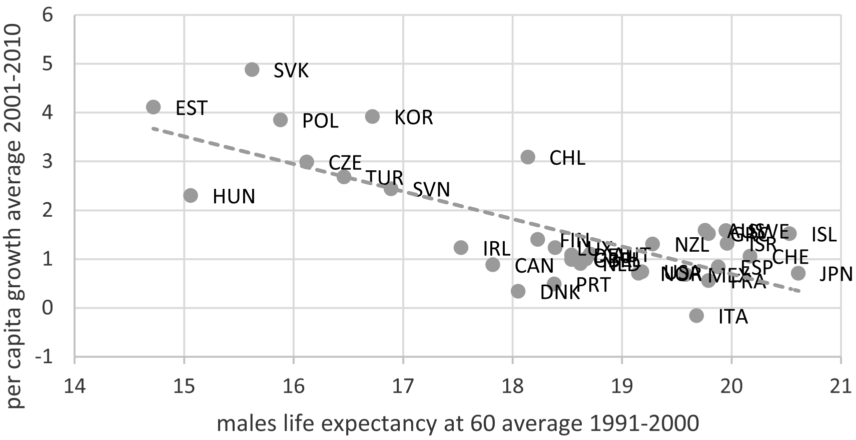
(b) Life expectancy and pension-GDP ratio

Note: Panel (a): The relationship between average pension-GDP ratio, 2001–2010, and average population growth rate, 1991–2000, for OECD countries. Panel (b): The relationship between average pension-GDP ratio, 2001–2010, and average male life expectancy at 60 years, 1991–2000, for OECD countries. Source: OECD.Stat (http://stats.oecd.org/). The original data of OECD show that the population growth rate in Germany in 1991 is 26.4%. This figure would reflect the unification of Germany. We eliminate this and compute the average population growth rate in Germany from 1992 to 2000.

Figure 2
Relationship between per Capita Growth and Population Growth and Life Expectancy



(c) Population growth and per capita GDP growth



(d) Life expectancy and per capita GDP growth

Note: Panel (a): The relationship between average per capita growth, 2001–2010, and average population growth rate, 1991–2000, for OECD countries. Panel (b): The relationship between average per capita growth, 2001–2010, and average male life expectancy at 60 years, 1991–2000, for OECD countries. Source: OECD.Stat (http://stats.oecd.org/).

with the present value of pensions, the politician is induced to offer lower pension benefits as life expectancy increases. This negative effect, peculiar to a no-annuity market, may outweigh the positive effect, depending on the political power of the elderly, and produce a ratio with a hump-shaped pattern.

Second, a lower population growth rate leads to a higher per capita growth rate. This result is consistent with the observed evidence in panel (a) of figure 2. However, higher life expectancy is shown to raise the growth rate, mainly because the young are incentivized to save more for their consumption in old age. To explain this contradictory finding, we focus on another aging factor, namely, the political power of the elderly. A rise in their power incentivizes the government to increase the public-pension burden on the young, thereby lowering the growth rate. Because higher life expectancy implies a larger share of the elderly in the population, the political power of the elderly is likely to be strengthened as life expectancy increases (OECD, 2006; Smets, 2012). Therefore, the negative correlation between life expectancy and per capita growth, which is observed for some countries, as depicted in panel (b) of figure 2, could be explained when the effects of life expectancy and the political power of the elderly are examined together.

Third, there is an intergenerational trade-off in terms of utility. The growth analysis shows that the growth rate is higher in the presence of an annuity market than in its absence. A higher growth rate generates more resources available for future generations, which work to improve their utility. However, the presence of an annuity market lowers public pension benefits, which, in turn, works to decrease lifetime consumption. This negative effect is relevant for all generations, while the positive growth effect is relevant for all but the initial generation. Therefore, the presence of an annuity market makes the initial generation worse off and future generations better off.

The present study is related to the literature on the political economy of public pensions. Earlier studies consider the political sustainability of public pensions (see, e.g., Grossman and Helpman, 1998; Cooley and Soares, 1999; Boldrin and Rustichini, 2000; Azariadis and Galasso, 2002; Forni, 2005; Mateos-Planas, 2008). Recently, the focus has changed to the political effect of population aging on pension provision. Examples are Gonzalez-Eiras and Niepelt (2008), Bassetto (2008), Chen and Song (2014), and Lancia and Russo (2016). However, their focus is mainly on a decline in the population growth rate. Increasing life expectancy, which is also a key factor for population aging, is abstracted away from their analyses.

Gonzalez-Eiras and Niepelt's (2012) work is, to the best of our knowledge, the first attempt to investigate the political effect of rising life expectancy on pensions and economic growth.[1] They conduct the analysis under an environment with perfect annuity markets. However, as reported in Rusconi (2008), the extent of annuitization varies across OECD countries, and some countries are still characterized by small annuity markets. Ono and Uchida (2016) overcome this problem by comparing the case of perfect annuity with the case of no annuity, but their analysis is confined to numerical simulation of the growth of human capital. By contrast, the present study, which analytically solves the model in both the presence and absence of annuity markets, shows aging effects consistent with observed data on OECD countries, providing welfare implications of annuities across generations.

The rest of the paper is organized as follows. Section 2 presents the model and characterizes an economic equilibrium. Section 3 characterizes a political equilibrium, investigates the effects of population aging on pensions and economic growth, and evaluates the welfare implications of annuities. Section 4 provides the concluding remarks. Proofs are provided in the appendix.

2. The Model and Economic Equilibrium

Consider an infinite-horizon economy composed of identical agents, perfectly competitive firms, and annuity markets. A new generation, called generation t, is born in each period $t = 0, 1, 2, ...$ Generation t is composed of a continuum of $N_t > 0$ identical agents. We assume that $N_t = (1+n)N_{t-1}$, that is, the net rate of population growth is $n > -1$.

2.1. Preferences and Utility Maximization

Agents live for a maximum of two periods, youth and old age. An agent dies at the end of youth with a probability of $1 - p \in (0, 1)$. The probability p represents life expectancy or longevity; both terms are used interchangeably in the following sections.[2] If an agent dies young, his or her annuitized wealth is transferred to the other agents, who live throughout old age, and his or her unannuitized wealth is unintentionally bequeathed to the children.

In youth, each agent is endowed with one unit of labor, which is supplied inelastically to firms, and each agent obtains wages. An agent in generation t divides wage w_t between own current consumption c_t^y; savings held as an

1 In addition, Zhang et al. (2003), Gradstein and Kaganovich (2004), and Kunze (2014) analyze the political effect of increasing life expectancy on economic growth. However, pensions are abstracted from their analyses. Their policy focus is on public education as an engine of economic growth.

2 We should note that the data on life expectancy in figures 1 and 2 present conditional life expectancy at age 60, while the parameter p represents the probability of being alive after retirement. The parameter p does not necessarily correspond to conditional life expectancy. However, we approximate life expectancy by using p for tractability of analysis.

annuity and invested into physical capital for consumption in old age, s_t; and tax payments as a proportion of his or her wage, $\tau_t w_t$, where τ_t is the period-t pension contribution rate. Thus, the budget constraint for a young agent in period t is

$$c_t^y + s_t \leq (1-\tau_t)w_t + u_t\,,$$

where u_t is the per capita unintentional bequest from generation $t-1$ to generation t.

If an agent is alive in old age, he or she consumes the returns from savings plus the public-pension benefit. The budget constraint for generation t in old age is

$$c_{t+1}^o \leq (R_{t+1} + \alpha_{t+1})s_t + b_{t+1}\,,$$

where c_{t+1}^o is consumption in old age and b_{t+1} is the public pension benefit. The return on savings is stated as the sum of the return of direct holdings of capital, R_{t+1}, and the return from annuity, α_{t+1}.

Let $\gamma \in \{0, 1\}$ denote the degree of annuitization in the economy. In particular, $\gamma = 0$ ($\gamma = 1$) implies the absence (presence) of annuity markets. If $\gamma = 0$, the unannuitized portion of an agent's wealth is distributed to his or her heirs as unintentional bequests. However, if $\gamma = 1$, the annuitized wealth is transferred via annuity markets to the agents who live throughout old age. Therefore, u_{t+1} satisfies $N_{t+1}u_{t+1} = N_t(1-\gamma)(1-p)R_{t+1}s_t$, or

$$u_{t+1} = \begin{cases} (1-p)R_{t+1}s_t/(1+n) & \text{if } \ \gamma = 0, \\ 0 & \text{if } \ \gamma = 1\,. \end{cases} \tag{1}$$

The return from annuity, α_{t+1}, satisfies $p\alpha_{t+1}s_t = \gamma(1-p)R_{t+1}s_t$, where the left-hand side denotes the aggregate payments to the agents who are alive in old age, while the right-hand side denotes return from annuity. Thus, α_{t+1} is given by

$$\alpha_{t+1} = \begin{cases} 0 & \text{if } \ \gamma = 0, \\ \frac{R_{t+1}}{p} & \text{if } \ \gamma = 1\,. \end{cases} \tag{2}$$

We focus on the two extreme cases, $\gamma = 0$ and 1, to demonstrate the role of annuity markets in a tractable way.

Agents consume private goods. We assume additively separable logarithmic preferences to obtain a closed-form solution. The utility of a young agent in period t is written as $\ln c_t^y + p\beta \ln c_{t+1}^o$, where $\beta \in (0, 1)$ is a discount factor. Thus, the expected-utility-maximization problem for a period-t young agent can be written as

$$\max_{\{c_t^y, c_{t+1}^o\}} \ln c_t^y + p\beta \cdot \ln c_{t+1}^o$$
$$\text{s.t. } c_t^y + s_t \leq (1-\tau_t)w_t + u_t,$$
$$c_{t+1}^o \leq (R_{t+1} + \alpha_{t+1})s_t + b_{t+1}$$
$$\text{given } \tau_t, w_t,\ b_{t+1},\ \alpha_{t+1}, \text{ and } R_{t+1}\,.$$

Solving the problem leads to the following consumption and saving functions:

$$c_t^y = \frac{1}{1+p\beta}\left[(1-\tau_t)w_t + u_t + \frac{b_{t+1}}{R_{t+1}+\alpha_{t+1}}\right],$$
$$c_{t+1}^o = \frac{p\beta(R_{t+1}+\alpha_{t+1})}{1+p\beta}\left[(1-\tau_t)w_t + u_t + \frac{b_{t+1}}{R_{t+1}+\alpha_{t+1}}\right],$$
$$s_t = \frac{p\beta}{1+p\beta}\left[(1-\tau_t)w_t + u_t - \frac{b_{t+1}}{p\beta(R_{t+1}+\alpha_{t+1})}\right].$$

In period 0, there are young agents in generation 0 and initial elderly agents in generation -1. Each agent in generation -1 is endowed with s_{-1} units of goods, earns return $R_0 s_{-1}$ plus pension benefit b_0, and consumes them. The initial elderly agents' measure is pN_{-1}. The utility of an agent in generation -1 is $(1-\theta)\ln c_0^o + \theta \ln g_0$.

2.2. Technology and Profit Maximization

There is a continuum of identical, perfectly competitive, profit-maximizing firms that produce output with a constant-returns-to-scale Cobb–Douglas production function, $Y_t = A_t(K_t)^\alpha(N_t)^{1-\alpha}$, where Y_t is aggregate output, A_t is the productivity parameter, K_t is aggregate capital, N_t is aggregate labor, and $\alpha \in (0,1)$ is a constant parameter representing capital share. The productivity parameter is assumed to be proportional to the aggregate capital per labor unit in the overall economy, that is, $A_t = A(K_t/N_t)^{1-\alpha}$. Capital investment thus involves a type of technological externality often used in theories of endogenous growth (see, e.g., Romer, 1986). Capital is assumed to fully depreciate within a period.

In each period t, a firm chooses capital and labor to maximize its profits, $\Pi_t = A_t(K_t)^\alpha(N_t)^{1-\alpha} - R_t K_t - w_t N_t$, where R_t is the rental price of capital and w_t is the wage rate. The firm takes these prices as given. The first-order conditions for profit maximization are given by

$$K_t : R_t = \alpha A_t(K_t)^{\alpha-1}(N_t)^{1-\alpha},$$
$$N_t : w_t = (1-\alpha)A_t(K_t)^\alpha(N_t)^{-\alpha}\,.$$

2.3. Government Budget Constraints

The government budget for pensions is assumed balanced in each period. Fiscal policy is determined through elections. A period-t budget constraint on pensions is $N_t \tau_t w_t = pN_{t-1} b_t$. Dividing both sides of the constraint by N_t, we obtain the per capita form of the government budget constraint:

$$\tau_t w_t = \frac{p}{1+n} b_t \,.$$

2.4. Economic Equilibrium

The market-clearing condition for capital is $K_{t+1} = N_t s_t$, which expresses the equality of total savings by young agents in generation t, $N_t s_t$, to the stock of aggregate physical capital. Dividing both sides by N_t leads to

$$(1+n)k_{t+1} = s_t \,,$$

where $k_t \equiv K_t / N_t$ is per capita capital.

Definition 1 An *economic equilibrium* is a sequence of prices, $\{w_t, R_t, \alpha_t\}_{t=0}^{\infty}$, allocations, $\{c_t^y, c_t^o, s_t, u_t\}_{t=0}^{\infty}$, capital stock $\{k_t\}_{t=0}^{\infty}$ with the initial condition $k_0 (> 0)$, and policies $\{\tau_t, b_t\}_{t=0}^{\infty}$, such that (i) utility is maximized with the budget constraints in youth and old age, (ii) profit is maximized, (iii) the government budget is constrained, and (iv) the annuity and capital markets clear.

Assuming productive externality, $A_t = A(K_t/N_t)^{1-\alpha}$, the first-order conditions for profit maximization are

$$R_t = R \equiv \alpha A \quad \text{and} \quad w_t = (1-\alpha) A k_t \,.$$

Using the saving function and the first-order conditions for profit maximization, we rewrite the capital-market-clearing condition as

$$(1+n)k_{t+1} = \frac{p\beta}{1+p\beta} \cdot \left[(1-\alpha)Ak_t - \frac{p}{1+n} b_t + (1-\gamma)(1-p)Rk_t - \frac{b_{t+1}}{p\beta\left(1+\frac{\gamma(1-p)}{p}\right)R} \right], \tag{3}$$

where $\gamma = 1 (= 0)$ if the annuity market is present (absent).

In an economic equilibrium, the indirect utility of a young agent in period t, V_t^y, and that of an elderly agent alive in period t, V_t^o, can be expressed as functions of government policy and capital stock:

$$V_t^y = (1+p\beta) \cdot \ln\left[(1-\tau_t)(1-\alpha)Ak_t + (1-\gamma)(1-p)Rk_t + \frac{b_{t+1}}{\left(1+\frac{\gamma(1-p)}{p}\right)R}\right] + \ln\frac{1}{1+p\beta} + p\beta\ln\frac{p\beta\left(1+\frac{\gamma(1-p)}{p}\right)R}{1+p\beta},$$

$$V_t^o = \ln\left(\left(1+\frac{\gamma(1-p)}{p}\right)R(1+n)k_t + b_t\right).$$

The first term of the young agent's indirect utility function corresponds to the utility of consumption in youth and old age, and the term of the elderly agent's indirect utility corresponds to the utility of consumption.

3. Political Equilibrium

This study assumes probabilistic voting to demonstrate the political mechanism. In each period, the government in power maximizes a political objective function. Formally, the political objective function in each period t is given by

$$\Omega_t = \omega p V_t^o + (1+n)V_t^y,$$

where ωp and $1+n$ are the relative weights of elderly and young agents, respectively. In particular, the parameter $\omega(>0)$ represents the political power of the elderly, which reflects the recent age gap in voter turnout in developed countries (OECD, 2006; Smets, 2012). The government's problem in period t is to maximize Ω_t subject to its budget constraints, given the state variable, k_t.[3]

The scope of this study is restricted to a stationary Markov-perfect equilibrium. Markov perfectness implies that outcomes depend only on the payoff-relevant state variable, that is, capital k. The stationary property implies that our focus is on equilibrium policy rules that do not depend on time. Therefore, the expected level of public pensions for the next period, b_{t+1}, is given by a function of the next period's stock of capital, $b_{t+1} = B(k_{t+1})$. Using recursive notation with x' denoting the next period x, we can define a stationary Markov-perfect political equilibrium in the present framework as follows.

3 An explicit micro foundation for this modeling is explained in Persson and Tabellini (2000, Chapter 3) and Acemoglu and Robinson (2005, Appendix). Song et al. (2012) outline the process to derive the political objective function under probabilistic voting in the framework of overlapping generations.

Definition 2 A *stationary Markov-perfect political equilibrium* is a set of functions, $\langle T, B \rangle$, where $T : \Re_{++} \to [0, 1]$ is a pension contribution rule, $\tau = T(k)$, and $B : \Re_{++} \to \Re_{+}$ is a public-pension rule, $b = B(k)$, such that the following conditions hold:

(i) the capital market clears:

$$(1+n)k' = \frac{p\beta}{1+p\beta} \cdot \left[(1 - T(k)) \cdot (1-\alpha)Ak + (1-\gamma)(1-p)Rk - \frac{B(k')}{p\beta\left(1+\frac{\gamma(1-p)}{p}\right)R} \right]; \tag{4}$$

(ii) $\langle T(k), B(k) \rangle = \arg\max \Omega(k, b, \tau, b')$ is subject to the expectation of future pensions, $b = B(k')$, the capital-market-clearing condition in (4), and the government budget constraint,

$$T(k)(1-\alpha)Ak = \frac{p}{1+n} \cdot B(k),$$

with a nonnegativity constraint, $b \geq 0$, where $\Omega(k, b, \tau, b')$ is defined by

$$\Omega(\cdot) \equiv \omega p \ln\left\{ \left(1 + \frac{\gamma(1-p)}{p}\right) R(1+n)k + b \right\} + (1+n)(1+p\beta)\ln\left\{ (1-\tau)(1-\alpha)Ak + (1-\gamma)(1-p)Rk + \frac{b'}{\left(1+\frac{\gamma(1-p)}{p}\right)R} \right\},$$

where irrelevant terms are omitted from the expression for $\Omega(\cdot)$.

The first condition describes agents' response to the government's choice of a tax, $T(k)$, under the expectation that future pensions will be set according to the rule $b' = B(k')$. The second condition states that the government chooses its fiscal policy to maximize its objective, subject to the government budget constraint, the capital-market-clearing condition representing the agents' response to the fiscal policy choice, and their expectation of future pensions, $b' = B(k')$. The solution to the government problem constitutes a stationary Markov-perfect equilibrium if $b = B(k)$. For example, suppose the expectation is given by $b' = \tilde{B} \cdot k'$, and the solution is $b = \hat{B} \cdot k$, where $\tilde{B}$ and $\hat{B}$ are constant. Then, the solution is said to constitute a stationary Markov-perfect equilibrium if $\tilde{B} = \hat{B}$.

3.1. Characterization of Political Equilibrium

Given that preferences are specified by the logarithmic utility function, we assume a linear policy function of public pensions for the next period, $B(k') = B_\gamma \cdot Ak'$, where $B_\gamma (> 0)$, $\gamma \in \{0, 1\}$, is a constant parameter. In this scenario, we solve the problem and determine the political equilibrium outcome as follows:

Proposition 1 There is a stationary Markov-perfect political equilibrium with $b > 0$ if

$$\gamma = 1 \text{ and } \alpha < \frac{1}{1 + (1 + p\beta)(1 + n)/\omega p}$$

or

$$\gamma = 0 \text{ and } \alpha < \frac{1}{p + (1 + p\beta)(1 + n)/\omega}.$$

For the case of $b > 0$ the policy function of public pensions is

$$B_\gamma Ak = \begin{cases} B_1 Ak & \text{if} \quad \gamma = 1, \\ B_0 Ak & \text{if} \quad \gamma = 0, \end{cases}$$

where

$$B_1 \equiv \frac{\omega(1 - \alpha) - (1 + p\beta)\frac{\alpha}{p}(1 + n)}{(1 + p\beta) + \frac{\omega p}{1+n}},$$

$$B_0 \equiv \frac{\omega\{(1 - \alpha) + (1 - p)\alpha\} - (1 + p\beta)\alpha(1 + n)}{(1 + p\beta) + \frac{\omega p}{1+n}} (> B_1).$$

Proof. See appendix 5.1.

The result in proposition 1 implies that for both cases of $\gamma = 1$ and 0, public pensions are more likely to be provided in the political equilibrium if the political power of the elderly (ω) is larger and/or the population growth rate (n) lower. Greater political power for the elderly attaches a larger weight to the utility of the elderly in the political objective function. This incentivizes the politician to provide higher pension benefits to the elderly. A smaller population growth rate implies lower savings per head for a given capital stock and, thus, lower consumption level for the elderly. To maintain their consumption level, the politician offers a larger pension benefit to the elderly.[4]

The effects of longevity (p) on the provision of public pensions differ between the two cases. In the presence of annuity markets ($\gamma = 1$), greater longevity has two opposing effects on pension provision. First, greater longevity leads to a lower rate of return from annuity and, thus, a lower consumption level in old age. To compensate for this loss of old-age consumption, the politician offers a larger pension benefit to the elderly. This is a positive effect on the public pension represented by the term α/p in the numerator of B_1. Second, greater longevity attaches a larger weight to the lifetime utility of the young. To improve their utility, the politician cuts the tax burden on the young by reducing public-pension provisions for the elderly. This is

4 There are additional effects on pension benefits. A lower population growth rate reduces the weight of the young and greater longevity increases the weight of the elderly in the political objective function. These work to increase the pension benefits for the elderly. However, this positive effect on pensions is offset by an increase in the replacement ratio, $p/(1 + n)$.

a negative effect on the public pension represented by the term $p\beta$ on the numerator of B_1. The condition $\alpha < [1 + (1 + p\beta)(1 + n)/\omega p]^{-1}$ implies that the former positive effect outweighs the latter negative effect when $\gamma = 1$.

In the absence of annuity markets ($\gamma = 0$), the negative effect remains, but the positive effect through the annuity returns disappears. Instead, there is an additional negative effect through the accidental bequest represented by the term $(1 - p)\alpha$ in the numerator of B_0. That is, an increase in longevity decreases the unintentional bequests and thereby creates a negative income effect on the young. This negative effect, accompanied by the negative effect in the absence of annuity markets, incentivizes the politician to cut the tax burden on the young and thereby reduce the pension benefit to the elderly. Therefore, in the absence of annuity markets, public pensions are less likely to be provided if longevity is higher.

3.2. Aging and Pensions

Based on the characterization of the political equilibrium, we now examine how the aging factors n, ω, and p affect public-pension provision.

Proposition 2 Consider an equilibrium with $b > 0$ demonstrated in proposition 1.

(i) The pension-GDP ratio increases with a lower population growth rate and greater political power for the elderly.
(ii) *Assume* $\gamma = 1$. The pension-GDP ratio increases with greater longevity.
(iii) Assume $\gamma = 0$. With greater longevity, the pension–GDP ratio increases if $\alpha \le \{\omega/(1+n)\}/\{(1+\beta) + \omega/(1+n)\}^2$, and shows a hump-shaped pattern otherwise.

Proof. See appendix 5.2.

To confirm proposition 2, we first compute the pension-GDP ratio, $pN_{t-1}b_t/Y_t = (p/(1+n)) \cdot (b/Ak)$, as follows:

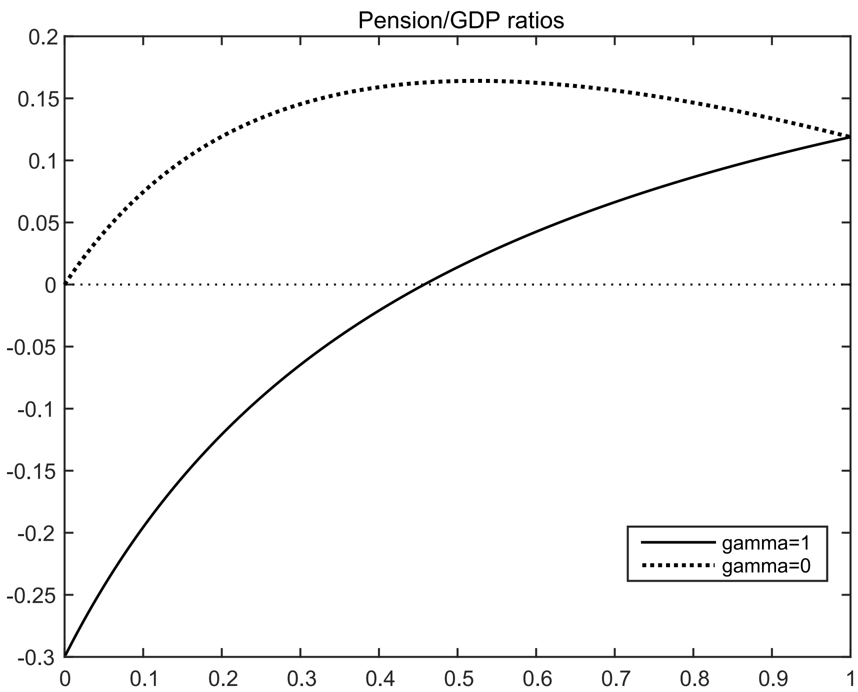
$$\frac{p}{1+n} \cdot \frac{b}{Ak} = \begin{cases} \dfrac{1}{\frac{1+p\beta}{\omega} \cdot \frac{1+n}{p} + 1} - \alpha & \text{if } \gamma = 1, \\ \dfrac{1}{\frac{1+p\beta}{\omega} \cdot \frac{1+n}{p} + 1} - \alpha p & \text{if } \gamma = 0 \,. \end{cases}$$

A lower population growth rate implies a smaller weight to the utility of the young, whereas greater political power for the elderly implies a larger weight to the utility of the elderly. This puts less value on the cost of public pensions for the young and more value on the benefit for the elderly. Therefore, the politician is incentivized to increase public-pension provision in response to a decrease in n and an increase in ω.

When $\gamma = 1$, the ratio is affected by greater longevity in two ways. First, greater longevity implies that the current young attach a larger weight to their utility of consumption in old age. They prefer saving for their own future consumption to public-pension spending on the currently elderly. This preference of the young incentivizes the government to cut current public-pension spending. Therefore, greater longevity has a negative effect on the pension-GDP ratio, as observed in the term $p\beta$.

Second, greater longevity leads to a higher dependency ratio, and thereby a higher pension-GDP ratio. This positive effect is observed in the term $(1+n)/p$. Therefore, greater longevity has two opposing effects on the ratio, but the result in proposition 1(ii) shows that the latter always outweighs the former in the political equilibrium if $\gamma = 1$. This result is consistent with the prediction of Gonzalez-Eiras and Niepelt's (2008) model based on a neoclassical growth framework. The present analysis indicates that their prediction also holds in an endogenous growth model with AK technology.

Figure 3
Effects of Longevity on Pension-GDP Ratio



Note: We assume each period lasts 30 years. Parameters are set at $\alpha = 0.3, \omega = 1.5, A = 8.8, \beta = (0.99)^{30}$ and $1+n = (1.006)^{30}$.

The solid curve in figure 3 illustrates a numerical example of the ratio when $\gamma = 1$.

When $\gamma = 0$, there is an additional negative effect represented by the term αp. Greater longevity implies that more weight is attached to the return from savings; agents increase savings as longevity increases. The politician considers this economic behavior in policy decision-making and, given that savings are perfectly substitutable with pensions, offers lower pension benefits as longevity increases. When $\gamma = 1$, this negative effect on pensions is offset by the decrease in the return from annuity, $\tilde{R} = R/p$. However, there is no such cancellation effect when $\gamma = 0$.

The term αp, representing the aforementioned negative effect, becomes larger as the interest rate, $R = \alpha A$, increases. In particular, if α is small, such that $\alpha \leq \{\omega/(1+n)\}/\{(1+\beta)+\omega/(1+n)\}^2$, the sum of the negative effects is outweighed by the positive effect; greater longevity leads to a higher pension-GDP ratio. However, if α is above the critical value, an increase in longevity produces an initial increase followed by a decrease in the ratio. The dotted curve in figure 3 illustrates a numerical example of the hump-shaped pattern of the ratio when $\gamma = 0$.

To recap, the pension-GDP ratio increases as longevity rises in the presence of an annuity market. In its absence, the ratio increases or exhibits a hump-shaped pattern as longevity rises. These different effects of longevity could be ascribed to the presence (or absence) of annuities. Therefore, the extent of annuitization may be an important feature in providing an explanation for the mixed evidence on the pension-GDP ratio observed for high-longevity countries in figure 1.

3.3. Aging and Economic Growth

Proposition 1 enables us to derive the growth rate of per capita capital, k'/k, and to investigate how the growth rate is affected by population aging. The following proposition summarizes the result.

Proposition 3 Consider an equilibrium with $b > 0$.

(i) The growth rate of per capita capital is

$$\frac{k'}{k} = \begin{cases} \left.\frac{k'}{k}\right|_{\gamma=1} \equiv \dfrac{(1+p\beta)A}{(1+n)\left\{(1+p\beta)+\frac{\omega p}{1+n}\right\}+\frac{\omega}{\beta p\cdot(\alpha/p)}} & \text{if} \quad \gamma = 1, \\ \left.\frac{k'}{k}\right|_{\gamma=0} \equiv \dfrac{(1+p\beta)A}{(1+n)\left\{(1+p\beta)+\frac{\omega p}{1+n}\right\}+\frac{\omega}{\beta p\cdot\alpha}} & \text{if} \quad \gamma = 0. \end{cases}$$

The growth rate is higher in the presence of an annuity market than in its absence.

(ii) The growth rate increases with a lower population growth rate, less political power for the elderly, and greater longevity: $\partial(k'/k)/\partial n < 0$, $\partial(k'/k)/\partial\omega < 0$, and $\partial(k'/k)/\partial p > 0$.

Proof. See appendix 5.3.

The growth rate of per capita capital is constant over time because the model exhibits a constant interest rate inherited from AK technology. In addition, the presence of an annuity market increases the growth rate. To understand the role of the annuity market, recall the capital-market-clearing condition in (3), which we reformulate using $b = B_\gamma Ak$, as

$$\frac{k'}{k} = \frac{\frac{p\beta}{1+p\beta}}{(1+n) + \frac{p\beta}{1+p\beta} \cdot \underbrace{\frac{B_\gamma}{p\beta\left(1 + \frac{\gamma(1-p)}{p}\right)\alpha}}_{(\#1)}} \cdot \left[(1-\alpha) - \underbrace{\frac{p}{1+n}}_{(\#2)} B_\gamma + \underbrace{(1-\gamma)(1-p)\alpha}_{(\#3)}\right] A. \tag{5}$$

The presence of an annuity market affects the growth rate in the following three ways. First, the per capita pension-GDP ratio, B_γ, is lower and the return from savings is higher in the presence of the annuity market than in its absence. This implies a lower present value of pension benefits, which incentivizes individuals to save more for future consumption. This is a positive effect of the annuity market on the growth rate, represented by (#1) in equation (5).

Second, the presence of an annuity market lowers the tax burden for pension provision. This creates a positive income effect on savings and the growth rate, represented by (#2). Finally, the unintentional bequest disappears in the presence of an annuity market. This produces a negative income effect represented by (#3). Overall, the positive effects of (#1) and (#2) dominate the negative effect of (#3). Therefore, the growth rate is higher in the presence of an annuity market than in its absence.

To observe the effect of the population aging factors on economic growth, let us recall the growth rate of per capita capital when $\gamma = 1$,

$$\left.\frac{k'}{k}\right|_{\gamma=1} = A \cdot \left[(1+n) + \frac{1}{1+p\beta}\left(\omega p + \frac{\omega}{\beta p \cdot (\alpha/p)}\right)\right]^{-1}.$$

Thus, the growth rate is affected by population aging via the terms $(1+n)$, $1+p\beta$, ωp, and $\omega/(\beta p \cdot (\alpha/p))$. When $\gamma = 0$, the term $\omega/(\beta p \cdot (\alpha/p))$ is replaced by $\omega/(\beta p\alpha)$.

Population growth, the political power of the old, and longevity have the following implications for economic growth via these terms. First, a lower population growth rate increases per capita capital equipment in the economy. This effect, which is observed in the term $(1+n)$, accelerates capital accumulation and economic growth. Second, greater political power for the elderly attaches a larger weight to the utility of the current elderly. This incentivizes the government to increase public-pension expenditure, resulting in a greater tax burden on the young. This effect, which is observed in the terms ωp and $\omega/(\beta p \cdot (\alpha/p))$, discourages saving by the young and impedes economic growth.

Greater longevity has the following effects on the growth rate via the terms $1+p\beta$, ωp, and $\omega/(\beta p \cdot (\alpha/p))$ when $\gamma = 1$. First, greater longevity attaches a larger weight to the lifetime utility of the young as represented by the term $p\beta$. The young save more for their consumption in old age in response to an increase in longevity. This creates a positive effect on economic growth. Second, greater longevity attaches a larger weight to the utility of the elderly, too. As described earlier in this subsection, this increases the tax burden on the young and, thus, has a negative effect on economic growth.

Third, greater longevity lowers the rate of return from annuity, $\tilde{R} = R/p$ and thus increases the present value of public-pension benefits. In response to this increase, individuals save less because pension benefits are perfect substitutes for savings. This negative effect on economic growth is observed in the term $\omega/(\beta p \cdot (\alpha/p))$ and is peculiar to the case of $\gamma = 1$. Thus far, the three effects imply that longevity has mixed growth effects, but the net effect is positive when $\gamma = 1$. The result is qualitatively unchanged when $\gamma = 0$, because the last negative effect disappears.

The model prediction regarding population growth is consistent with the observation in panel (a) of figure 2, but the prediction regarding longevity is inconsistent with the observation in panel (b); the evidence shows that higher life expectancy results in a lower growth rate. This inconsistency is resolved when we focus on another aging factor, namely, the political power of the elderly. Greater longevity implies a larger share of the elderly in the population, which might, in turn, lead to larger political power for the elderly (OECD, 2006; Smets, 2012). Given the negative effect of the power on economic growth, we might argue that the negative correlation between life expectancy and per capita economic growth observed in panel (b) of figure 2 could be explained when the effects of longevity and the political power of the elderly are examined together.

3.4. Welfare Analysis

The previous subsection shows that the growth rate is higher in the presence of an annuity market than in its absence. A higher growth rate generates more resources available for future generations. This implies that future generations are better off in the presence of an annuity market. However, pension benefits in the presence of an annuity market are lower than in its absence, as shown in proposition 1. This works to decrease the present value of lifetime income and consumption. Hence, the present generation is probably worse off in the presence of an annuity market than in its absence.

In order to investigate such a possibility, we compare the indirect utility functions when $\gamma = 1$ and $\gamma = 0$ for a given k as follows:

$$
\begin{aligned}
V^y|_{\gamma=1} &\gtreqless V^y|_{\gamma=0} \\
&\Leftrightarrow p\beta \ln(1/p) + (1+p\beta)\ln p + (1+p\beta)\ln\left(k'/k\big|_{\gamma=1}\right)^t \\
&\gtreqless (1+p\beta)\ln\left(k'/k\big|_{\gamma=0}\right)^t
\end{aligned}
\tag{6}
$$

From this condition, we obtain the following result.

Proposition 4 Consider an equilibrium with $b > 0$.

(i) For generation 0, the expected utility of the young is higher in the absence of an annuity market than in its presence.

(ii) Along the equilibrium path with $k'/k > 1$, there is some $T(> 1)$ such that the expected utility of generation $t(\geq T)$ is higher in the presence of an annuity market than in its absence.

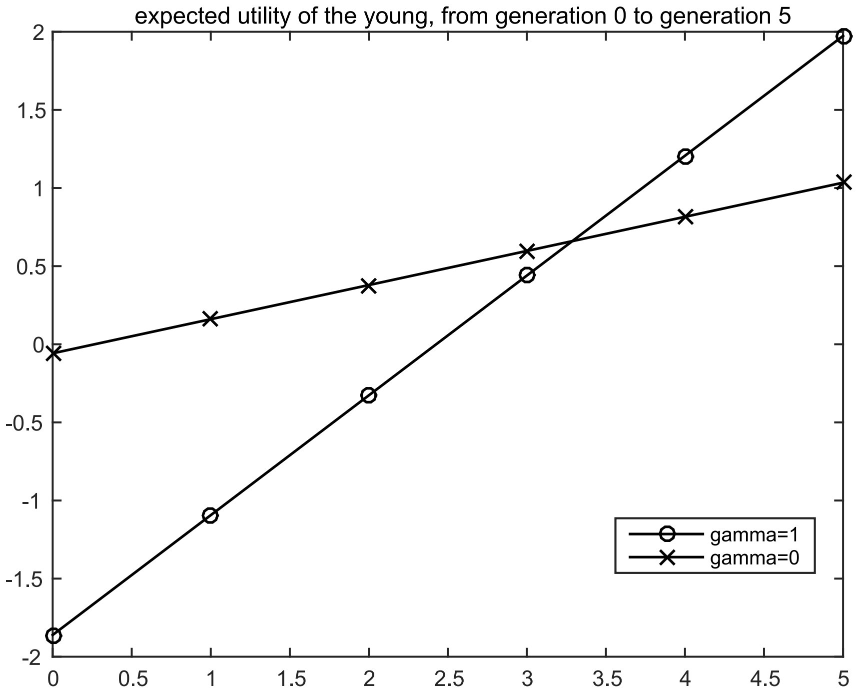
Proof. See appendix 5.4.

To understand the argument in proposition 4, let us look at the condition in (6). The first term on the left-hand side, $p\beta\ln(1/p)$, shows a positive effect for the annuity market. The presence of an annuity market increases the return from saving, which works to increase old-age consumption. However, pension benefits are lower in the presence of an annuity market than in its absence, as demonstrated in proposition 1. This implies that the annuity market has a negative effect on utility, as observed in the second term on the left-hand side, $(1+p\beta)\ln p$. The net result of these two opposing effects through the interest rate is negative.

The third term on the left-hand side, $(1+p\beta)\ln(k'/k|_{\gamma=1})^t$, and the term on the right-hand side, $(1+p\beta)\ln(k'/k|_{\gamma=0})^t$, show the lifetime income (i.e., lifetime consumption), which is affected by the growth rate. Given the result in proposition 3, we find that lifetime income is larger when $\gamma = 1$ than when $\gamma = 0$. The growth effect works from generation 1 onward, and the difference in lifetime income increases over time.

Overall, the presence of an annuity market creates a negative effect through pension benefits and a positive effect through the growth rate. The growth effect is irrelevant for generation 0, which is thus better off in the absence of an annuity market than in its presence. However, the positive growth effect becomes relevant from generation 1 onward and increases

Figure 4
Expected Lifetime Utility of the Young from Generation 0 *to Generation* 5.



Note: Parameters are set at $\alpha = 0.3, \omega = 1.5, A = 8.8, p = 0.7, \beta = (0.99)^{30}$, and $1 + n = (1.006)^{30}$. The initial condition is $k_0 = 1$.

over time. Therefore, there is some $T(> 0)$ such that the expected utility of generation $t(\geq T)$ is higher in the presence of an annuity market than in its absence. A numerical example is demonstrated in figure 4.

Before closing this section, we briefly consider how public-pension provision by itself affects welfare across generations. To examine this issue, compare an economic equilibrium without public pensions and a political equilibrium with public pensions as demonstrated in proposition 1. The elderly in the initial period are better off with public pensions, because they have more resources for consumption. However, public pensions provide individuals less incentive to save, resulting in a lower economic growth rate than that without public pensions. This in turn lowers the income, and therefore the consumption and utility, of future generations. Therefore, public pensions create a utility trade-off between the current and future generations.

4. Concluding Remarks

This study has attempted to examine how an aging population affects voting on pension expenditure, and how this expenditure in turn affects economic growth. To address these issues, we used an endogenous-growth overlapping-generation model in which pension expenditure is financed by a tax on the working young. The expenditure is determined via probabilistic voting that captures the intergenerational conflict caused by the three factors of population aging – a decline in the population growth rate, a rise in life expectancy, and an increase in the political power of the elderly.

We considered two alternative cases, in which an annuity market is either present or absent, and showed the following. First, the pension-GDP ratio increases as life expectancy increases in the presence of an annuity market. However, the ratio may show a hump-shaped pattern in the absence of an annuity market.[5] Second, the growth rate is increased by a lower population growth rate, less political power of the elderly, and greater longevity. However, when the longevity and political power of the elderly are examined together, greater longevity could be associated with a lower growth rate. These results are consistent with the observed evidence in OECD countries.

To evaluate the role of the annuity market, we compared the two cases, the presence and absence of an annuity market, in terms of growth and welfare. We showed that (i) the growth rate is higher in the presence of an annuity market than in its absence; and (ii) due to this growth effect, future generations are better off in the presence of an annuity market, but the current generation cannot benefit from future growth. These results suggest that the development of annuity markets is beneficial from the viewpoint of economic growth, but implies an intergenerational trade-off in terms of utility.

5. Appendix

5.1. Proof of Proposition 1

Suppose that public pensions are provided in equilibrium in the next period, $b' > 0$. Assume a linear policy function of public pensions for the next period, $B(k') = B_\gamma A k'$, where $B_\gamma (> 0)$ is a constant parameter. Given this assumption and the government budget constraint, the capital-market-clearing condition

5 In other words, greater longevity is less likely to increase the pension-GDP ratio in the absence of an annuity market than in its presence. In Ono (2016), it is shown that this property also holds when we alternatively assume a neoclassical production function and allow for endogenous interest rates.

in definition 2 becomes

$$(1+n)k' = \frac{p\beta}{1+p\beta}$$
$$\cdot \left[(1-\alpha)Ak - \frac{p}{1+n}B(k) + (1-\gamma)(1-p)Rk - \frac{B_\gamma A k'}{p\beta\left(1+\frac{\gamma(1-p)}{p}\right)R} \right],$$

which is rewritten as

$$k' = \frac{\frac{p\beta}{1+p\beta}}{(1+n)+\frac{p\beta}{1+p\beta}\cdot\frac{B_\gamma A}{p\beta\left(1+\frac{\gamma(1-p)}{p}\right)R}}$$
$$\cdot \left[(1-\alpha)Ak - \frac{p}{1+n}B(k) + (1-\gamma)(1-p)Rk \right]. \tag{7}$$

Using (7) and the assumption $B(k') = BAk'$, we write the present value of lifetime income as

$$(1-\alpha)Ak - \frac{p}{1+n}B(k) + (1-\gamma)(1-p)Rk + \frac{B_\gamma A}{\left(1+\frac{\gamma(1-p)}{p}\right)R}k'$$
$$= \left[1 + \frac{B_\gamma A}{\left(1+\frac{\gamma(1-p)}{p}\right)R} \cdot \frac{\frac{p\beta}{1+p\beta}}{(1+n)+\frac{p\beta}{1+p\beta}\cdot\frac{B_\gamma A}{p\beta\left(1+\frac{\gamma(1-p)}{p}\right)R}} \right]$$
$$\times \left[(1-\alpha)Ak - \frac{p}{1+n}B(k) + (1-\gamma)(1-p)Rk \right]. \tag{8}$$

Equations (7) and (8) enable us to write the political objective function as follows:

$$\Omega = \omega p \ln\left(\left(1+\frac{\gamma(1-p)}{p}\right) R(1+n)k + B(k) \right)$$
$$+ (1+n)(1+p\beta)\ln\left[(1-\alpha)Ak - \frac{p}{1+n}B(k) + (1-\gamma)(1-p)Rk \right],$$

where the terms unrelated to political decisions are omitted from the expression.

The first-order condition with respect to $B(k)$ is given by

$$B(k): \frac{\omega p}{\left(1+\frac{\gamma(1-p)}{p}\right)R(1+n)k + B(k)}$$
$$= \frac{(1+n)(1+p\beta)\frac{p}{1+n}}{(1-\alpha)Ak - \frac{p}{1+n}B(k) + (1-\gamma)(1-p)Rk}.$$

Recalling that $R = \alpha A$, this expression is reduced to

$$B(k) = \frac{\omega\left\{(1-\alpha)+(1-\gamma)(1-p)\alpha\right\} - (1+p\beta)\left(1+\frac{\gamma(1-p)}{p}\right)\alpha(1+n)}{(1+p\beta)+\frac{\omega p}{1+n}} \cdot Ak\,,$$

or

$$B(k) = \begin{cases} B_{\gamma=1}Ak & \text{if } \gamma = 1, \\ B_{\gamma=0}Ak & \text{if } \gamma = 0\,, \end{cases}$$

where $B_{\gamma=1}$ and $B_{\gamma=0}$ are defined by

$$B_{\gamma=1} \equiv \frac{\omega(1-\alpha)-(1+p\beta)\frac{\alpha}{p}(1+n)}{(1+p\beta)+\frac{\omega p}{1+n}},$$

$$B_{\gamma=0} \equiv \frac{\omega\left\{(1-\alpha)+(1-p)\alpha\right\}-(1+p\beta)\alpha(1+n)}{(1+p\beta)+\frac{\omega p}{1+n}}\,.$$

Therefore, we obtain $b > 0$ if

$$B_{\gamma=1} > 0 \Leftrightarrow \omega > \frac{\alpha(1+p\beta)(1+n)}{(1-\alpha)p} \text{ when } \gamma = 1,$$

$$B_{\gamma=0} > 0 \Leftrightarrow \omega > \frac{\alpha(1+p\beta)(1+n)}{(1-\alpha)+(1-p)\alpha} \text{ when } \gamma = 0\,.$$

■

5.2. Proof of Proposition 2

Using the policy function $B(\cdot)$ in proposition 1, the pension-GDP ratio, $pN_{t-1}b_t/Y_t = (p/(1+n)) \cdot (b/Ak)$, is computed as follows:

$$\frac{p}{1+n}\cdot\frac{b}{Ak} = \begin{cases} \frac{p}{1+n}\cdot\frac{\omega-\alpha\left\{\omega+(1+p\beta)(1+n)/p\right\}}{(1+p\beta)+\omega p/(1+n)} & \text{if } \gamma = 1, \\ \frac{p}{1+n}\cdot\frac{\omega-\alpha\left\{p\omega+(1+p\beta)(1+n)\right\}}{(1+p\beta)+\omega p/(1+n)} & \text{if } \gamma = 0\,. \end{cases}$$

That is,

$$\frac{p}{1+n}\cdot\frac{b}{Ak} = \begin{cases} \frac{1}{\frac{(1+p\beta)(1+n)}{\omega p}+1} - \alpha & \text{if } \gamma = 1, \\ \frac{1}{\frac{(1+p\beta)(1+n)}{\omega p}+1} - \alpha p & \text{if } \gamma = 0. \end{cases}$$

This expression shows that the ratio is decreasing in n and increasing in ω. In addition, the ratio is increasing in p if $\gamma = 1$.

To determine the effect of p when $\gamma = 0$, we take the first and second derivatives with respect to p and obtain

$$\partial\left[\frac{p}{1+n}\cdot\frac{b}{Ak}\bigg|_{\gamma=0}\right]/\partial p = \frac{\omega(1+n)}{[(1+p\beta)(1+n)+\omega p]^2} - \alpha,$$

$$\partial^2\left[\frac{p}{1+n}\cdot\frac{b}{Ak}\bigg|_{\gamma=0}\right]/\partial p^2 < 0\,.$$

The ratio is strictly concave in p with

$$\partial\left[\frac{p}{1+n}\cdot\frac{b}{Ak}\bigg|_{\gamma=0}\right]/\partial p\bigg|_{p=0} = \frac{\omega}{1+n} - \alpha,$$

$$\partial\left[\frac{p}{1+n}\cdot\frac{b}{Ak}\bigg|_{\gamma=0}\right]/\partial p\bigg|_{p=1} = \frac{\omega/(1+n)}{\left\{(1+\beta)+\omega/(1+n)\right\}^2} - \alpha\,.$$

The result suggests that there are two threshold values of α, namely $\omega/(1+n)$ and $\omega/(1+n)\{(1+\beta)+\omega/(1+n)\}^2$, such that

$$\partial\left[\frac{p}{1+n}\cdot\frac{b}{Ak}\bigg|_{\gamma=0}\right]/\partial p < 0\ \forall p\in(0,1) \text{ if } \frac{\omega}{1+n} \le \alpha,$$

$$\partial\left[\frac{p}{1+n}\cdot\frac{b}{Ak}\bigg|_{\gamma=0}\right]/\partial p > 0\ \forall p\in(0,1) \text{ if } \frac{\omega/(1+n)}{\left\{(1+\beta)+\omega/(1+n)\right\}^2} \ge \alpha\,.$$

Recall that when $\gamma = 0$, $b > 0$ holds if $\alpha < [p + (1+p\beta)(1+n)/\omega]^{-1}$ (Proposition 1). After some manipulation, we find that the following relation holds:

$$\frac{\omega/(1+n)}{\left\{(1+\beta)+\omega/(1+n)\right\}^2} < \frac{1}{p+(1+p\beta)(1+n)/\omega} < \frac{\omega}{1+n}\,.$$

This relation indicates that the threshold value $\omega/(1+n)$ is irrelevant for the current case. Therefore, we can conclude that

$$\partial\left[\frac{p}{1+n}\cdot\frac{b}{Ak}\bigg|_{\gamma=0}\right]/\partial p > 0\ \forall p\in(0,1) \text{ if } \alpha \le \frac{\omega/(1+n)}{\left\{(1+\beta)+\omega/(1+n)\right\}^2},$$

$$\partial\left[\frac{p}{1+n}\cdot\frac{b}{Ak}\bigg|_{\gamma=0}\right]/\partial p\bigg|_{p=0} > 0 \text{ and } \partial\left[\frac{p}{1+n}\cdot\frac{b}{Ak}\bigg|_{\gamma=0}\right]/\partial p\bigg|_{p=1} < 0$$

$$\text{if } \frac{\omega/(1+n)}{\left\{(1+\beta)+\omega/(1+n)\right\}^2} < \alpha < \frac{1}{p+(1+p\beta)(1+n)/\omega}\,.$$

■

5.3. Proof of Proposition 3

(i) Recall the capital-market-clearing condition in definition 1. With the use of the government budget constraint, $\tau(1-\alpha)Ak = pb/(1+n)$, the condition is reformulated as

$$k' = \frac{\frac{p\beta}{1+p\beta}}{(1+n) + \frac{p\beta}{1+p\beta}\cdot\frac{B_\gamma}{p\beta\left(1+\frac{\gamma(1-p)}{p}\right)\alpha}} \cdot \left[(1-\alpha)Ak - \frac{p}{1+n}B(k) + (1-\gamma)(1-p)\alpha Ak\right].$$

Substituting the policy function derived in proposition 1, we reformulate the above expression as presented in proposition 3(i). Direct comparison leads to $k'/k|_{\gamma=1} > k'/k|_{\gamma=0}$.

(ii) The expressions in proposition 3(i) indicate that

$$\partial\, k'/k\big|_{\gamma=1}/\partial\omega < 0, \quad \partial\, k'/k\big|_{\gamma=1}/\partial n < 0,$$
$$\partial\, k'/k\big|_{\gamma=0}/\partial\omega < 0, \quad \partial\, k'/k\big|_{\gamma=0}/\partial n < 0\,.$$

To find the effect of p, we first reformulate the expressions as

$$\left.\frac{k'}{k}\right|_{\gamma=1} = A\cdot\left[(1+n) + \frac{\omega}{1+p\beta}\left(p+\frac{1}{\beta\alpha}\right)\right]^{-1}, \tag{9}$$

$$\left.\frac{k'}{k}\right|_{\gamma=1} = A\cdot\left[(1+n) + \frac{\omega}{1+p\beta}\left(p+\frac{1}{p\beta\alpha}\right)\right]^{-1}. \tag{10}$$

The derivative of the term $\frac{\omega}{1+p\beta}\left(p+\frac{1}{\beta\alpha}\right)$ in (9) with respect to p is

$$\partial\left\{\frac{\omega}{1+p\beta}\left(p+\frac{1}{\beta\alpha}\right)\right\}/\partial p = \frac{1}{(1+p\beta)^2}\left(1-p(1-\beta)-\frac{1}{\beta\alpha}\right) < 0\,.$$

The derivative of the term $\frac{\omega}{1+p\beta}\left(p+\frac{1}{p\beta\alpha}\right)$ in (10) with respect to p is

$$\partial\left\{\frac{\omega}{1+p\beta}\left(p+\frac{1}{p\beta\alpha}\right)\right\}/\partial p = \frac{1}{(1+p\beta)^2}\left(1-\frac{1}{\beta\alpha p^2}-\frac{2}{\alpha p}\right) < 0\,.$$

Therefore, $\partial\, k'/k|_\gamma/\partial p > 0$ holds for both cases. ∎

5.4. Proof of Proposition 4

Let $V^y|_{\gamma=1}$ and $V^y|_{\gamma=0}$ denote the lifetime utility functions of the young when $\gamma = 1$ and $\gamma = 0$, respectively. For a given k, the functions are

$$\begin{aligned}
V^y|_{\gamma=1} =& (1+p\beta)\ln\left[(1-\alpha)Ak - \frac{p}{1+n}B_{\gamma=1}Ak\right] \\
&+ (1+p\beta)\ln\frac{B_{\gamma=1}}{\alpha/p}\cdot\frac{\frac{p\beta}{1+p\beta}}{(1+n)+\frac{p\beta}{1+p\beta}\cdot\frac{B_{\gamma=1}}{p\beta\alpha/p}} + \ln\frac{1}{1+p\beta} \\
&+ p\beta\ln\frac{p\beta}{1+p\beta}\cdot\frac{\alpha}{p}A, \\
V^y|_{\gamma=0} =& (1+p\beta)\ln\left[(1-\alpha)Ak - \frac{p}{1+n}B_{\gamma=0}Ak + (1-p)\alpha Ak\right] \\
&+ (1+p\beta)\ln\frac{B_{\gamma=0}}{\alpha}\cdot\frac{\frac{p\beta}{1+p\beta}}{(1+n)+\frac{p\beta}{1+p\beta}\cdot\frac{B_{\gamma=0}}{p\beta\alpha}} + \ln\frac{1}{1+p\beta} \\
&+ p\beta\ln\frac{p\beta}{1+p\beta}\cdot\alpha A\,.
\end{aligned}$$

We first reformulate the first term in each expression as follows:

$$\begin{aligned}
&(1+p\beta)\ln\left[(1-\alpha)Ak - \frac{p}{1+n}B_{\gamma=1}Ak\right] \\
&\qquad = (1+p\beta)\ln\frac{1+p\beta}{(1+p\beta)+\frac{\omega p}{1+n}}A\left(k'/k\big|_{\gamma=1}\right)^t k_0, \\
&(1+p\beta)\ln\left[(1-\alpha)Ak - \frac{p}{1+n}B_{\gamma=0}Ak + (1-p)\alpha Ak\right] \\
&\qquad = (1+p\beta)\ln\frac{1+p\beta}{(1+p\beta)+\frac{\omega p}{1+n}}A\left(k'/k\big|_{\gamma=0}\right)^t k_0\,.
\end{aligned}$$

Second, we compare the second term in each expression and obtain

$$\begin{aligned}
&(1+p\beta)\ln\frac{B_{\gamma=1}}{\alpha/p}\cdot\frac{\frac{p\beta}{1+p\beta}}{(1+n)+\frac{p\beta}{1+p\beta}\cdot\frac{B_{\gamma=1}}{p\beta\alpha/p}} \\
&\qquad\qquad \gtreqless (1+p\beta)\ln\frac{B_{\gamma=0}}{\alpha}\cdot\frac{\frac{p\beta}{1+p\beta}}{(1+n)+\frac{p\beta}{1+p\beta}\cdot\frac{B_{\gamma=0}}{p\beta\alpha}} \\
&\qquad\qquad \Leftrightarrow (1+p\beta)\ln p \gtreqless (1+p\beta)\ln 1 = 0\,.
\end{aligned}$$

Given the result thus far, we obtain

$$
\begin{aligned}
V^y|_{\gamma=1} &\gtrless V^y|_{\gamma=0} \\
&\Leftrightarrow (1+p\beta)\ln\left(k'/k\big|_{\gamma=1}\right)^t + p\beta\ln(1/p) \\
&\quad + (1+p\beta)\ln p \gtrless (1+p\beta)\ln\left(k'/k\big|_{\gamma=0}\right)^t \\
&\Leftrightarrow (1+p\beta)\ln\left(k'/k\big|_{\gamma=1}\right)^t + \ln p \gtrless (1+p\beta)\ln\left(k'/k\big|_{\gamma=0}\right)^t .
\end{aligned}
$$

If $t = 0$, then we obtain $V^y|_{\gamma=1} < V^y|_{\gamma=0}$ because $\ln p < 0$. For $t \geq 1$, we have $k'/k|_{\gamma=1} > k'/k|_{\gamma=0}$, as demonstrated in proposition 3. Therefore, there is some $T(> 1)$ such that $V^y|_{\gamma=1} > V^y|_{\gamma=0}$ for $t \geq T$. ■

References

Abel, A. B. (1985), Precautionary Saving and Accidental Bequests, American Economic Review 75, 777–791.

Acemoglu, D., and Robinson, J. A. (2005), Economic Origins of Dictatorship and Democracy, Cambridge University Press, New York, NY.

Azariadis, C., and Galasso, V. (2002), Fiscal Constitutions, Journal of Economic Theory 103, 255–281.

Bassetto, M. (2008), Political Economy of Taxation in an Overlapping-Generations Economy, Review of Economic Dynamics 11, 18–43.

Boldrin, M., and Rustichini, A. (2000), Political Equilibria with Social Security, Review of Economic Dynamics 3, 41–78.

Chen, K., and Song, Z. (2014), Markovian Social Security in Unequal Societies, Scandinavian Journal of Economics 116, 982–1011.

Cooley, T. F., and Soares, J. (1999), A Positive Theory of Social Security Based on Reputation, Journal of Political Economy 107, 135–160.

Forni, L. (2005), Social Security as Markov Equilibrium in OLG Models, Review of Economic Dynamics 8, 178–194.

Gonzalez-Eiras, M., and Niepelt, D. (2008), The Future of Social Security, Journal of Monetary Economics 55, 197–218.

Gonzalez-Eiras, M., and Niepelt, D. (2012), Ageing, Government Budgets, Retirement, and Growth, European Economic Review 56, 97–115.

Gradstein, M., and Kaganovich, M. (2004), Aging Population and Education Finance, Journal of Public Economics 88, 2469–2485.

Grossman, G., and Helpman, E. (1998), Intergenerational Redistribution with Short-Lived Governments, Economic Journal 108, 1299–1329.

Kunze, L. (2014), Life Expectancy and Economic Growth, Journal of Macroeconomics 39, 54–65.

Lancia, F., and Russo, A. (2016), Public Education and Pensions in Democracy: A Political Economic Theory, Journal of the European Economic Association 14, 1038–1073.

Lindbeck, A., and Weibull, J. (1987), Balanced Budget Redistribution and the Outcome of Political Competition, Public Choice 52, 273–297.

Mateos-Planas, X. (2008), A Quantitative Theory of Social Security without Commitment, Journal of Public Economics 92, 652–671.

OECD (2006), Society at a Glance: OECD Social Indicators – 2006 Edition, OECD Publishing, Paris.

OECD (2011), OECD Factbook 2011–2012: Economic, Environmental and Social Statistics, OECD Publishing, Paris.

Ono, T. (2016), Aging, Pensions, and Growth, Osaka University, Discussion Papers in Economics and Business No. 14-17-Rev.2.

Ono, T., and Uchida, Y. (2016), Pensions, Education, and Growth: A Positive Analysis, Journal of Macroeconomics 48, 127–143.

Persson, T., and Tabellini, G. (2000), Political Economics: Explaining Economic Policy, MIT Press, Cambridge, MA.

Romer, Paul M. (1986), Increasing Returns and Long run Growth, Journal of Political Economy 94, 1002–1037.

Rusconi, R. (2008), National Annuity Markets; Features and Implications. Working Papers on Insurance and Private Pensions, No. 24, OECD Publishing, Paris.

Sheshinski, E., and Weiss, Y. (1981), Uncertainty and Optimal Social Security Systems, Quarterly Journal of Economics 96, 189–206.

Smets, K. (2012), A Widening Generational Divide? The Age Gap in Voter Turnout through Time and Space, Journal of Elections, Public Opinion and Parties 22, 407–430.

Song, Z., Storesletten, K., and Zilibotti, F. (2012), Rotten Parents and Disciplined Children: A Politico-economic Theory of Public Expenditure and Debt, Econometrica 80, 2785–2803.

Zhang, J., Zhang, J., and Lee, R. (2003), Rising Longevity, Education, Savings, and Growth, Journal of Development Economics 70, 83–101.

The Nature of Returns to Scale in Aggregate Production with Public Intermediate Goods

Johannes Pauser*

Received 20 January 2015; in revised form 13 March 2016; accepted 10 October 2016

This paper examines returns to scale in aggregate production when public inputs are used in the production process of firms. The conditions for productive efficiency are derived for commonly used public input specifications, where a rise in the number of firms or in the employment of private factors causes congestion. The analysis shows how the returns to scale varies with the congestion technology, the degree of congestion, and the returns to scale in public input production. While the conditions for productive efficiency differ for two commonly used public input specifications, aggregate production may exhibit increasing returns even for fully congestible public inputs in both cases.

Keywords: public inputs, returns to scale, congestion

JEL classification: H 40, H 54

1. Introduction

Public inputs (public intermediate goods) in production are put to use in important economic areas, such as international trade, growth theory, and international tax competition for foreign direct investments. The results of economic modelling in these research areas depend to a large extent upon the exact specification and properties of the aggregate production function, for instance, whether there are increasing returns to scale when a public factor is employed in the private sector in addition to the private factors of production, labour and capital. With a production function that is linear homogeneous in private inputs, firm profits are zero in the competitive equilibrium, while public input provision generates firm profits if the production function exhibits constant returns to scale in all factors, including the public factor. In the literature, the former specification is also referred to

* TU Dortmund University, D-44221 Dortmund (johannes.pauser@tu-dortmund.de).
I would like to thank Volker Arnold and Lutz Altenburg for their helpful support during the research. I am also indebted to the participants of seminars at the University of Hagen and of the FAU/IAB-Seminar Macroeconomics and Labor Markets for comments and suggestions. In addition, I thank Alfons Weichenrieder and an anonymous referee, whose comments contributed significantly to improve the article.

FinanzArchiv 73 (2017), 190–212 doi: 10.1628/001522117X14877521353474
ISSN 0015-2218

as factor-augmenting public input and the latter as firm-augmenting public input.[1]

The sensitivity of results (and their policy implications) to the nature of the returns to scale is well-known, for instance, in models on international tax competition, where equilibrium public good provision levels resulting from tax competition may depend on the returns to scale in the production function. Matsumoto (1998) has examined under what circumstances the optimal provision of a public input differs in an international tax competition setting between the factor-augmenting type and the firm-augmenting type of the public input. Returns to scale in production are also critical for the conclusions drawn in the analysis of Martínez and Sánchez-Fuentes (2011) who examine the optimal provision of public inputs, and the sensitivity of the results to the technology, taxation and consumer preferences. The specification of the production technology is, in addition, crucial in the literature on trade theory (e.g. Abe, 1990; Altenburg, 1992) and in models on endogenous growth (e.g. Barro, 1990; Glomm and Ravikumar, 1994).

In the empirical literature, results on the returns to scale in aggregate production are mixed. Cole and Ohanian (1999) argue that this may to some extent occur as a result of difficulties in measuring aggregate returns to scale. Levy (1990) estimates a long-term aggregate production function and finds slight but statistically significant increasing returns to scale in the U.S. for the post-war period (1948–1983). With an application of the production function approach (Aschauer, 1989) on U.S. data, Batina (1999) rejects the hypotheses of constant returns to scale in private inputs in a sensitivity analysis of the estimation of the long run effect of public capital on output in all of eight examined cases. The same analysis is also able to reject constant returns to scale in all inputs in seven cases.

In order to derive meaningful policy implications, economic studies usually require the specification of the production function of an industry or the economy rather than of a single firm. In general, the production function of an industry provides the functional relationship between efficiently employed input factors and the resulting maximum output in that industry (e.g. Fisher, 1969; Shephard, 1970). In theoretical work, however, issues related to the specification of an aggregate production function, including the nature of

1 Both types of public input specifications are introduced and discussed in literature by Henderson (1974), Hillman (1978), and McMillan (1979). Some authors have argued that the firm-augmenting public input, which is non-rival on the firm-level, has problematic theoretical implications and may be of limited empirical importance (e.g., Hillman, 1978; Feehan, 1989). In a recent paper, Colombier (2008) re-examined efficiency in the provision of a firm-augmenting input under perfect competition and he concluded that the firm-augmenting public input cannot be provided efficiently by governments.

the returns to scale in aggregate production, have received little attention.[2] There is still not much research with an explicit focus on productive efficiency in the economy, especially when one allows also for the realistic possibility that the productive public input is subject to congestion. On the other side, the assumption that publicly-provided goods are congestible is common in economic theory, and there is a broad consensus that a large part of the (productive) infrastructure exhibits at least some degree of congestion (e.g. Stiglitz, 1988; Barro and Sala-I-Martin, 1992; Glomm and Ravikumar, 1994; Gramlich, 1994). Feehan (1989) and Colombier and Pickhardt (2002, 2005) have provided an overview of public input specifications and of the aggregate production function using homogenous production functions, and they have contributed significantly to the understanding of the role of public input specifications for the derivation of policy implications. For the derivation of the nature in the returns to scale in aggregate production with an endogenous number of firms, this paper assumes instead that returns to scale are variable with first increasing and then decreasing returns to scale.[3]

Such as the specification of an aggregate production function, the specification of the congestion technology is crucial in economic modelling. It is usually introduced at the beginning of the implementation of models, and the implications to be derived from models often depend on the chosen specification of the technology. Section 2 introduces and analysis two common public input specifications in the literature: One, in which the public factor used in the production process of firms is congestible by an increasing number of firms, and a second specification, where the public factor is congestible by an increasing employment of private factors of production, labour and (or) capital. The aim of the subsequent two sections is then to analyse the nature of the returns to scale in aggregate production for two common public input specifications in the social optimum. As the determination of the optimal number of firms (and the efficient size of a single firm) in an industry is crucial for the derivation of the maximum output in that industry, the model solves simultaneously for the optimal amounts of primary factors of production in the public and in the private sector, and for the optimal firm size. The issue of the optimal number of firms in an industry has initially been addressed by Hicks (1939) and Henderson (1974).

2 This is partly because the aggregation of production functions is fraught with difficulties. The corresponding work on this issue is often referred to as 'aggregation literature' and there is still no sound theoretical foundation of aggregate production functions. Some authors have argued that a purely technological definition of an aggregate production function, which is common in neoclassical theory, is inappropriate. Felipe and Fisher (2003) review the recent literature on this topic.

3 The assumption of variability in the returns to scale in the production function of firms is also due to some well-known problems arising from a priori homogeneity assumptions on the production function (cf. Henderson, 1974; Arnold, 1980).

Assuming non-congestible public inputs and variable returns to scale in production, Arnold (1980) and Kohn (1985) have derived a condition for the optimal number of firms in an industry as an additional condition for productive efficiency, and Arnold (1980) has found increasing returns to scale in aggregate production. Mason and Polasky (1997) have shown in a dynamic setting that a socially optimal output cannot be obtained with a fixed number of firms. In their approach, the optimal industry size is determined as that one that produces the socially optimal output level. More recently, Chang et al. (2007) have distinguished their work on optimal fiscal policies in the presence of congestion and market imperfections from the endogenous growth literature that assumes a fixed number of firms. In their paper congestion is caused by an endogenously determined number of firms and the aggregate amount of capital. The present paper shares the view of these studies in that it assumes that the size and number of firms is an important determinant in aggregate production and is to be determined endogenously.

As a consequence, next to the modified Samuelson-condition for an efficient provision of public inputs, which determines the efficient allocation of primary factors in the public and in the private sector, the efficiency conditions for an optimal size of a firm are derived when a congestible public factor is employed by the private sector. Building on these efficiency conditions, the approach is able to solve for the nature of the returns to scale in the aggregate production function of the economy.

The results are useful for work that considers congestible public inputs in production and does not fix the number of firms, because they can provide more clarity on the appropriate specification of the production technology in this case. For instance, the derived condition for the efficient size of a representative firm requires that the production function exhibits increasing returns to scale, and that returns to scale are decreasing in private inputs alone, if the public input in production is congestible in the sense that the service derived by a single firm diminishes as the number of firms increases. In that case two frequently considered specifications of production functions in literature, i.e., constant returns to scale in all inputs and constant returns to scale in primary factors (excluding the public factor), are not consistent with production efficiency unless for the polar cases with full or no congestion. The latter specification is sometimes argued to be, in general, the more plausible specification for production technologies that involve public inputs (e.g. Hillman, 1978; McMillan, 1979; Feehan, 1989, 1998).[4] The former specification has been considered, for instance, in Matsumoto (2000) for

4 In the literature it often appears as if a choice has to be made between the two specifications. This view is not shared in the present work. As should become clear from the subsequent analysis, inconsistence with productive efficiency is mainly a result of *assuming* a certain degree of homogeneity for the production function.

public inputs that are congestible in the number of firms. In his model, the production technology exhibits constant returns to scale in all factors and congestion externalities are incorporated into a cost function or, in a variation of the model, taken into the production function as in the present paper (Matsumoto, 2000, pp. 245–246, footnotes 5 and 6). For the latter specification the subsequent analysis finds that the technology exhibits increasing returns to scale when the output level of the representative firm is efficient.

If congestion is, in contrast, a cause of the increased employment of private inputs (e.g. capital and labour), production efficiency requires that the technology exhibits constant returns to scale in private inputs (for a constant service derived from infrastructure provision). In addition to the two well-known specifications of a production function when the publicly-provided input in production is either of the "pure" or of the "unpaid factor" type (cf. e.g. Feehan, 1998), the paper suggests therefore a third specification, in which the nature of returns to scale is dependent on the type and degree of congestion. For instance, returns to scale in private inputs is decreasing and a negative function of the congestion parameter and the output elasticity of the public input if the public input is at least partly congestible by the number of firms. The paper finds further that a constant returns to scale production technology is not consistent with the derived conditions of productive efficiency for both specifications of congestion technologies reviewed in this paper, unless for publicly-provided private inputs. This implies that also the firm-augmenting public input, which assumes no congestion on the firm level (e.g. Kunce and Shogren, 2005) is not consistent with the derived conditions for productive efficiency.

The structure of the paper is as follows: The subsequent section introduces and analyses two common public input specifications subject to congestion. Section 3 discusses the production conditions in an economy consisting of a public and a private sector with an endogenous number of firms. In section 4, the conditions for productive efficiency are analysed for both public input specifications subject to congestion, and for pure public inputs. The final section summarises the results and concludes the paper.

2. Public Inputs and Congestion

Productive public inputs in production are used collectively by the private sector. The properties of public inputs, however, rarely match those of true public goods as for productive infrastructures the main features of public goods, non-excludability and non-rivalry, are not fulfilled. The exclusion of additional users is possible for a large number of productive infrastructure, and there are numerous examples where the quality of (or service derived from) the infrastructure diminishes for each user if the number of users increases. Roads, for instance, can be used without congestion up to a certain number of vehicles on a particular stretch of the road only.[5] Other examples of productive public goods that exhibit at least some degree of congestion are seaports, airports, bridges, higher education, legal courts or communication infrastructure (e.g., the internet). However, productive infrastructure is rarely fully congestible, like a private good, either. This is why Samuelson (1969, p. 108) recognising that most of the goods that can be used collectively are subject to some degree of congestion, removed his former strict separation of private goods and pure public goods: "A public good is one that enters two or more persons' utility. What are we left with? Two poles and a continuum between? No. With a knife-edge pole of the private-good case, and with all the rest of the world in the public-good domain by virtue of involving some 'consumption externality'."

With respect to the term *user* of a public input, one might typically think of a firm, as a public input is usually modelled as a productive factor in the production function of firms, which makes them the primary users of the public factor. Therefore, some articles refer to congestion that is caused by an increasing number of firms. However, as public inputs increase the marginal productivity of primary factors, capital and labour,[6] and therefore the 'quality' of these factors, primary inputs themselves can be interpreted as 'users' of the public input. As a consequence, increasing the employment of private factors of production may cause overcrowding in the sense that the service derived from a public input diminishes for an increasing employment of private factors.

The first specification assumes that the service from the public input diminishes with an increasing number of firms (e.g. Feehan, 1998; Matsumoto, 2000):

$$\widetilde{B} = \frac{B}{m^{\gamma}}; \quad 0 \leq \gamma \leq 1 \,. \tag{1}$$

Within specification (1), $\widetilde{B}$ is the quantity of the public factor available to a single firm in the industry and B is the total quantity of the public factor provided by the government. A variable number of firms operating in the industry is denoted with m and all firms are assumed to be identical.[7] The

5 The literature on transportation economics provides a functional relationship between the average travel time and the travel density, where the latter is the quotient of the traffic volume and the capacity of a stretch of a road (cf. e.g. Mohring, 1976).

6 For instance, in the factor-augmenting case, the productivity gain from the public input provision is derived entirely from the increase in marginal productivities of capital and labour.

7 The assumption of identical firms can be justified by assuming that firms use all available information regarding the production process, and therefore employ an identical production technology.

parameter γ denotes the degree of rivalry of B between firms, and therefore within the industry. As can be gathered from (1), for congestible public inputs, the higher γ, the more the supply with the public factor declines for each firm when the number of firms increases. The two poles of a pure public factor and a publicly-provided private factor are obtained for $\gamma = 0$ and $\gamma = 1$.[8] In the first case, we obtain $\widetilde{B} = B$, and the amount of the public factor available to a single firm does not diminish if the number of firms increases. In the second case ($m\widetilde{B} = B$), a proportional increase in the number of firms (m) and the quantity of the public factor (B) leaves the public factor available per firm ($\widetilde{B}$) unchanged.[9]

In the second specification, overcrowding is the result of an increasing employment of private factors of production, labour and/or capital in literature. In the subsequent analysis the quantity of the public input available for each firm is

$$\widetilde{B} = \frac{B}{A^{\alpha}K^{\beta}}; \quad 0 \leq \alpha \leq 1; \quad 0 \leq \beta \leq 1\,, \tag{2}$$

where A and K are the total amounts of primary factors employed in the private sector, and α and β are the degrees of congestion. This rather general specification allows for the derivation of more special cases of congestion, where the public factor becomes congested by one private factor only or by some combination of both factors (e.g. Buettner, 1999). In its original form, specification (2) can be traced back to Boadway and Flatters (1982), who tie congestion to labour only. For $\alpha = \beta = 0$, the public factor is non-congestible, like a pure public good. However, in case that at least one parameter is different from zero, an increasing employment of labour and/or capital reduces the amount of the public factor available to each firm. The case of a fully congestible public input, in the sense that a proportional increase in the congestion-causing factors, labour (A) and capital (K), and

8 To be precise, as another feature of pure public goods, an exclusion of additional users of the public good is impossible.

9 The concept of congestibility is therefore similar to congestible public consumption goods. Public consumption goods are usually referred to as fully congestible if a proportional increase in the number of people that consume the good and the quantity of the good leave the public good available per person constant. For congestible public inputs as specified in (1), the situation $\gamma > 1$ implies that the decline in the public factor available per firm ($\widetilde{B}$) resulting from a one percent increase in the number of firms (m) exceeds one percent (and therefore an increase in the public factor (B) by more than one percent is required for $\widetilde{B}$ to remain constant). This becomes obvious from the congestion elasticity $\gamma = -\frac{\partial \widetilde{B}}{\partial m}\frac{m}{\widetilde{B}}$. Examples for productive public factors for which the degree of congestion is above that one of a fully congestible good ($\gamma > 1$) are rare. It may, however, be realistic for local public goods (see, Edwards, 1990).

the quantity of the public factor (B) leaves the public input available per firm ($\widetilde{B}$) unchanged, is then obtained for $\alpha + \beta = 1$.[10]

The use of public input specifications subject to congestion as specified above is, probably, most numerous in the literature on endogenous growth theory, where productive government spending has been frequently considered since the work of Barro (1990). Most of the subsequent work indeed builds on this growth model, and a large part of the studies allows for the possibility that the publicly-provided factor is subject to some form of congestion. In reviewing this literature, one encounters different specifications of the congestion technology and there seems to be no systematic pattern between the scale elasticity of the production function and the specification of congestion.[11] In addition to other specifications, Barro and Sala-I-Martin (1992), for instance, consider congestion that is caused by an increasing employment of aggregate capital (equation (14) in their third version of the model). In detail, the authors assume a publicly-provided private good, which makes their specification identical to the knife-edge case in the present paper, where $\beta = 1$ and $\alpha = 0$ in specification (2). With respect to the production technology, constant returns to scale in all inputs is assumed. Glomm and Ravikumar (1994, 1997) assume a public factor that is congestible by the aggregate employment of capital and labour, exactly as in specification (2) of the present paper, and their production function exhibits constant returns to scale in the two private inputs (for a constant amount of the public input available per firm). Eicher and Turnovsky (2000), who address the relationship between congestion and growth, combine congestion caused by the number of firms and by the aggregate employment of private factors. In addition, their congestion technology distinguishes between three broad categories. The first specification is a non-rival public factor, which is obtained from specifications (1) and (2) with $\gamma = 0$ and $\alpha + \beta = 0$, respectively. Their case of pure relative congestion relates to specification (1) if $\gamma > 0$, while a similar case to pure aggregate congestion results from specification (2) if $\alpha + \beta > 0$.[12] With respect to the production technology, constant returns to scale in pri-

10 For $\alpha + \beta > 1$ a one percent increase in A and K would decrease $\widetilde{B}$ by more than one percent $\left(\alpha + \beta = -\frac{\partial \widetilde{B}}{\partial (m\widetilde{A})}\frac{(m\widetilde{A})}{\widetilde{B}} - \frac{\partial \widetilde{B}}{\partial (m\widetilde{K})}\frac{(m\widetilde{K})}{\widetilde{B}}\right)$. In analogy to $\gamma \leq 1$, $\alpha + \beta \leq 1$ is assumed in the analysis.

11 This issue is discussed in detail in section 4.

12 It is common in endogenous growth models to assume that individuals are households and producers. In contrast, the remainder of the present analysis explicitly focuses on the supply side of the economy. Given this difference, the similarity between relative and aggregate congestion in many growth models (e.g. Eicher and Turnovsky, 2000; Gómez, 2008) and congestion functions (1) and (2) becomes obvious if one assumes that the aggregate stock of capital in the growth literature (the number of individuals times the individual capital stock) is equal to the aggregate capital stock in the present analyses, which is the the number of firms (m) times the capital stock per firm ($\widetilde{K}$). The similarity be-

vate factors and increasing returns in all three factors, including the public factor, is assumed. The congestion technology (and modifications thereof) introduced by Eicher and Turnovsky (2000) has been frequently taken up in the later literature. The congestion model is identical to the one employed in Eicher and Turnovsky (2000), for instance, in Gómez (2008). However, the latter work allows for the possibility that the individual production function exhibits either constant or diminishing returns to scale in private inputs.[13] The endogenous growth models of Dioikitopoulos and Kalyvitis (2008) and Ott and Soretz (2008) incorporate congestion of the public factor that is solely caused by the number of agents (firms), which is typically referred to as relative congestion in the endogenous growth literature. Both articles assume a constant returns to scale production function for the individual plant.

3. The Supply Side of the Economy

In order to derive the properties of aggregate production in an industry, we examine the supply side of the economy and assume that there are two sectors: a private sector that produces a single final good X with m identical firms operating in this industry, and a public sector producing the public factor B. The private good X is produced with the primary factors of production, labour A and capital K, and the public factor B.[14] The latter is provided to the firms by the government, and may be subject to congestion according to (1) or (2). In the private sector, the output of the X-industry equals the output of a representative firm multiplied by the number of firms: $X = \widetilde{X}m$. The output per firm $\widetilde{X}$ is produced by the employment of the primary factors of production, labour $\widetilde{A}$ and capital $\widetilde{K}$, and with the help of the public factor $\widetilde{B}$ available per firm. Thus, output of a representative firm is:

$$\widetilde{X} = \widetilde{X}(\widetilde{A}, \widetilde{K}, \widetilde{B}), \tag{3}$$

where marginal products of all three inputs are assumed to be positive and diminishing. For the production technology, the literature usually assumes

tween relative congestion and specification (1) becomes obvious, for instance, in Ott and Soretz (2008, p. 610, footnote 12).

13 Similarly, Fisher and Turnovsky (1998) decide to leave open the question whether there are constant returns in private factors alone or constant returns in all inputs. The only restriction in Fisher and Turnovsky (1998), where congestion is caused by the number of agents (firms), is non-increasing returns to scale in the private inputs of production, capital and labour.

14 The analysis becomes more complex if, as it has been assumed in some articles, there is a second private sector in which public factors of production are employed. It is well known that in the case of increasing returns to scale, the problem of a non-convex production set may arise and result in multiple equilibria and corner solutions. For a discussion of these issues, see, for instance, Tawada (1980) and Altenburg (1987).

constant returns to scale in all factors or constant returns to scale in the primary factors of production, labour and capital. In addition, different assumptions regarding the specification of the public input available to firms, i.e., its degree of publicness, are made. The objective of the present analysis is the derivation of the extent of the returns to scale in aggregate production for different public input specifications when there are no a priori assumptions on the scale elasticity of the production function. Instead, variability to returns to scale with the scale of output is assumed with first increasing and then decreasing returns to scale (cf. e.g. Arnold, 1980; Kohn, 1985).

Total amounts of primary inputs in the economy, A_0 and K_0, are fixed, and fully employed in the private or in the public sector:

$$A_0 = A_B + m\widetilde{A}; \quad K_0 = K_B + m\widetilde{K}\,. \tag{4}$$

$A \equiv m\widetilde{A}$ and $K \equiv m\widetilde{K}$ are the quantities of primary factors employed in the X-industry, and A_B and K_B are the respective amounts employed for the production of the public input. As stated by Kaizuka (1965), it is reasonable to assume that the public factor is produced with primary factors of production only:

$$B = B(A_B, K_B)\,. \tag{5}$$

For the production of the public input, usually a constant returns to scale production technology is assumed in the literature. However, alternative assumptions regarding the nature of the returns to scale for the production technology (5) are conceivable, as productive infrastructure also tends to exhibit economies of scale in production. Aschauer (1989), for instance, stresses that this is one of the reasons for government intervention in some inputs provided to private production, such as the acquisition and distribution of water. In a more recent paper, Matsumoto and Feehan (2010) account for the possibility of scale economies in public input production for the analysis of public input provision in an international tax competition setting. In the subsequent analysis we allow for increasing returns to scale in public input production and define:[15]

$$\varepsilon_{B/\lambda} \equiv \frac{\partial B}{\partial A_B}\frac{A_B}{B} + \frac{\partial B}{\partial K_B}\frac{K_B}{B} \tag{6}$$

for the scale elasticity in production of the public factor, where for $\varepsilon_{B/\lambda} > 1$ ($\varepsilon_{B/\lambda} = 1, \varepsilon_{B/\lambda} < 1$) returns to scale in public production are increasing (constant, decreasing).

15 In general, there are varying definitions of the returns to scale in public production in literature. Duncombe and Yinger (1993) analyse different concepts of returns to scale in the public sector and provide an overview of the literature.

4. Productive Efficiency

Using similar assumptions for the production side of the economy, Kaizuka (1965) was the first to derive the conditions for an efficient production in the economy. The derived Samuelson-conditions for the optimal provision of public inputs were discussed again in Sandmo (1972), Henderson (1974), and in some later articles. If also the number of firms in an industry is considered an endogenous variable, for a complete set of efficiency conditions, the quantity of primary inputs of production and the number of firms m must be determined simultaneously to maximise:[16]

$$\begin{aligned} \max \quad & X = m\widetilde{X}(\widetilde{A}, \widetilde{K}, \widetilde{B}) \\ \text{s.t.} \quad & A_0 = A_B + m\widetilde{A} \\ & K_0 = K_B + m\widetilde{K} \\ & \widetilde{B} = \frac{B(A_B, K_B)}{m^\gamma}, \end{aligned} \tag{7}$$

where public input specification (1) has been used in the maximisation problem allowing for congestibility of the public factor that is caused by an increasing number of firms. The conditions for an efficient production in the X-industry can be derived with help of the Lagrangian:

$$\begin{aligned} L(\widetilde{A}, \widetilde{K}, \widetilde{B}, m, A_B, K_B) &= m\widetilde{X}(\widetilde{A}, \widetilde{K}, \widetilde{B}) + \lambda_1(A_0 - m\widetilde{A} - A_B) \\ &\quad + \lambda_2(K_0 - m\widetilde{K} - K_B) + \lambda_3\left(\widetilde{B} - \frac{B(A_B, K_B)}{m^\gamma}\right). \end{aligned} \tag{8}$$

The first-order conditions are (for an interior maximum):[17]

$$\frac{\partial L}{\partial \widetilde{A}} = m\frac{\partial \widetilde{X}}{\partial \widetilde{A}} - \lambda_1 m = 0 \tag{9}$$

$$\frac{\partial L}{\partial \widetilde{K}} = m\frac{\partial \widetilde{X}}{\partial \widetilde{K}} - \lambda_2 m = 0 \tag{10}$$

$$\frac{\partial L}{\partial m} = \widetilde{X} - \lambda_1\widetilde{A} - \lambda_2\widetilde{K} + \lambda_3\gamma B m^{-\gamma-1} = 0 \tag{11}$$

$$\frac{\partial L}{\partial A_B} = -\lambda_1 - \lambda_3 m^{-\gamma}\frac{\partial B}{\partial A_B} = 0 \tag{12}$$

$$\frac{\partial L}{\partial K_B} = -\lambda_2 - \lambda_3 m^{-\gamma}\frac{\partial B}{\partial K_B} = 0 \tag{13}$$

$$\frac{\partial L}{\partial \widetilde{B}} = m\frac{\partial \widetilde{X}}{\partial \widetilde{B}} + \lambda_3 = 0\,. \tag{14}$$

16 For an analysis with pure public inputs, see Arnold (1980) and Kohn (1985).

17 It is assumed that $m \geq 1$ is an integer, and that the second-order conditions for a maximum are satisfied.

4.1. Modified Samuelson-Conditions for the Efficient Provision of Public Inputs

From (9), (12), and (14) one obtains the modified Samuelson-condition:[18]

$$\frac{\partial \widetilde{X}}{\partial \widetilde{A}} = m^{1-\gamma} \frac{\partial \widetilde{X}}{\partial \widetilde{B}} \frac{\partial B}{\partial A_B}\,, \tag{15a}$$

or, because of $\partial \widetilde{B}/\partial B = m^{-\gamma}$:

$$\frac{\partial \widetilde{X}}{\partial \widetilde{A}} = m \frac{\partial \widetilde{X}}{\partial \widetilde{B}} \frac{\partial \widetilde{B}}{\partial B} \frac{\partial B}{\partial A_B}\,. \tag{15b}$$

According to efficiency conditions (15), *direct* marginal productivity of a private factor, i.e., through its employment in a firm (left-hand side (LHS) of equation (15)), must be equal to the *indirect* marginal (collective) productivity of the private factor, i.e., when the private factor is used for the production of the public input (right-hand side (RHS) of equation (15)). For a congestible public input, indirect marginal productivity of a private factor is the lower, the higher the degree of congestion, as less than the full amount of the public input is available to a single firm ($\partial \widetilde{B}/\partial B = m^{-\gamma} < 1$). For $\gamma = 1$, (15) degenerates to the provision rule for publicly-provided private goods. In contrast, for pure public inputs, the full amount of the public factor is an input to all firms regardless of the size of the industry, so that one obtains the provision rule for pure public goods with $\gamma = 0$. In addition, from (9), (10), (12), and (13) one obtains equality of the marginal rate of technical substitution in the private and in the public sector for public input specification (1).[19]

If congestion is defined as in (2), the modified Samuelson-condition for an efficient provision of the public input reads:[20]

$$\frac{\partial \widetilde{X}}{\partial \widetilde{A}} = m \frac{\partial \widetilde{X}}{\partial \widetilde{B}} (m\widetilde{A})^{-\alpha} (m\widetilde{K})^{-\beta} \frac{\partial B}{\partial A_B} + m \frac{\partial \widetilde{X}}{\partial \widetilde{B}} B\alpha (m\widetilde{A})^{-\alpha-1} (m\widetilde{K})^{-\beta}\,. \tag{16a}$$

Because of $\partial \widetilde{B}/\partial B = (m\widetilde{A})^{-\alpha}(m\widetilde{K})^{-\beta}$ and $\partial \widetilde{B}/\partial (m\widetilde{A}) = -B\alpha (m\widetilde{A})^{-\alpha-1} (m\widetilde{K})^{-\beta}$, (16a) can also be expressed by:

$$\frac{\partial \widetilde{X}}{\partial \widetilde{A}} = m \frac{\partial \widetilde{X}}{\partial \widetilde{B}} \frac{\partial \widetilde{B}}{\partial B} \frac{\partial B}{\partial A_B} - m \frac{\partial \widetilde{X}}{\partial \widetilde{B}} \frac{\partial \widetilde{B}}{\partial (m\widetilde{A})}\,. \tag{16b}$$

18 See Kaizuka (1965, p. 120, (3)) and Sandmo (1972, p. 153, (17)) for a modified representation of the efficiency conditions, and for $\gamma = 0$. If the X-industry consists of a single firm only, conditions (15) are equal to the efficiency condition derived by Kaizuka. With (10), (13), and (14), a similar condition can be derived for the production factor capital.

19 This can be stated formally as: $\frac{\partial \widetilde{X}}{\partial \widetilde{A}} / \frac{\partial \widetilde{X}}{\partial \widetilde{K}} = \frac{\partial B}{\partial A_B} / \frac{\partial B}{\partial K_B}$.

20 Condition (16) is derived in the Appendix. Again, a similar condition can be derived for the production factor capital.

Efficiency condition (16b) obviously differs from efficiency condition (15b), which is obtained in the case that congestion is caused by an increasing number of firms. According to conditions (16), allocation of private inputs between the private and the public sector is efficient, if direct marginal productivity equals indirect marginal productivity[21] plus the loss in production in all firms that occurs due to an increased employment of private factors and the resulting congestion of the public factor. This loss in production is the higher, the higher the degrees of congestion α and β, and it is captured by the second term on the RHS of efficiency conditions (16). In contrast to public input specification (1), equality of the marginal rate of technical substitution in the private and in the public sector does not necessarily hold for public input specification (2).[22]

4.2. Conditions for the Efficient Output Level of a Firm

In addition to the conditions for an efficient provision of public inputs as derived above, a complete set of efficiency conditions requires the derivation of the efficient size of a firm, i.e., the efficient output level of a firm in the long run. If the intermediate good is subject to congestion within the industry according to (1), solving (9), (10), and (14) for $\lambda_1, \lambda_2,$ and λ_3 and inserting the results into (11) yields:

$$\widetilde{X} - \frac{\partial \widetilde{X}}{\partial \widetilde{A}}\widetilde{A} - \frac{\partial \widetilde{X}}{\partial \widetilde{K}}\widetilde{K} - \gamma B m^{-\gamma} \frac{\partial \widetilde{X}}{\partial \widetilde{B}} = 0\,. \tag{17a}$$

Rearranging, and substitution of (1) and $\partial \widetilde{B}/\partial m = -\gamma B m^{-\gamma-1}$ yields:

$$\frac{\partial \widetilde{X}}{\partial \widetilde{A}}\frac{\widetilde{A}}{\widetilde{X}} + \frac{\partial \widetilde{X}}{\partial \widetilde{K}}\frac{\widetilde{K}}{\widetilde{X}} = 1 - \gamma \frac{\partial \widetilde{X}}{\partial \widetilde{B}}\frac{\widetilde{B}}{\widetilde{X}} = 1 + \frac{\partial \widetilde{X}}{\partial \widetilde{B}}\frac{\partial \widetilde{B}}{\partial m}\frac{m}{\widetilde{X}}\,. \tag{17b}$$

The efficient size of a firm is obtained for that point of production, for which the sum of production elasticities of private factors equals the term on the RHS of (17b). As a result, only for a non-congestible public input, the sum of production elasticities of private factors equals one, which is the

21 As stated for the derivation of condition (15), the marginal productivity of private inputs when used for the production of the public input (indirect marginal productivity) is the lower, the higher the degree(s) of congestion. Note that the RHS of (16) consists of this term only, if the public factor is non-congestible by labour ($\alpha = 0$), or the public factor is non-congestible by both private inputs ($\alpha = \beta = 0$).

22 For congestion technology (2) one obtains from the first-order conditions for a maximum:
$$\frac{\frac{\partial \widetilde{X}}{\partial \widetilde{A}}}{\frac{\partial \widetilde{X}}{\partial \widetilde{K}}} = \frac{\frac{\partial B}{\partial A_B} - \frac{\partial \widetilde{B}}{\partial m\widetilde{A}} / \frac{\partial \widetilde{B}}{\partial B}}{\frac{\partial B}{\partial K_B} - \frac{\partial \widetilde{B}}{\partial m\widetilde{K}} / \frac{\partial \widetilde{B}}{\partial B}}.$$
As a result, equal marginal rates of technical substitution are only obtained if
$$\frac{\frac{\partial \widetilde{X}}{\partial \widetilde{A}}}{\frac{\partial \widetilde{X}}{\partial \widetilde{K}}} = \frac{\frac{\partial B}{\partial A_B}}{\frac{\partial B}{\partial K_B}} = \frac{\frac{\partial \widetilde{B}}{\partial m\widetilde{A}}}{\frac{\partial \widetilde{B}}{\partial m\widetilde{K}}}$$
holds in addition.

condition for an efficient size of a firm in the case of pure public inputs. For a congestible public input ($\gamma > 0$), the sum of production elasticities of primary factors is below one, and it is a negative function of the degree of congestion (γ) times the production elasticity of the public input available to firms $[(\partial\widetilde{X}/\partial\widetilde{B}) \cdot (\widetilde{B}/\widetilde{X})]$.[23] As a result, for the efficient size of a firm, and for congestible public factors as specified in (1), the production function of the firm exhibits decreasing returns to scale in private factors of production.[24]

We now turn to the case that the public input is congestible by private inputs, so that (2) is an input in the production function of firms (3). Similar calculations as used for the derivation of condition (17) yield:[25]

$$\frac{\partial\widetilde{X}}{\partial\widetilde{A}}\frac{\widetilde{A}}{\widetilde{X}} + \frac{\partial\widetilde{X}}{\partial\widetilde{K}}\frac{\widetilde{K}}{\widetilde{X}} = 1\,. \tag{18}$$

Therefore, if congestion is caused by an increasing employment of primary factors, the efficient firm size is obtained for that point of production, for which the sum of production elasticities of private factors equals one (regardless of the degree of congestion). This is a major difference to the case that congestion is caused by an increasing number of firms, in which (17) must hold for the efficient firm size. Condition (18) is also identical to the condition for an efficient firm size for pure public inputs. In both cases for congestibility of the public factor according to (2) and for pure public factors, the production function of the firm exhibits constant returns to scale in private factors (assuming a constant service derived from infrastructure provision) for the long-run efficient output level of a firm. Also, in this case output is fully exhausted by the marginal products of labour and capital times the quantities of this private factors, i.e., no rent accrues to the firms from public input provision.[26] The reason that condition (18) is, in addition to pure public inputs, also obtained for this specification of congestibility lies in the allocation of private inputs between the X- and the B-industry according to efficiency condition (16): Efficiency in the allocation of private factors between the public and the private sector requires accounting for the loss in

23 It can also be taken from the RHS of (17b) that the product of γ and the production elasticity of $\widetilde{B}$ is equal to the percentage contraction in output due to an increase in the number of firms by one percent, and the resulting decline in the public factor available to firms $[(\partial\widetilde{X}/\partial\widetilde{B}) \cdot (\partial\widetilde{B}/\partial m) \cdot (m/\widetilde{X}) < 0]$.

24 It can be shown that this is also true for a combination of both congestion technologies, as sometimes assumed in endogenous growth models with public capital (see section 2).

25 Condition (18) is derived in the Appendix.

26 Note that in case of congestion that is caused by the number of firms, the optimal firm size is, in contrast, not obtained in a competitive economy with free entry and exit for firms that pay private factors according to their marginal products. It may, however, be enforceable if firms are charged an appropriate price for the provision of the public input. One must admit that it would be difficult to enforce the optimal firm size through forced reorganisations such as a merger or break-up of firms.

production due to congestion that is caused by the private factors of production. As a result, (direct) marginal productivities (and production elasticities) of labour and capital are higher compared to congestion technology (1).

4.3. Scale Elasticity of the Aggregate Production Function

The efficiency conditions derived above are used to derive the nature of the returns to scale in the aggregate production function. Solving the first order conditions (9)–(14) for the optimal quantities of the choice variables, and inserting them in the objective function yields the indirect objective function $X = X(A_0, K_0)$. Applying the envelope theorem on the indirect objective function yields:

$$\frac{\partial X}{\partial A_0} = \frac{\partial L}{\partial A_0}; \qquad \frac{\partial X}{\partial K_0} = \frac{\partial L}{\partial K_0}. \tag{19}$$

The scale elasticity of the aggregate production function ($\varepsilon_{X/\lambda}$) equals the sum of production elasticities of labour and capital:

$$\varepsilon_{X/\lambda} = \frac{\partial X}{\partial A_0}\frac{A_0}{X} + \frac{\partial X}{\partial K_0}\frac{K_0}{X} \equiv \varepsilon_{X/A_0} + \varepsilon_{X/K_0}, \tag{20}$$

where for $\varepsilon_{X/\lambda}$ greater (equal, less) than one, the aggregate production function exhibits increasing (constant, decreasing) returns to scale. For the production elasticity of A_0, we derive with (19), (8) and (4):

$$\varepsilon_{X/A_0} = \lambda_1 \frac{m\widetilde{A} + A_B}{m\widetilde{X}}. \tag{21}$$

Substitution from (9), (12) and (14) for λ_1 yields:[27]

$$\varepsilon_{X/A_0} = \frac{\partial \widetilde{X}}{\partial \widetilde{A}}\frac{\widetilde{A}}{\widetilde{X}} + \frac{\partial \widetilde{X}}{\partial \widetilde{B}}\frac{\widetilde{B}}{\widetilde{X}}\frac{\partial B}{\partial A_B}\frac{A_B}{B}. \tag{22}$$

The production elasticity of K_0 is obtained in analogy to (22):

$$\varepsilon_{X/K_0} = \frac{\partial \widetilde{X}}{\partial \widetilde{K}}\frac{\widetilde{K}}{\widetilde{X}} + \frac{\partial \widetilde{X}}{\partial \widetilde{B}}\frac{\widetilde{B}}{\widetilde{X}}\frac{\partial B}{\partial K_B}\frac{K_B}{B}. \tag{23}$$

Using (20) and the production elasticities of A_0 and K_0 from above, we can derive the scale elasticity of the aggregate production function as the sum of production elasticities of the primary factors of production, labour ($\widetilde{A}$) and capital ($\widetilde{K}$), and the production elasticity of the public input ($\widetilde{B}$) times the sum of production elasticities of labour (A_B) and capital (K_B) employed in the public sector:

$$\varepsilon_{X/\lambda} = \frac{\partial \widetilde{X}}{\partial \widetilde{A}}\frac{\widetilde{A}}{\widetilde{X}} + \frac{\partial \widetilde{X}}{\partial \widetilde{K}}\frac{\widetilde{K}}{\widetilde{X}} + \frac{\partial \widetilde{X}}{\partial \widetilde{B}}\frac{\widetilde{B}}{\widetilde{X}} \cdot \left(\frac{\partial B}{\partial A_B}\frac{A_B}{B} + \frac{\partial B}{\partial K_B}\frac{K_B}{B}\right). \tag{24}$$

27 For the derivation of (22) it has been used that $B \cdot (\partial\widetilde{B}/\partial B) = \widetilde{B}$.

Finally, using the definition for the returns to scale in public input production (6), one obtains for the scale elasticity of the aggregate production function for an efficient size of the firm according to (17):[28]

$$\varepsilon^{*}_{X/\lambda} = 1 + \frac{\partial \widetilde{X}}{\partial \widetilde{B}} \frac{\widetilde{B}}{\widetilde{X}} \cdot (\varepsilon_{B/\lambda} - \gamma)\,, \tag{25a}$$

which can also be stated as:

$$\varepsilon^{*}_{X/\lambda} = 1 + \varepsilon_{B/\lambda} \cdot \frac{\partial \widetilde{X}}{\partial \widetilde{B}} \frac{\widetilde{B}}{\widetilde{X}} + \frac{\partial \widetilde{X}}{\partial \widetilde{B}} \frac{\partial \widetilde{B}}{\partial m} \frac{m}{\widetilde{X}}\,. \tag{25b}$$

The results can be summed up as follows: If returns to scale in public input production ($\varepsilon_{B/\lambda}$) exceed (are equal to, are below) the degree of rivalry of the public input, and for an efficient size of the industry according to (17), the aggregate production function exhibits increasing (constant, diminishing) returns to scale.[29] For increasing returns to scale in public production, the aggregate production function exhibits increasing returns to scale even for fully congestible public inputs ($\gamma = 1$), and would only exhibit non-increasing returns to scale for the extreme situation, in which a one percent increase in the number of firms (m) results in a decline in the public factor available to firms ($\widetilde{B}$) by more than one percent ($\gamma > 1$). If constant returns to scale in public input production are assumed, the aggregate production function exhibits increasing returns to scale except for the case of a publicly-provided private good, for which the scale elasticity equals one.

Turning now to the case that congestion is caused by an increasing employment of private factors according to (2). It is straightforward to show that in this case the aggregate production function exhibits increasing returns to scale, unless the public factor is highly congestible by an additional employment of private factors of production, labour and/or capital. More precisely, for an efficient size of a firm according to (18), we have for the scale elasticity of the aggregate production function:[30]

$$\varepsilon^{*}_{X/\lambda} = 1 + \frac{\partial \widetilde{X}}{\partial \widetilde{B}} \frac{\widetilde{B}}{\widetilde{X}} \cdot (\varepsilon_{B/\lambda} - \alpha - \beta)\,. \tag{26a}$$

In analogy to the derivation of (25b), (26a) can be transformed into:

$$\varepsilon^{*}_{X/\lambda} = 1 + \varepsilon_{B/\lambda} \cdot \frac{\partial \widetilde{X}}{\partial \widetilde{B}} \frac{\widetilde{B}}{\widetilde{X}} + \frac{\partial \widetilde{X}}{\partial \widetilde{B}} \frac{\partial \widetilde{B}}{\partial m\widetilde{A}} \frac{m\widetilde{A}}{\widetilde{X}} + \frac{\partial \widetilde{X}}{\partial \widetilde{B}} \frac{\partial \widetilde{B}}{\partial m\widetilde{K}} \frac{m\widetilde{K}}{\widetilde{X}}\,. \tag{26b}$$

28 $\varepsilon^{*}_{X/\lambda}$ is the scale elasticity of the aggregate production function for an efficient size of each firm, while $\varepsilon_{X/\lambda}$ determined in (20) is the general scale elasticity of the aggregate production function.

29 As stated in section 3, it is assumed that the production elasticity of the public factor $\widetilde{B}$ is positive, i.e., public input provision is productivity-enhancing.

30 Condition (26) is derived in the Appendix.

Regarding the nature of the returns to scale in the aggregate production function, we can derive the following result: If the sum of the degrees of congestion $\alpha + \beta$ is below (equal to, greater than) the scale elasticity in production of the public input, the aggregate production function exhibits increasing (constant, diminishing) returns to scale. Assuming increasing returns to scale in public input production, non-increasing returns to scale in aggregate production may occur only if the public factor is more than fully congestible by private factors ($\alpha + \beta > 1$).[31]

The results indicate that the nature of returns to scale in aggregate production with congestible public inputs is driven by two separate sources. The first is increasing returns to scale in public input production and the second is increasing returns to scale in consumption of the public input. With respect to the first source, Matsumoto and Feehan (2010) have provided some evidence, that the assumption of scale economies in public input production is plausible. The literature is, however, less clear on the second source. Although there is a broad agreement that public goods are subject to some degree of congestion, little is known about the extent of the congestion externality, especially when the public factor under consideration is a productive input in production. Besides other factors, the extent of "publicness" of the public factor may depend on the type of input considered (e.g., certain productive infrastructure), and on the level of government (e.g., state or local) at which the public input is provided. Although methodological issues have been raised in some studies, median voter demand studies tend to indicate that local public goods are subject to a rather high degree of congestion, close to publicly-provided private goods (e.g. Edwards, 1990; Reiter and Weichenrieder, 1997). In order for an aggregate production function to exhibit increasing returns to scale, minor economies of scale in consumption is sufficient, while high degrees of congestion will, of course, limit the scope for scale economies.

The results can provide a useful input for work that seeks to specify an aggregate production function with a congestible public factor, and where the number of firms is considered endogenous such as, for instance, in the literature on fiscal competition. The derived conditions for productive efficiency in this section indicate that the production technology exhibits increasing

31 Non-increasing returns to scale may therefore only occur in the situation, in which a one percent increase in labour and capital decreases the public factor available to the firm by more than one percent. This would be the case, for instance, if a one percent increase in both factors decreases the public factor available per firm by more than 0.5 percent each. If, as sometimes assumed in literature, congestion is caused by one factor (labour or capital) only, a one percent increase in this factor is required to decrease the public input per firm by more than one percent for the possibility of non-increasing returns to scale to occur.

returns to scale (with decreasing returns in private inputs) if congestion is caused by the number of firms, rather than constant returns to scale as assumed, for instance, in Matsumoto (2000).[32] Despite the common assumption of congestible public factors in production, it is important to point out that results do not directly carry over to endogenous growth models that fix the number of agents in the economy, and that assume that agents are consumers and producers.[33]

In addition to the well-known specification of an aggregate production function in the two polar cases of congestion, i.e., constant returns to scale in private and constant returns to scale in all inputs (e.g. Feehan, 1998), the present paper postulates that in a third specification of a production function with a congestible public input, the sum of production elasticities of primary factors is dependent on the congestion specification (and the degree of congestion), exactly as determined by the conditions for the optimal firm size (see equations (17) and (18)). With respect to the congestion specification, there is still a lack of knowledge what exactly drives congestion of productive public factors, i.e., more emphasise is needed on the specification of the congestion technology. There are, in principle, more alternatives to specify congestion than in the case of public consumption goods because in addition to firms, private inputs of production have been identified as the users of the public factor. As the production and congestion technologies have been shown to be interconnected, care should be taken in choosing both specifications independently or making certain a-priori assumptions on the nature of returns to scale in production.

32 Observe that the specification of the production technology in Matsumoto (2000, p. 245, footnote 5) considers congestion as introduced in equation (1) of the present paper.

33 Following Barro (1990), a fixed number of households that act as consumer-producers is assumed, for instance, in Futagami and Shibata (1993), Glomm and Ravikumar (1994, 1997), Fisher and Turnovsky (1998), Dioikitopoulos and Kalyvitis (2008), Eicher and Turnovsky (2000), and Gómez (2008). In contrast, Barro and Sala-I-Martin (1992), Turnovsky (1996) and Ott and Soretz (2008) do not explicitly emphasise on the number of firms or consider alternatives, where the number of households can differ from the number of firms. Imposing the (additional) assumption that the number of firms is variable and the entry for firms is free, one can derive the condition for the efficient firm size in these models. Specifications that combine relative congestion with a production technology that exhibits constant returns to scale in primary inputs would then obviously not be in line with the additional requirement of an efficient firm size for intermediate degrees (no polar case) of congestion (e.g. Eicher and Turnovsky, 2000; Ott and Soretz, 2008; Dioikitopoulos and Kalyvitis, 2008). On the other side, some studies (e.g. Barro and Sala-I-Martin, 1992; Glomm and Ravikumar, 1994, 1997) are consistent with the derived conditions for productive efficiency as their specification of the production technology is also the one that emerges when the size of a representative firm is optimal.

5. Concluding Remarks

The paper has derived conditions for productive efficiency in an industry on the basis of two alternative public input specifications subject to congestion, and the assumption of a variable number of firms in an industry. Depending on the congestion technology and the degree of congestion of the public factor, modified Samuelson-conditions for an efficient provision of congestible public inputs have been presented. The conditions for an optimal size of a firm have been found to differ with respect to the inspected types of congestion. Accounting for all conditions for productive efficiency, the approach was able to solve for the nature of the returns to scale in the aggregate production function as a result of the maximisation process in an industry. The analysis shows how the returns to scale exactly varies with the congestion technology, the degree of congestion, and the returns to scale in public input production. As all of these factors may vary for different productive infrastructures, the specification of the aggregate production function should also be case sensitive to the public input under consideration. Making the usual assumptions on the returns to scale may therefore be problematic, as for an inconsistent specification of the aggregate production function output is not maximised, and the economy is not producing on the production possibility frontier.

As a main result of the analysis, the aggregate production function of an industry can be shown to exhibit non-increasing returns to scale for a rather limited spectrum of productive inputs subject to congestion. This may only occur when the public factor is fully congestible by the number of firms or by private factors of production, which has been argued to be the case if the (sum of) degree(s) of congestion of the congestion-causing factor(s) equals one. However, although more or less congestible, productive public factors that are publicly-provided private goods cover only a limited spectrum of all conceivable cases (Samuelson's 'knife-edge pole' for public consumption goods). Still, for increasing returns to scale in public input production, the aggregate production function exhibits increasing returns to scale even in this case, as the productivity gain from public input provision more than offsets the loss in production due to congestion. As a major difference between both public input specifications, the production function exhibits decreasing returns to scale in private factors, when the public factor is congestible by the number of firms, whereas the production technology exhibits constant returns to scale in private factors (for a constant service derived from the public factor) if congestion is caused by the private factors of production. As a further difference, modified Samuelson-conditions differ for both public input specifications, and the marginal rates of technical substitution in the private and in the public sector are not necessarily equal if congestion is caused by the private factors of production.

6. Appendix

6.1. Derivation of Conditions (16) and (18)

Start with the Lagrangian:

$$L(\widetilde{A}, \widetilde{K}, \widetilde{B}, m, A_B, K_B) = m\widetilde{X}(\widetilde{A}, \widetilde{K}, \widetilde{B}) + \lambda_1(A_0 - m\widetilde{A} - A_B) + \lambda_2(K_0 - m\widetilde{K} - K_B) + \lambda_3\left[\widetilde{B} - \frac{B(A_B, K_B)}{(m\widetilde{A})^{\alpha}\,(m\widetilde{K})^{\beta}}\right]. \tag{27}$$

The first-order conditions for a maximum are:

$$\frac{\partial L}{\partial \widetilde{A}} = m\frac{\partial \widetilde{X}}{\partial \widetilde{A}} - \lambda_1 m + \lambda_3 \alpha B(m\widetilde{A})^{-\alpha-1}(m\widetilde{K})^{-\beta} m = 0 \tag{28}$$

$$\frac{\partial L}{\partial \widetilde{K}} = m\frac{\partial \widetilde{X}}{\partial \widetilde{K}} - \lambda_2 m + \lambda_3 \beta B(m\widetilde{K})^{-\beta-1}(m\widetilde{A})^{-\alpha} m = 0 \tag{29}$$

$$\frac{\partial L}{\partial m} = \widetilde{X} - \lambda_1\widetilde{A} - \lambda_2\widetilde{K} + \lambda_3 B[\alpha(m\widetilde{A})^{-\alpha-1}\widetilde{A}(m\widetilde{K})^{-\beta} + \beta(m\widetilde{A})^{-\alpha}\widetilde{K}(m\widetilde{K})^{-\beta-1}] = 0 \tag{30}$$

$$\frac{\partial L}{\partial A_B} = -\lambda_1 - \lambda_3(m\widetilde{A})^{-\alpha}(m\widetilde{K})^{-\beta}\frac{\partial B}{\partial A_B} = 0 \tag{31}$$

$$\frac{\partial L}{\partial K_B} = -\lambda_2 - \lambda_3(m\widetilde{A})^{-\alpha}(m\widetilde{K})^{-\beta}\frac{\partial B}{\partial K_B} = 0 \tag{32}$$

$$\frac{\partial L}{\partial \widetilde{B}} = m\frac{\partial \widetilde{X}}{\partial \widetilde{B}} + \lambda_3 = 0\,. \tag{33}$$

From (28), (31), and (33) one obtains conditions (16a) and (16b). Solving (28) and (29) for λ_1 and λ_2, insertion in (30), and rearranging yields (18).

6.2. Derivation of Condition (26)

In order to derive the scale elasticity of the aggregate production function, we apply (19)–(21) from above, together with (27). From (28) and (33) we have for λ_1:

$$\lambda_1 = \frac{\partial \widetilde{X}}{\partial \widetilde{A}} - m\frac{\partial \widetilde{X}}{\partial \widetilde{B}}\alpha B(m\widetilde{A})^{-\alpha-1}(m\widetilde{K})^{-\beta}\,. \tag{34}$$

From (31) and (33), one obtains:

$$\lambda_1 = m\frac{\partial \widetilde{X}}{\partial \widetilde{B}}(m\widetilde{A})^{-\alpha}(m\widetilde{K})^{-\beta}\frac{\partial B}{\partial A_B}\,. \tag{35}$$

Inserting (34) and (35) in (21) yields:

$$
\begin{aligned}
\varepsilon_{X/A_0} &= \frac{\partial \widetilde{X}}{\partial \widetilde{A}} \frac{\widetilde{A}}{\widetilde{X}} - m \frac{\partial \widetilde{X}}{\partial \widetilde{B}} \alpha B (m\widetilde{A})^{-\alpha-1} (m\widetilde{K})^{-\beta} \frac{\widetilde{A}}{\widetilde{X}} \\
&+ m \frac{\partial \widetilde{X}}{\partial \widetilde{B}} (m\widetilde{A})^{-\alpha} (m\widetilde{K})^{-\beta} \frac{\partial B}{\partial A_B} \frac{A_B}{m\widetilde{X}} .
\end{aligned} \tag{36}
$$

And, after rearranging:

$$
\varepsilon_{X/A_0} = \frac{\partial \widetilde{X}}{\partial \widetilde{A}} \frac{\widetilde{A}}{\widetilde{X}} + \frac{\partial \widetilde{X}}{\partial \widetilde{B}} \frac{B}{\widetilde{X}} (m\widetilde{A})^{-\alpha} (m\widetilde{K})^{-\beta} \left(\frac{\partial B}{\partial A_B} \frac{A_B}{B} - \alpha \right) . \tag{37}
$$

Because of (2), the production elasticity of A_0 is:

$$
\varepsilon_{X/A_0} = \frac{\partial \widetilde{X}}{\partial \widetilde{A}} \frac{\widetilde{A}}{\widetilde{X}} + \frac{\partial \widetilde{X}}{\partial \widetilde{B}} \frac{\widetilde{B}}{\widetilde{X}} \left(\frac{\partial B}{\partial A_B} \frac{A_B}{B} - \alpha \right) . \tag{38}
$$

For the production elasticity of K_0, we derive:

$$
\varepsilon_{X/K_0} = \frac{\partial \widetilde{X}}{\partial \widetilde{K}} \frac{\widetilde{K}}{\widetilde{X}} + \frac{\partial \widetilde{X}}{\partial \widetilde{B}} \frac{\widetilde{B}}{\widetilde{X}} \left(\frac{\partial B}{\partial K_B} \frac{K_B}{B} - \beta \right) . \tag{39}
$$

Summing up both elasticities according to (20) yields:

$$
\varepsilon_{X/\lambda} = \frac{\partial \widetilde{X}}{\partial \widetilde{A}} \frac{\widetilde{A}}{\widetilde{X}} + \frac{\partial \widetilde{X}}{\partial \widetilde{K}} \frac{\widetilde{K}}{\widetilde{X}} + \frac{\partial \widetilde{X}}{\partial \widetilde{B}} \frac{\widetilde{B}}{\widetilde{X}} \left(\frac{\partial B}{\partial A_B} \frac{A_B}{B} + \frac{\partial B}{\partial K_B} \frac{K_B}{B} - \alpha - \beta \right) . \tag{40}
$$

For an efficient size of the industry as determined by (18), the scale elasticity of the aggregate production function can be expressed by (26).

References

Abe, K. (1990), A Public Input as a Determinant of Trade, Canadian Journal of Economics 23, 400–407.

Altenburg, L. (1987), Production Possibilities with a Public Intermediate Good, Canadian Journal of Economics 20, 715–734.

Altenburg, L. (1992), Some Trade Theorems with a Public Intermediate Good, Canadian Journal of Economics 25, 310–332.

Arnold, V. (1980), Steigende Skalenerträge bei der Bereitstellung öffentlicher Zwischenprodukte, FinanzArchiv/Public Finance Analysis 38, 256–273.

Aschauer, D. A. (1989), Is Public Expenditure Productive?, Journal of Monetary Economics 23, 177–200.

Barro, R. (1990), Government Spending in a Simple Model of Endogenous Growth, Journal of Political Economy 98, 103–126.

Barro, R., and Sala-I-Martin, X. (1992), Public Finance in Models of Economic Growth, Review of Economic Studies 59, 645–661.

Batina, R. G. (1999), On the Long Run Effect of Public Capital on Aggregate Output: Estimation and Sensitivity Analysis, Empirical Economics 24, 711–717.

Boadway, R. W., and Flatters, F. R. (1982), Efficiency and Equalization Payments in a Federal System of Governments: A Synthesis and Extension of Recent Results, Canadian Journal of Economics 15, 613–633.

Buettner, T. (1999), Determinants of Tax Rates in Local Capital Income Taxation: A Theoretical Model and Evidence from Germany, FinanzArchiv/Public Finance Analysis 56, 363–388.

Chang, J., Hung, H., Shieh, J., and Lai, C. (2007), Optimal Fiscal Policies, Congestion and Over-Entry, Scandinavian Journal of Economics 109, 137–151.

Cole, H. L., and Ohanian, L. E. (1999), Aggregate Returns to Scale: Why Measurement is Imprecise, Federal Reserve Bank of Minneapolis Quarterly Review 23, 19–28.

Colombier, C. (2008), Efficiency in Public Infrastructure Provision – A Theoretical Note, Journal of Economic Studies 35, 528–543.

Colombier, C., and Pickhardt, M. (2002), Public Inputs – Some Clarifications and a Taxonomy, in: Barens, I., and Pickhardt, M. (Eds.), Die Rolle des Staates in der Oekonomie – Finanzwissenschaftliche Perspektiven, Metropolis, Marburg, 253–286.

Colombier, C., and Pickhardt, M. (2005), A Note on Public Input Specifications, International Advances in Economics Research 11, 13–18.

Dioikitopoulos, E. V., and Kalyvitis, S. (2008), Public Capital Maintenance and Congestion: Long-Run Growth and Fiscal Policies, Journal of Economic Dynamics & Control 32, 3760–3779.

Duncombe, W., and Yinger, J. (1993), An Analysis of Returns to Scale in Public Production, with an Application to Fire Protection, Journal of Public Economics 52, 49–72.

Edwards, J. H. Y. (1990), Congestion Function Specification and the Publicness of Local Public Goods, Journal of Urban Economics 27, 80–96.

Eicher, T., and Turnovsky, S. J. (2000), Scale, Congestion and Growth, Economica 67, 325–346.

Feehan, J. P. (1989), Pareto-Efficiency with Three Varieties of Public Input, Public Finance 44, 237–48.

Feehan, J. P. (1998), Public Investment: Optimal Provision of Hicksian Public Inputs, Canadian Journal of Economics 31, 693–707.

Felipe, J., and Fisher, F. M. (2003), Aggregation in Production Functions, What Applied Economists Should Know, Metroeconomica 54, 208–262.

Fisher, F. M. (1969), The Existence of Aggregate Production Functions, Econometrica 37, 553–577.

Fisher, W. H., and Turnovsky, S. J. (1998), Public Investment, Congestion, and Private Capital Accumulation, Economic Journal 108, 399–413.

Futagami, M. Y., K. and Shibata, A. (1993), Dynamic Analysis of an Endogenous Growth Model with Public Capital, Scandinavian Journal of Economics 95, 607–625.

Glomm, G., and Ravikumar, B. (1994), Public Investment in Infrastructure in a Simple Growth Model, Journal of Economic Dynamics and Control 18, 1173–1187.

Glomm, G., and Ravikumar, B. (1997), Productive Government Expenditures and Long-Run Growth, Journal of Economic Dynamics and Control 21, 183–204.

Gómez, M. A. (2008), Fiscal Policy, Congestion, and Endogenous Growth, Journal of Public Economic Theory 10, 595–622.

Gramlich, E. M. (1994), Infrastructure Investment: A Review Essay, Journal of Economic Literature 18, 1176–1196.

Henderson, J. V. (1974), A Note on the Economics of Public Intermediate Inputs, Economica 41, 322–327.

Hicks, J. R. (1939), The Foundations of Welfare Economics, Economic Journal 49, 696–712.

Hillman, A. L. (1978), Symmetries and Asymmetries Between Public Input and Public Good Equilibria, Public Finance 33, 267–279.

Kaizuka, K. (1965), Public Goods and Decentralization of Production, Review of Economics and Statistics 47, 118–120.

Kohn, R. E. (1985), A General Equilibrium Analysis of the Optimal Number of Firms in a Polluting Industry, Canadian Journal of Economics 18, 247–354.

Kunce, M., and Shogren, J. F. (2005), On Efficiency of Decentralized Environmental Regulation, Journal of Regulatory Economics 28, 129–140.

Levy, D. (1990), Aggregate Output, Capital, and Labor in the Post-War U.S. Economy, Economics Letters 33, 41–45.

Martínez, D., and Sánchez-Fuentes, A. J. (2011), The Optimal Policy in the Provision of Public Inputs: To What Extent do Technology, Taxation and Preferences Matter?, Journal of Economic Policy Reform 14, 259–271.

Mason, C. F., and Polasky, S. (1997), The Optimal Number of Firms in the Commons: A Dynamic Approach, Canadian Journal of Economics 30, 1143–1160.

Matsumoto, M. (1998), A Note on Tax Competition and Public Input Provision, Regional Science and Urban Economics 28, 465–473.

Matsumoto, M. (2000), A Tax Competition Analysis of Congestible Public Inputs, Journal of Urban Economics 48, 242–259.

Matsumoto, M., and Feehan, J. P. (2010), Capital-Tax Financing and Scale Economies in Public-Input Production, Regional Science and Urban Economics 40, 116–121.

McMillan, J. (1979), A Note on the Economics of Public Intermediate Goods, Public Finance 34, 293–299.

Mohring, H. (1976), Transportation Economics, Ballinger, Cambridge, MA.

Ott, I., and Soretz, S. (2008), Growth Strategies: Fiscal Versus Institutional Policies, Economic Modelling 25, 605–622.

Reiter, M., and Weichenrieder, A. (1997), Are Public Goods Public? A Critical Survey of the Demand Estimates for Local Public Services, FinanzArchiv/Public Finance Analysis 54, 374–408.

Samuelson, P. A. (1969), Pure Theory of Public Expenditure and Taxation, in: Margolis, J., and Guitton, H. (Eds.), Public Economics: An Analysis of Public Production and Consumption and Their Relations to the Private Sector, Basingstoke, London, 98–123.

Sandmo, A. (1972), Optimality Rules for the Provision of Collective Factors of Production, Journal of Public Economics 1, 149–157.

Shephard, R. W. (1970), Theory of Cost and Production Functions, Princeton University Press, Princeton.

Stiglitz, J. E. (1988), Economics of the Public Sector, Norton and Company, New York, NY.

Tawada, M. (1980), The Production Possibility Set with Public Intermediate Goods, Econometrica 48, 1005–1012.

Turnovsky, S. J. (1996), Optimal Tax, Debt, and Expenditure Policies in a Growing Rconomy, Journal of Public Economics 60, 21–44.

Manipulating Fiscal Forecasts: Evidence from the German States

Björn Kauder, Niklas Potrafke, Christoph Schinke*

Received 22 February 2015; in revised form 13 May 2016; accepted 4 January 2017

We examine whether German state governments manipulated fiscal forecasts before elections. Our data set includes three fiscal measures over the period 1980–2014. The results do not show that electoral motives influenced fiscal forecasts in West German states. By contrast, East German state governments underestimated spending in preelection years (compared to other years) by about 0.20 percent of GDP, tax revenues by about 0.36 percent of GDP, and net lending by 0.30 percent of GDP. Predicting low levels of spending and tax revenues, East German state governments thus underestimated the size of government in preelection years.

Keywords: fiscal forecasts, electoral cycles, East and West Germany

JEL classification: H 68, E 32, E 62

1. Introduction

Governments prepare forecasts on tax revenues, spending, and deficits. Most realizations do not, of course, meet the forecast values. An important question is whether fiscal forecast errors simply result from unforeseeable circumstances, or whether forecast errors are tantamount to manipulation by governments. There are political incentives towards manipulation. In times of an approaching election, for example, governments may use fiscal forecasts to boost reelection prospects (the political business-cycle theory asserts that politicians use expansionary policies before elections).[1] By manipulat-

* Kauder: Ifo Institute, Ifo Center for Public Finance and Political Economy and University of Munich, Poschingerstr. 5, D-81679 Munich, Germany (kauder@ifo.de); Potrafke: Ifo Institute, Ifo Center for Public Finance and Political Economy and University of Munich, Poschingerstr. 5, D-81679 Munich, Germany (potrafke@ifo.de); Schinke: Ifo Institute, Ifo Center for Public Finance and Political Economy, Poschingerstr. 5, D-81679 Munich, Germany (christophschinke@hotmail.com). We would like to thank Christian Breuer, Robert Lehmann, Felix Rösel, and Christian Thater for their helpful comments. Dominik Hecker, Lion Henrich, Julia Richenhagen, and Julia Tubes provided excellent research assistance. Financial support by the Fritz Thyssen Foundation is gratefully acknowledged.

1 On how electoral motives influence fiscal policy, see, for example, Berger and Woitek (1997), de Haan and Klomp (2013), Efthyvoulou (2012), Katsimi and Sarantidis (2012),

FinanzArchiv 73 (2017), 213–236 doi: 10.1628/001522117X14877521353438
ISSN 0015-2218

ing tax revenue, spending, or deficit forecasts, parties that champion tax cuts or increased spending wish to convey the impression that policies they advocate are fundable. Voters endorsing such reforms may then be inclined to reconsider their vote. Against the background of the political business-cycle theories, the hypothesis to be tested is clear-cut: governments are overoptimistic and sugarcoat fiscal forecasts before elections by predicting too high tax revenues and too low spending and deficits.

Scholars examine whether electoral motives and government ideology influence fiscal forecasts. We discuss related studies in section 2 and for now focus on Germany. Fiscal forecasts at the German *federal* level were biased towards overoptimism over the period 1968–2003: deficit forecasts were lower before elections; deficit, tax, and spending forecasts were lower under right-wing governments (Heinemann, 2006). For short-term tax revenue forecasts over the period 1971–2013, the results of Buettner and Kauder (2015) are not indicative of a bias, electoral cycles, or an influence of government ideology; the government influenced the revenue forecasts, however, by providing the underlying GDP forecast and revenue estimates of tax-law changes. Medium-term tax revenue forecasts between 1968 and 2012 were biased upwards, in particular after the German reunification (Breuer, 2015). For the West German *states* over the period 1992–2002, the results of Bischoff and Gohout (2010) do not give rise to the conclusion that electoral motives and government ideology influenced tax projections. Tax projections were greater, however, the more voters disliked incumbent parties.

Our contribution is twofold. We examine whether politicians manipulated spending, tax revenue, and net lending forecasts at the German state level. We also investigate differences in strategic manipulation of fiscal variables between East and West German state governments. The results show that in preelection years (compared to other years) East German state governments underestimated spending by about 0.20 percent of GDP, tax revenues by 0.36 percent of GDP, and net lending by 0.30 percent of GDP. Predicting low levels of spending and tax revenues, East German state governments thus underestimated the size of government. In contrast, the results do not show that electoral motives influenced fiscal forecasts in West German states.

2. Related Literature

Experts investigate the quality of forecasts in terms of precision and accuracy, as measured, for example, by the standard deviation of the forecast error.

Klomp and de Haan (2013), Lane (2003), Seitz (2000), and Shi and Svensson (2006). See Debrun et al. (2009) and Wyplosz (2008) on fiscal councils. On election-motivated policies in the German states see, for example, Tepe and Vanhuysse (2009), Mechtel and Potrafke (2013), and Kauder et al. (2016).

In OECD countries, the timing of forecasts, uncertainty about GDP growth rates, and independence of forecasting institutions from government were shown to influence the accuracy of revenue forecasts (Buettner and Kauder, 2010). In U.S. states, forecast accuracy increased with independent forecasting agencies and decreased when there was a dominant political party (Deschamps, 2004; Bretschneider et al., 1989). Revenue forecast accuracy also increased when states employed politically appointed and merit-selected forecasters (Krause et al., 2006).

Testing the precision and accuracy of forecasts refers to the forecasting techniques. To test whether governments manipulate forecasts before elections, experts examine the rationality of forecasts in terms of unbiasedness and efficiency, as measured, for example, by the relative forecast error (see Keane and Runkle, 1989 and 1990; Nordhaus, 1987; Holden and Peel, 1990). Do individual factors give rise to overly optimistic or overly pessimistic (and hence *biased*) forecasts? Do forecasters incorporate all relevant information available at the time of the forecast preparation (*efficiency*)?

Many empirical studies investigated the rationality of fiscal forecasts in cross-country analyses or in individual countries.[2] In member states of the European Union, budget balance forecasts were overoptimistic before elections (Brück and Stephan, 2006; Merola and Pérez, 2013; Pina and Venes, 2011). The results of von Hagen (2010), however, do not corroborate election-year effects. Budget forecasts were also too optimistic during boom periods and when the budget deficit was high (Frankel, 2011; Frankel and Schreger, 2013). Jonung and Larch (2006) portray the nexus between growth forecasts and budget balances and suggest that having independent forecasts may avoid political biases (see also Beetsma et al., 2009). In OECD countries, electoral motives do not appear to have influenced fiscal balance reviews (Cimadomo, 2012; Jong-A-Pin et al., 2012). Left-wing governments, however, produced more optimistic revenue forecasts than right-wing governments (Jochimsen and Lehmann, in press).

In the United States (federal level), evidence suggests that revenue forecasts of the Office of Management and Budget (OMB) and the Congressional Budget Office (CBO) were not biased, spending and thus deficits were underestimated, and forecast revisions were serially correlated; biases were larger under Republican administrations (Auerbach, 1999; Blackley and DeBoer, 1993; Campbell and Ghysels, 1995; Plesko, 1988). In U.S. states, revenue forecasts were shown to be unbiased but inefficient (Mocan and Azad, 1995). Revenue forecasts for election years, however, were shown to be overly optimistic (Boylan, 2008). Conservatives were overoptimistic in

2 See Kyobe and Danninger (2005) and Schroeder and Wasylenko (1989) for revenue forecasting in low-income countries. Leal et al. (2008) discuss “lessons and challenges” from fiscal forecasting in the European Union.

forecasting sales tax revenues in years without tax increases (Bretschneider and Gorr, 1992). The results of Cassidy et al. (1989) do not suggest that government ideology influenced forecast errors. In three U.S. states, forecasts were shown to be downward biased (Feenberg et al., 1989).

In Belgian municipalities, two-party governments were more optimistic in forecasting tax revenues than single-party governments (Goeminne et al., 2008). In Swiss cantons, revenue forecasts were more pessimistic under left-wing finance ministers than under right-wing finance ministers (Chatagny, 2015), and pessimistic revenue forecasts were shown to reduce spending and thus fiscal deficits (Chatagny and Soguel, 2012).[3] Also in the United Kingdom, political factors influenced revenue forecasts (Paleologou, 2005).

The mixed evidence on forecasting performance produced by the individual studies corroborates that exploring political determinants of fiscal forecast errors is a worthwhile endeavor. Whether German state governments manipulated fiscal forecasts remains an undetermined empirical question.

3. Institutional Backdrop

3.1. Budget Rules

The German constitution prescribes in Article 109 that the states are autonomous and independent from the federal level in setting up their budgets. In 2009, the so-called debt brake was introduced, providing that state budgets should in principle be balanced without borrowing as of 2020. Exceptions can be made for business fluctuations, natural disasters, and other cases of emergency, if specific rules describe how credits are repaid. State governments can decide on whether they want to comply with the debt brake earlier and how a balanced budget is to be reached (see, for example, Potrafke et al., 2016). It is unclear, however, whether there will be sanctions if a state fails to consolidate its budget until 2020 (Fuest and Thöne, 2013). To be sure, the federal debt brake does not make any prescriptions for the states' fiscal policies until 2019. Since 2009, twelve states have introduced debt brakes at the state level.

Most states' constitutions require that borrowing be warranted by a law. Borrowing must moreover not exceed spending for investment; exceptions are only possible to maintain the "overall economic equilibrium." Many states, however, have disregarded the law so that borrowing exceeded investment.

3 See Chatagny and Siliverstovs (2015) on the rationality of tax revenue forecasts under asymmetric loss functions.

3.2. Projections of Fiscal Figures

The Federal Minister of Finance Franz Josef Strauß (Christian Social Union – CSU) and his successor Alex Möller (Social Democratic Party – SPD) introduced medium-term planning in 1968 at both the federal and the state level. Medium-term plans are set up in the budgeting process and include fiscal forecasts for the current and the following four years (see also Lübke, 2008). Forecast figures include, among others, spending, tax revenues, and net lending. Though states also receive transfers from the federal level and from the other states via the financial equalization scheme, tax revenues are the most important source of revenue. Tax revenue forecasts are prepared by the independent tax revenue forecast group (*Arbeitskreis Steuerschätzungen*) on the federal level. The subcommittee on regionalization calculates how much tax revenues may accrue to the individual states. The state governments adjust these figures for reasons such as the timing of the budgetary process, economic development of the state, or tax reforms.

For some years in individual states, medium-term plans are not available, because in some cases state governments passed a budget for two years, and thus published medium-term plans only every other year. We focus on the most important figures, referring to years t and $t+1$, because governments' budget plans are based on the forecasts for the years t and $t+1$.

3.3. State Elections

Elections in the German states take place every five years. The only exceptions are Hamburg and Bremen, where they take place every four years. In the past, other states held elections every four years. Parliaments may also call early elections. Out of 109 elections in our sample, 11 were early elections. In most states, voters cast two votes in a personalized proportional representation system. The first vote determines which candidate is to obtain the direct mandate in one of the electoral districts, with a relative majority. With the second vote, voters select an individual party. The parties obtain a number of the seats in parliament that corresponds to the party's second-vote share. Candidates voted into the parliament with the first vote (direct mandate) obtain their seats first. Candidates from party lists obtain the remaining seats.

4. Empirical Analysis

4.1. Descriptive Statistics

We use the fiscal forecasts from 1980 to 2014 for West German states and from 1996 to 2014 for East German states as published by the ministries of

Table 1a

Descriptive Statistics for All States

Forecast errors (in percent of *ex post* state GDP)	Obs.	ME	RMSE	Min.	Max.
Total spending, year t	398	0.011	0.376	−1.916	1.601
Total spending, year $t+1$	389	−0.062	0.415	−2.597	1.230
Total spending, year $t+2$	374	−0.124	0.516	−2.728	1.380
Total spending, year $t+3$	358	−0.148	0.674	−2.602	2.566
Total spending, year $t+4$	343	−0.135	0.820	−2.690	2.271
Tax revenues, year t	405	−0.032	0.398	−1.549	1.217
Tax revenues, year $t+1$	390	−0.012	0.508	−1.549	1.387
Tax revenues, year $t+2$	375	0.111	0.728	−1.732	1.866
Tax revenues, year $t+3$	359	0.268	0.867	−1.960	2.656
Tax revenues, year $t+4$	344	0.443	0.965	−1.764	2.314
Net lending, year t	399	−0.216	0.517	−2.407	1.510
Net lending, year $t+1$	390	−0.151	0.733	−2.358	6.281
Net lending, year $t+2$	375	−0.096	0.908	−3.609	6.227
Net lending, year $t+3$	359	0.012	0.910	−3.646	3.900
Net lending, year $t+4$	344	0.114	0.886	−3.766	3.675
***Ex post* realizations (in percent of state GDP)**	**Obs.**	**Mean**	**Std. Dev.**	**Min.**	**Max.**
Total spending	450	14.685	5.284	8.571	30.239
Tax revenues	450	8.505	1.258	6.363	11.821
Net lending	450	−1.013	1.130	−6.692	2.156
Unemployment rate	474	10.873	4.466	2.300	22.100
GDP growth rate (nominal)	474	3.168	2.637	−10.000	10.900
State government ideology (left)	474	0.525	0.453	0.000	1.000
Last forecast before election	474	0.207	0.405	0.000	1.000
Last forecast before election that induced regime change	474	0.084	0.278	0.000	1.000
Last forecast before election that did not induce regime change	474	0.122	0.328	0.000	1.000
Education level	144	13.285	3.452	8.096	25.876
Unemployment rate relative to state average	474	0.986	0.325	0.438	2.216
Fiscal rule	474	0.074	0.262	0.000	1.000

Note: ME = mean error; RMSE = root-mean-square error.

finance in the individual states. We exclude fiscal forecasts from the East German states before 1996 and from Berlin between 1990 and 1995, because of the German reunification. Table 1a shows descriptive statistics for all states. A positive (negative) forecast error indicates that the expected value of a fiscal variable was overstated (understated) compared to the ex post realization. Average forecast errors for total spending and tax revenues for the same year and the next year were less than 0.07 percent of GDP. Average forecast errors for net lending were larger: net lending for the same year and the next year was underestimated, respectively, by 0.22 percent of GDP and by 0.15 percent of GDP on average. The root-mean-square error of forecasts for the same year is 0.38 percent of GDP for total spending, 0.40 for tax revenues, and 0.52 for net lending. Root-mean-square errors increase as the forecast horizon increases. Tables 1b and 1c show descriptive statistics separately for East German states and West German states.

Table 1b
Descriptive Statistics for East German States

Forecast errors (in percent of *ex post* state GDP)	Obs.	ME	RMSE	Min.	Max.
Total spending, year t	104	0.139	0.520	−1.916	1.601
Total spending, year $t+1$	98	−0.002	0.562	−2.597	1.230
Total spending, year $t+2$	92	−0.181	0.638	−2.728	1.380
Total spending, year $t+3$	86	−0.243	0.830	−2.602	2.566
Total spending, year $t+4$	79	−0.220	1.006	−2.690	2.271
Tax revenues, year t	106	−0.029	0.520	−1.549	1.217
Tax revenues, year $t+1$	100	−0.050	0.648	−1.549	1.387
Tax revenues, year $t+2$	94	0.092	0.939	−1.732	1.866
Tax revenues, year $t+3$	88	0.260	1.147	−1.960	2.079
Tax revenues, year $t+4$	81	0.514	1.296	−1.764	2.314
Net lending, year t	105	−0.357	0.720	−2.407	1.380
Net lending, year $t+1$	99	−0.248	1.143	−2.358	6.281
Net lending, year $t+2$	93	−0.187	1.453	−3.609	6.227
Net lending, year $t+3$	87	−0.010	1.444	−3.646	3.900
Net lending, year $t+4$	80	0.084	1.424	−3.766	3.675
Ex post **realizations (in percent of state GDP)**	**Obs.**	**Mean**	**Std. Dev.**	**Min.**	**Max.**
Total spending	124	22.255	3.537	16.095	30.239
Tax revenues	124	10.000	1.036	7.058	11.608
Net lending	124	−1.120	1.648	−6.692	2.156
Unemployment rate	124	16.088	4.024	4.300	22.100
GDP growth rate (nominal)	124	2.346	2.108	−4.400	8.200
State government ideology (left)	124	0.504	0.380	0.000	1.000
Last forecast before election	124	0.218	0.414	0.000	1.000
Last forecast before election that induced regime change	124	0.137	0.345	0.000	1.000
Last forecast before election that did not induce regime change	124	0.081	0.273	0.000	1.000
Education level	54	13.380	4.381	8.653	25.876
Unemployment rate relative to state average	124	1.345	0.164	0.963	1.671
Fiscal rule	124	0.113	0.318	0.000	1.000

Note: ME = mean error; RMSE = root-mean-square error.

Figure 1 shows the forecast errors for three fiscal measures in years t and $t+1$. We distinguish between the last fiscal forecast before a state election (in light gray) and other forecasts (in dark gray). We call the last fiscal forecast before a state election the "preelection forecast" henceforth, as opposed to "other forecasts." Whiskers describe 95-percent confidence intervals. Total spending was always underestimated, except in forecasts for year t in other years. Forecasts of total spending before elections and in other years appear to differ. Tax-revenue forecast errors were quite small and similar before elections and in other years. Net lending was always underestimated, i.e., deficits were lower than predicted. The difference between forecast errors before elections and in other years hardly ever attains statistical significance.

The results may differ between East and West German states because institutions have developed differently between 1949 and 1990, and institutional differences may influence fiscal forecasts after the reunification. Figure 2

Table 1c
Descriptive Statistics for West German States

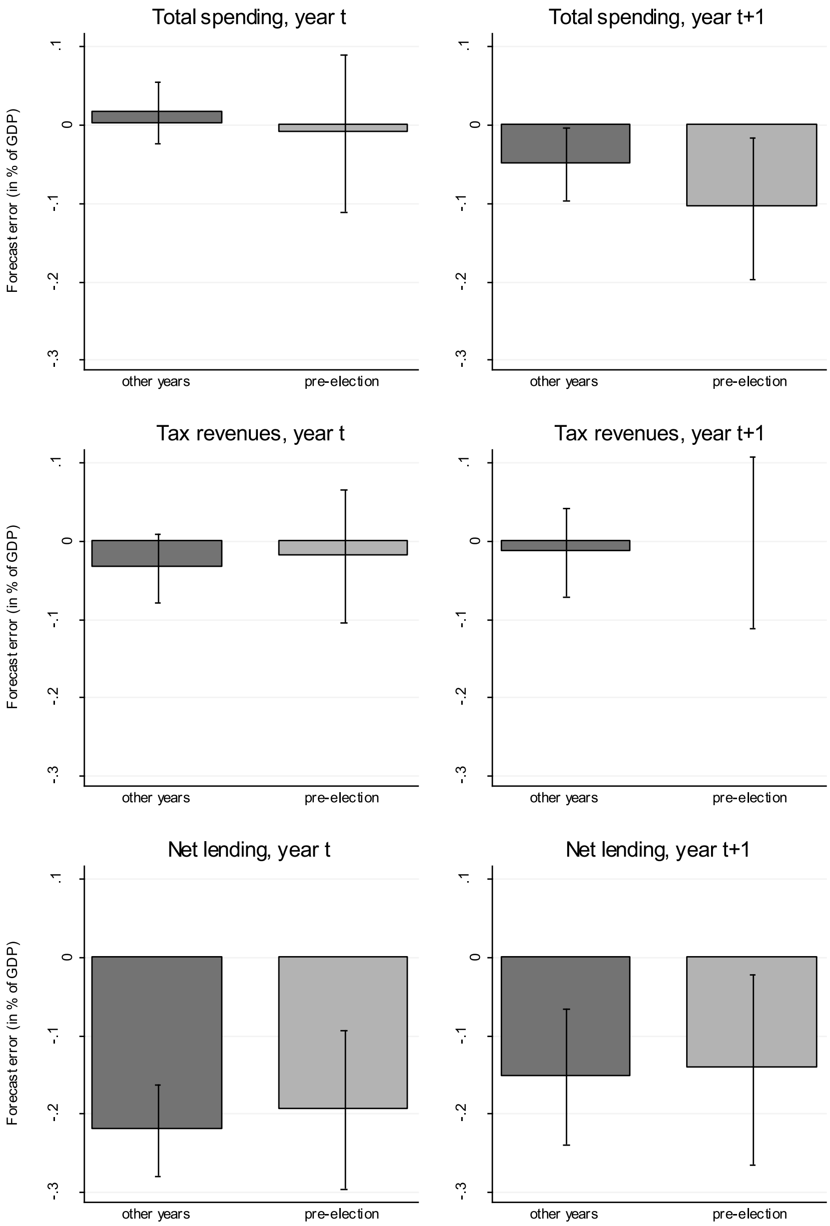
Forecast errors (in percent of *ex post* state GDP)	Obs.	ME	RMSE	Min.	Max.
Total spending, year t	294	−0.035	0.298	−1.596	1.511
Total spending, year $t+1$	291	−0.082	0.351	−1.734	0.888
Total spending, year $t+2$	282	−0.105	0.469	−2.065	1.262
Total spending, year $t+3$	272	−0.117	0.615	−2.582	1.397
Total spending, year $t+4$	264	−0.109	0.756	−2.646	1.524
Tax revenues, year t	299	−0.033	0.346	−1.330	1.077
Tax revenues, year $t+1$	290	0.001	0.450	−1.330	1.264
Tax revenues, year $t+2$	281	0.117	0.644	−1.619	1.814
Tax revenues, year $t+3$	271	0.271	0.756	−1.678	2.656
Tax revenues, year $t+4$	263	0.421	0.839	−1.641	2.299
Net lending, year t	294	−0.165	0.412	−2.106	1.510
Net lending, year $t+1$	291	−0.118	0.524	−2.325	1.777
Net lending, year $t+2$	282	−0.066	0.635	−2.883	1.962
Net lending, year $t+3$	272	0.019	0.657	−2.626	2.628
Net lending, year $t+4$	264	0.123	0.645	−2.366	2.488
Ex post **realizations (in percent of state GDP)**	**Obs.**	**Mean**	**Std. Dev.**	**Min.**	**Max.**
Total spending	326	11.805	1.911	8.571	18.032
Tax revenues	326	7.936	0.777	6.363	11.821
Net lending	326	−0.973	0.855	−4.784	1.008
Unemployment rate	350	9.025	2.869	2.300	18.300
GDP growth rate (nominal)	350	3.460	2.745	−10.000	10.900
State government ideology (left)	350	0.533	0.477	0.000	1.000
Last forecast before election	350	0.203	0.403	0.000	1.000
Last forecast before election that induced regime change	350	0.066	0.248	0.000	1.000
Last forecast before election that did not induce regime change	350	0.137	0.344	0.000	1.000
Education level	90	13.227	2.776	8.096	24.118
Unemployment rate relative to state average	350	0.858	0.268	0.438	2.216
Fiscal rule	350	0.060	0.238	0.000	1.000

Note: ME = mean error; RMSE = root-mean-square error.

shows the results separately for East and West German states. In many cases, the difference between preelection forecast errors and other forecast errors was larger in East German states than in West German states.[4] In East German states, forecast errors were mostly lower before elections than in other years. Forecast errors for total spending in year t (year $t+1$) were on average 0.11 percent of GDP (0.23 percent of GDP) lower before elections than in other years. The difference between the total-spending forecast errors before elections and in other years for the next year in the East attains statistical significance at the 10-percent level. Forecast errors of tax revenues in year t (year $t+1$) were on average 0.01 percent of GDP (0.02 percent of GDP) lower (higher) before elections than in other years. Forecast errors

4 Eastern firms also predict their productivity less accurately than Western firms (Triebs and Tumlinson, 2013).

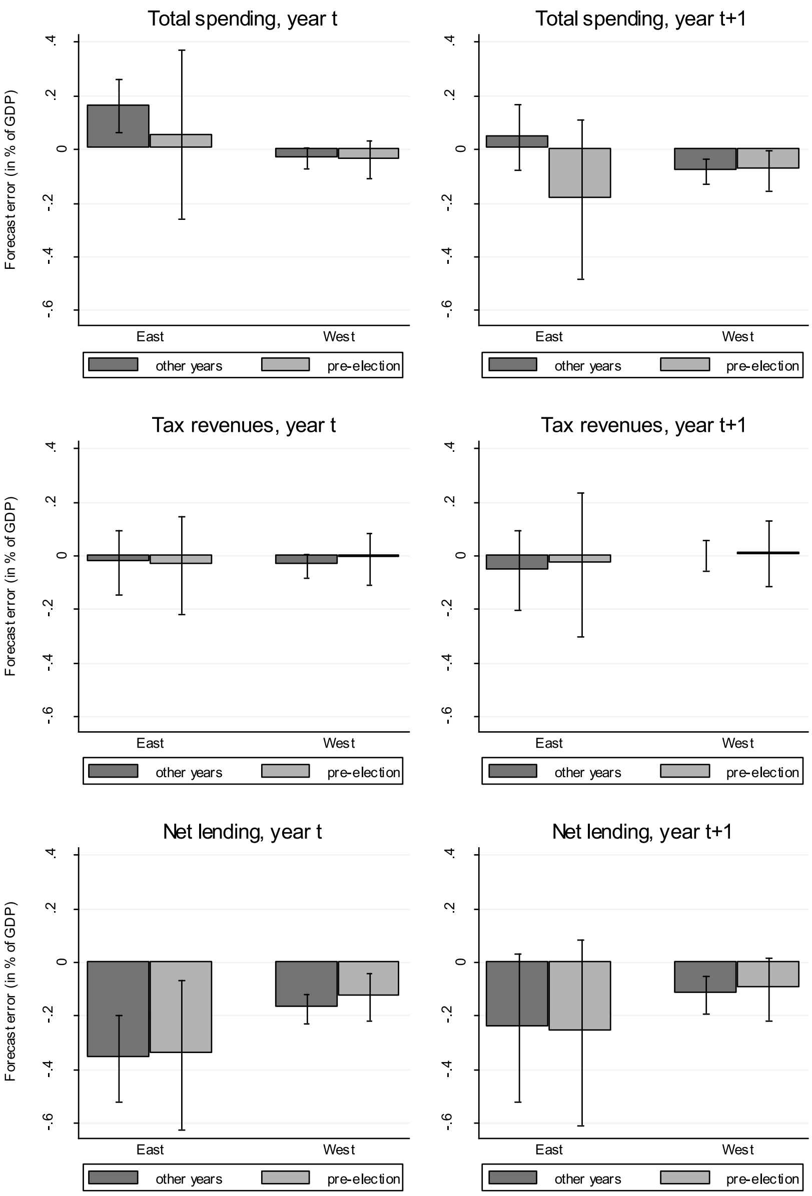
Figure 1
Forecast Errors in Preelection Years and Other Years



Note: The differences between preelection and other years do not turn out to be statistically significant. Whiskers describe 95-percent confidence intervals.

Figure 2

Forecast Errors by Region in Preelection Years and Other Years

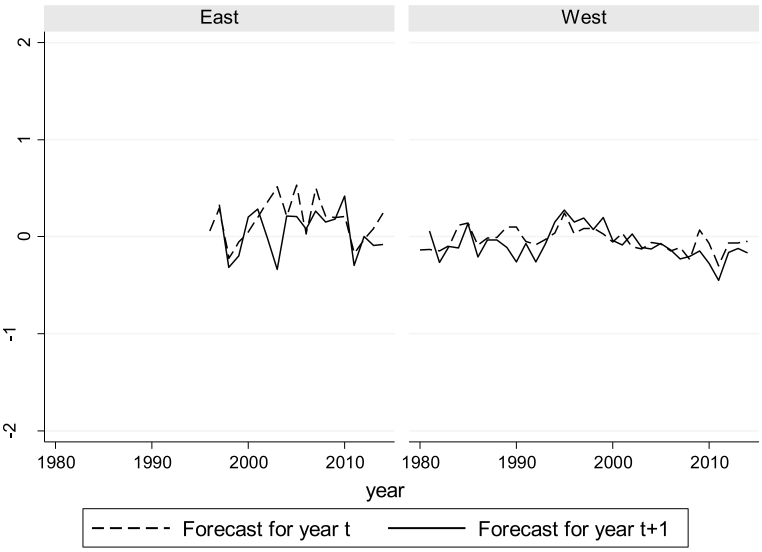


Note: The difference between preelection and other years is statistically significant at the 10-percent level for total spending in year $t+1$ in East German states. Whiskers describe 95-percent confidence intervals.

of net lending in year t (year $t+1$) were on average 0.01 percent of GDP (0.02 percent of GDP) higher (lower) before elections than in other years.

Figures 3a–c show how the forecast errors for the three fiscal measures in years t and $t+1$ evolved over time. Because the uncertainty differs, forecast errors for year t are in absolute values smaller than forecast errors for year $t+1$. Forecast errors in absolute values are larger in East German states, in particular for tax revenues and net lending.

Figure 3a
Total-spending Forecast Errors, 1980–2014



4.2. Empirical Strategy

The basic empirical model has the following form:

$$
\begin{aligned}
\textit{Forecast error}_{ijkt} &= \beta_{jk}\,\textit{Preelection}_{it} + \Sigma_l\,\delta_{jkl}\,X_{ilt} \\
&\quad + \varepsilon_{jk}\textit{Forecast error}_{ijkt-1} + \eta_{ijk} + \tau_{jkt} + u_{ijkt} \\
&\text{with } i = 1, ..., 16, \quad j = 1, ..., 3, \quad k = 0, 1, l = 1, ..., 3, \\
&\quad t = 1980, ..., 2014,
\end{aligned}
$$

where $\textit{Forecast error}_{ijkt}$ describes the difference between forecast and realized value for forecast type j (total spending, tax revenues, and net lending) relative to realized GDP with forecast horizon k (0 or 1) in state i in period t.

Figure 3b
Tax-revenue Forecast Errors, 1980–2014

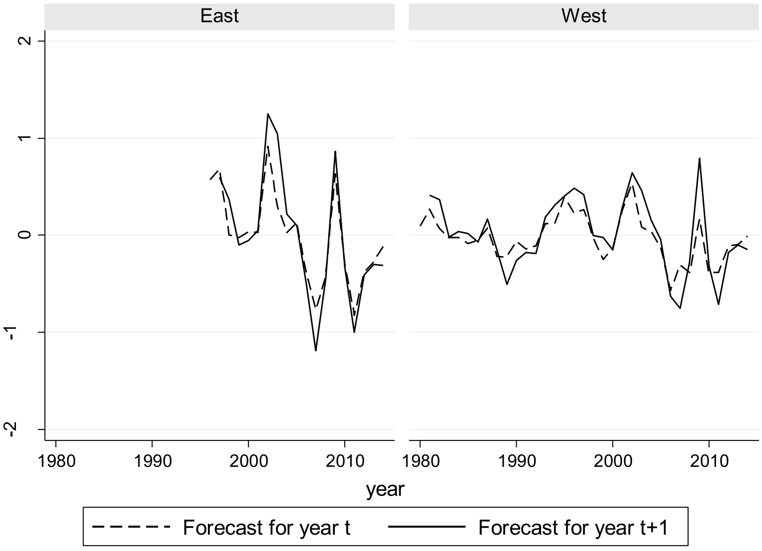
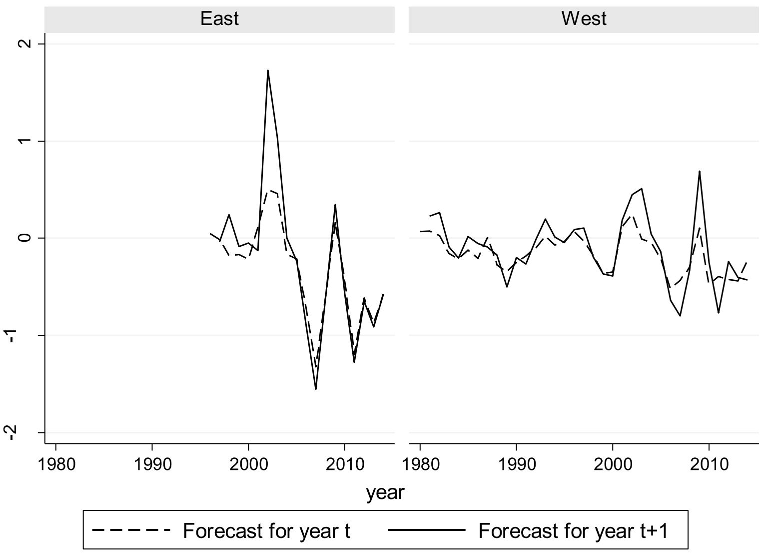


Figure 3c
Net-lending Forecast Errors, 1980–2014



The dummy variable $Preelection_{it}$ assumes the value 1 when the forecast was the last forecast issued before a regular state election (predetermined elections are exogenous explanatory variables). $\Sigma_l X_{ilt}$ contains three control variables. We include the ideological orientation of the government.[5] We also include the unemployment rate to take account of different incentives to manipulate forecasts in economically good and bad times.[6] Finally, we include the variable whose forecast error we consider as a share of realized GDP from one period ago, to control for mean reversion. $Forecast\ error_{ijkt-1}$ describes the lagged dependent variable to control for autocorrelation of forecast errors. η_i is a fixed state effect, τ_t is a fixed time effect, and u_{it} is the error term.

We estimate fixed-effects models with standard errors robust to heteroskedasticity (Huber-White sandwich standard errors – Huber, 1967; White, 1980). Including the lagged dependent variable gives rise to Nickell bias (Nickell, 1981), which is however small (1/T).

4.3. Regression Results

Table 2 shows the results for all states. Column (1) shows the coefficient estimates for the forecast of total spending for the same year (the preelection year), and column (2) shows the results for the next year (the election year). The number of observations decreases as the forecast horizon increases. The coefficient of the election variable and the coefficient of the government ideology variable do not turn out to be statistically significant. The coefficient of the lagged forecast error is significant in columns (1) and (2). The numerical meaning of the coefficient in column (1) is that when the lagged forecast error increases by 1 percent of GDP, the current forecast error increases by 0.32 percent of GDP. The coefficient of the lagged unemployment rate lacks statistical significance. Columns (3) to (6) show the results for tax revenues and net lending. The coefficient of the election variable does not turn out to be statistically significant in any specification. The coefficient of the government ideology variable is statistically significant in column (5). The

5 We distinguish between left-wing and right-wing governments on a left-right scale by using the variable *Left*, which takes on the value 1 in periods when a left-wing government was in office (SPD without a coalition partner, or SPD in a coalition with the Greens, the left-wing party Die Linke, or the Free Democratic Party (FDP)), 0.5 when a center government was in office (coalition of the Christian Democratic Union (CDU) with the SPD or the Greens, or with the Greens and the FDP), and 0 when a right-wing government was in office (CDU/CSU without a coalition partner or in a coalition with the FDP). On ideology-induced policymaking in the German states see, for example, Oberndorfer and Steiner (2007) and Potrafke (2011).

6 Inferences do not change when we use the GDP growth rate instead of the unemployment rate.

numerical meaning of the coefficient is that under left-wing governments, net-lending forecast errors are higher by 0.1 percentage points of GDP than those under right-wing governments. The coefficient of the lagged realization of net lending is statistically significant in column (5).

Table 2

Fixed-effects Regressions with Standard Errors Robust to Heteroskedasticity (Huber-White Sandwich Standard Errors) – All States

	(1) Total spending forecast error, year t	(2) Total spending forecast error, year $t+1$	(3) Tax revenue forecast error, year t	(4) Tax revenue forecast error, year $t+1$	(5) Net lending forecast error, year t	(6) Net lending forecast error, year $t+1$
Preelection forecast	−0.083 (0.050)	−0.076 (0.061)	−0.029 (0.026)	−0.000 (0.031)	0.025 (0.046)	0.008 (0.054)
State government ideology (left)	0.054 (0.043)	0.126 (0.093)	−0.007 (0.038)	−0.009 (0.050)	0.104* (0.058)	0.066 (0.086)
Realization of j (t−1)	0.009 (0.025)	0.015 (0.030)	0.019 (0.053)	−0.018 (0.053)	−0.060* (0.034)	−0.025 (0.059)
Unemployment rate ($t-1$)	0.010 (0.019)	−0.009 (0.011)	0.009 (0.018)	0.021 (0.017)	0.006 (0.020)	0.023 (0.020)
Forecast error ($t-1$)	0.311*** (0.048)	0.188* (0.091)	0.068 (0.054)	0.012 (0.057)	0.086 (0.064)	0.224 (0.171)
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	346	332	351	337	346	332
Groups	16	16	16	16	16	16
Within R^2	0.227	0.167	0.535	0.735	0.380	0.429
Overall R^2	0.270	0.149	0.483	0.697	0.355	0.404

Note: Robust standard errors in parentheses (Huber-White sandwich standard errors); * $p<0.10$, *** $p<0.01$.

We estimate our basic empirical model separately for the East and West German states. Table 3 shows the results for East German states (excluding Berlin).[7] The coefficient of the preelection variable is negative and statistically significant for total spending in year t and year $t+1$ (columns 1 and 2), tax revenues in year t (column 3), and net lending in year t (column 5). The numerical meaning of the coefficient in column (1) is that in preelection years, total spending is underestimated by 0.20 percent of GDP (compared to other years). Tax revenues are underestimated by 0.36 percent of GDP in preelection years (column 3); net lending is underestimated by 0.30 percent

7 We cannot distinguish the East German part of Berlin from the West German part of Berlin. We therefore include Berlin only in the regressions for all 16 states.

of GDP in preelection years (column 5). Note that the forecast errors for government spending and tax revenues do not add up to the forecast error for net lending. The discrepancy arises from forecast errors for revenues from sources other than taxes, such as transfers from the federal level, revenues from state-owned companies, capital receipts, fees, and fines. Yet, taxes are the most important source of revenues in all states. The coefficient of state government ideology is statistically significant for total spending in year t and year $t+1$ and for tax revenues in year t. The numerical meaning of the coefficient in column (1) is that under left-wing governments, total spending is overestimated by 0.66 percent of GDP more than under right-wing governments.

Table 3

Fixed-effects Regressions with Standard Errors Robust to Heteroskedasticity (Huber-White Sandwich Standard Errors) – East German States

	(1) Total spending forecast error, year t	(2) Total spending forecast error, year $t+1$	(3) Tax revenue forecast error, year t	(4) Tax revenue forecast error, year $t+1$	(5) Net lending forecast error, year t	(6) Net lending forecast error, year $t+1$
Preelection forecast	−0.198*	−0.552***	−0.362**	−0.251	−0.296***	−0.099
	(0.083)	(0.102)	(0.114)	(0.209)	(0.059)	(0.202)
State government ideology (left)	0.660*	0.686**	0.254**	0.077	0.244	−0.222
	(0.307)	(0.243)	(0.091)	(0.070)	(0.533)	(0.390)
Realization of j $(t-1)$	−0.073	0.073	1.014**	−0.489	0.071	0.503
	(0.171)	(0.071)	(0.273)	(0.330)	(0.205)	(0.243)
Unemployment rate $(t-1)$	−0.097	−0.134	−0.256**	0.035	−0.221	−0.264***
	(0.063)	(0.063)	(0.061)	(0.064)	(0.133)	(0.020)
Forecast error $(t-1)$	0.113	0.054	0.052	−0.078	−0.064	0.445*
	(0.149)	(0.043)	(0.096)	(0.211)	(0.224)	(0.209)
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	64	59	65	60	64	59
Groups	5	5	5	5	5	5
Within R^2	0.363	0.500	0.783	0.905	0.657	0.872
Overall R^2	0.311	0.507	0.402	0.808	0.406	0.509

Note: Robust standard errors in parentheses (Huber-White sandwich standard errors); $^{*}p<0.10$, $^{**}p<0.05$, $^{***}p<0.01$.

Table 4 shows the results for West German states. The coefficient of the election variable lacks statistical significance in all specifications.

We also used forecast errors for total spending, tax revenues, and net lending in years $t+2$, $t+3$, and $t+4$ as dependent variables. The coefficient of the election variable does not turn out to be statistically significant in any specification, except for net lending at the $t+3$-years forecast horizon, which in preelection years (compared to other years) is underestimated by 0.47 percent of GDP in East German states (results not shown).

Table 4

Fixed-effects Regressions with Standard Errors Robust to Heteroskedasticity (Huber-White Sandwich Standard Errors) – West German States

	(1) Total spending forecast error, year t	(2) Total spending forecast error, year $t+1$	(3) Tax revenue forecast error, year t	(4) Tax revenue forecast error, year $t+1$	(5) Net lending forecast error, year t	(6) Net lending forecast error, year $t+1$
Preelection forecast	−0.052	−0.016	−0.013	0.021	0.039	0.019
	(0.046)	(0.034)	(0.015)	(0.024)	(0.049)	(0.044)
State government ideology (left)	−0.009	0.098	−0.013	−0.003	0.151	0.062
	(0.043)	(0.062)	(0.038)	(0.049)	(0.092)	(0.072)
Realization of j $(t-1)$	0.072**	0.178***	0.041	−0.011	0.022	−0.007
	(0.028)	(0.033)	(0.060)	(0.071)	(0.034)	(0.050)
Unemployment rate $(t-1)$	0.038	−0.004	0.014	0.009	−0.022	−0.007
	(0.024)	(0.036)	(0.020)	(0.020)	(0.029)	(0.027)
Forecast error $(t-1)$	0.261***	0.133	0.029	0.039	0.179	−0.071
	(0.038)	(0.097)	(0.092)	(0.058)	(0.110)	(0.110)
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	261	253	265	257	261	253
Groups	10	10	10	10	10	10
Within R^2	0.255	0.315	0.539	0.721	0.355	0.493
Overall R^2	0.126	0.0398	0.423	0.682	0.348	0.404

Note: Robust standard errors in parentheses (Huber-White sandwich standard errors); ** $p<0.05$, *** $p<0.01$.

4.4. Robustness Tests

We submitted all of our results to several robustness tests. In our baseline model, we included fixed time effects. We tested whether inferences change when we do not include fixed time effects, but include the deviation between the GDP forecast of the Federal government as underlying the official revenue forecasts and the actual GDP, to measure economic uncertainty (at the national level). Inferences regarding the election variable do not change. When we do not include a lagged dependent variable in the regressions, inferences do not change either.

We have included other control variables. Inferences regarding the election variable do not change when we include variables measuring the level of education of voters (percent of population above 15 years with university degree),[8] the state unemployment rate relative to the German average, or a variable that assumes the value one when a state has a fiscal rule (debt brake) included in the constitution or in the state budget code.

The results may depend on including irregular elections. The only irregular election in East Germany was in Berlin in 2001. Berlin is not included in the regressions reported in table 3. There were 10 irregular elections in West Germany over the period 1980–2014. Inferences for West Germany do not change when we include the irregular elections.

Realizations of fiscal variables after changes in government may be less predictable than realizations after elections that did not give rise to changes in government. There were 43 regular elections that were followed by a change in government ideology, and 61 regular elections that were not followed by a change in government ideology. Replicating the results for the 16 states (table 2) confirms that before elections that induced changes in government ideology (compared to other years), total spending for the next year was underestimated by 0.20 percent of GDP, and tax revenues for the same year were underestimated by 0.07 percent of GDP (both coefficients are statistically significant at the 5-percent level). Replicating the results for West Germany (table 4) confirms that before elections that induced changes in government ideology (compared to other years), tax revenues for the same year were underestimated by 0.05 percent of GDP (the coefficient is statistically significant at the 5-percent level). Because of the limited number of observations, we cannot investigate subsamples in East Germany.

We run placebo tests and replace the preelection variable with dummy variables for other years. When we use a dummy variable for election years and reestimate table 3, the coefficient of the election-year variable is negative and statistically significant in columns (2) and (5). When we use a dummy variable measuring a two-year distance to the next election and reestimate table 3, the coefficient of the dummy variable always lacks statistical significance. When we use a dummy variable measuring a three-year distance to the next election and reestimate table 3, the coefficient of the dummy variable is positive and statistically significant in columns (2) and (3).

We reestimated our regression models for the West German states for the period 1996–2014, i.e., the same period that we examine for the East German states. Inferences regarding the coefficients of the election variable do not change. In particular, the results still do not show a bias in forecasts before elections.

8 Data on education levels in individual states is only available over the period 2005–2014.

We reestimated our regression models for the period 1992–2002 in the West German states to compare our results more closely with Bischoff and Gohout (2010). Our results also do not show (a) that tax revenue forecasts were biased in preelection years or (b) that state government ideology influenced tax revenue forecast errors for the next year.

Forecast errors may have increased during the financial and debt crisis. When we exclude the crisis years 2008 and 2009, we find that in the full sample (replicating table 2) spending for the same year was underestimated by 0.08 percent of GDP in preelection years (compared to other years). The coefficient is significant at the 10-percent level. On replicating table 3 (East Germany), the inferences do not change, except for column (1), where the election variable lacks statistical significance. Replicating table 4, the results still do not show that fiscal forecasts were biased in West Germany in preelection years.

When we exclude individual years, one at a time, we find that the main findings for the East German states are robust. The election variable does not turn out to be statistically significant in column (1) when we exclude the year 1998, 2001, 2005, 2007, 2008, 2010, or 2013, nor in columns (1) and (3) when we exclude the year 2003. The coefficients of the election variable, however, remain negative throughout all specifications.

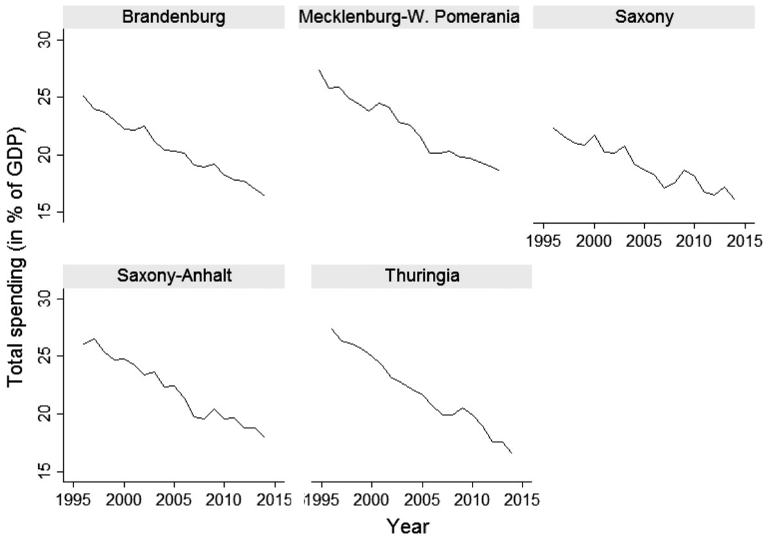
The city states Bremen and Hamburg may differ from other West German states. We reestimated the regressions for all states and for the West German states, excluding Bremen and Hamburg. The inferences regarding the election variable do not change.

Jackknife tests in which we exclude an individual state, one at a time, corroborate that the main findings generalize to most states. In the sample including the East German states, the election variable lacks statistical significance in column (1) when we exclude Brandenburg or Saxony, in columns (1) and (5) when we exclude Mecklenburg-Western Pomerania, and in columns (1), (2), and (5) when we exclude Saxony-Anhalt. When we exclude Thuringia, the election variable does not turn out to be statistically significant in columns (1), (3), and (5). While standard errors increase when we exclude individual states, the coefficients of the election variable remain negative throughout all specifications.

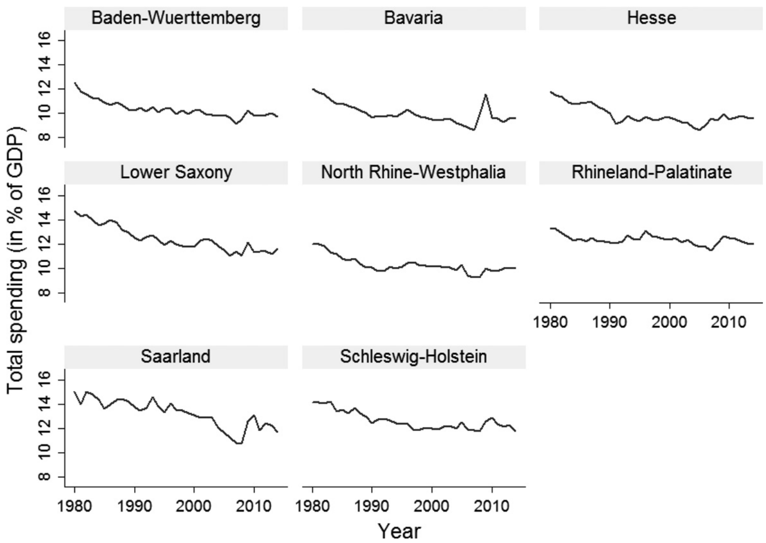
5. Conclusion

Our findings do not indicate that electoral motives influenced fiscal forecasts in West German states, a result that corroborates previous findings of Bischoff and Gohout (2010). By contrast, in preelection years (compared to other years), East German state governments underestimated spending

Figure 4
Total Spending by State, 1980–2014

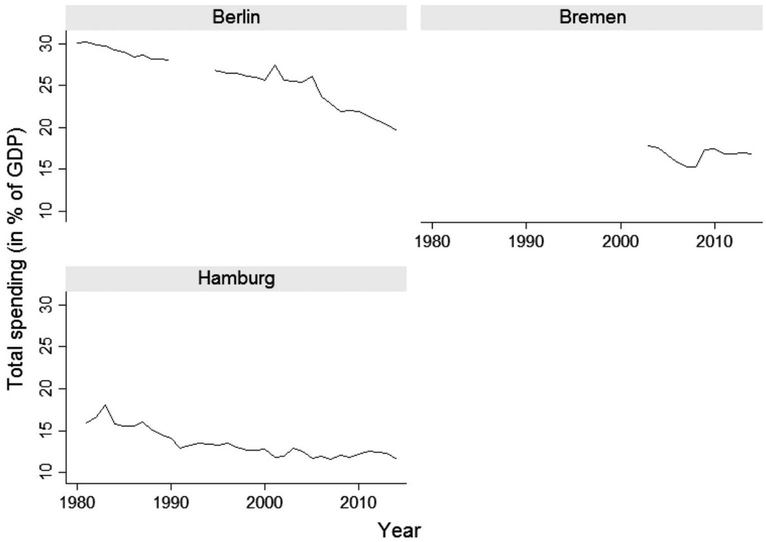


(i) East Germany



(ii) West Germany

Figure 4
Continued.



(iii) City-States

by about 0.20 percent of GDP, tax revenues by 0.36 percent of GDP, and net lending by 0.30 percent of GDP. East German state governments were thus overoptimistic regarding spending and net lending, and overpessimistic regarding tax revenues. Our prediction that governments sugarcoat all three fiscal forecasts by being overoptimistic before elections cannot be corroborated. Predicting low levels of spending and tax revenues, East German state governments rather underestimated the size of government and overestimated their ability to decrease the size of government.

Why is it that East German state governments underestimated the size of government and West German state governments did not?[9] It is well known that the communist experience in Eastern Germany between 1949 and 1990 influenced social norms and attitudes towards government differently from the market-based system in the West (Alesina and Fuchs-Schündeln, 2007; Brosig-Koch et al., 2011). Many studies describe differences between East and West Germans regarding cooperation and solidarity behavior (Ockenfels and Weimann, 1999; Brosig-Koch et al., 2011), individual preferences for social policies and redistribution (Corneo, 2004; Alesina and Fuchs-Schündeln, 2007), and inequality of wages, income, and consumption (Fuchs-Schündeln

9 Previous studies have shown that ideology-induced policies differed in East and West German states (Tepe and Vanhuysse, 2014; Kauder and Potrafke, 2013; Potrafke, 2013).

et al., 2010). We cannot, however, test whether differences in social norms and attitudes towards government give rise to our results. We propose an alternative explanation. At the time of the reunification, Chancellor Helmut Kohl promised “blossoming landscapes” in East Germany, describing a quick convergence in economic prosperity. The size of government in East German states is however still larger than in West German states, some convergence since the 1990s notwithstanding (figure 4). We conjecture that East German state governments wanted to pretend convergence to the West German states by using forecasts in preelection years as a low-cost signaling device. East German politicians may well believe that promising a size of government similar to that in Western states is valued by voters, the stronger preferences for redistribution in East Germany notwithstanding (note that redistribution is a federal task and that the largest share in state government spending is staff spending). Whether voters reward such promises remains however as an open question for further research.

References

Alesina, A., and Fuchs-Schündeln, N. (2007), Good-Bye Lenin (or Not?): The Effect of Communism on People’s Preferences, American Economic Review 97, 1507–1528.

Auerbach, A. J. (1999), On the Performance and Use of Government Revenue Forecasts, National Tax Journal 52, 767–782.

Beetsma, R., Giuliodori, M., and Wierts, P. (2009), Planning to Cheat: EU Fiscal Policy in Real Time, Economic Policy 24, 753–804.

Berger, H., and Woitek, U. (1997), Searching for Political Business Cycles in Germany, Public Choice 91, 179–197.

Bischoff, I., and Gohout, W. (2010), The Political Economy of Tax Projections, International Tax and Public Finance 17, 133–150.

Blackley, P. R., and DeBoer, L. (1993), Bias in OMB’s Economic Forecasts and Budget Proposals, Public Choice 76, 215–232.

Boylan, R. T. (2008), Political Distortions in State Forecasts, Public Choice 136, 411–427.

Bretschneider, S. I., and Gorr, W. L. (1992), Economic, Organizational, and Political Influences on Biases in Forecasting State Sales Tax Receipts, International Journal of Forecasting 7, 457–466.

Bretschneider, S. I., Gorr, W. L., Grizzle, G., and Klay, E. (1989), Political and Organizational Influences on the Accuracy of Forecasting State Government Revenues, International Journal of Forecasting 5, 307–319.

Breuer, C. (2015), On the Rationality of Medium-Term Tax Revenue Forecasts: Evidence from Germany, Jahrbücher für Nationalökonomie und Statistik 235, 22–40.

Brosig-Koch, J., Helbach, C., Ockenfels, A., and Weimann, J. (2011), Still Different after All These Years: Solidarity Behavior in East and West Germany, Journal of Public Economics 95, 1373–1376.

Brück, T., and Stephan, A. (2006), Do Eurozone Countries Cheat with their Budget Deficit Forecasts?, Kyklos 59, 3–15.

Buettner, T., and Kauder, B. (2010), Revenue Forecasting Practices: Differences across Countries and Consequences for Forecasting Performance, Fiscal Studies 31, 313–340.

Buettner, T., and Kauder, B. (2015), Political Biases despite External Expert Participation? An Empirical Analysis of Tax Revenue Forecasts in Germany, Public Choice 164, 287–307.

Campbell, B., and Ghysels, E. (1995), Federal Budget Projections: A Nonparametric Assessment of Bias and Efficiency, Review of Economics and Statistics 77, 17–31.

Cassidy, G., Kamlet, M. S., and Nagin, D. S. (1989), An Empirical Examination of Bias in Revenue Forecasts by State Governments, International Journal of Forecasting 5, 321–331.

Chatagny, F. (2015), Incentive Effects of Fiscal Rules on the Finance Minister's Behavior: Evidence from Revenue Projections in Swiss Cantons, European Journal of Political Economy 39, 184–200.

Chatagny, F., and Siliverstovs, B. (2015), Evaluating Rationality of Level and Growth Rate Forecasts of Direct Tax Revenues under Flexible Loss Function: Evidence from Swiss Cantons, Economics Letters 134, 65–68.

Chatagny, F., and Soguel, N. C. (2012), The Effect of Tax Revenue Budgeting Errors on Fiscal Balance: Evidence from the Swiss Cantons, International Tax and Public Finance 19, 319–337.

Cimadomo, J. (2012), Fiscal Policy in Real Time, Scandinavian Journal of Economics 114, 440–465.

Corneo, G. (2004), Wieso Umverteilung? Einsichten aus ökonometrischen Umfrageanalysen, in: Genser, B. (Ed.), Finanzpolitik und Umverteilung, Duncker & Humblot, Berlin, 55–88.

Debrun, X., Hauner, D., and Kumar, M. S. (2009), Independent Fiscal Agencies, Journal of Economic Surveys 23, 44–81.

De Haan, J., and Klomp, J. (2013), Conditional Political Budget Cycles: A Review of Recent Evidence, Public Choice 157, 387–410.

Deschamps, E. (2004), The Impact of Institutional Change on Forecast Accuracy: A Case Study of Budget Forecasting in Washington State, International Journal of Forecasting 20, 647–657.

Efthyvoulou, G. (2012), Political Budget Cycles in the European Union and the Impact of Political Pressures, Public Choice 153, 295–327.

Feenberg, D. R., Gentry, W., Gilroy, D., and Rosen, H. S. (1989), Testing the Rationality of State Revenue Forecasts, Review of Economics and Statistics 71, 300–308.

Frankel, J. (2011), Over-optimism in Forecasts by Official Budget Agencies and its Implications, Oxford Review of Economic Policy 27, 536–562.

Frankel, J., and Schreger, J. (2013), Over-optimistic Official Forecasts and Fiscal Rules in the Eurozone, Review of World Economics 149, 247–272.

Fuchs-Schündeln, N., Krueger, D., and Sommer, M. (2010), Inequality Trends for Germany in the Last Two Decades: A Tale of Two Countries, Review of Economic Dynamics 13, 103–132.

Fuest, C., and Thöne, M. (2013), Durchsetzung der Schuldenbremse in den Bundesländern, FiFo Köln.

Goeminne, S., Geys, B., and Smolders, C. (2008), Political Fragmentation and Projected Tax Revenues: Evidence from Flemish Municipalities, International Tax and Public Finance 15, 297–315.

Heinemann, F. (2006), Planning or Propaganda? An Evaluation of Germany's Medium-term Budgetary Planning, FinanzArchiv/Public Finance Analysis 62, 551–578.

Holden, K., and Peel, D. A. (1990), On Testing for Unbiasedness and Efficiency of Forecasts, Manchester School 58, 120–127.

Huber, P. J. (1967), The Behavior of Maximum Likelihood Estimates under Nonstandard Conditions, Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, 221–233.

Jochimsen, B., and Lehmann, R. (in press), On the Political Economy of National Tax Revenue Forecasts: Evidence from OECD Countries, Public Choice.

Jong-A-Pin, R., Sturm, J. E., and de Haan, J. (2012), Using Real-Time Data to Test for Political Budget Cycles, CESifo Working Paper No. 3939.

Jonung, L., and Larch, M. (2006), Improving Fiscal Policy in the EU: The Case for Independent Forecasts, Economic Policy 21, 491–534.

Katsimi, M., and Sarantidis, V. (2012), Do Elections Affect the Composition of Fiscal Policy in Developed Established Democracies?, Public Choice 151, 325–362.

Kauder, B., Krause, M., and Potrafke, N. (2016), Electoral Cycles in MPs' Salaries: Evidence from the German States, CESifo Working Paper No. 6028.

Kauder, B., and Potrafke, N. (2013), Government Ideology and Tuition Fee Policy: Evidence from the German States, CESifo Economic Studies 59, 628–649.

Keane, M. P., and Runkle, D. E. (1989), Are Economic Forecasts Rational? Quarterly Review of the Federal Reserve Bank of Minneapolis 13, 26–33.

Keane, M. P., and Runkle, D. E. (1990), Testing the Rationality of Price Forecasts: New Evidence from Panel Data, American Economic Review 80, 714–735.

Klomp, J., and de Haan, J. (2013), Political Budget Cycles and Election Outcomes, Public Choice 157, 245–267.

Krause, G. A., Lewis, D. E., and Douglas, J. W. (2006), Political Appointments, Civil Service Systems, and Bureaucratic Competence: Organizational Balancing and Executive Branch Revenue Forecasts in the American States, American Journal of Political Science 50, 770–787.

Kyobe, A., and Danninger, S. (2005), Revenue Forecasting – How is it done? Results from a Survey of Low-Income Countries, IMF Working Paper No. 05/24.

Lane, P. R. (2003), The Cyclical Behaviour of Fiscal Policy: Evidence from the OECD, Journal of Public Economics 87, 2661–2675.

Leal, T., Pérez, J. J., Tujula, M., and Vidal, J.-P. (2008), Fiscal Forecasting: Lessons from the Literature and Challenges, Fiscal Studies 29, 347–386.

Lübke, A. (2008), Medium-term Financial Planning in the Federal Republic of Germany, Presupuesto y Gasto Público 51, 133–144.

Mechtel, M., and Potrafke, N. (2013), Electoral Cycles in Active Labour Market Policies, Public Choice 156, 181–194.

Merola, R., and Pérez, J. J. (2013), Fiscal Forecast Errors: Governments versus Independent Agencies? European Journal of Political Economy 32, 285–299.

Mocan, H. N., and Azad, S. (1995), Accuracy and Rationality of State General Fund Revenue Forecasts: Evidence from Panel Data, International Journal of Forecasting 11, 417–427.

Nickell, S. (1981), Biases in Dynamic Models with Fixed Effects, Econometrica 49, 1417–1426.

Nordhaus, W. D. (1987), Forecasting Efficiency: Concepts and Applications, Review of Economics and Statistics 69, 667–674.

Oberndorfer, U., and Steiner, V. (2007), Generationen- oder Parteienkonflikt? Eine empirische Analyse der deutschen Hochschulausgaben, Perspektiven der Wirtschaftspolitik 8, 165–183.

Ockenfels, A., and Weimann, J. (1999), Types and Patterns: An Experimental East-West-German Comparison of Cooperation and Solidarity, Journal of Public Economics 71, 275–287.

Paleologou, S. M. (2005), Political Manoeuvrings as Sources of Measurement Errors in Forecasts, Journal of Forecasting 24, 311–324.

Pina, Á. M., and Venes, N. M. (2011), The Political Economy of EDP Fiscal Forecasts: An Empirical Assessment, European Journal of Political Economy 27, 534–546.

Plesko, G. A. (1988), The Accuracy of Government Forecasts and Budget Projections, National Tax Journal 41, 483–501.

Potrafke, N. (2011), Public Expenditures on Education and Cultural Affairs in the West German States: Does Government Ideology Influence the Budget Composition? German Economic Review 12, 124–145.

Potrafke, N. (2013), Government Ideology and Economic Freedom across the German States, Regional Studies 47, 433–449.

Potrafke, N., Riem, M., and Schinke, C. (2016), Debt Brakes in the German States: Governments' Rhetoric and Actions, German Economic Review 17, 253–275.

Schroeder, L., and Wasylenko, M. (1989), Public Sector Forecasting in the Third World, International Journal of Forecasting 5, 333–345.

Seitz, H. (2000), Fiscal Policy, Deficits and Politics of Subnational Governments: The Case of the German Laender, Public Choice 102, 183–218.

Shi, M., and Svensson, J. (2006), Political Budget Cycles: Do they Differ across Countries and why?, Journal of Public Economics 90, 1367–1389.

Tepe, M., and Vanhuysse, P. (2009), Educational Business Cycles – The Political Economy of Teacher Hiring across German States, 1992–2004, Public Choice 139, 61–82.

Tepe, M., and Vanhuysse, P. (2014), A Vote at the Opera? The Political Economy of Public Theatres and Orchestras in the German States, European Journal of Political Economy 36, 254–273.

Triebs, T., and Tumlinson, J. (2013), Learning Capitalism the Hard Way – Evidence from Germany's Reunification, NBER Working Paper No. 19209.

Von Hagen, J. (2010), Sticking to Fiscal Plans: The Role of Institutions, Public Choice 144, 487–503.

White, H. (1980), A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity, Econometrica 48, 817–838.

Wyplosz, C. (2008), Fiscal Policy Councils: Unlovable or Just Unloved? Swedish Economic Policy Review 15, 173–192.