Contents

FinanzArchiv 74 (2018)
ISSN 0015-2218
doi: 10.1628/fa-2018-0006

Editorial Note

This first issue of volume 74 is a Festschrift in honor of Hans-Werner Sinn, who celebrates his 70th birthday on 7 March 2018. As a scientist and as a public intellectual, Hans-Werner Sinn has contributed to a wide range of topics in the economic policy debate. Sometimes he initiated the topics of this debate, and always he significantly influenced its direction. This applies most prominently to his early work on German reunification, as well as his work on the financial crisis, the European sovereign debt crisis, and global climate problems. His economic policy work is based on his own theoretical contributions, which he published in the first half of his scientific career, and whose economic relevance and powerful influence may extend to future decades.

Good examples of this are his work on corporate limited liability and its influence on investment behavior, and his work on the intertemporal functioning of markets for exhaustible resources. The former work was a fundamental exploratory tool for understanding the great financial crisis after 2007; the latter work was fundamental to his engagement with global climate policy and the deeper understanding of the essential relationship between public interventions in the fossil-fuel market and the dynamics of greenhouse-gas emissions. For all of this variety of topics, which far exceed public finance in the narrower sense, public finance was and is his "home ground": Fiscal and economic policy must be grounded in the individual preferences of the citizens. Politics must create institutions and establish conditions of economic policy within which the free action of economic agents leads to an optimum in the sense of the Paretian welfare theory. He is, in this respect, in the tradition of other great welfare-state thinkers like Richard Musgrave or Paul Samuelson.

As a researcher, he has followed this agenda over many decades as a chaired professor of public finance and economics, as president of the Ifo Institute and as CEO of CESifo, as a member of advisory boards and academies of science, in many other offices, and not least as a university teacher and thesis advisor to a large group of graduate students, with remarkable influence in all these activities. He also played an important role as Associate Editor from 1993 to 2013 at *FinanzArchiv/Public Finance Analysis*, where this collection of works is presented.

FinanzArchiv 74 (2018), 1–3
ISSN 0015-2218
doi: 10.1628/fa-2018-0007

We edited this collection of papers as a small token of appreciation to Hans-Werner Sinn on the occasion of his 70th birthday. As a teacher and advisor, he shaped our way of thinking in economics. This Festschrift recognizes his outstanding achievements in the field of public finance and economic policy. It brings together nine contributions from a set of leading international scholars in the field, who were and continue to be Hans-Werner's companions, fellow combatants, and friends. They are almost all part of the international CESifo research network, which was developed by him for more than 25 years and now has more than 1000 academic members throughout the world. Some of the topics covered take up longstanding research interests of Hans-Werner Sinn:

- Michael Keen combines optimal bailout policy with the question of how to tax banks optimally to deal with externalities in the real economy that derive from a breakdown of financial institutions.
- Robin Boadway, jointly with Motohiro Sato and Jean-François Tremblay, looks at the incentives of subfederal jurisdictions and identifies a tendency towards overextraction of natural resources.
- Rick van der Ploeg takes on the *green paradox* idea coined by Sinn. In his view, the possibility of renewables speeds up extraction of oil, but the possibility of larger amounts of resources that are not extracted even in the long term implies that there may be an offsetting effect to the green paradox.
- Another environmental issue is addressed by Peter Birch Sørensen. He looks at the optimal Pigovian tax in a circular economy that gives proper incentives for recycling.
- Justina Klimaviciute and Pierre Pestieau take on the question of how to design a system of long-term nursery care that meets demands for sustainability and allows positive bequests.
- Health-care issues are also the focus in the paper by Manuel Flores and Barbara Wolfe. They are looking at the effect of health problems in early life on later lifetime income and ask how this health effect depends on the health-care system.
- Joel Slemrod, jointly with William C. Boning, looks at the role of firms in optimal taxation and reviews recent contributions that warn against omitting important features of real-world firms in economic models of taxation.
- Robert Haveman, in a paper with Kathryn Wilson, revisits the development of the U.S. poverty rate and investigates to what extent this rate has been affected by a change in demographics.
- Finally, Assaf Razin and Efraim Sadka theoretically analyze immigration and its effect on redistribution.

The editors thank all the authors for their contributions and all the referees for their most helpful and prompt reviews.

Kai A. Konrad (Guest Editor), Ronnie Schöb (Editor), Marcel Thum (Guest Editor), Alfons J. Weichenrieder (Editor)

Bank Taxes, Bailouts and Financial Crises

Michael Keen*

Received 17 September 2017; in revised form 16 December 2017; accepted 17 December 2017

Following the Great Financial Crisis, more than a dozen countries adopted innovative bank taxes as part of their response. This paper characterizes, calibrates and discusses Pigovian taxes on bank borrowing to address externalities associated with either the collapse of systemic financial institutions or, to prevent that, public guarantees to bail out their creditors. It also characterizes optimal bailout policy, differentiating between circumstances in which the government can and cannot commit. Building on the analysis for a representative bank, it considers the implications for corrective taxation of various aspects of bank heterogeneity, connectedness, and asymmetries of information.

Keywords: bank taxation, Pigovian taxation, financial crisis

JEL classification: H 21, G 21

1. Introduction

The Great Financial Crisis that began in 2007 has left a long trail, both practical and intellectual. One central policy response has been the Basel III program of reform of the regulation and supervision of the financial sector. Less noted, but no less innovative, has been a fundamental reconsideration of the tax treatment of the financial sector. At one level – and perhaps ultimately most importantly – this has meant recognizing that pre-existing tax distortions

* International Monetary Fund, Washington DC 20431, USA (mkeen@imf.org). This is a substantially revised version of (parts of) Keen (2011a). For comments and suggestions on that earlier version I am grateful to Alan Auerbach, Stijn Claessens, Carlo Cottarelli, Giovanni Dell'Ariccia, Ruud de Mooij, Michael Devereux, Gregorio Impavido, Luc Laeven, Ben Lockwood, Jack Mintz, Michael Moore, Ceyla Pazarbasioglu, Lev Ratnovksi, Dan Shaviro, Joel Slemrod, Beatrice Weder di Mauro, Alfons Weichenrieder and a referee. Views expressed here should not be attributed to the IMF, its staff, Executive Board or its Management. A deeper acknowledgement is due to Hans-Werner Sinn. The basic topic of this paper – the public policy aspects of the potentially problematic incentive structures faced by banks – is just one of the many to which he has brought his characteristic blend of penetrating insight, forceful argument and sheer good writing (Sinn (2003, 2010)). For this, and his warm generosity and enthusiasm, we are all much the better.

doi: 10.1628/fa-2018-0001
ISSN 0015-2218

may be costlier than had generally been supposed, notably[1] the bias towards debt finance inherent in most corporate tax systems (providing deductions for interest but not the return to equity). All that, however, was well-known to the public finance community before the crisis. The more novel issue raised was whether new types of tax instruments, applied specifically to discourage socially excessive leverage and risk-taking by financial institutions, might have a constructive role to play in limiting the likelihood of and social damage from financial failures.[2] A charge of this type was proposed in a report to G20 by the IMF (2010).[3] But action ran far ahead of analysis. In 2011, for instance, the U.K. introduced a levy on (essentially) banks' uninsured debt obligations, "... to encourage banks to move away from risky funding models that threaten the stability of the financial sector and wider economy" (HM Revenue & Customs, 2010). Within a few years, France, Germany, Sweden and others – a total of fourteen member states, as well as three non-EU members – had introduced some special charge on financial institutions.

These bank taxes/levies[4] are a wholly novel development in tax policy.[5] They differ in significant ways in rationale and design. One strand of thought has stressed their potential role in financing, ex ante, the bank resolution and other costs likely to be experienced in future crises.[6] This is conceptually distinct from, though often conflated with, another role for such taxes, as stressed, for instance, by the U.K.: the idea of such taxes as playing a purposive role in addressing externalities emanating from financial stress and failure.[7] Indeed the idea of Pigovian taxes on the financial sector has become increasingly

1 But not only. Another long-standing concern is the exemption of most financial services under the value added tax: see for example IMF (2101a), Keen et al. (2016). Shackelford et al. (2010) also discuss aspects of financial sector taxation in light of the crisis.

2 The links between this debt bias issue and the corrective taxes that are the focus of this paper are taken up in the concluding section.

3 Background papers, along with the report itself, are in Claessens et al. (2010).

4 For brevity, we speak of 'banks' and 'bank taxes' throughout, though in both principle and practice the issues extend beyond narrowly-defined banks, and the term 'levies' is also often used when emphasis is placed on the rationale for a charge as a user fee to cover ex ante the costs of cleaning up post-crisis.

5 Taxation in the shape of deposit insurance has long been familiar in the sector, but this has generally been seen as targeted at potentially ill-informed retail depositors rather than as addressing the systemic risks at issue in the crisis. Another tax innovation coming out of the crisis was the deployment of bonus taxes, aimed at addressing both distributional concerns and incentives for excessive risk-taking: see for instance (on theory and evidence respectively) Besley and Ghatak (2013) and Von Ehrlich and Radulescu (2017). Short-lived, and in some cases explicitly temporary, these are not considered here.

6 This is the rationale, notably, for the contributions to the Single Resolution Fund (SRF) of the EU, to be made by members of the banking union, established in 2016. These are of essentially the form proposed in IMF (2010).

7 Distinct charges to these distinct ends could even co-exist. Austria, for example, reportedly indicated an intention to impose both the bank charge introduced in 2011 and mandatory contributions to the SRF.

common currency, one recent example being the appearance of such a tax on shadow banks in recent reform proposals of the Federal Reserve Bank of Minneapolis (2016). These innovations in practical policy have taken place, however, without (beyond a few important exceptions noted later) much precision as to how such taxes should be structured or calibrated.

The aim of this paper, which originated in the heady days of the crisis, is to sketch how bank taxes might be designed to address what the crisis reminded us can be strongly adverse external effects from the collapse, and potential collapse, of systemic financial institutions,[8] along with associated issues relating to bailout policy.

In their detail, the mechanics of these effects, operating both across financial institutions and between the financial and nonfinancial sectors, are complex and varied. The former includes, for instance, the effects of firesale externalities as distressed asset sales by one institution lead to price reductions that jeopardize the solvency of others: this is a pecuniary externality that has real effects as a consequence of incompleteness of markets and regulatory and other constraints. They include too information spillovers (as bad news about one institutions is taken as cause for concern regarding others). The latter include the likelihood that sharper credit constraints will limit opportunities open to nonfinancial businesses.[9] Their ultimate consequences, however, have been clear. During the Great Financial Crisis, many governments, lacking tools to resolve systemically important institutions in an orderly fashion, faced the dilemma of either letting such collapse occur, allowing these externalities full play and accepting the economic disruption that would imply, or, instead, committing sufficient public funds for bailouts to avert this damage, but in so doing trigger another kind of externality, and potential inequities, by creating an expectation of future bailouts: the 'too big to fail; syndrome.' Or, of course, doing some of both.

The social costs associated with either course of action have proved very high. Laeven and Valencia (2016) estimate the median cumulative output loss in the four years following banking crises in advanced economies to be around 33 percent of GDP; Basel Committee on Banking Supervision (BCBS, 2010) find a median cumulative output loss of 63 percent of initial GDP, and a mean of over 100 percent. In narrower fiscal terms, governments' exposures at the height of the Great Financial Crisis were huge: through guarantees and the like, the advanced G-20 economies committed to making an average of

8 There is of course also evidence of positive externalities from well-functioning financial systems (Levine, 2005), but these are not at issue here.

9 Systemic externalities originating in the financial sector are reviewed in Bank of England (2009) and Wagner (2010).

25 percent of GDP available for support operations.[10] And in normal times, of course, the expectation of bailout manifests itself in reduced borrowing costs that can amplify any inherent tendency toward excessive leverage of financial institutions.

This dilemma is at the core of this paper. The central aim is to characterize and explore, in a series of settings, the optimal design of corrective taxes on bank borrowing (to discourage inappropriately low capital ratios) in the presence of this inherent dilemma: between, on one hand, incurring the collapse externality associated with failure of systemic institutions or, on the other, incurring a 'bailout' externality by providing resources to bail out creditors should such institutions become distressed. In relation to the latter, a key and very practical question is whether or not the government can credibly commit to its bailout policy; both possibilities, and their implications for corrective taxation, are examined.

Several post-crisis papers have taken up aspects of the use of Pigovian taxes in the financial sector.[11] Closest to the analysis here is Acharya et al. (2016), which, as discussed further below, arrives by a different route at an optimal tax formula that differs from but has strong similarities to that in Proposition 1 below; from there, however, they turn to their principal concert of characterizing and measuring the contribution of individual financial institutions to systemic risk, while the analysis here explores further the issues of tax design and optimal bailout policy. The recent literature has also taken up the fundamental issue as to the relative merits of taxation and regulation in this context. These are discussed, for example, in IMF (2010), Keen (2011a, b) and Coulter et al. (2014). These issues,[12] however, would require a fuller treatment that is

10 As of mid-2014 (no more recent data of this kind seem to be available), public support actually extended to the financial sector since the Great Financial Crisis in a selection of advanced countries averaged 7.4 percent of GDP (with a high of 41 percent in Ireland), of which a little over one-third had been recovered (IMF (2014), table 1.4).

11 Less formal arguments to the same effect are in Shin (2010) and in the proposal of Perotti and Suarez (2009) for a corrective tax on maturity mismatch that would be, in effect, largely a tax on short-term debt. Several papers have argued for corrective taxation of unsecured borrowing on other grounds. In Huang and Ratnovski (2009), for instance, it serves to reduce banks' funding reliance on creditors with such high seniority that they may impose inefficient liquidation in response to noisy signals on the institution's prospects; in Jeanne and Korinek (2010) it serves to discourage borrowing that increases asset prices and so, in the presence of collateral constraints, amplifies volatility by allowing others to borrow more too; see also Bianchi and Mendoza (2010) and Korinek (2009). As noted by Korinek (2009), the characterization of optimal policy in terms of Pigovian taxation in these papers is as an analytical convenience, with corresponding regulation seen as just as good a way to implement it.

12 A further option for dealing with externalities is through ex post liability (Shavell, 2011) – this is not especially attractive in the context of financial crises, however, since institutions contributing to them may by then no longer exist.

possible here, and so are not taken up in what follows – although, as will be seem, some of the results below are relevant to it.

One area in which knowledge has advanced significantly since the financial crisis is the empirical importance of the tax issues addressed here. At that point, it was not even clear whether – given the capital requirements that they face – taxation had any significant impact on the financing decisions of banks. Now it is clear that it does: banks generally hold a capital buffer above those requirements, leaving clear scope for tax effects, as shown by De Mooij and Keen (2016) and Hemmelgarn and Teichmann (2014). And indeed there is evidence that the recent bank taxes have themselves had an appreciable effect: see Devereux et al. (2013).[13] It is worth noting too that while capital requirements are now higher, and of higher quality, under Basel III, controversy continues as to whether they are adequate.[14]

The plan of the paper is as follows. The next section sets out a model of a representative bank whose decisions determine its own risk of failure, and formalize and explores the collapse and bailout externalities. Section 3 then characterizes and calibrates the optimal corrective tax in that context. Recalling the policy interest in using bank charges to provide ex ante financing for ex post resolution, it also asks whether the revenue raised by such a charge can be expected to be adequate to meet expected bailout costs. Section 4 then characterizes optimal bailout policy and its implications for corrective taxation, drawing an important distinction between the cases in which the government can and cannot commit to its bailout policy. Section 5 extends the analysis to settings with multiple and heterogeneous banks, dealing first with the case in which banks are unconnected, before turning to that in which they are connected through inter-bank deposits and finally considering the implications of asymmetries of information between banks and government, Section 6 concludes.

2. Banking and Systemic Externalities

This section develops a basic model of a single bank,[15] with endogenous failure risk, that allows an initial characterization and exploration of the policy dilemma raised above: the choice between allowing a failed institution to collapse, or, to avoid the damage this would cause, bailing out its creditors.

13 Buch et al. (2017), however, find little evidence of a sizable impact from the German levy.

14 The highest capital requirement under Basel III, for systemically important banks, is 15.5 percent (relative to risk-weighted assets); there is also a leverage ratio (relative to unweighted assets) of 3 percent. In contrast, the influential book by Admati and Hellwig (2013), for instance, argues for a leverage ratio in the order of 25–30 percent, and Federal Reserve Bank of Minneapolis (2016) recommends one of 15 percent.

15 Or many identical ones.

2.1. The Bank

The 'bank' has equity capital in an amount K, taken throughout as given,[16] and chooses how much to borrow, B and lend, L ('loans'), with

$$L = K + B. \tag{1}$$

It offers creditors a rate of return of ρ (inclusive of return of principal), the determination of which is considered below. The return on loans (also inclusive of principal), $r \geq 0$, is stochastic; its distribution, described by density ϕ and (twice continuously differentiable) distribution function Φ, is taken as given, with $\phi(r) > 0$ for all $r > 0$. Denoting the risk-free return by ζ, it is also assumed throughout that

$$\int_{-\infty}^{\infty} (r - \zeta)\phi(r)dr = E[r - \zeta] > 0. \tag{2}$$

Loans are thus expected to yield more than the safe return: this ensures that they are socially desirable and that banks are willing to borrow in order to make them.

The assumed exogeneity of the distribution of returns means that the bank has no choice as to the riskiness of its assets.[17] This does not mean, however, that it has no risk-taking decision to make. To the contrary, this assumption serves to focus attention on the most fundamental of any bank's risk decisions: that of how large a risk to accept that the return on its assets will prove so low, relative to its capital base and promises to creditors, that the bank fails and its equity is wiped out.

Such failure arises if and only if the bank is unable to meet its obligations to creditors in full, even by fully exhausting equity K. This happens[18] if and only if

$$rL < \rho B. \tag{3}$$

Defining the capital ratio $k \equiv K/L$ (and using (1)) this defines a critical return on loans of

$$R \equiv \rho(1 - k) \tag{4}$$

below which failure occurs. All else equal, failure is thus less likely the lower is the interest rate at which the bank borrows and the higher is its capital ratio

16 Keen (2011a) considers briefly the implications of adding an upward-sloping supply of bank capital.

17 An alternative model of bank behavior (for a different purpose) that does incorporate a decision as to the riskiness of the bank's assets is set out in de Mooij and Keen (2016); see also Devereux et al. (2013).

18 To see that equity is in this case wiped out, recall that the interest terms include repayment of principal: the rL term thus includes in effect full use of equity to pay creditors.

(or the lower its leverage $b \equiv B/L = 1-k$). The probability of failure is zero if and only if $k=1$ or, equivalently, $b=0$.

In the event of bankruptcy, bank owners are assumed to incur costs – beyond the loss of their equity – of δK. These might be literal bankruptcy costs, a loss of ego rents (as in Dewatripont and Tirole, 1993) or loss of franchise value of the bank (Hellman et al., 2000).

The government levies a per unit tax on the bank's borrowing at rate τ, implying a tax charge of $\tau B = \tau(L-K)$. This, we assume for simplicity, is payable and paid in full whether or not the bank fails.[19] The choice of τ is a central concern in what follows.

Bank owners (and, later, creditors and the government) are risk-neutral. Normalizing relative to the (fixed) amount of equity capital K, their problem is thus to choose the capital ratio k to maximize the payoff to the bank's owners, which is given by

$$\pi \equiv -\Phi(R)\delta + \int_R^\infty \left\{ r\left(\frac{1}{k}\right) - \rho\left(1-\frac{1}{k}\right)\right\}\phi(r)\,dr - \tau\left(\frac{1}{k}-1\right), \tag{5}$$

where $\Phi(R)$ is the probability of failure and the truncation in the integral reflects the operation of limited liability (the full consequences of which evidently depend on the return ρ required by the bank's creditors, to which we turn in a moment). We refer to π as the bank's (after-tax) profits, though it also reflects the bankruptcy costs δ that may in part be non-pecuniary. To focus solely on Pigovian taxation as a policy instrument, there are no capital requirements and no tax-induced debt bias of the kind that, as noted in the Introduction, is inherent in most corporate tax systems.[20]

It remains to characterize the determination of ρ, the rate at which the bank borrows. One approach would be to assume creditors to be myopic, taking no account of the possibility that the bank may be unable to repay them: this is the archetypal view of small depositors and provides one rationale for deposit insurance. More at issue in the crisis, however, was the behavior of large and uninsured wholesale depositors, more naturally assumed to be sufficiently sophisticated (and well-informed) to take the possibility of failure into account in their lending decisions. That is the assumption made here.[21]

19 This assumption – implicit already in the failure condition (3) – avoids complications relating to the priority accorded to such obligations that are not of the essence to the issues at stake.

20 De Mooij and Keen (2016) analyze debt bias in the presence of capital requirements.

21 As John et al. (1991) and Sinn (2003, 2010) stress, limited liability reacquires importance when information is asymmetric between the bank and its creditors. Though the assumption the lenders are fully informed requires quite a leap of faith, asymmetric information of this type is not considered here, so as to focus on the inefficiencies associated with collapse and bailout.

In forming their expectations, creditors are therefore assumed to take account of any prospect of their being bailed out by the government if the bank itself cannot meet its obligations to them. This possibility is characterized by a parameter $\mu \in [0,1]$ that can be interpreted as either the probability that all creditors will be fully protected (in the sense of receiving the return ρ) by the government or – the language used below – the extent to which each will be protected (with $1-\mu$ being, conversely, the extent of the haircut each will take). Creditors are assumed throughout to take μ as given, and it assumed for now – this will be relaxed later – that the government's commitment to this bailout policy is fully credible.

Given their alternative of simply investing at the risk-free rate ζ, and assuming a competitive loan market, creditors will thus require a return ρ such that

$$\zeta = \rho\{1-\Phi(R)\} + \mu\rho\Phi(R) + \left(\frac{1-\mu}{1-k}\right)\int_{-\infty}^{R} r\phi(r)dr, \tag{6}$$

where the first term on the right reflects full payment of ρ on its borrowing of B if the bank does not fail, the second the extent of the bailout of creditors if it does, and the third[22] that creditors not bailed out in the event of failure receive only the residual value of the bank's assets. Notwithstanding the limited liability of the bank's owners, creditors thus receive an expected return equal to the risk-free rate, through some combination of an elevated return when the bank does not fail and in injection of public funds when it does.

Recalling that $R = \rho(1-k)$, equation (6) defines the rate of return on borrowing as a function $\rho(k,\mu)$, routine comparative statics showing that[23] ρ_k and ρ_μ are both strictly negative (except, for the former, when $\mu = 1$, in which case the bank can borrow at the safe rate).[24] This is as expected: the less likely is failure, and the higher is the probability of bailout, the lower is the rate at which the creditors will be willing to lend to the bank. From this, the critical return at which the bank fails is given by

$$R(k,\mu) = \rho(k,\mu)(1-k) \tag{7}$$

22 The lower limit of the integral is taken to be $-\infty$ in expressions like this, even though r is assumed non-negative in all realizations, as a reminder that integration is over a range that includes failure of the bank. Use is also made in this third term of the implication of (1) that $L/B = 1/(1-k)$).

23 Derivatives are indicated by subscripts for functions of several variables.

24 Explicitly:

$$\rho_k = -\left(\frac{(1-\mu)S}{(1-k)^2[1-\Phi(1-\mu)]}\right)$$
$$\rho_\mu = -\left(\frac{kS}{(1-k)[1-\Phi(1-\mu)]}\right). \tag{N.1}$$

with both R_k and R_μ strictly negative.

Solving (6) for $\rho\{1-\Phi(R)\}$ and substituting into (5), the bank's objective function, regarded as a function of its capital ratio and the two policy parameters, can be written as

$$\pi(k,\tau,\mu) = -\Phi[R(k,\mu)]\delta + \int_{-\infty}^{\infty}\left\{r\left(\frac{1}{k}\right) - \zeta\left(1-\frac{1}{k}\right)\right\}\phi(r)dr + \mu S(k,\mu) - \tau\left(\frac{1}{k}-1\right), \tag{8}$$

where

$$S(k,\mu) \equiv \rho(k,\mu)\Phi[R(k,\mu)]\left(\frac{1}{k}-1\right) - \left(\frac{1}{k}\right)\int_{-\infty}^{R(k,\mu)} r\phi(r)dr, \tag{9}$$

$$= \left(\frac{1}{k}\right)\int_{-\infty}^{R(k,\mu)}(R(k,\mu)-r)r\phi(r)dr, \tag{10}$$

the final step following from $R = \rho(1-k)$.

The expected payoff to bank owners thus comprises three components. The first is the expected private cost of failure. This arises whether or not the government bails out creditors: equity is assumed to be wiped out whenever the bank is unable to meet its obligations, any bailout applying only to creditors. The second component is the value that the bank would have if there were simply unlimited liability and no possibility of bailout (recalling that creditors are then compensated for the risk of failure by a sufficiently high return ρ).[25] The third component is the expected value of the bailout, μS, and of interest to owners not because they themselves will be rescued but because it reduces the rate at which they can borrow while the bank is in operation. This term is central in what follows, and merits closer attention.

2.2. The Bailout Externality

This third component of the bank's maximand in (8), $\mu S(k,\mu)$, is the value to the bank – conversely, the revenue cost to the government – of expected public support (topping up the residual value of assets) to pay off creditors (expressed per unit of equity. K). This term thus captures the implicit subsidy

25 To see this, note that the return to bank owners given limited liability is

$$\int_R^{\infty}\left\{r\left(\frac{1}{k}\right) - \rho\left(\frac{1}{1-k}\right)\right\}\phi(r)dr.$$

Setting $\mu = 0$ in (6), when there is no possibility of bailout the return to creditors is

$$\rho = \frac{\zeta}{1-\Phi(R)} - \left(\frac{1}{(1-k)(1-\Phi(R))}\right)\int_{-\infty}^{R} r\phi(r)dr.$$

Combining these two gives the second term in (8).

from the prospect of bailout: the 'too big to fail' subsidy. It is in itself a transfer, but inefficiencies will arise from the actions of bank owners to exploit it (as well as from any distortionary taxes levied to pay for it). We refer to these inefficiencies as the *bailout externality*.[26]

The properties of the bailout subsidy $\mu S(k,\mu)$ will be important in what follows and of interest in themselves. It is immediate from (10) that, for all $k<1$, S is strictly positive, while differentiating in (10) gives[27]

$$kS_k = -S + R_k\Phi(R) < 0 \tag{11}$$

so that (since that $R_k < 0$) the bailout subsidy is strictly decreasing in the capital ratio: banks that choose a safer capital ratio stand to benefit less from the prospect of being bailed out. One might expect it also to be convex in k; differentiating again, this will indeed be the case under the plausible condition that $R_{kk} \geq 0$ (though this will not be assumed in what follows). it can be shown too that – also as one would expect – the bailout subsidy is strictly increasing in the probability of bailout out,[28] μ.

To provide some sense of the possible magnitude of the bailout subsidy, denote by $\rho' \equiv \rho(k,0)$ the return that the bank would pay if there were no prospect of bailout. Setting $\mu=0$ in (6), this is given by

$$\zeta = \rho'\{1-\Phi'\} + \left(\frac{1}{1-k}\right)\int_{-\infty}^{R'} r\phi(r)dr, \tag{12}$$

where R' and Φ' denote the corresponding critical return and risk of failure. Comparing this with (6), evaluated at the same capital ratio but for any bailout probability μ, gives

$$S(k,\mu) = \{\rho'(1-\Phi'\} - \rho\{1-\Phi\})b + \left(\frac{1}{k}\right)\int_{R}^{R'} r\phi(r)dr. \tag{13}$$

The subsidy is thus closely related to $\rho'-\rho$, which is the reduction in the bank's borrowing rate consequent on the possibility of bail out. Estimates have put this in the range of 10–50 basis points, and commonly around 20.[29] Equa-

26 Kocherlakota (2010) refers to this as a 'risk externality'.

27 It is assumed that $k \in (0,1)$ at all private and social optima considered.

28 Differentiating (10) with respect to μ gives

$$\begin{aligned} S_\mu &= \left(\frac{1}{k}\right)R_\mu\Phi \\ &= -\left(\frac{S\Phi}{1-\Phi(1-\mu)}\right) \end{aligned}$$

the second equality following on noting that $R_\mu = \rho_\mu(1-k)$ and on using equation (N.1) of footnote 24 above. Hence

$$S+\mu S_\mu = \frac{(1-\Phi)S}{1-\Phi(1-\mu)} > 0.$$

29 See Baker and McArthur (2009), Haldane (2009) and, for a review and extension of the evidence, IMF (2010).

tion (13) also shows, however, that is not an exact measure of the value of the bailout subsidy, ignoring as it does the two considerations that the benefit of the lower borrowing rate applies only when the bank does not fail, and that residual assets are lower in the event of default when there is some degree of bailout (since default than occurs at a lower return on assets), which reduces the cost to the government of making creditors whole and so reduces the value of the bailout subsidy).

2.3. The Bank's Choice of Capital Ratio

Differentiating in (8), the necessary condition on the bank's choice of capital ratio is

$$\pi_k(k,\tau,\mu) \equiv -\phi(R)\delta R_k - \left(\frac{1}{k}\right)^2 E[r-\zeta] + \mu S_k(k) + \tau\left(\frac{1}{k}\right)^2 = 0. \quad (14)$$

The bank thus trades off the beneficial effects of a higher capital ratio in reducing the chances of failure and cutting its tax bill against the adverse effects through contracting a profitable loan portfolio and reducing the value of the implicit bailout subsidy.

Satisfaction of the second order condition, however, is not assured by the assumptions made so far.[30] That the bank's maximand may not be globally concave is analytically inconvenient. The potential that it implies for small changes in the environment to lead to large shifts between stable equilibria has potential implications for the wider issue of choosing between tax and regulatory approaches that are discussed in Keen (2011b). For present purposes, however, we set these difficulties aside and take the second order condition to be satisfied, in which case (14) is easily seen to define the privately optimal capital ratio as a decreasing function $k(\tau)$ of the rate of bank taxation. Substituting this into (8) then gives maximized profits $\pi[k(\tau),\tau]$.

30 This requires negativity of

$$\pi_{kk} = -\Psi\delta + 2\left(\frac{1}{k}\right)^3 E[r-\zeta] + \mu S_{kk} - \tau 2\left(\frac{1}{k}\right)^3$$

where $\Psi \equiv \phi'(R_k)^2 + \phi R_{kk}$. It is reasonable to suppose the density of returns to be increasing in the low tail associated with failure, and the intuitively plausible assumption that $R_{kk} \geq 0$ then ensures that $\Omega > 0$. The second order condition can be satisfied, however, only if the effects through this and the tax-related term are strong enough to outweigh two others. The first of these reflects the mechanical feature that increasing the capital ratio requires a smaller reduction in profitable loans the higher is the initial ratio. The second reflects the likely convexity of the bailout subsidy discussed above.

3. Optimal Corrective Taxation with a Representative Bank

One of the two externalities highlighted in the introduction – the bailout externality – is directly present in the bank's optimization problem, it is time now to introduce the other.

3.1. The Collapse Externality

Failure of the bank is assumed to generate wider social costs, additional to the cost δK borne by owners and the fiscal cost S. These arise, however, only to the extent that its creditors are not bailed out, reflecting the common rationale for bailout as a circuit breaker, forestalling the operation of external effects from a bank's failure that trigger wider failures and damage to the real economy.[31] The protection of creditors of AIG, Bear Sterns and RBS, for example, reflected fear of the wider fallout from their collapse; the collapse of Lehman, on the other hand, was an instance of the damage that unmitigated failure of a systemically important institution can cause. As noted in the Introduction, the mechanisms by which the distress and failure of financial institutions give rise to these wider social costs of collapse are many and diverse, Here, however, we simply take the social harm from unmitigated bank failure to be of the form $(1-\mu)\Delta K$, with the magnitude of the *collapse externality* $\Delta \geq 0$ taken as exogenous. While a complete treatment would model these costs explicitly, this would add another level of complexity; the simpler approach here of taking the social cost of the externality as parametric – also taken by Acharya et al (2016) – has the merit of clarity and sharpness.[32]

3.2. The Optimal Tax on Bank Borrowing

The government, we assume, attaches full weight to the costs incurred by owners in the event of failure. It also faces costs of $\lambda \geq 0$ in raising the revenue needed to finance any bailouts, reflecting not only deadweight losses in the wider tax system but also perhaps a lesser social value of transfers to bank creditors than of transfers from general taxpayers. Again normalizing relative

31 In practice, the decision to bail out an institution may of course reflect considerations other than the avoidance of wider social damage, such as regulatory capture, the impact of lobbying or the view of large financial institutions as 'national resources': see for instance Shull (2010).

32 Proportionality in K is readily relaxed: the possibility, perhaps more plausible, that social costs are proportional to the volume of loans or of deposits, for instance, can be encompassed in the more general supposition that collapse costs (per unit of equity) are of the form $\Lambda(k)$, with $\Delta'(k) < 0$.

to equity, the government thus evaluates policy by the objective function

$$w(k,\tau,\mu) \equiv \pi(k,\tau,\mu) - \Phi(R(k,\mu))(1-\mu)\Delta - (1+\lambda)\mu S(k,\mu) + \tau\left(\frac{1}{k} - 1\right), \quad (15)$$

so adjusting the private after-tax profits of the bank by taking account of the revenue it receives from the bank tax and the impact of the two externalities at work: the collapse externality, and the fiscal cost associated with the bailout externality.[33]

Differentiating in (15) with respect to τ, taking account of the impact k_τ on the bank's choice of capital ratio (and the envelope property $\pi_k = 0$) gives

$$\frac{dw}{d\tau} = -\left((1-\mu)\Delta\phi(R)R_k + (1+\lambda)\mu S_k + \tau\left(\frac{1}{k}\right)^2\right)k_\tau. \quad (16)$$

Hence, since $k_\tau < 0$:

Proposition 1 The optimal tax on bank borrowing is characterized by

$$\tau = (1-\mu)\tau_C + \mu\tau_B \geq 0, \quad (17)$$

where

$$\tau_C \equiv -\Delta\phi(R)R_k k^2 > 0 \quad (18)$$

$$\tau_B \equiv -(1+\lambda)S_k k^2 > 0. \quad (19)$$

The optimal tax is thus a very straightforward weighted average of two corrective terms, each addressed to one of the possible consequences of failure identified above, the weights reflecting the likelihood of bail out. Each reflects an element of social gain from an increase in the capital ratio (with the multiplication by k^2 translating this into a tax on leverage).[34]

The first component in Proposition 1, τ_C, is addressed to the collapse externality, with each small increase in the capital ratio induced by the tax reducing the probability of collapse by $-\phi(R)R_k$. Strict negativity of R_k ensures that this component is strictly positive.

The second component, τ_B, counteracts the bank's incentive to increase the bailout subsidy by setting a lower capital ratio than it otherwise would, an aspect of corrective policy stressed, for example, in the informal treatments of Weder di Mauro (2010) and Kocherlakota (2010). From (11), this term too is

33 There is, it should be noted, an element of schizophrenia here, if λ is to be interpreted as to some degree reflecting the deadweight loss of the wider tax system. For then revenue from the bank tax, τb, should also be weighted by λ in (15). Allowing for this would simply introduce a revenue-raising motive for the bank tax, which seems a second-order concern for present purposes relative to the fiscal challenges often posed by bank bail outs.

34 Recall that $b = 1/(1-k)$.

strictly positive. Its magnitude depends, however, on the extent of deadweight loss (or distributional angst) associate with financing the bailout: the greater this is, the higher, as one would expect, is the optimal tax on bank borrowing.

In their Proposition 1, Acharya et al. (2016) arrive, within a different setting, at a similar optimal tax formula that also comprises two additive components. One, corresponding roughly to τ_B, relates to the cost of government guarantees and is seen as related to idiosyncratic risk. The second, closer to the collapse component τ_C, relates the externalities associated with systemic failure arising from capital shortfall across the entire financial sector to the expected shortfall of the institution conditional on such system-wide shortfall: the institution's 'systemic expected shortfall (SES). The central focus in Acharya et al. (2016) is then on the empirical exploration of the determinants of SES. Here, however, we pursue further the public finance perspective, exploring the tax design and other implications (including for bailout policy) of, and to that end extend, the simpler but more direct characterization established above.

One clear public finance concern is with the revenue raised by the optimal corrective tax. A common rationale offered for the bank levies introduced in the wake of the financial crisis,[35] as noted at the outset, was to meet the fiscal costs of dealing with bank failures. While this is a conceptually quite different rationale from the Pigovian objective of changing behavior that is the main concern here, it is thus of interest that (the proof being in appendix 7.1):

Proposition 2 The revenue raised by the component τ_B of the optimal tax on bank borrowing is at least as great as the fiscal cost of bailout, μS.

Put differently, the optimal corrective tax to address the bailout externality is higher than that needed for an insurance-type charge to meet the expected fiscal cost of bailout. Intuitively, this follows from the likely convexity of S in k noted above (though that is not required for the result): the marginal damage from reducing the capital ratio exceeds the average damage.

It is natural too to wonder how large the optimal corrective described in Proposition 1 might be. Some rough calculations[36] can give a sense of possible orders of magnitude of each of its two components.

Consider first the optimal tax on borrowing to address the collapse externality. Estimates of the probability of crisis $\Phi[R(k,\mu)]$ (reviewed in Annex 2 of BCBS (2010)) suggest it to be strongly convex in the capital ratio. Thus $-\phi(R(k,\mu))R_k(k,\mu) \geq \Phi(R(k,\mu))/(1-k)$, and so a lower bound on the

35 And now reflected in the structure of the EU's SRF.
36 What follows are not, it should be stressed, simulations.

corrective tax related to the collapse externality is given by[37]

$$\tau_C \geq \left(\frac{\overline{\Delta}\Phi(R(k,\mu))k^2}{\overline{K}(1-k)} \right) \tag{20}$$

where $\overline{K} \equiv K/GDP$ and $\overline{\Delta} \equiv \Delta/GDP$ denote respectively bank capital and collapse costs in percent of GDP. For the U.K., which has a large banking sector, bank capital is around 10 percent of GDP; and reasonable figures for cumulative output loss from systemic collapse might be, recalling the discussion above, between 63 and 100 percent of GDP. For the five largest banks in the United Kingdom, BCBS (2010) reports estimated (annual) probabilities of failure $\Phi(k)$ at various capital ratios (calculated using a Bank of England model). These are shown in the first two rows of Table 1. Using these values, the third and fourth rows of Table 1 report the implied upper bound on τ_C at different capital ratios and for the two alternative collapse costs $\overline{\Delta}$.

Two features stand out. The first is that the (lower bound on the) optimal tax is in some cases quite large, certainly much larger than those generally adopted or envisaged: about 50 basis points in the more extreme of the circumstances shown. In the U.K. for instance, the bank levy was initially levied at 7.5 basis points, and peaked in 2015 at 21 basis points. There is, it should be noted, an important difficulty of interpretation here, in that the model is of a single institution while the collapse cost estimates refer to the wider financial system. For an institution that is indeed systemically important, of course, the distinction is moot. Nevertheless, one might expect losses from isolated failures – to the extent that those can be imagined – to be less than those from the system as a whole. This naturally reduces the optimal tax, though it plausibly remains significant: if damage is only 25 percent of GDP, for instance, it falls to 12 basis points at a capital ratio of 6 percent. The second and still more striking aspect of the results in Table 1 is the very strong variation of the optimal tax with the capital ratio (reflecting that of the probability of crisis): even with potential output costs of 100 percent of GDP, the optimal tax is negligible at a capital ratio of 12 percent. The implication is that the optimal borrowing tax is likely to be highly nonlinear, increasing rapidly as capital ratios fall so low as to markedly increase the likelihood of crisis.

Consider now the corrective taxation in respect of the bailout externality. Taking the extreme case in which $\mu = 1$, an upper bound on the optimal tax on borrowing is shown in appendix 7.2 to be given by[38]

$$\tau_B \leq (1+\lambda)\zeta\Phi(k)k^2. \tag{21}$$

37 Here regarding Φ as a function $\Phi(R(k))$ of k.
38 The error reflects the residual value of the bank's assets in the event of its failure.

The final two rows of Table 1 tabulate values of this approximation assuming a risk-free interest rate ζ of 3 percent and at two values of the marginal excess cost of raising public revenue λ: the lower of these, at 0.25, reflects common estimates to be found in the literature, while the higher, at unity, would be appropriate if bailouts could be financed by lump sum taxation but – not wholly at odds with the views expressed by many – zero social value was attached to the benefit that bank owners derived from being bailed out.

The implied corrective tax aimed at the bailout externality clearly looks much smaller than that directed at collapse. But it is not trivial, being around the order of magnitude of the initial bank levy in the U.K. It seems hard on these very simple calculations, however, to justify the final value of the U.K. levy without appeal to concerns with bank collapse.

Table 1

Approximating the Optimal Corrective Tax Components

	Capital ratio, k			
	6	8	10	12
Probability of crisis (annual), $\Phi(k)$	12.8	2.6	0.9	0
Collapse externality, $\tau_C\,(\mu=0)$:				
$\Delta = 63$	31	11	5	0
$\Delta = 100$	49	19	9	0
Bailout externality, $\tau_C\,(\mu=1)$:				
$\lambda = 0.25$	6	4	2	0
$\lambda = 1$	9	7	3	0

Note: Tax rates are in basis points, rest in percent. The ratio of bank capital to GDP, $\overline{K}$, is taken to be 10 percent, and the risk-free return, ζ, to be 1.03 (recalling that returns in the analysis are inclusive of return of principal). The reported tax rates are (approximations to) the rates that would be optimal if, given the assumed parameter values, they induced banks to choose the capital ratio indicated.

4. Optimal Bailout Policy

The question also arises as to the government's optimal bailout policy: the choice, that is, of the extent to which it credibly commits to bail out creditors in the event of bank failure. The banks' owners, of course, always prefer bail out to be as complete as possible, since it enables them to borrow at a lower rate, which reduces the chances of their being wiped out and increases their profits when they are not. This can be seen on differentiating in (8) and using the envelope property $\pi_k = 0$ to find

$$\frac{d\pi}{d\mu} = -\delta\phi R_\mu + (\mu + S_\mu) \geq 0 \tag{22}$$

with non-negativity following from the earlier observations that R is strictly decreasing and μS is strictly increasing in μ. Differentiating in (15), using (14) and again that $\pi_k = 0$, the impact on social welfare of an increased likelihood of bailout is then given by

$$\frac{dw}{d\mu} = -[\delta + (1-\Phi)\Delta]\phi R_\mu + \Delta\Phi - \lambda(S + \mu S_\mu). \tag{23}$$

This is a straightforward trade-off between, on one hand, the benefits that a more extensive bailout brings in reducing expected collapse and bankruptcy costs (by allowing the bank to borrow more cheaply) and, on the other, the expected fiscal cost of bailout. This runs starkly counter to the connotation of bailouts as something uniformly undesirable. Apart from the distortionary (or distributional) costs of financing them, bailouts are simply a transfer that enables a socially desirable expansion of the bank's loan portfolio by providing a guarantee to its creditors and so reducing its borrowing costs. If there are no fiscal costs associated with bailing out ($\lambda = 0$) – unlikely, but an important benchmark – then (23) implies that complete bailout is a first-best instrument for addressing the inefficiencies associated with the possibility of failure – with, of course, an appropriately high tax rate. And even when there is some social cost to financing bailouts, so that $\lambda > 0$, it will generally not be optimal for the government to commit to never bail out.

Summarizing so far:

Proposition 3a If the government can commit, then full bailout ($\mu = 1$) is optimal if $\lambda = 0$. And while it is necessary for less than full bailout to be optimal that $\lambda > 0$, this is not sufficient.

It is perhaps more plausible to suppose, however, that the government cannot credibly commit to its bail out policy, but must decide whether or not to bail out creditors conditional on the realization of r. When this is below the critical level R, its options are to either allow unmitigated failure, resulting in social costs of ΔK, or to bail out, incurring costs of $\lambda(\rho B - rL)$ in raising tax revenue to top up the residual value of assets so as to leave creditors whole. So bailout is ex post optimal if only if

$$r \geq \tilde{R}(k) \equiv \tilde{\rho}(1-k) - \left(\frac{k\Delta}{\lambda}\right). \tag{24}$$

where $\tilde{\rho}$ denotes the rate at which the bank can borrow in these circumstances. The government will thus bailout only if the failure is not too spectacular: when $r < \tilde{R}$, the bank is 'too big to bail'.[39]

39 The practical importance of this possibility is stressed by Demirgüç-Kunt and Huizinga (2013).

Taking account of the implications for the bank's borrowing rate, which is now determined not by (6) but by

$$\zeta B = \rho(k)\{1-\Phi(\tilde{R})\} + \left(\frac{1}{1-k}\right)\int_{-\infty}^{\tilde{R}} r\phi(r)dr, \tag{25}$$

and proceeding as in deriving (8) above it is straightforward to show that both the payoff to the bank and the objective of the government differ in the no-commitment case only in that μS is replaced by

$$\tilde{S}(k) \equiv \frac{1}{k}\int_{\tilde{R}}^{R} (R-r)\phi(r)dr, \tag{26}$$

where $R(k) = \tilde{\rho}(1-k) > \tilde{R}$. This simply recognizes that the bank is now bailed out only in the more circumscribed circumstances in which the return on its loans is below R but above $\tilde{R}$.

Proceeding as for Proposition 1, the optimal bailout-related corrective tax is in the no-commitment case then given by $-(1+\lambda)\tilde{S}_k k^2$, where from (26)

$$k\tilde{S}_k = -\tilde{S} + R_k\{\Phi(R) - \Phi(\tilde{R})\} - \tilde{R}_k(R-\tilde{R})\phi(\tilde{R}). \tag{27}$$

This is evidently more complex than the analogous expression for the commitment case, equation (11). To see the key difference, recall that in the commitment case an increase in the capital ratio reduces the value of the bailout subsidy by making it less likely that the bank will fail and hence less likely that any bailout will come into play. In the no commitment case, in contrast a higher capital ratio makes it cheaper for the government to bail out creditors, and hence more likely that it will choose to do so – which acts in the direction of the bank's choosing a *higher* capital ratio than otherwise.

This new consideration has a striking implication: the optimal bailout related component τ_B may be strictly negative. That is, the incentive for the bank to lower its borrowing costs by making itself more salvageable could be so strong that the optimal corrective policy is to tax not borrowing, but equity capital. A sufficient condition for this is that $\Phi(R)$ be concave between $\tilde{R}$ and R: not what one would normally suppose, but enough – given too that clearly $\tilde{S}_k < 0$ for sufficiently low $\tilde{R}$ (which effectively takes us back to the commitment case) – to establish that the optional corrective tax in the no commitment case can take either sign. More precisely:

Proposition 3b When the government cannot commit, the optimal bailout-related corrective tax is $-(1+\lambda)\tilde{S}_k k^2$. Its sign is ambiguous, a sufficient condition for it to be strictly negative being that $\Phi(R)$ is concave between $\tilde{R}$ and R.

Proof. See appendix 7.3.

5. Heterogeneous Banks

The assumption of a representative bank is a reasonable first pass at thinking about either banks that are so systemically important that what happens to them is effectively all that matters or so small that that their impact on others can be ignored. But this is clearly a restrictive view, and this section sets about relaxing it.

5.1. Unconnected Banks

Suppose now that there are two banks, A and B, distinguished by superscripts. They are unconnected in the sense that they do not transact with one another; the returns that they earn on their loans, however, may be correlated. Both are as described in Section II, identical except perhaps in the marginal distributions and realizations of the returns earned on their loans and perhaps in the social damage that their collapse would cause. The joint distribution of the returns on the loans they make is denoted by $\Phi(r^A, r^B)$. Each bank is treated in the same way by the government, so each faces the same optimization problem, which is as above. They may choose different capital ratios because they differ in the distribution of the return on the loans that they make (or, perhaps, in bankruptcy costs).

From the social perspective, however, a range of possibilities now arise as to whether neither bank, both banks, or only one bank fails – which may have quite different external effects. This could arise in terms of both the collapse and the bailout externality. For the latter, it could be that a rising marginal cost of public funds makes it more than twice as expensive to bail out two banks as it is to bail out one. Here, however, we focus on non-linearity in relation to the costs of bank failure. This can be captured by distinguishing between the collapse costs Δ^A and Δ^B associated with failure of only one or other bank and a collapse cost of Δ^{AB} when both fail. The assumption that

$$\Delta^{AB} > \Delta^A + \Delta^B \tag{28}$$

then captures the idea that some additional social cost arises when both banks fail beyond those associated with the isolated failure of each.

Recalling (15), and assuming for simplicity that the extent of bailout μ is the same for isolated and simultaneous failures, the government's maximand is now

$$w = \sum_{i=A,B} \omega^i(k^i) + (1-\mu)\Omega(k^A, k^B) \tag{29}$$

where, denoting marginal distribution and densities of r^i by Φ^i and ϕ^i,

$$\omega^i(k^i) \equiv \Phi^i(R(k^i,\mu))\delta^i + \int_{-\infty}^{\infty} \left\{ r\left(\frac{1}{k_i}\right) - \zeta\left(\frac{1}{k^i} - 1\right) \right\} \phi^i dr \\ - \lambda\mu S(k_i,\mu) + \tau\left(\frac{1}{k^i} - 1\right) \tag{30}$$

and

$$\Omega(k^A,k^2) \equiv \Delta^A \int_0^{R(k^A)} \int_{R(k^B)}^{\infty} \phi(r^A,r^B) dr^A dr^B \\ + \Delta^B \int_0^{R(k^B)} \int_{R(k^A)}^{\infty} \phi(r^A,r^B) dr^B dr^A \\ + \Delta^{AB} \int_0^{R(k^A)} \int_0^{R(k^B)} \phi(r^A,r^B) dr^B dr^A \tag{31}$$

the three terms of which correspond to failure only of A (which arises when r^A falls short of the critical R^A but r^B exceeds R^B), of B only, and of both banks.

The only change to the government's problem thus arises through the more complex structure of expected collapse costs. To see the implications, note that differentiating in (31) shows the effect on expected collapse costs of a small increase in k_A – which will drive the level of the corrective tax – to be

$$\frac{\partial\Omega}{\partial k^A} = \left\{ \Delta^A \int_{R(k^B)}^{\infty} \phi(R(k^A),r^B) dr^B - \Delta^B \int_0^{R(k^B)} \phi(R(k^A),r^B) dr^B \right. \\ \left. + \Delta^{AB} \int_0^{R(k^B)} \phi(R(k^A),r^B) dr^B \right\} R_k^A. \tag{32}$$

The interpretation here is that an increase in A's capital ratio makes it less likely that only A will fail (the first term on the right), and (hence) also less likely that both banks will fail (the third term), but makes it more likely that only B will fail (the second). Rearranging this and denoting by $\Phi^{r^B|r^A}$ the conditional distribution of r^B, it is then straightforward to show, proceeding as in deriving Proposition 1, that:

Proposition 4 With two unconnected banks, the optimal corrective tax on borrowing by bank A is given by $\tau = (1-\mu)\tau_C + \mu\tau_B \geq 0$ where

$$\tau_C \equiv \{\Delta^A + (\Delta^{AB} - \Delta^A - \Delta^B)\Phi^{r^B|r^A}(R^B|R^A)\}\phi^A(R^A)R_k^A(k^A)^2 > 0 \tag{33}$$

while τ_B remains as in (19). That for bank B is symmetric.

Comparing with (18) of Proposition 1, so long as there is some systemic loss from a failure of both banks in the sense of (28) (and there is some probability that B will fail when A is at the cusp of failure) the optimal corrective tax on each bank is thus unambiguously greater than it would be if each were the only bank in existence.

This is so, importantly, whatever the joint distribution of banks' returns: the additional tax component is positive, in particular, whether the correlation in the banks' returns is positive or negative. The nature of the correlation does, however, affect the magnitude of this additional component of the corrective tax. With independent returns, it becomes simply $(\Delta^{AB}-\Delta^{A}-\Delta^{B})\Phi^{B}(R^{B})$. A positive covariance between the banks' returns would be expected to increase this component, leading to a greater corrective tax, since bank B is then more likely to fail when A is on the cusp of failure. If returns are jointly normal distributed, for instance, it can be shown that $\Phi^{r^B|r^A}(R^B|R^A)$ is increasing in their covariance so long as $R^i < E[r^i]$, for $i = A$ and B. In the limit, when the returns of the two banks are perfectly correlated (so that there is no chance of only one failing) the analysis reduces to that of a single bank as above, with each bank optimally taxed at a rate reflecting the social cost Δ^{AB} of their both collapsing.

5.2. Connected Banks

One key source of systemic importance as the concept emerged from the crisis is that of interconnectedness: the idea that the distress or failure of one institution directly increases the likelihood of distress or failure for others. And one important source of such contagion, analyzed for instance by Allen and Gale (2000), is inter-bank lending.

Imagine then that there are again two banks modeled as in Section II, but now with bank B borrowing from bank A, but not conversely. Then B acquires systemic importance in the sense that its failure, an inability to repay its creditors will make failure of bank A more likely. This evidently makes the analysis far more complex, but one likely conclusion seems clear: relative to the characterization of the optimal tax on unconnected banks in Proposition 4, with connected banks the corrective tax on borrowing by the systemic bank B in relation to the collapse externality will include an additional term capturing the increase in the likelihood of A's failure conditional on B's failure.

What quickly becomes clear, however, is that there is in this case an additional and potentially important tax instrument to consider, beyond that on bank borrowing in general: one on inter-bank borrowing. Addressing the richer possibilities that thus arise with connected banks is an important but difficult task that is not attempted here.

5.3. Asymmetric Information

The assumption so far has been that the bank and government have the same information on how the banks' capital ratio affects its payoff and the probability of its failure. In practice, banks are likely to have superior information as to their own circumstances that affect these relationships – such as the riskiness of their asset positions, the quality of their managers, and their willingness to accept risk.

To see the possible implications of this for optimal tax policies – and their link with regulatory ones too, it will turn out – suppose now, adopting a stylized version of the model above, that the bank's maximand is of the form $\pi(k,a)$, strictly concave in k, and where a – referred to as 'efficiency' though this is not the only interpretation – is some exogenous characteristic known by the bank but not observed by the policy maker. Greater efficiency leads to higher profits, so $\pi_a > 0$, and this effect is assumed to be stronger at lower levels of the capital ratio (perhaps because of the greater difficulty of monitoring the larger volume of loans this implies): thus, as a single-crossing condition, $\pi_{ka} < 0$. The privately optimal capital ratio for a bank of type a, $\overline{k}(a)$, is thus defined by

$$\pi_k(\overline{k},a) = 0 \tag{34}$$

and is decreasing in a. Social welfare, in contrast, is given by $\pi(k,a) - \theta(k,a)$, where – again following the broad structure of the model above, while shedding its details – the term θ can be thought of as an amalgam of collapse and bailout externalities; it is assumed that $\theta_k < 0$, $\theta_{kk} > 0$ and $\theta_{ka} \geq 0$. The first best capital ratio for type a, implicitly defined by

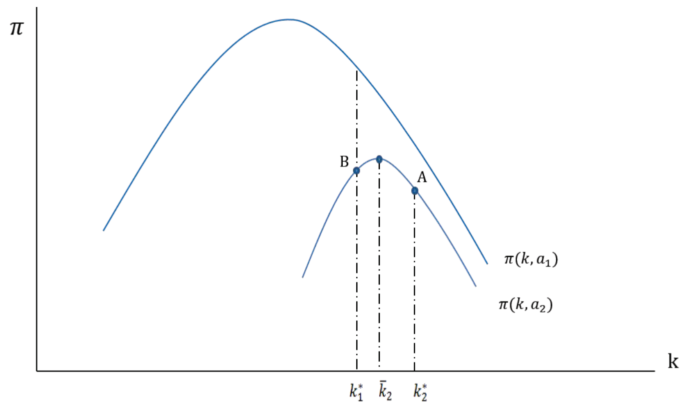
$$\pi_k(k^*,a) = \theta_k(k^*,a) < 0, \tag{35}$$

is then readily shown to be decreasing in a. Clearly too $k^*(a) > \overline{k}(a)$, so that, as in the simple model of Section II, the privately optimal capital ratio is, for any type, lower than is socially desirable.

Supposing there to be just two possible efficiency types, with $a_1 > a_2$, it follows that the more efficient bank has the lower first best capital ratio: in obvious notation, $k_1^* < k_2^*$. The question is how this allocation can be implemented. Figure 1 illustrates, showing the payoff each type as a function of k.

Importantly, regulation alone – in the form of a minimum capital requirements – cannot implement the first best, because of the self-selection constraints that need to be respected. If regulation takes the form of a minimum capital requirement, this would have to be set at k_1^* in order to place the high efficiency bank at the appropriate capital ratio. But then the low efficiency bank is unrestricted in the neighborhood of its first best k_2^*, where its profits

Figure 1
Bank Payoffs with Differing Efficiencies



are strictly decreasing in k, and so will not choose that first-best: as drawn in figure 1, it will instead choose its own private optimum $\overline{k}_2$.

An alternative strategy is to offer banks a choice between capital ratios k_1^* and k_2^*. The difficulty then arises that the less efficient bank may prefer the low capital ratio intended for the more efficient type to that intended for itself:[40] this is the case in figure 1, the payoff to the low efficiency type 2 being higher at point B than at A.

The figure also suggests, however, that the first can be implemented by levying a tax T on any bank choosing the lower capital ratio, intended for the higher efficiency bank, that is high enough to deter the low efficiency from mimicking the high efficiency bank – but not so large as to induce the high efficiency bank to switch to the higher capital ratio that is socially appropriate for the low efficiency bank. This means finding an amount T such that the self-selection constraints

$$\pi(k_2^*, a_2) \geq \pi(k_1^*, a_2) - T \tag{36}$$

$$\pi(k_1^*, a_1) - T \geq \pi(k_2^*, a_1) \tag{37}$$

are both satisfied.[41] That such a T can be found is established in appendix 7.4, giving

40 Note that in the absence of any tax the self-selection constraint will never bite for the high ability bank, since convexity of π in k means that that $\pi(k, a_1)$ is decreasing in k above $\overline{k}_1$, and $k_2^* > k_1^* > \overline{k}_1$.

41 It is assumed that $\pi(k_2^*, a_2) \geq 0$, so that the participation constraint for the low efficiency type is met at its socially optimal capital ratio. It is then straightforward to show that the participation constraint for the more efficient type will also be met if the self-selection constraint (15) is met.

Proposition 5 The first-best can be implemented by offering banks the choice between (k_1^*, T) and $(k_2^*, 0)$, where

$$T \equiv \max\{\pi(k_1^*, a_2) - \pi((k_2^*, a_2), 0\} \geq 0. \tag{38}$$

Implementation can thus be achieved by offering banks a menu that allows them to choose a capital ratio lower than some norm only on payment of an appropriate tax.[42] This result points too towards an integration of regulatory and tax policies, being equivalent to setting a minimum capital requirement as the norm, with the option of choosing a lower level conditional on payment of tax. (Or, equivalently, to setting a minimum capital ratio but providing an appropriate tax reduction or subsidy to banks choosing a higher ratio). The nonlinear tax schemes to which Proposition 5 thus points are potentially complex – but not obviously any more so than the differentiated capital requirements under Basel III.

6. Concluding

Before the Great Financial Crisis, the presumption of tax policy makers was that banks should be taxed in essentially the same way as all other businesses. The externalities associated with bank failures were something for regulators and supervisory authorities to worry about and take care of. While recalibration of regulatory and supervisory oversight has been a primary policy response to the financial sector externalities so painful during the crisis, the emergence of bank taxes is a marked departure from that prior view of an essentially passive role for taxation. Beyond drawing a general analogy with Pigovian taxes, however, relatively little formal attention was paid to the design of such taxes from an explicitly corrective perspective. The aim in this paper has been to go some way to providing such an analysis.

For the benchmark case of a single representative bank, the analysis here establishes the optimal corrective tax on bank borrowing as a weighted average of two components, with the weights reflecting the probability that collapse will be averted by bailing out creditors. One component addresses the collapse externality (and so is weighted by the probability of collapse). In the case of a representative bank, this takes a simple and predictable form: bank borrowing is taxed at rate equal to the product of the impact of higher leverage on the probability of failure and the social damage that failure would cause. The other component of the optimal corrective tax is addressed to the social costs of guaranteeing bank creditors in order to avoid collapse. This depends not only on the marginal cost of public funds – which shapes the social cost of taxpayer support – but, in more complex ways, or the extent to which the

42 Boyer and Kempf (2017) arrive at a similar result.

government can credibly commit to such bailout. When the government can commit, the extent of that commitment can itself be seen as a choice variable – and it has been seen here that when the marginal cost of public funds is low, it may indeed be optimal to wholly insure bank creditors. Importantly, given the fee-type rationale sometimes given for bank taxation. the bailout component – mitigating the incentive for the bank to take on excessive risk in the expectation of public support in the event of failure – is optimally set above the insurance-like level that would recoup the expected fiscal cost of future bailouts. When the government cannot commit to its bailout policy, and may lack the resources to bail out creditors in the event of failure, this component of the optimal corrective charge may well be lower, since the bank itself then has an incentive to limit its borrowing so as to make it relatively cheap for the government to bail out its creditors.

Additional considerations arise with heterogenous banks. When banks are not directly connected but differ in the distributions of their returns, then – whatever the correlation between these returns – the optimal corrective tax is higher than in the case of a representative bank so long as there are additional social costs from simultaneous failures. Connectedness through one-way interbank loans, the analysis here also suggests, points to a still higher corrective tax on the depositing bank – an outcome, importantly, that no single-rate bank tax (or uniform capital requirement) can achieve. A similar conclusion emerges from the analysis of asymmetric information, with the optimal policy implemented by offering banks a menu that involves a higher tax charge on those that, being more efficient, wish to operate with a lower capital ratio.

The focus of this paper is in important respects narrow. Technically, there is much that is left open by the analysis here. Inter-connectedness, in particular, raises complex issues of both definition and measurement – as explored, for instance, in Acharya et al. (2016) and Adrian and Brunnermeier (2016) – as well as, potentially, a richer set of tax instruments than the simple charge on bank borrowing considered here. Nor, more fundamentally, has the paper taken up the relative merits of taxation and regulation. Much of the analysis could be interpreted, indeed, in terms of defining optimal capital requirements. What then does emerge, however, is the inadequacy of applying the same capital requirements to heterogeneous banks – and indeed Basel III steps ways from that, with the introduction of supplementary requirements for systemically important institutions. Intellectually at least – and therefore perhaps, at some point, in practical terms too – the question of whether that differentiation is best achieved by tax or regulatory measures (which has hardly been raised, for instance, in the context of inter-connectedness), remains open.

The discussion has been narrow too in terms of practical policy priorities. Given current tax policies, the question of whether Pigovian taxes on bank barrowing are appropriate is very much second order. The first order issue is the systemic bias towards debt finance inherent in most corporate tax systems. The Pigovian question is whether borrowing should be penalized. But existing debt bias means that it is now inherently, and very extensively, tax-favored. Bank taxes could in principle be a way to offset that bias in the financial sector. But that would require such taxes to be levied at a much higher rate than observed in practice: at a borrowing rate of 5 percent, for example, undoing the effects of deductibility at a corporate tax rate of 20 percent would require a tax on borrowing of 100 basis points. Recent work suggests that the social costs of this debt bias may well be high: De Mooij et al. (2014), for example, put the gain in expected output from eliminating the debt bias associated a corporate tax rate of 28 percent at up to 12 percent of GDP. It is, at the very least, perverse that regulatory measures designed to discourage excessive bank borrowing are combined with tax systems that do the exact opposite. Dealing with this remains the central challenge in fixing the tax treatment of the financial sector.

7. Appendix

7.1. Proof of Proposition 2

The revenue raised by the component τ_B of the optimal tax on borrowing is $-(1+\lambda)S_k k^2 B$, while the cost of bailout is SK. Since $\lambda \geq 0$ and $B/K = (1-k)/k$, it therefore suffices to show that $-k(1-k)S_k > S$. For this, note from (11) that

$$-k(1-k)S_k = (1-k)S - (1-k)R_k\Phi \tag{39}$$

$$= (1-k)S + R\Phi - (1-k)^2\rho_k\Phi \tag{40}$$

$$= S + \int_{-\infty}^{R} r\phi dr - (1-k)^2\rho_k\Phi > S, \tag{41}$$

where the second equality uses the implication of (4) that $R_k = \rho_k(1-k) - \rho$, and the third follows from (10).

7.2. Derivation of Equation (21)

With $\mu = 1$, (6) implies that $\rho = \zeta$, and so from (9)

$$S = \zeta\Phi[\zeta(1-k)]\left(\frac{1-k}{k}\right) - \left(\frac{1}{k}\right)\int_{-\infty}^{\zeta(1-k)} r\phi(r)dr. \tag{42}$$

Differentiating this with respect to k and canceling terms gives

$$S_k = -\zeta\Phi(R) + \left(\frac{1}{k}\right)^2 \int_{-\infty}^{\zeta(1-k)} r\phi(r)dr \tag{43}$$

from which, the integral term being non-negative and recalling the definition of τ_C in (19), the inequality in (21) follows.

7.3. Proof of Proposition 3b

Note first that, subtracting and adding $\tilde{R}_k\{\Phi(R)-\Phi(\tilde{R})\}$, (27) can be written as

$$k\tilde{S}_k = -\tilde{S} + (R_k - \tilde{R}_k)\{\Phi(R)-\Phi(\tilde{R})\} + \tilde{R}_k\{\Phi(R) - \Phi(\tilde{R}) - (R-\tilde{R})\phi(\tilde{R})\}. \quad (44)$$

From (26),

$$k\tilde{S} = \int_{\tilde{R}}^{R} (R-\tilde{R})\phi(r) + \int_{\tilde{R}}^{R} (\tilde{R}-r)\phi(r)dr \quad (45)$$

$$= (R-\tilde{R})\{\Phi(R)-\Phi(\tilde{R})\} + \int_{\tilde{R}}^{R} (\tilde{R}-r)\phi(r)dr \quad (46)$$

and hence, since $R-\tilde{R} = k(R_k - \tilde{R}_k)$,

$$-S = -(R_k - \tilde{R}_k)\{\Phi(R)-\Phi(\tilde{R})\} + \left(\frac{1}{k}\right)\int_{\tilde{R}}^{R} (r-\tilde{R})\phi(r)dr. \quad (47)$$

Substituting this into (44) gives

$$k\tilde{S}_k = \tilde{R}_k\{\Phi(R)-\Phi(\tilde{R}) - (R-\tilde{R})\phi(\tilde{R})\} + \left(\frac{1}{k}\right)\int_{\tilde{R}}^{R} (r-\tilde{R})\phi(r)dr. \quad (48)$$

Since the final term of the right of (48) is strictly positive for $R > \tilde{R}$ and concavity of Φ over this range implies that $\Phi(R)-\Phi(\tilde{R}) < (R-\tilde{R})\phi(\tilde{R})$, it suffices to show that $\rho_k < 0$ and hence that $\tilde{R}_k < 0$.

For this, dividing by B and differentiating in (25) gives, on canceling terms and rearranging

$$\rho_k = -\left(\frac{1}{(1-\Phi(\tilde{R})(1-k)^2}\right)\int_{-\infty}^{\tilde{R}} r\phi(r)dr < 0 \quad (49)$$

as required.

7.4. Proof of Proposition 5

There are two possibilities.

The first is that $T = \pi(k_1^*, a_2) - \pi(k_2^*, a_2) > 0$. In this case, in the absence of any tax, the self-selection constraint on the low efficiency type would bind:

so (36) must hold with equality. To see that (37) holds, note that since $k_1^* < k_2^*$ and $\pi_{ka} < 0$,

$$\int_{k_1^*}^{k_2^*} \{\pi_k(k,a_2) - \pi_k(k,a_1)\} dk > 0. \tag{50}$$

Hence

$$\pi(k_2^*,\ a_2) - \pi(k_1^*,a_2) > \pi(k_2^*,a_1) - \pi(k_1^*,a_1) \tag{51}$$

or

$$-T > \pi(k_2^*,a_1) - \pi(k_1^*,a_1) \tag{52}$$

which gives (37).

The second possibility is that $\pi(k_1^*,a_2) - \pi(k_2^*,a_2) < 0$, and hence $T = 0$. In this case it is immediate that (36) holds. And (37) holds because $k_2^* > k_1^* > \overline{k}_1$ (so that k_2^* is further along the downward-sloping part of $\pi(k,a_1)$ than is k_1^*).

References

Acharya, V., Pedersen, L., Philippon, T., and Richardson, M. (2016), Measuring Systemic Risk, Review of Financial Studies 30, 2–47.

Admati, A., and Hellwig, M. (2013), The Bankers' New Clothes: What's Wrong with Banking and What to Do about It, Princeton University Press, Princeton.

Adrian, T., and Brunnermeier, M. (2016), CoVaR, American Economic Review 106, 1705–1741.

Allen, F., and Gale, D. (2000), Financial Contagion, Journal of Political Economy 106, 1–33.

Bank of England (2009), The Role of Macroprudential Policy, Discussion Paper.

Baker, D., and McArthur, T. (2009), The Value of the "Too Big to Fail" Big Bank Subsidy, Center for Economic and Policy Research, Washington, D.C.

Besley, T., and Ghatak, M. (2013), Bailouts and the Optimal Taxation of Bonus Pay, American Economic Review 103, 163–167.

Bianchi, J., and Mendoza, E. (2010), Overborrowing, Financial Crises and 'Macro-Prudential' Taxes, NBER Working Paper 16091.

Basel Committee on Banking Supervision (2010), An Assessment of the Long-Term Impact of Stronger Capital and Liquidity Requirements, http://www.bis.org/publ/bcbs173.htm (Access Date: 2017-12-09).

Boyer, P. C., and Kempf, H. (2017), Regulatory Arbitrage and the Efficiency of Bank Regulation, Journal of Financial Intermediation, 1–17.

Buch, C. M., Hilberg, B., and Tonzer, L. (2016), Taxing Banks: An Evaluation of the German Bank Levy, Journal of Banking and Finance 72, 52–66.

Claessens, S., Keen, M., and Pazarbasioglu, C. (2010), Financial Sector Taxation: The IMF's Report to the G-20 and Background Material, https://www.imf.org/external/np/seminars/eng/2010/paris/pdf/090110.pdf (Access Date: 2017-12-09).

Coulter, B., Mayer, C., and Vickers, J. (2014), Taxation and Regulation of Banks to Manage Systemic Risk, in: De Mooij, R., and Nicodeme, G. (Eds), Taxation and Regulation of the Financial Sector, MIT Press, Boston, 67–88.

De Mooij, R., and Keen, M. (2016), Debt, Taxes and Banks, Journal of Money, Credit and Banking 48, 5–33.

De Mooij, R., Keen, M., and Orihara, M. (2014), Taxation, Bank Leverage and Financial Crises, in: De Mooij, R., and Nicodeme, G. (Eds), Taxation and Regulation of the Financial Sector, MIT Press, Boston, 235–260.

Demirgüç-Kunt, A., and Huizinga, H. (2013), Are Banks Too Big to Fail or Too Big to Save? International Evidence from Equity Prices and CDS Spreads, Journal of Banking and Finance 37, 875–894.

Devereux, M. P., Johannesen, N., and Vella, J. (2013), Can Taxes Tame Banks? Evidence from European Bank Levies, Centre for Business Taxation Working Paper no. 1325, Oxford University.

Dewatripont, M., and Tirole, J. (1993), The Prudential Regulation of Banks, MIT Press, Cambridge, MA.

Federal Reserve Bank Of Minneapolis (2016), The Minneapolis Plan to End Too Big To Fail, https://www.minneapolisfed.org/~/media/files/publications/studies/endingtbtf/the-minneapolis-plan/the-minneapolis-plan-to-end-too-big-to-fail-2016.pdf?la=en (Access Date: 2017-12-09).

Haldane, A. (2009), Banking on the State, 12th Annual Federal Reserve of Chicago International Banking Conference, www.bankofengland.co.uk/publications/speeches/2009/speech409.pdf (Access Date: 2017-12-09).

Hellman, T. F., Murdock, K. C., and Stiglitz, J. E. (2000), Liberalization, Moral Hazard in Banking and Prudential Regulation: Are Capital Requirements Enough?, American Economic Review 90, 147–165.

HM Revenue & Customs (2010), Bank Levy, http://www.hmrc.gov.uk/budget-updates/autumn-tax/tiin1065.htm (Access Date: 2017-12-09).

Hemmelgarn, T., and Teichmann, D. (2014), Tax Reforms and the Capital Structure of Banks, International Tax and Public Finance 21, 645–693.

Huang, R., and Ratnovski, L. (2009), The Dark Side of Wholesale Funding, IMF, mimeo.

International Monetary Fund (2010), A Fair and Substantial Contribution by the Financial Sector: Final Report for the G-20, www.imf.org/external/np/g20/pdf/062710b.pdf (Access Date: 2017-12-09).

International Monetary Fund (2014), Back to Work: How Fiscal Policy Can Help, Fiscal Monitor, http://www.imf.org/en/Publications/FM/Issues/2016/12/31/Back-To-Work-How-Fiscal-Policy-Can-Help (Access Date: 2017-12-09).

Jeanne, O., and Korinek, A. (2010), Managing Credit Booms and Busts: A Pigouvian Taxation Approach, Johns Hopkins University, mimeo.

John, K., John, T., and Senbet, L. (1991), Risk-Shifting Incentives of Depository Institutions: A New Perspective on Federal Deposit Insurance Reform, Journal of Banking and Finance 15, 895–915.

Keen, M. (2011a), The Taxation and Regulation of Banks, IMF Working Paper 11/206.

Keen, M. (2011b), Rethinking the Taxation of the Financial Sector, CESifo Economic Studies 57, 1–24.

Keen, M., Krelove, R., and Norregaard, J. (2016), The Financial Activities Tax, Canadian Tax Journal 64, 398–400.

Kocherlakota, N. (2010), Taxing Risk and the Optimal Regulation of Financial Institutions, Economic Policy Paper, Federal Reserve Bank of Minneapolis.

Korinek, A. (2009), Systemic Risk-Taking: Amplification Effects, Externalities, and Regulatory Responses, University of Maryland, mimeo.

Laeven, L., and Valencia, F. (2016), Systemic Banking Crises Database, IMF Economic Review 61, 55–83.

Levine, R. (2005), Finance and Growth: Theory and Evidence, University of Minnesota, mimeo.

Perotti, E., and Suarez, J. (2009), Liquidity Insurance for Systemic Crises, CEPR Policy Insight 31.

Shackelford, D. A., Shaviro, D., and Slemrod, J. (2010), Taxation and the Financial Sector, National Tax Journal 63, 781–806.

Shavell, S. (2011), Corrective Taxation Versus Liability, American Economic Review: Papers and Proceedings 10, 273–276.

Shin, H. S. (2010), Non-Core Liabilities Tax As a Prudential Tool, policy memo, http://www.princeton.edu/~hsshin/www/NonCoreLiabilitiesTax.pdf (Access Date: 2017-12-09).

Shull, B. (2010), Too Big to Fail in a Financial Crisis, Hunter College, CUNY, Working Paper No. 601.

Sinn, H.-W. (2003), Risk-Taking, Limited Liability, and the Competition of Bank Regulators, FinanzArchiv / Public Finance Analysis 59, 305–329.

Sinn, H.-W. (2010), Casino Capitalism, Oxford University Press, Oxford.

Von Ehrlich, M., and Radulescu, D. (2017), The Taxation of Bonuses and Its Effect on Executive Compensation and Risk-Taking: Evidence from the UK Experience, Journal of Economics & Management Strategy 26, 712–731.

Wagner, W. (2010), In the Quest of Systemic Externalities: A Review of the Literature, CESifo Economic Studies 56, 96–111.

Weder di Mauro, B. (2010), Taxing Systemic Risk: Proposal for a Systemic Risk Charge and a Systemic Risk Fund, University of Mainz, mimeo.

Natural Resource Extraction in a Federation

Robin Boadway, Motohiro Sato, and Jean-François Tremblay*

Received 03 August 2017; in revised form 14 September 2017; accepted 06 November 2017

We analyze a natural resource extraction problem in a two-region economy with mobile labour. One region produces manufacturing goods while the other produces agriculture and extracts a non-renewable resource. Manufacturing production exhibits increasing returns-to-scale if the production level is sufficiently high. There are multiple long-run equilibrium labour allocations towards which the economy may converge. Under decentralized resource management, a tendency to over-extract the resource relative to the federal optimum makes convergence to a low-income equilibrium more likely. Optimal extraction from the federation's perspective satisfies a modified Hotelling's rule that takes into account the impact of resource extraction on manufacturing production.

Keywords: natural resource extraction, decentralization, inter-regional mobility

JEL classification: H 70, Q 32, Q 33

1. Introduction

It is well-known that increases in a country's resource wealth do not necessarily translate into sustained growth and higher living standards (e.g. Sachs and Warner, 2001; van der Ploeg, 2011). Natural resource wealth can turn into a curse if resource extraction results in a reallocation of production factors away from sectors with high productivity growth, or if the rents derived from non-renewable resource exploitation are not optimally invested in other productive assets such as public infrastructure. In this paper, we examine how the potential for a resource curse may be exacerbated in a federal setting with decentralized natural resource management and inter-regional labour mobility.

There are various mechanisms by which natural resource exploitation can have a negative impact on aggregate production. Krugman (1987, 1991) and Sachs and Warner (1999) examined how a resource boom can lower aggre-

* Boadway: Queen's University, Kingston, K7L 3N6, Canada (boadwayr@econ.queensu.ca); Sato: Hitotsubashi University, Tokyo, 186-8601, Japan (satom@econ.hit-u.ac.jp); Tremblay: University of Ottawa, Ottawa, K1N 6N5, Canada (Jean-Francois.Tremblay@uottawa.ca).
Prepared for the Festschrift in honour of Hans-Werner Sinn. We explore themes to which Hans-Werner made important contributions early in his career, including natural resource extraction and dynamic policy analysis. We have benefited from insightful comments by the referee.

FinanzArchiv 74 (2017), 34–51
ISSN 0015-2218
doi: 10.1628/001522118X15109346479939

gate productivity growth by attracting factors of production away from sectors where production involves learning-by-doing or increasing returns-to-scale. Sachs and Warner (2001) provide empirical evidence consistent with this explanation of the resource curse. Corden and Neary (1982) showed how a resource boom may reduce the competitiveness of traded-goods sectors by inducing a real appreciation, ultimately leading to lower aggregate production. A vast literature, recently surveyed by van der Ploeg (2011), has examined how natural resource exploitation may reduce aggregate growth because of rent-seeking behaviour, corruption and conflict. Our analysis focuses on additional sources of inefficiencies that can arise because of decentralization in the management of natural resource exploitation. Our paper is related to Raveh (2013) who examined how labour mobility might alleviate Dutch disease effects in a federal setting, although his analysis focuses mainly on tax competition incentives.

Our analysis is also related to the fiscal federalism literature that focuses on efficiency in the allocation of labour across regions (e.g. Flatters et al., 1973; Boadway and Flatters, 1982; Gordon, 1983; Albouy, 2012), and on the existence of multiple equilibrium allocations of labour in the presence of agglomeration effects (e.g. Mitsui and Sato, 2001; Baldwin and Krugman, 2004; Bucovetsky, 2005). We contribute to this literature by introducing a dynamic non-renewable resource extraction problem in a federal setting with labour mobility.

We consider a dynamic two-region model with a natural resource sector, a manufacturing sector and an agricultural sector. The manufacturing sector is located in one region whereas the natural resource and agricultural sectors are located in the other. The manufacturing sector exhibits increasing returns-to-scale and requires public infrastructure provided by the regional government. The rate of extraction of the non-renewable resource is controlled by the government of the resource region. There is labour mobility across regions, although migration requires time so per capita incomes are only gradually equalized across regions. In this setting, the analysis shows that there are multiple equilibrium allocations of labour towards which the economy may converge in the long-run. Initial conditions with respect to the stock of resources and the allocation of labour across regions determines the equilibrium towards which the economy will converge. An increase in the price of the natural resource or a decrease in the share of resource rents captured by producers tend to shrink the set of initial conditions under which the economy converges to the high-income equilibrium. Under decentralization, the extraction decision of the government of the resource region is distorted relative to the constrained federal optimum, which tends to make convergence to the inefficient long-run equilibrium more likely. We derive a modified Hotelling rule for optimal resource extraction that takes into account the fact that resource extraction shifts labour away from the manufacturing sector thereby diluting the gains from economies of scale.

2. The Model

There are two regions M and R, each specializing in different types of production. Region M produces manufacturing goods using one of two possible technologies: a traditional one with constant returns to scale or a modern one with increasing returns. The modern technology requires public infrastructure and is only adopted if the manufacturing sector reaches a minimum size. Public infrastructure is provided by the regional government and is financed with a labour income tax. Manufacturing goods are tradable. In region R, there is an agricultural sector that operates under a constant returns technology, as a well as a natural resource sector.[1] Agricultural output is tradable across regions but non-tradable internationally. The natural resource is non-renewable and all resource production is sold on international markets. Resource extraction is controlled by the government of region R, for example, by issuing permits. The economy is assumed to be small so manufacturing goods and natural resources are sold at given world prices. There is imperfect interregional mobility in the sense that a reallocation of labour across regions requires time. However, labour is perfectly mobile between the traditional and modern technology in region M and between agriculture and natural resources in region R.

2.1. Manufacturing Sector in Region M

The production structure in manufacturing follows that found in Krugman (1991), Sachs and Warner (1999) and Murphy et al. (1989). Under the traditional technology, manufacturing production at time t is $\tilde{X}_t = \mu L_t^M$, where L_t^M is the amount of labour in region M. With perfect competition in the labour market and the price of manufacturing goods normalized to unity, the wage rate of workers under the traditional technology equals the marginal product of labour, i.e. $\tilde{w}_t^M = \mu$.

Under the modern technology, final manufacturing goods X_t are produced using a continuum of intermediate goods as inputs according to

$$X_t = \left(\int^{N_t} \left(x_t^i\right)^{\sigma} di \right)^{\frac{1}{\sigma}} G_t^{\alpha} \tag{1}$$

1 The structure of the economy with a manufacturing sector exhibiting increasing returns to scale and a constant returns to scale agricultural sector is similar to that in Krugman (1991). Sachs and Warner (1999) consider a similar structure although they do not refer to the two sectors as manufacturing and agriculture.

where x_t^i is the ith intermediate good. The range of intermediate goods spans the interval $[0, N_t]$, and the number of producers N_t is determined endogenously. G_t is the level of public infrastructure provided in region M, and $0 < \sigma$, $\alpha < 1$. There is monopolistic competition among producers of intermediate goods and instantaneous free entry.

Demand for intermediate goods at time t will solve:

$$\max_{\{x_t^i\}} \quad \left(\int^{N_t} \left(x_t^i\right)^\sigma di\right)^{\frac{1}{\sigma}} G_t^\alpha - \int^{N_t} p_t^i x_t^i di$$

where p_t^i is the price of the ith intermediate good. The solution to this problem gives:

$$x_t^i = \left(\int^{N_t} \left(x_t^i\right)^\sigma di\right)^{\frac{1}{\sigma}} G_t^{\frac{\alpha}{1-\sigma}} \left(p_t^i\right)^{\frac{-1}{1-\sigma}} \tag{2}$$

Thus, the demand for x_t^i is increasing in infrastructure G_t and decreasing in price p_i.

The production of intermediate goods uses only labour. The amount of labour required to produce x_t^i units of intermediate good i is:

$$\ell_t^i = a x_t^i + b \tag{3}$$

Given the presence of the fixed cost b, average costs in the production of each intermediate goods are declining. The problem of intermediate goods producers is $\max_{\{p_t^i\}} \quad p_t^i x_t^i - w_t^M \ell_t^i$, where w_t^M is the wage rate. Using (2) and (3) to substitute for x_t^i and ℓ_t^i, and noting that producers take $\int^{N_t} \left(x_t^i\right)^\sigma di$ as given, the solution to this problem yields:

$$p_t^* = \frac{a}{\sigma} w_t^M \tag{4}$$

All intermediate goods have the same equilibrium price and therefore $x_t^i = x_t$ and $\ell_t^i = \ell_t$ for all i. Using (3) and (4), the profit of each intermediate good producer π_t is:

$$\pi_t = p_t x_t - w_t \ell_t = \left(\frac{1-\sigma}{\sigma} a x_t - b\right) w_t^M$$

Free entry of intermediate good producers implies that profits are zero, leading to:

$$x_t = \overline{x} = \frac{\sigma}{1-\sigma} \cdot \frac{b}{a} \tag{5}$$

With the manufacturing sector producing under the modern technology, equilibrium in the labour market is such that total demand for labour by intermediate good producers equals total labour supply in region M, i.e. $L_t^M = N_t(a\overline{x} + b) = N_t(b/1-\sigma)$. Therefore, the number of intermediate good producers at time t is:

$$N_t = \frac{1-\sigma}{b} L_t^M \tag{6}$$

Substituting (4) and (6) in (2), using $x_t^i = \overline{x}$ and solving for the wage rate, we get:

$$w_t^M(L_t^M, G_t) = DG_t^\alpha \left(L_t^M\right)^{\frac{1-\sigma}{\sigma}} \quad \text{where} \quad D \equiv \frac{\sigma}{a}\left(\frac{1-\sigma}{b}\right)^{\frac{1-\sigma}{\sigma}} \tag{7}$$

Thus, the wage rate is increasing in the size of the labour force, reflecting economies of scale in production. Since the level of production of each intermediate good is fixed, an increase in the size of the labour force results in an increase in the variety of intermediate goods. In turn, given that intermediate goods are complementary in the production of final goods, an increase in the variety of intermediate goods raises the productivity of all intermediate goods in final goods production. Higher productivity raises the price of intermediate goods and the wage rate.

With $x_t^i = \overline{x}$, final goods production is:

$$\begin{aligned} X_t &= (N_t \overline{x}^\sigma)^{\frac{1}{\sigma}} G_t^\alpha = \left(\frac{1-\sigma}{b}\right)^{\frac{1}{\sigma}} \overline{x} G_t^\alpha \left(L_t^M\right)^{\frac{1}{\sigma}} \\ &= DG_t^\alpha \left(L_t^M\right)^{\frac{1}{\sigma}} = w_t^M(L_t^M, G_t) L_t^M \end{aligned} \tag{8}$$

Final goods production equals total wages, and therefore, profits of final goods producers equal zero, as do those of intermediate goods producers. Producers in the final goods sector take as given the number of varieties of intermediate goods. From their perspective, production exhibits constant returns to scale. As a result, the costs of purchasing intermediate goods in equilibrium fully exhaust the value of production. Since intermediate goods production uses only labour, the value of final goods will be equal to total wages.

The government in region M levies a wage tax at rate τ_M to finance investment in public infrastructure. Assume for simplicity that government current expenditures are restricted to equal current revenues. Then, the government budget constraint is $G_t = \tau_M w_t^M L_t^M = \tau_M X_t$. Using this, (7) and (8) can be rewritten as:

$$w_t^M = D^{\frac{1}{1-\alpha}} \tau_M^{\frac{\alpha}{1-\alpha}} \left(L_t^M\right)^\gamma, \quad \text{and} \quad X_t = D^{\frac{1}{1-\alpha}} \tau_M^{\frac{\alpha}{1-\alpha}} \left(L_t^M\right)^{\gamma+1} \tag{9}$$

where $\gamma \equiv 1/(\sigma(1-\alpha)) - 1$. Since $0 < \alpha,\ \sigma < 1$, we have $\gamma > 0$, so w_t^M will increase with L_t^M, and X_t is convex in L_t^M. As well, we assume that $\gamma < 1$.

The technology used in manufacturing production in region M will be the one under which the productivity of workers is highest, which in turn depends on the scale of production. Thus, we have the following:

Lemma 1 Manufacturing production will operate under the modern technology if

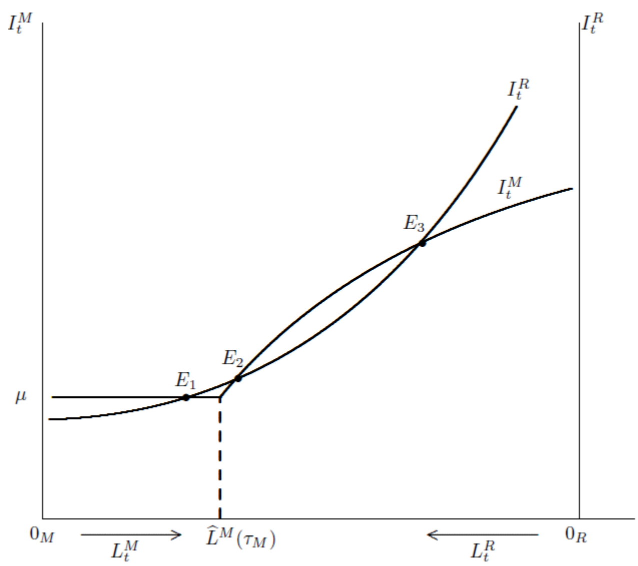
$$(1-\tau_M) w_t^M = K\left(L_t^M\right)^\gamma \geq \mu, \quad \text{where} \quad K \equiv (1-\tau_M) D^{\frac{1}{1-\alpha}} \tau_M^{\frac{\alpha}{1-\alpha}} \tag{10}$$

Let $\widehat{L}_t^M = \widehat{L}^M(\tau_M)$ be the size of the labour force in region M at which the condition above holds with equality.[2] The after-tax income of residents in region M is therefore:

$$I_t^M(L_t^M, \tau_M) = \begin{cases} \mu & \text{if} \quad L_t^M < \widehat{L}_M \\ K\left(L_t^M\right)^\gamma & \phantom{\text{if}} \quad L_t^M \geq \widehat{L}_M \end{cases} \tag{11}$$

Since $\gamma < 1$, I_t^M is concave in L_t^M for $L_t^M > \widehat{L}_M$. Note that if the manufacturing sector operates under the traditional technology, no infrastructure is needed so the wage tax rate is zero. The level of after-tax income in region M as a function of the size of the labour force is illustrated by the curve I_t^M in figure 1.

Figure 1
Interregional Allocation of Labour



2 Note that the after-tax wage rate appears on the left side of the condition above, rather than simply the wage rate, since the modern technology will only be adopted if workers are willing to move to the modern sector and that requires that the after-tax wage rate be greater than the marginal product of labour under the traditional technology.

2.2. Agricultural Sector in Region R

Labour supply in region R, L_t^R, is divided between employment in agriculture, L_t^A, and natural resources, L_t^N. Total population is normalized to 1, so $1 - L_t^M = L_t^R = L_t^A + L_t^N$. Production of agricultural output A_t exhibits constant returns to scale according to:

$$A_t = L_t^A = L_t^R - L_t^N.$$

Residents of both regions derive utility from agricultural goods and manufacturing goods according to $u_t^j = X_t^j + v\left(A_t^j\right)$, with $v'\left(A_t^j\right) > 0$, $v''\left(A_t^j\right) < 0$ and $v'(0) \to \infty$, for $j = M, R$. Let w_t^A denote the wage rate in the agricultural sector. It will be equal to the price of agricultural goods P_t^A. The budget constraint of consumers in region j can be written as $X_t^j + P_t^A A_t^j = I_t^j$ where I_t^j denotes disposal income in region j. Utility maximization yields equal per capita consumption of agricultural goods in each region satisfying $P_t^A = v'(A_t^*)$, and therefore:

$$w_t^A = v'(A_t^*) = v'(L_t^R - L_t^N) \tag{12}$$

For convenience, we assume that $v(A_t) = H \ln(A_t)$ with $H > 0$, so $v'(A_t) = H/A_t$. Since per capita consumption of agricultural goods will be equalized across regions, the migration equilibrium will not depend on agricultural output.

2.3. Natural Resource Sector in Region R

The extraction of natural resources uses labour and manufacturing goods as inputs. The process is assumed to require a fixed amount of labour L_t^N per unit of extraction Z_t, so $L_t^N = Z_t$, and an amount of the manufacturing good that depends on the remaining stock of the resource according to:

$$X_t^N = \phi(S_t) Z_t \equiv C(S_t, Z_t)$$

where S_t denotes the stock of natural resources at time t, $\phi'(S_t) < 0$ so the cost of extraction increases as the stock is depleted, and $\dot{S}_t = -Z_t$.[3]

Let P_t^N denote the unit price of the natural resource. We assume that it increases exogenously over time at a constant rate. Since labour is freely mobile between the agricultural and natural resource sectors, both located in region R, the wage rate in the natural resource sector will equal that in the agricultural sector, so the wage rate in region R is given by $w_t^R = w_t^A$. The rent generated from the extraction of natural resources Π_t is:

$$\Pi_t = P_t^N Z_t - w_t^R L_t^N - \phi(S_t) Z_t = P_t^N Z_t - w_t^R Z_t - \phi(S_t) Z_t \tag{13}$$

3 Our main results do not rely critically on the assumption that the extraction cost depend on the remaining stock. Alternatively, we could consider a stock-independent cost function that would be convex in the instantaneous extraction rate.

Π_t will be shared between labour in region R and resource producers as specified below.

3. Equilibrium under Decentralization

In this section, we characterize the level of infrastructure investment, the level of natural resource extraction and the allocation of labour under decentralization. In setting their policies, we assume for simplicity that regional governments take as given the allocation of labour across regions. They do not foresee the impact of their policies on migration. Considering forward-looking governments with respect to migration would complicate the analysis considerably without adding much insight. Our main objective is to examine the distortion in the extraction rate chosen by the region R government relative to that which is optimal from the perspective of the federation. The extraction rate will be distorted because the resource region has no incentive to take into account the impact of resource extraction on labour productivity in the manufacturing region. If we assumed that the government of the resource region took into account the impact of its policy on labour allocation, the distortion would be amplified. The government of region R does however anticipate the impact of resource extraction on resource depletion.

3.1. Infrastructure Investment in Region M

Regional governments choose their policies to maximize total after-tax income in the current period.[4] If the manufacturing sector uses the modern technology, the problem of the region M government, using (9) and $\gamma = 1/(\sigma(1-\alpha))-1$ is:

$$\max_{\tau_M}(1-\tau_M)w_t^M L_t^M = (1-\tau_M)D^{\frac{1}{1-\alpha}}\tau_M^{\frac{\alpha}{1-\alpha}}\left(L_t^M\right)^{\frac{1}{\sigma(1-\alpha)}}$$

The solution to this problem gives $\tau_M^* = \alpha$, so the optimal tax rate is independent of the allocation of labour. Using the government budget constraint, we have $G_t^* = \alpha X_t$.

3.2. Natural Resource Extraction in Region R

The region R government values equally rents to labour and to producers, who are also assume to be residents of region R, and sets the extraction level

4 Since the allocation of population is taken as given by regional governments, choosing policies to maximize total income or per capita income is equivalent.

to maximize the discounted flow of total regional income anticipating the impact of extraction on the natural resource stock. Using (12) and (13), period-t income is:

$$Y_t^R = w_t^R L_t^R + \Pi_t = P_t^N Z_t - \phi(S_t)Z_t + v'(L_t^R - Z_t)(L_T^R - Z_t) \tag{14}$$

Using $v'(A_t) = H/A_t = H/(L_t^R - Z_t)$ and $\dot{S}_t = -Z_t$, the Lagrangian of the government problem is:

$$\mathcal{L}(Z_t, S_t) = \int e^{-\rho t}\big[P_t^N Z_t - \phi(S_t)Z_t + H\big]dt - \int \lambda_t[Z_t + \dot{S}_t]dt \tag{15}$$

Assumption $v'(A_t) = H/A_t$ implies that income in the agricultural sector is independent of the allocation of labour. Noting that

$$\begin{aligned}\int \lambda_t[Z_t + \dot{S}_t]dt &= \int \lambda_t Z_t dt + \int \lambda_t \dot{S}_t dt \\ &= \int \lambda_t Z_t dt + \lambda_t S_t \,|_0^\infty - \int \dot{\lambda}_t S_t dt\end{aligned} \tag{16}$$

and using the transversality condition $\lambda_t S_t \,|_0^\infty = 0$, (15) can be rewritten as:

$$\mathcal{L}(Z_t, S_t) = \int e^{-\rho t}\big[P_t^N Z_t - \phi(S_t)Z_t + H\big]dt - \int \lambda_t Z_t dt + \int \dot{\lambda}_t S_t dt$$

The first-order conditions are:

$$\frac{\partial \mathcal{L}}{\partial Z_t} = e^{-\rho t}\big[P_t^N - \phi(S_t)\big] - \lambda_t = 0 \tag{17}$$

$$\frac{\partial \mathcal{L}}{\partial S_t} = -e^{-\rho t}\phi'(S_t)Z_t + \dot{\lambda}_t = 0 \tag{18}$$

From (17), the discounted increase in income resulting from higher extraction equals the shadow value of one unit of the resource. Equivalently, the marginal value of extracting one more unit of resource now is set equal to the shadow value of a unit of stock. By (18), the stock of natural resources evolves such that the discounted increase in the cost of extraction equals the change in the shadow value of a unit of stock. Let $Z^*(P_t^N, S_t)$ denote the solution to this problem. Using $Z^*(P_t^N, S_t)$, total income in region R equals:

$$Y_t^{R^*} = P_t^N Z_t^* - \phi(S_t)Z_t^* + H \equiv Y^{R^*}(P_t^N, S_t) \tag{19}$$

To provide additional interpretation for the solution of the government problem in region R, combine (17), (18) and (19) to derive a version of Hotelling's rule:[5]

$$\frac{\dot{Y}_z^{R,t}}{Y_z^{R,t}} = \rho + \frac{C_s(S_t, Z_t)}{Y_z^{R,t}} \tag{22}$$

5 Rearrange (17) and differentiate with respect to t to obtain $\rho e^{\rho t}\lambda_t + e^{\rho t}\dot{\lambda}_t = \dot{P}_t^N - \phi'(S_t)\dot{S}_t$. Substituting (17) and (18) on the left-hand side gives:

$$\rho\big[P_t^N - \phi(S_t)\big] + \phi'(S_t)Z_t = \dot{P}_t^N - \phi'(S_t)\dot{S}_t \tag{20}$$

Eq. (22) indicates that, on the chosen extraction path, the rate of change in the net benefits of extracting the resource from the perspective of region R equals the rate of time preference plus the effect of depleting the resource stock on the cost of extraction.

A proportion $0 \leq \theta \leq 1$ of the rent is assumed to be taxed by the regional government and shared equally among labour located in region R. The remaining proportion $1-\theta$ accrues to resource producers. The parameter θ is assumed to be determined exogenously. Note that our results do not require that producers capture part of the rent. They will hold in the special case where $\theta = 1$ so the entire rent is captured by labour. The per capita income of the resident workers in region R is:

$$I_t^R = w_t^R + \theta \frac{\Pi_t}{L_t^R} = (1-\theta)v'(L_t^R - Z_t) + \frac{\theta}{L_t^R} Y^R(P_t^N, S_t)$$

where the last equality follows from $\Pi_t = Y_t^R - w_t^R L_t^R$ by (13) and (19) and $w_t^R = v'(L_t^R - Z_t)$ by (12). Using $v'(A_t) = H/(L_T^R - Z_t)$, this expression for I_t^R can be written:

$$\begin{aligned} I_t^R &= (1-\theta)\frac{H}{1-L_t^M - Z(P_t^N, S_t)} + \frac{\theta}{1-L_t^M} Y^R(P_t^N, S_t) \\ &\equiv I^R(1-L_t^M, P_t^N, S_t, \theta) \end{aligned} \tag{23}$$

From (23), we can readily verify that $\partial I_t^R / \partial L_t^M > 0$ and $\partial^2 I_t^R / \partial \left(L_t^M\right)^2 > 0$. Hence, we have the following proposition.

Proposition 1 Assuming that i) $v(A) = Hln(A)$, and that ii) the government of region R chooses the extraction rate to maximize regional income taking as given the allocation of population across regions $L_t^R = 1 - L_t^M$, then I_t^R is increasing and convex in L_t^M.

Per capita income in region R, I_t^R, as a function of L_t^M is depicted in figure 1 for given values of P_t^N, S_t and θ.

3.3. Interregional Allocation of Labour

At any point in time, there will be a migration flow towards the region with the highest per capita disposable income, as specified below. The long-run equilibrium allocation of labour will be such that incomes are fully equalized

Differentiate Y_t^{R*} in (19) by Z_t to give $Y_z^{R,t}$ and then by t to get $\dot{Y}_z^{R,t} = \dot{P}_t^N - \phi'(S_t)\dot{S}_t$. Note also that

$$\frac{\partial C(S_t, Z_t)}{\partial S_t} \equiv C_s(S_t, Z_t) = \phi'(S_t) Z_t \tag{21}$$

Using these expressions, rewrite (20) as $\rho Y_z^{R,t} + C_s(S_t, Z_t) = \dot{Y}_z^{R,t}$, which gives (22).

across regions. However, for any given resource stock S_t, there can be multiple equilibria, as illustrated in figure 1. Depending on the initial distribution of labour, the economy may converge to equilibrium E_1 where the manufacturing sector in region M operates under the traditional technology, or to equilibrium E_3 in which the manufacturing sector uses the modern technology. In contrast to equilibria E_1 and E_3 which are stable, the equilibrium denoted by E_2 is unstable. Starting at E_2, a small increase in L_t^M will induce further migration towards region M and convergence to allocation E_3. Similarly, a small decrease in L_t^M from E_2 will lead to convergence to E_1.

Per capita income is higher in both regions at E_3 than at E_1. This reflects the fact that the manufacturing sector uses the modern technology at E_3. The higher productivity of labour that results from increasing returns-to-scale in manufacturing leads to higher per capita income in both regions.

Transitional dynamics will involve a migration flow towards the region where the level of utility, or disposable income, is highest. However, labour mobility is assumed to be imperfect in the sense that the migration of workers requires time so that disposable income will not be equalized instantaneously. This is captured by assuming that the flow of migration towards region M at time t is equal to the following:

$$\dot{L}_t^M = \eta\left(I_t^M - I_t^R\right) \tag{24}$$

where η reflects the speed of adjustment in the inter-regional allocation of labour. Using (11) and (23) for I_t^M and I_t^R, (24) becomes:

$$\dot{L}_t^M = \begin{cases} \eta\left[\mu - I_t^R(1-L_t^M, S_t, P_t^N, \theta)\right] \\ \quad \equiv \Omega_0(L_t^M, S_t, P_t^N, \theta) & \text{if} \quad L_t^M < \widehat{L}_M \\ \eta\left[K\left(L_t^M\right)^\gamma - I_t^R(1-L_t^M, S_t, P_t^N, \theta)\right] \\ \quad \equiv \Omega_1(\tau_M, L_t^M, S_t, P_t^N, \theta) & \phantom{\text{if}} \quad L_t^M \geq \widehat{L}_M \end{cases} \tag{25}$$

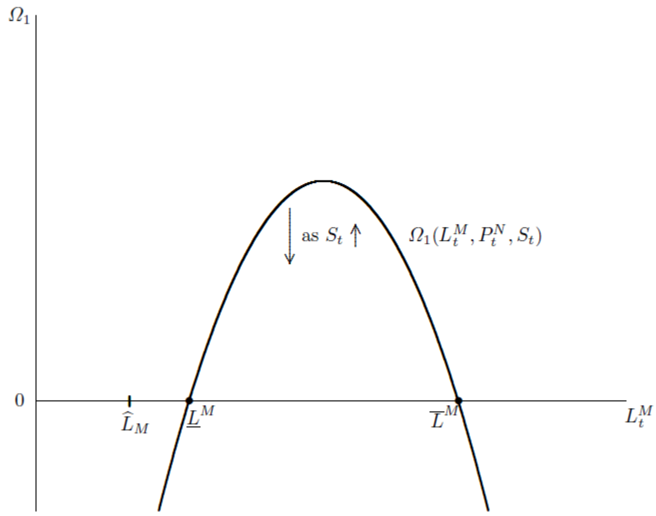
Differentiating $\Omega_1(\cdot)$ in (25) and using (11), we obtain:

$$\frac{\partial^2 \Omega_1}{\partial (L_t^M)^2} = \eta\left[\frac{\partial^2 I_t^M}{\partial (L_t^M)^2} - \frac{\partial^2 I_t^R}{\partial (L_t^M)^2}\right] < 0 \tag{26}$$

where the sign follows from noting that I_t^M is concave in L_t^M by (11) and I_t^R is convex in L_t^M by Proposition 1. Therefore, $\Omega_1(\cdot)$ is concave in L_t^M as depicted in figure 2.

For a given stock of natural resource S_t, migration towards region M will be positive ($\Omega_1 > 0$) if the population of region M is within the interval $[\underline{L}^M, \overline{L}^M]$, where we have assumed that $\underline{L}^M > \widehat{L}^M$. For values of L_t^M below $\underline{L}^M$, economies of scale in manufacturing are relatively small (or nonexistent if $L_t^M < \widehat{L}^M$) so the wage rate in region M is relatively low. For values of L_t^M above $\overline{L}^M$, L_t^R is relatively low so the wage rate is relatively

Figure 2
Migration Flow Towards Region M



high. Thus, for values of L_t^M outside of $[\underline{L}^M, \overline{L}^M]$, there is a flow of migration towards region R.

The bounds of this interval, $\underline{L}^M$ and $\overline{L}^M$, are the solutions to $\Omega_1(\tau_M, L_t^M, S_t, P_t^N, \theta) = 0$. Using the expression that characterizes Ω_1 above, we can derive:

$$\frac{\partial \underline{L}_t^M}{\partial S_t} > 0, \quad \text{and} \quad \frac{\partial \overline{L}_t^M}{\partial S_t} < 0$$

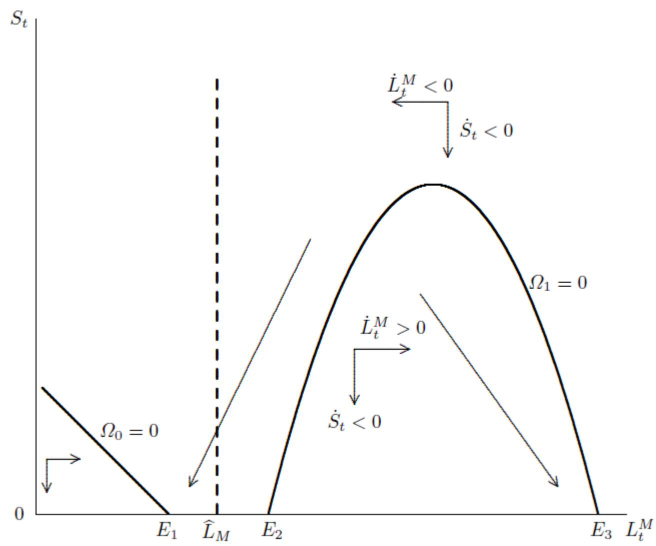
This implies that a higher level of the natural resource stock S_t reduces the size of the interval $[\underline{L}^M, \overline{L}^M]$ for which migration towards region M is positive and shifts the curve $\Omega_1(\tau_M, L_t^M, S_t, P_t^N, \theta)$ downwards in the (L_t^M, Ω_1)-space. Thus, we have the following:

Proposition 2 Assuming that $\big(1/(\sigma(1-\alpha))\big) - 1 \equiv \gamma < 1$, the following holds:

i. $\Omega_1(\theta_M, L_t^M, S_t, P_t^N, \theta)$ is concave in L_t^M; and
ii. $\partial \underline{L}_t^M / \partial S_t > 0$ and $\partial \overline{L}^M / \partial S_t < 0$, so an increase in S_t reduces the size of the interval $[\underline{L}^M, \overline{L}^M]$ for which migration towards region M is positive.

The values of $\underline{L}^M$ and $\overline{L}^M$ correspond to a specific level of the stock S_t. By varying S_t we can trace out all combinations of S_t and L_t^M for which $\Omega_1(\tau_M, L_t^M, S_t, P_t^N, \theta) = 0$. The curve of $\Omega_1(\tau_M, L_t^M, S_t, P_t^N, \theta) = 0$ is represented in the (L_t^M, S_t)-space in figure 3. The curve of $\Omega_0(L_t^M, S_t, P_t^N, \theta) = 0$ and transitional dynamics are also shown in figure 3.

Figure 3
Transitional Dynamics



For initial combinations of L_t^M and S_t located below the $\Omega_1 = 0$ curve, there will be a positive migration flow towards region M and the economy will converge in the long-run towards E_3. If the initial combination of L_t^M and S_t lies above the $\Omega_1 = 0$ curve, the economy may converge towards E_1 or E_3 depending on whether the transition path crosses the $\Omega_1 = 0$ curve or not. The economy is more likely to converge to E_1 where the manufacturing sector uses the traditional technology if the initial stock of natural resource and the initial proportion of population located in region R are relatively high.

Using (23), it is straightforward to show that $\partial\Omega_1(\cdot)/\partial P_t^N < 0$. An increase in the price of natural resources will increase the rent captured by the residents of region R, both because the rent per unit of extraction increases and because a higher extraction rate will be chosen. In turn, this will induce a larger migration flow towards region R for any given value of the stock of natural resources S_t and any initial allocation of labour (L_t^M, L_t^R). This also implies that if the price of natural resource increases, the set of initial conditions over (L_t^M, S_t) under which the economy converges to equilibrium E_3 in the long-run will become smaller. In the short-run however, that is, for a given inter-regional allocation of labour, total income in the federation will increase.

From (23), we can also readily verify that $\partial\Omega_1(\cdot)/\partial\theta < 0$. Since the government of region R is assumed to set the extraction rate to maximize total regional income, the share of the rent captured by labour θ does not affect the extraction rate. However, it does affect the incentives to migrate towards region R. Thus, an increase in θ tends to shrink the set of initial conditions over (L_t^M, S_t) for which the economy converges to E_3, and vice-versa. In fact, for sufficiently low values of θ, equilibrium E_1 may not exist at all. To see this, simply note that for a given extraction rate, per capita income in region R decreases with the value of θ. As a result, as θ decreases, the curve labeled I_t^R in figure 1 will shift downwards. For a sufficiently large shift, equilibria E_1 and E_2 may disappear, leading to a unique equilibrium.

The following proposition summarizes the main results of this section.

Proposition 3 Under decentralization, the economy exhibits the following properties:

i. There exists multiple equilibrium allocations of labour each characterized by equal per capita disposable income in both regions;
ii. In the high-income equilibrium, the manufacturing sector uses the modern technology and generates increasing returns-to-scale;
iii. An increase in the price of natural resources shrinks the set of initial conditions over (L_t^M, S_t) for which the economy converges to the high-income equilibrium E_3. Total income in the federation increases in the short run but may decrease in the long run; and
iv. If the share of resource rents captured by producers increases, the set of conditions over (L_t^M, S_t) for which the economy converges to the high-income equilibrium becomes larger.

4. The Constrained Federal Optimum

The constrained social optimum from the perspective of the federation is defined by the tax rate in region M and the path of natural resource extraction in region R that maximize the discounted flow of aggregate income in both regions. We characterize a constrained optimum in which extraction efficiency is achieved but not necessarily full efficiency in the allocation of labour, as discussed below. Total income for both regions in period t is:

$$Y_t = K\left(L_t^M\right)^{\gamma+1} + P_t^N Z_t - \phi(S_t) Z_t + H \tag{27}$$

The first term on the right side of (27) is total after-tax income in region M, the second and third terms correspond to total income in the resource sector including the share of rents captured by producers, and H is total income in the agricultural sector. Because of the specific form assumed for $v(A)$, total income in the agricultural sector is independent of L^A. We therefore fix L^A at some arbitrary level, which implies that $dL_t^M + dZ_t = 0$.

The Lagrangian for the federal problem, given the constraint $\dot{S}_t = -Z_t$, is:

$$\mathcal{L}(\tau_M, Z_t, S_t) = \int e^{-\rho t}\left[K\left(L_t^M\right)^{\gamma+1} + P_t^N Z_t - \phi(S_t)Z_t + H\right]dt - \int \lambda_t [Z_t + \dot{S}_t]dt$$

Using (16) and $\lambda_t S_t \,|_0^\infty = 0$, this can be rewritten as:

$$\mathcal{L}(\cdot) = \int e^{-\rho t}\left[K\left(L_t^M\right)^{\gamma+1} + P_t^N Z_t - \phi(S_t)Z_t + H\right]dt - \int \lambda_t Z_t dt + \int \dot{\lambda}_t S_t dt$$

Note that this problem is not subject to the migration condition. Since L^A is taken as given, $dL_t^M = -dZ_t$, so the time-path of Z_t determines the interregional labour allocation. It is straightforward to verify that the optimal tax rate is $\tau_M^* = \alpha$, the same as under decentralized policy-making. The first-order conditions on Z_t and S_t are (using $dL_t^M = -dZ_t$):

$$\frac{\partial \mathcal{L}}{\partial Z_t} = e^{-\rho t}\left[P_t^N - \phi(S_t) - K(\gamma+1)\left(L_t^M\right)^{\gamma}\right] - \lambda_t = 0 \tag{28}$$

and (18) characterizing the evolution of the natural resource stock. Eq. (28) indicates that extraction at time t is set so that the discounted increase in aggregate income in region R that results from increasing the extraction rate minus the reduction in total income in region M resulting from the associated decrease in labour (since $dL_t^M = -dZ_t$) is set equal to the shadow value of a unit of natural resources at time t.

Let $Z^{**}(P_t^N, S_t)$ denote the solution to this problem. By comparing (28) to (17), it is clear that the extraction decision of the government of region R under decentralization is distorted relative to the constrained federal optimum. Condition (28) includes an additional term, $K(\gamma+1)\left(L_t^M\right)^{\gamma} > 0$, corresponding to the impact of increasing Z_t, and the associated reallocation of labour towards region R, on the marginal product of labour in region M. The fact that the government of region R does not take this effect into account in its extraction decision implies that there is a tendency to over-extract the resource relative to the constrained federal optimum. In turn, this tendency to over-extract under decentralization makes it more likely that the economy converges to the low-income equilibrium in the long-run.

As under decentralization, we can combine (28) and (18) to derive a modified version of Hotelling's rule characterizing the constrained federal optimum:[6]

$$\frac{\dot{Y}_z^t}{Y_z^t} = \rho + \frac{C_s(S_t, Z_t)}{Y_z^t} \tag{31}$$

Eq. (31) is similar to the Hotelling's rule derived under decentralization, although here the net marginal benefit of extracting takes into account the reduction in the marginal product of labour in region M that results from shifting labour to the resource sector. It is therefore a modified version of the standard Hotelling's rule since it integrates the interregional externality present in our model.

Note that in addition to the inefficiency of extraction under decentralization, the equilibrium is also subject to migration inefficiencies. Because of agglomeration externalities in the manufacturing sector, there will be too little labour located in region M even in the high-income equilibrium (i.e. at E_3 in figure 1). That will be the case even in the absence of any resource extraction. Moreover, as long as workers capture part of the natural resource rent, there will be incentives for rent-seeking migration towards region R, which will further distort the allocation of labour. These inefficiencies are exacerbated by the extraction inefficiency. The constrained federal optimum characterized above achieves extraction efficiency, but not migration efficiency.

The main results of this section are as follows.

Proposition 4

i. The extraction decision of the government of region R is distorted relative to the constrained federal optimum, leading to a tendency to over-extract the resource under decentralization; and
ii. In the constrained federal optimum, the inter-regional allocation of labour L_t^M and the stock of natural resources S_t evolve over time according to (31), $\dot{S}_t = -Z_t$, and in the long-run $S_t \to 0$, $Z_t \to 0$ and $L_t^M \to 1 - L_t^A$.

6 The derivation is similar to that of equation (22). Differentiating (28) with respect to t, and substituting (28) and (18) in the resulting equation gives:

$$\rho\left[P_t^N - \phi(S_t) - K(\gamma+1)\left(L_t^M\right)^\gamma\right] + \phi'(S_t)Z_t = \dot{P}_t^N - \phi'(S_t)\dot{S}_t - K(\gamma+1)\gamma\left(L_t^M\right)^{\gamma-1}\dot{L}_t^M \tag{29}$$

Differentiating (27) by Z_t to give Y_z^t and then by t, we obtain:

$$\dot{Y}_z^t = \dot{P}_t^N - \phi'(S_t)\dot{S}_t - K(\gamma+1)\gamma\left(L_t^M\right)^{\gamma-1}\dot{L}_t^M \tag{30}$$

Using (21) and (30) in (29) gives (31).

5. Conclusion

This paper analyzed a resource extraction problem in a federal setting with decentralized natural resource management and inter-regional labour mobility. The analysis showed that there are multiple equilibrium allocations of labour towards which the economy may converge in the long-run. Under decentralization, the government of the resource region tends to set an inefficiently high level of extraction relative to the constrained federal optimum, which makes convergence to the low-income equilibrium more likely. In contrast, the optimal extraction path from the perspective of the whole federation takes into account the impact of resource extraction on manufacturing production.

Two extensions would be worth pursuing. First, we could examine how a central government might intervene to induce the socially optimal levels of extraction and migration. Migration is inefficient because productivity in manufacturing is increasing in the number of workers and this benefit is not internalized in wages. Rents obtained by workers in the resource-producing region exacerbates this externality. The conventional remedy for migration inefficiency is a system of equalizing interregional transfers deployed by the federal government. Federal policies to correct the inefficiency of regional resource extraction are more controversial. In principle, the federal government could impose a tax on resource extraction to internalize this externality, though this would entail federal interference with regional resource extraction policies. The federal government could also affect migration and resource extraction indirectly by spending on infrastructure in the manufacturing region which would increase labour productivity and lead to greater employment in manufacturing.

Second, we could examine the incentives that the resource region might have to use some of the resource rents to invest in infrastructure in order to develop a manufacturing sector. Doing so would contribute to diversifying the resource region's economy but would tend to dilute economies of scale in the manufacturing region with potentially adverse effects on aggregate income in the federation.

References

Albouy, D. (2012), Evaluating the Efficiency and Equity of Federal Fiscal Equalization, Journal of Public Economics 96, 824–839.

Baldwin, R., and Krugman, P. (2004), Agglomeration, Integration, and Tax Harmonization, European Economic Review 48, 1–23.

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

Bucovetsky, S. (2005), Public Input Competition, Journal of Public Economics 89, 1763–1787.

Corden, W. M., and Neary, J.P. (1982), Booming Sector and De-industrialisation in a Small Open Economy, Economic Journal 92, 825–848.

Flatters, F., Henderson, V., and Mieszkowski, P. (1974), Public Goods, Efficiency, and Regional Fiscal Equalisation, Journal of Public Economics 3, 99–112.

Gordon, R. (1983), An Optimal Taxation Approach to Fiscal Federalism, Quarterly Journal of Economics 97, 567–586.

Krugman, P. (1987), The Narrow Moving Band, the Dutch Disease, and the Competitive Consequences of Mrs. Thatcher: Notes on Trade in the Presence of Dynamic Scale Economies, Journal of Development Economics 27, 41–55.

Krugman, P. (1991), Increasing Returns and Economic Geography, Journal of Political Economy 99, 483–499.

Mitsui, K., and Sato, M. (2001), Ex Ante Free Mobility, Ex Post Immobility, and Time-Consistent Policy in a Federal System, Journal of Public Economics 82, 445–460.

Murphy, K., Shleifer, A., and Vishny, R. (1989), Industrialization and the Big Push, Journal of Political Economy 97, 1003–1026.

Raveh, O. (2013), Dutch Disease, Factor Mobility, and the Alberta Effect: The Case of Federations, Canadian Journal of Economics 46, 1317–1350.

Sachs, J., and Warner, A. (1999), The Big Push, Natural Resource Booms and Growth, Journal of Development Economics 59, 43–76.

Sachs, J., and Warner, A. (2001), The Curse of Natural Resources, European Economic Review 45, 827–838.

van der Ploeg, F. (2011), Natural Resources: Curse or Blessing?, Journal of Economic Literature 49, 366–420.

Breakthrough Renewables and the Green Paradox

Frederick van der Ploeg*

Received 22 August 2017; in revised form 27 September 2017; accepted 04 October 2017

We show how a fossil fuel monopoly responds to a carbon-free substitute becoming available at some uncertain point in the future if demand is isoelastic and variable extraction costs are zero but upfront exploration investment costs have to be made. Before the breakthrough, oil reserves are depleted too rapidly; afterwards, the oil depletion rate drops and the oil price jumps up by discrete amounts. Subsidizing green R&D to speed up the breakthrough speeds up oil extraction before the breakthrough, but more oil is left in situ as exploration investment is lower. The latter can offset the Green Paradox effect.

Keywords: regime shift, green R&D, Green Paradox

JEL classification: D 81, H 20, Q 31, Q 38

1. Introduction

The idea that well-intended climate policy may have undesirable unintended consequences has received a lot of attention during last decade due to the seminal contribution 'Public Policies against Global Warming' by Hans-Werner Sinn (Sinn, 2008a,b,c), which has found its origin in a paper co-authored with Ngo van Long almost a quarter century earlier (Long and Sinn, 1985). This contribution, like so many other of Sinn's contributions on topics as diverse as German unification, the "bazar" economy, immigration into Europe and the financial crisis in Europe, has important policy implications and has spawned a huge literature (e.g., Gerlagh, 2011; Hoel, 2010; Grafton et al., 2010, 2012; van der Ploeg and Withagen, 2012b) which has been surveyed (e.g., van der Werf and di Maria, 2012; van der Ploeg and Withagen, 2015; Pittel et al., 2014). One way of stating Sinn's pioneering insight is that by levying a steeply rising carbon tax or subsidizing the use of renewables, owners of oil wells and gas fields anticipate capital losses on their underground reserves. They

* University of Oxford, VU University Amsterdam, CESifo and CEPR, Manor Road Building, Oxford OX1 3UQ, United Kingdom (rick.vanderploeg@economics.ox.ac.uk). I gratefully acknowledge the support from the BP funded Oxford Centre for the Analysis of Resource Rich Economies. This essay has been prepared for a special issue of *FinanzArchiv* to mark Hans-Werner Sinn's 70^{th} birthday. I have benefited from the helpful comments of the editor Ronnie Schöb and an anonymous referee.

FinanzArchiv 74 (2018), 52–70
ISSN 0015-2218
doi: 10.1628/001522118X15101422148687

are therefore encouraged to extract and sell their oil and gas reserves more quickly rather than waiting when their operations will have become less profitable. This will exacerbate carbon emissions and global warming in the short run. This counterintuitive result has been coined the Green Paradox and is the intertemporal variant of the spatial notion of import leakage (e.g., van der Ploeg, 2016). However, if oil extraction becomes more costly as fewer reserves are left, the total amount of oil extracted from the earth is endogenous and not all oil reserves are necessarily fully exhausted. Over time, oil will become less attractive relative to the carbon-free backstop. Hence, a rising schedule for the carbon tax or a renewables subsidy makes it more attractive to keep more oil reserves in the crust of the earth and thus limit cumulative carbon emissions and peak global warming. This offsets and can overturn the Green Paradox, both in terms of green welfare and total welfare (van der Ploeg and Withagen, 2012a).

Our objective is to provide an alternative rationale for the Green Paradox not to hold. To make our case as stark as possible, we abstract from stock-dependent extraction costs and allow instead for initial outlays on exploration investment that determine the initial stock of oil reserves (cf., Gaudet and Laserre, 1988; Daubanes and Laserre, 2012). This also gives two margins: how quickly to extract oil and how much oil in total to extract from the earth or equivalently how much carbon emissions in total will result. We then argue that the prospect of a radical, low-cost breakthrough in the invention and bringing to the market of a carbon-free substitute (e.g., fusion energy) induces oil to be pumped up more rapidly. As a result, carbon is more quickly emitted into the atmosphere and thus global warming is exacerbated. These effects are less strong if the carbon-free backstop is a worse substitute for oil (cf., Grafton et al., 2012). At the moment the carbon-free substitute becomes available, oil use jumps down by a discrete amount and the oil price jumps up by a discrete amount unless the cost reduction of renewables and the degree of substitutability is large enough in which case the oil price jumps down. From then on, the rate of decline in the rate of oil depletion and the rate of increase in the oil price follow Hotelling paths, albeit starting from a lower level of oil reserves than if there would have been no hazard of a cheaper substitute coming to the market. This inefficiency is stronger if the risk of discovery and drop in the price of the carbon-free energy substitute are higher. Once the cheap carbon-free substitute is on the market, the rate of oil depletion follows the Hotelling rule. Uncertainty about timing of the breakthrough causes inefficiencies, not the breakthrough itself.

However, the prospect of cost-effective renewables becoming available at some random future moment implies also that exploration investment is curbed and thus that the total stock of available oil reserves diminishes. This inefficiency in exploration investment is a manifestation of the hold-up problem (e.g., Rogerson, 1992; Holmström and Roberts, 1998). It reduces the total of carbon emitted into the atmosphere and thus curbs global warming. Subsidizing green R&D to bring forward the expected time of the introduction of breakthrough renewables leads to more rapid oil extraction before the breakthrough, but more oil is left in situ as exploration investment will be lower. The latter offsets and can even reverse the Green Paradox.

To highlight the inefficiencies from the eventual arrival of breakthrough carbon-free substitutes in the most striking manner, we suppose iso-elastic fossil fuel demand and zero variable resource extraction costs. This is a useful and analytically convenient benchmark, since the monopolistic resource extraction problem is efficient under these two assumptions (cf., Stiglitz, 1976).

The idea of a discrete change in demand resulting from a breakthrough technology occurring at some unknown date in the future goes back a long time (e.g., Dasgupta and Heal, 1974; Dasgupta and Stiglitz, 1981) and has recently been used to argue that innovation unsupported by carbon pricing can lead to runaway global warming worse if the carbon cycle contains strong positive feedback effects (Winter, 2014). Our contribution is to give a tractable analysis of the effects of a breakthrough in renewables technology for the path of oil extraction and exploration investment and investigate the robustness of the Green Paradox in this context.

Our model is closely related to the ones found in the literatures on potential machine failure (Kamien and Schwartz, 1971), nationalization, expropriation and confiscation risk (Long, 1975; Bohn and Deacon, 2000; Laurent-Luchetti and Santaguni, 2012; van der Ploeg, 2017), collapses of the resource stock and changes in system dynamics (regime shifts) in pollution control (e.g., Cropper, 1976; Heal, 1984; Clarke and Reed, 1994; Tsur and Zemel, 1996; Naevdal, 2006; Polaski et al., 2011; de Zeeuw and Zemel, 2012), the effects on the speed of resource extraction of uncertainty about the time at which a resource cartel is broken up (Benchekroun et al., 2006), and the interplay between political risk and foreign investment (e.g., Cherian and Perotti, 2001).

Section 2 presents a tractable model of oil extraction and exploration investment by a monopolistic owner of oil reserves faced with the possible arrival of breakthrough renewables. It faces a constant hazard of a breakthrough at some unknown future date. Section 3 derives the optimal oil depletion and price paths before and after the breakthrough. Section 4 characterizes the solution and gives illustrative simulations. Section 5 shows that a higher chance of a renewables breakthrough depresses exploration investment. Section 6 discusses climate policy and the Green Paradox. Section 7 summarizes results and offers suggestions for further research.

2. The Model

We suppose that the economy needs two types of energy, viz. fossil fuel or oil for short, O, and renewables, R, which are imperfect substitutes in energy demand. Renewable energies such as solar and wind energy are getting cheaper all the time, but suffer from the problem of intermittence and thus rely on cost-effective storage (e.g., pumping water on top of hill or mountain or very efficient batteries). As long as the problem of storage is not solved, fossil fuel and renewable energy will be imperfect substitutes. Our assumption contrasts with the usual assumption of a perfect substitute (called a backstop source of energy), which is analytically convenient but less realistic.

Oil has zero extraction cost, but needs investment outlays I which lead to proven initial oil reserves S_0. The price of oil is endogenous and denoted by p. The breakthrough occurs at time $T > 0$ and calendar time is denoted by t. Before the breakthrough ($t < T$), renewables are infinitely elastically supplied at cost $\tilde{b}(t) = b$. After the breakthrough ($t \geq T$), they are supplied at cost $\tilde{b}(t) = b - \Delta$ where $0 < \Delta \leq b$. The monopolistic owner of the oil reserves chooses its level of exploration investment and extraction path to maximize the present value of its profits,

$$\underset{O,I}{\text{Max}}\, E\left[\int_0^\infty p(t)O(t)e^{-rt}dt\right] - qI \tag{1}$$

subject to the oil depletion equations,

$$\dot{S}(t) = -O(t),\ \forall t \geq 0, \quad S(0) = S_0 > 0, \quad \int_0^\infty O(t)dt \leq S_0, \tag{2}$$

the oil exploration investment schedule,

$$S_0 = \Theta(I), \quad \Theta' > 0, \quad \Theta'' < 0, \tag{3}$$

the oil demand schedule,

$$O(t) = \Upsilon p(t)^{-\varepsilon} b^{\sigma},\ 0 \leq t < T, \quad O(t) = \Upsilon p(t)^{-\varepsilon}(b-\Delta)^{\sigma},\ \forall t \geq T, \tag{4}$$

and the probability that the breakthrough technology occurs in the interval ending at time t,

$$\Pr(T \leq t) = 1 - \exp(-ht),\ \forall t \geq 0, \quad h \geq 0, \tag{5}$$

where S, q, I and r denote the stock of oil reserves, the price of oil exploration investment, the volume of oil exploration investment and the market interest rate, respectively. The price of oil exploration investment (q) and the market rate of interest (r) are exogenously determined on world markets and constant over time. The concavity of $\Theta(.)$ ensures decreasing returns to exploration investment. The own price elasticity of oil demand (ϵ) exceeds unity, so that marginal oil revenue is positive.[1] With the demand function (4), marginal revenue is always finite and oil reserves are fully exhausted asymptotically. Oil and renewables are supposed to be gross substitutes, so that the constant cross price elasticity of oil demand (σ) is positive. The inverse demand function for oil is given by $p = (\Upsilon \tilde{b}^{\sigma}/O)^{1/\varepsilon} \equiv p(O, \tilde{b})$.

The probability that the breakthrough technology has not taken place before time t is $\Pr(T > t) = \exp(-ht)$. The breakthrough will occur definitely as this probability tends to zero as time tends to infinity. The expected time for the breakthrough to occur is $E[T] = 1/h$. The exponential distribution has a constant hazard rate h, so that the conditional probability that the breakthrough does not take place for another t years given that the breakthrough has not already taken place in the first s years is the same as the initial probability that the breakthrough does not take place for another t years: $\Pr(T > s+t \,|\, T > s) = \Pr(T > t)$, $\forall s, t \geq 0$.

3. Optimal Oil Depletion Paths Before and After the Renewables Breakthrough

Using the principle of dynamic programming, we work backward in time and first solve the deterministic problem from unknown time of the breakthrough, T, onwards when the cheap carbon-free substitute arrives on the market, then solve the more interesting stochastic problem of oil extraction before the substitute has arrived on the market, and finally solve for the optimal level of exploration investment. We denote the problems of oil extraction *after* and *before* the breakthrough technology with superscripts A and B, respectively, and solve them for given S_0 in the rest of this section and characterize the outcomes in section 4. Section 5 then solves for the optimal level of exploration investment I and initial reserves S_0.

3.1. After the Breakthrough

Marginal oil revenue must equal the oil scarcity rent, λ, which according to the Hotelling rule must rise at a rate equal to the market interest rate, r:

$$(1 - 1/\varepsilon)\, p^A = \lambda, \quad \dot{\lambda}/\lambda = r. \tag{6}$$

It follows from (6) and the iso-elastic demand schedule (4) with zero extraction costs that the oil price and depletion paths are efficient despite the oil

1 Aggregate oil demand is relatively inelastic, but the relevant elasticity for an individual oil-producing firm is much higher as it cannot easily manipulate the price without losing market share.

owner being a monopolist:

$$\dot{p}^A/p^A = r > 0, \quad \dot{O}^A/O^A = -\varepsilon r < 0. \tag{7}$$

The intuition behind this rule is that the return on taking a marginal barrel out of the ground and investing it (i.e., the return on assets r) must equal the return on keeping this marginal barrel in the ground (i.e., the capital gains on underground reserves). Using (7) in (2), we solve for the optimal paths of oil depletion, oil reserves and (using (7)) the oil price after the breakthrough:

$$\begin{aligned} O^A(t) &= \varepsilon r S(t) = \varepsilon r e^{-\varepsilon r(t-T)} S(T), \\ 0 < S^A(t) &= e^{-\varepsilon r(t-T)} S(T) \leq S(T) < S_0, \\ p^A(t) &= e^{r(t-T)} \left[\frac{(b-\Delta)^\sigma \Upsilon}{\varepsilon r S(T)} \right]^{1/\varepsilon}, \quad \forall t > T. \end{aligned} \tag{8}$$

Equations (8) imply that fossil fuel reserves are asymptotically fully depleted. They also indicate that a lower cost of the backstop ($\Delta > 0$) pushes down the oil price, especially if the backstop is a good substitute for oil (high σ), but does not affect the path of oil depletion rates except for depressing the final stock of oil (see outcomes before the breakthrough). Substituting (8) into (1), we get the present value of profits of the oil firm after the breakthrough technology comes to market (i.e., the value function after the regime shift):

$$V^A(S(t), b-\Delta) = \left[\frac{(b-\Delta)^\sigma \Upsilon}{\varepsilon r} \right]^{1/\varepsilon} S(T)^{1-1/\varepsilon}, \quad \forall t \geq T. \tag{9}$$

A future breakthrough ($\Delta > 0$) reduces the cost of the substitute and thus curbs the future price of oil. As a result, the present value of oil profits is lower. Oil profits after the breakthrough are high if remaining reserves at the time of the breakthrough are high.

3.2. Before the Breakthrough

The Hamilton-Jacobi-Bellman equation for the optimization problem before the breakthrough is:

$$\begin{aligned} &\max_{O^B} \left[p(O^B, b) O^B - V_S^B(S, b, \Delta, h) O^B \right] \\ &\quad - h\left[V^B(S, b, \Delta, h) - V^A(S, b-\Delta) \right] = r V^B(S, b, \Delta, h), \end{aligned} \tag{10}$$

where V^B (S, b, Δ, h) denotes the value function (i.e., the present value of profits to go excluding the cost of the initial outlay on exploration investment) before the breakthrough (see appendix for a mathematical derivation). Equation (10) states that maximum oil rents *minus* the expected loss in value terms of the carbon-free substitute coming to market must equal the return from investing proceeds at the market rate of interest. The maximization of oil rents in (10) requires marginal oil revenue to be set to the marginal value of in situ oil reserves:

$$(1-1/\varepsilon)p^B(t) = V_S^B(S(t),b,\Delta,h), \quad 0 \le t < T. \tag{11}$$

Using (4) and (11), we obtain the optimal oil depletion rate before the regime shift:

$$O^B(t) = \Upsilon b^\sigma \left(\frac{V_S^B(S(t),b,\Delta,h)}{1-1/\varepsilon}\right)^{-\varepsilon}, \quad 0 \le t < T. \tag{12}$$

Upon substitution of (11) and (12) into (10), we obtain:

$$\frac{1}{\varepsilon}\left(\frac{V_S^B(S,b,\Delta,h)}{1-1/\varepsilon}\right)^{1-\varepsilon} \Upsilon b^\sigma - h\big[V^B(S,b,\Delta,h) - V^A(S,b-\Delta)\big] \\ = rV^B(S,b,\Delta,h). \tag{13}$$

To solve (13), we guess the value function $V^B(S,b,\Delta,h) = KS^{1-1/\varepsilon}$, substitute it with the post-shift value function (9) into (13), and use the method of undetermined coefficients to solve for K. It then turns out that $K = K(b-\Delta,h)$ must satisfy the nonlinear equation:

$$\frac{1}{\varepsilon}\Upsilon b^\sigma K^{1-\varepsilon} + h\left[\frac{(b-\Delta)^\sigma \Upsilon}{\varepsilon r}\right]^{1/\varepsilon} = (r+h)K. \tag{14}$$

Using the resulting value function in (11) and using the oil demand function (4), we get:

$$p^B(t) = K(b-\Delta,h)S(t)^{-1/\varepsilon}, \quad O^B(t) = L(b,\Delta,h)S(t), \quad 0 \le t < T. \tag{15}$$

where $L(b,\Delta,h) \equiv K(b-\Delta,h)^{-\varepsilon}\Upsilon b^\sigma$. Solving for the time paths from (15) and (2), we obtain:

$$p^B(t) = e^{Lt/\varepsilon} K S_0^{-1/\varepsilon}, \quad O^B(t) = Le^{-Lt}S_0, \quad S^B(t) = e^{-Lt}S_0, \\ 0 \le t < T. \tag{16}$$

The first part of (16) implies a distortion in the Hotelling rule, since the capital gains on keeping a barrel in the ground is now depressed by the risk of capital losses due to the breakthrough of the substitute so that the rule becomes $(\dot{p}^B/p^B) - (L-\varepsilon r)/\varepsilon = r$ with $L > \varepsilon r$ as we will see.

4. Characterization of the Solution: Aggressive Oil Depletion

To understand the solution more fully, we characterize the function $K = K(b-\Delta,h)$. The benchmark corresponds to a zero hazard rate. If there is no chance of a breakthrough, $h=0$, so that (14) gives $K(b-\Delta,0)=(\Upsilon b^{\sigma}/\varepsilon r)^{1/\varepsilon}$ and thus $O^A(t) = O^B(t) = \varepsilon r S(t)$, $\forall t \geq 0$. Further, $L = \varepsilon r$ and thus the oil price rises at the Hotelling rate, $\dot{p}^B/p^B = r$. Conversely, if the expected time of the breakthrough is imminently small, $h \to \infty$, (14) indicates that $K \to \left[\frac{\Upsilon(b-\Delta)^{\sigma}}{\varepsilon r}\right]^{1/\varepsilon}$ and $L \to \varepsilon r\left(\frac{b}{b-\Delta}\right)^{\sigma} > \varepsilon r$. The rate of oil depletion is thus faster than the Hotelling rate if the breakthrough is imminent. If the breakthrough is so radical that it leads to zero (marginal) cost of the carbon-free substitute ($b=\Delta$), (14) gives $K(0,h)=[\Upsilon b^{\sigma}/\varepsilon(r+h)]^{1/\varepsilon} < K(b-\Delta,0), \forall h>0$, and thus $L(b,b,h)=\epsilon(r+h)$. The possibility of oil being made completely obsolete thus depresses expected profits to go for any stock of oil reserves. Clearly, it also leads to more aggressive depletion of oil reserves than under efficient depletion as $\epsilon(r+h) > \epsilon r$.

Total differentiation of equation (14) yields:

$$dK = \frac{-\frac{K}{\varepsilon h}(L-\varepsilon r)dh + \frac{h\sigma}{\varepsilon(b-\Delta)}\left[\frac{(b-\Delta)^{\sigma}\Upsilon}{\varepsilon r}\right]^{1/\varepsilon} d(b-\Delta)}{r+h+(1-1/\varepsilon)\Upsilon b^{\sigma} K^{-\varepsilon}}. \tag{14'}$$

Since $L=\epsilon r$ if $h=0$ and $L>\epsilon r$ if $h\to\infty$, (14') implies $K_h(b-\Delta,h)<0$, $L_h(b,\Delta,h)>0$, $\forall h>0$. This reflects that a higher probability of a renewables technology breakthrough reduces the expected profit to go for the oil well owner, lifts up the path for the oil depletion rate, and depresses the oil price path before the shift. It follows that $\varepsilon r < L < \varepsilon(r+h)$ for any $0<h<\infty$. Also, if $h>0$, $K_{b-\Delta}(b-\Delta,h)>0$, $\forall \Delta \in (0,1)$ from (14'). A bigger size of the climate calamity thus curbs profits to and makes oil depletion more aggressive.

Suppose that the breakthrough occurs at date T. We know from (16) that just before we have $O^B(T-) = Le^{-LT}S_0$and $p^B(T-) = e^{LT/\varepsilon}(LS_0)^{-1/\varepsilon}$. Using $S^A(T) = S^B(T) = e^{-LT}S_0$ in (8), we get:

$$\begin{gathered} O^A(T+) = \varepsilon r e^{-LT}S_0 < O^B(T-) = Le^{-LT}S_0, \\ p^A(T+) = \left[\frac{(b-\Delta)^{\sigma}\Upsilon}{\varepsilon r S_0}\right]^{1/\varepsilon} e^{LT/\varepsilon} > p^B(T-) = \left(\frac{b^{\sigma}\Upsilon}{LS_0}\right)^{1/\varepsilon} e^{LT/\varepsilon} \\ \text{iff } \left(\frac{b}{b-\Delta}\right)^{\sigma} > \frac{L}{\varepsilon r}. \end{gathered} \tag{17}$$

We thus arrive at the following proposition.

Proposition 1 After the breakthrough the oil depletion rate and oil reserves decline at the rate ϵr and the oil price rises at the Hotelling rate r with the corresponding time paths given by (8). Before the breakthrough the oil depletion rate and oil reserves decline too rapidly at the rate $L \equiv K(b-\Delta,h)^{-\varepsilon}\Upsilon b^{\sigma} > \varepsilon r$ and the oil price rises too rapidly at the rate $L/\epsilon > r$ with the time paths given by (16) where $K = K(b-\Delta,h)$, $K_{b-\Delta} > 0$, $K_h < 0$ solves (14). At the time of the breakthrough, there is a discrete increase fall in oil extraction. If renewables enjoy a big enough cost reduction and are a good enough substitute, the oil price falls by a discrete amount.

Anticipation of a future breakthrough in renewables technology boosts the initial oil depletion rate and depresses the initial oil price, especially if the chance of a breakthrough occurring and the expected cost reduction are high. Whilst the breakthrough technology is not on the market, the oil depletion rate falls and the oil price rises too rapid and may even cross their efficient paths. Once the breakthrough technology is on the market, the oil depletion rate jumps down and the oil price jumps down by a discrete amount if the breakthrough yields a big enough cost reduction and renewables are a good enough substitute, else the oil price jumps up by a discrete amount (see (17)). From then on oil depletion and oil prices follow Hotelling paths, but starting out from an inefficiently low level of oil reserves.

Initially, the path for oil depletion exceeds the efficient path and the oil price path is below the efficient Hotelling path, since $L > \varepsilon r$. However, if the realized date of the breakthrough is distant enough, the pre-breakthrough depletion rate can fall below and the oil price path can fall above the efficient paths.

4.1. Benchmark: Certainty-equivalent Outcome

As benchmark we also calculate outcomes if the breakthrough in renewables technology is introduced *with certainty* at the expected date of the breakthrough, i.e., at time $T = 1/h$. After the breakthrough equations (8) hold. Since there cannot be a discontinuity in the price path at time $1/h$ and oil prices follow a Hotelling path, the initial oil price is $p(0) = e^{-r/h}\left[\frac{(b-\Delta)^{\sigma}\Upsilon}{\varepsilon r S(1/h)}\right]^{1/\varepsilon}$. Equation (4) gives $O(0) = \varepsilon r e^{\varepsilon r/h}\left(\frac{b}{b-\Delta}\right)^{\sigma} S(1/h)$and so $O(t) = \varepsilon r e^{-\varepsilon r(t-1/h)}\left(\frac{b}{b-\Delta}\right)^{\sigma} S(1/h), \forall t \in [0,1/h]$. Putting this into (2) and integrating, we get the stock of oil at the expected time of breakthrough: $S(1/h) = \frac{S_0}{1+(e^{\varepsilon r/h}-1)\left(b/(b-\Delta)\right)^{\sigma}} \leq e^{-\varepsilon r/h} S_0 \leq S_0$. The certainty-equivalent oil price, depletion and reserves paths before and after the breakthrough can thus be calculated:

$$S(t)=\frac{1+\left[e^{\varepsilon r(1/h-t)}-1\right]\left(\frac{b}{b-\Delta}\right)^{\sigma}}{1+(e^{\varepsilon r/h}-1)\left(\frac{b}{b-\Delta}\right)^{\sigma}}S_0,\ t\in[0,1/h],$$
$$S(t)=\frac{e^{\varepsilon r(1/h-t)}S_0}{1+(e^{\varepsilon r/h}-1)\left(\frac{b}{b-\Delta}\right)^{\sigma}},\ t\geq 1/h. \tag{18}$$

Obviously, the path of oil depletion rates is unaffected by the cost of renewables if there is no breakthrough. However, a renewables breakthrough ($\Delta > 0$) at the known time $1/h$ induces a lower stock of oil reserves at time $1/h$ and thus before the breakthrough oil depletion occurs at a faster rate than after the breakthrough, especially if renewables are a good substitute for oil (high σ). The path for oil prices satisfies the Hotelling arbitrage principle and is efficient given that the breakthrough is certain to occur at time $1/h$. The efficient paths for oil depletion rates and oil reserves corresponding to a constant cost of renewables of either b or $b-\Delta$ from time zero onwards are identical; these paths are unaffected by the breakthrough. However, the efficient oil price paths corresponding to cost b and to cost $b-\Delta$ rise at the market rate of interest r and are, respectively, above and below the certainty-equivalent path of oil prices.

4.2. Simulation of the Impact of Expected Breakthrough in Renewables Technology

To illustrate Proposition 1, figure 1 offers some illustrative simulations of our model. We set the own price elasticity of oil to $\epsilon = 2$, the cross price elasticity of oil to $\sigma = 1$ and autonomous oil demand to $\Upsilon = 1$. We set the interest rate to 4 % per annum, $r = 0.04$. The hazard rate for the breakthrough is set to $h = 0.1$, so the expected time for the breakthrough to arrive is 10 years. Hence, $0.08 = \epsilon r < L = \epsilon(r + h) < 0.28$. The cost of renewables is set to 100 before the breakthrough and to 20 after the breakthrough, so $b = 100$ and $\Delta = 80$. Finally, the initial stock of oil reserves is set to $S_0 = 1000$.

Figure 1 gives various time paths for the price oil, the oil depletion rate and oil reserves. The *certainty-equivalent* paths(solid lines) correspond to the situation where the market believes that he breakthrough occurs at the expected date of arrival, which corresponds to the inverse of the hazard rate, $1/h$, or ten years (see section 4.1). The time paths for the outcome where the market takes full account of uncertainty about the future date of the breakthrough correspond to different realizations of the date of the breakthrough, e.g., 10 and 25 year (short and long dashed lines). Figure 1 also shows two *efficient* time paths (dotted lines), which given zero extraction costs and isoelastic demand imply that oil prices must rise at a rate equal to the rate of interest. These correspond to the optimal outcome under a social planner with perfect foresight and where the breakthrough occurs either immediately or never. In the latter case, the price path is higher due to the higher cost of the substitute for oil. Note that the oil depletion rate and reserves are exactly the same for both these two efficient outcomes. Hotellling pricing also occurs in the certainty-equivalent outcome, but the price paths takes on an intermediate position.

With the parameters set to our chosen values, we find that the solution to (14) is $K = 0.802$ and thus that $L = 0.155$. The speed of oil depletion, 0.155, is thus almost twice as high as the speed after the breakthrough, $\epsilon r = 0.08$. Figure 1 shows simulations with realized times of the breakthrough technology occurring at times 10 and 25 by long dashes and short dashes, respectively. We compare these with the certainty-equivalent paths (solid) and the efficient paths (dotted) when the cost of renewable energy either immediately falls or never falls. The initial oil price if there is never a breakthrough is 0.0354 and if there is an immediate breakthrough in renewables technology it is 0.0158. From then on oil prices follow a Hotelling path in each of these two cases. As already mentioned, the paths for oil depletion rates and reserves do not depend on whether there is never or an immediate breakthrough. The certainty-equivalent path starts off with an oil price in between, 0.0283, and then also follows a Hotelling path. Oil depletion is affected by the certainty of a breakthrough at some future date: until the breakthrough reserves are depleted at a rapid rate and therefore at a lower rate after the breakthrough.

Not knowing the date of the breakthrough also speeds up the rate of oil extraction before the breakthrough compared with the certainty-equivalent (and a fortiori the efficient) path. This means that initially oil depletion is higher and oil prices lower than in the certainty-equivalent path, but after some time as a consequence of the faster rate of oil depletion oil depletion is lower and oil prices higher than in the certainty-equivalent outcome. At the moment the breakthrough comes to market, both the rate of oil depletion and oil prices jump down and thereafter continue along their Hotelling paths, albeit from an inefficient base. If the cost reduction would have not been so substantial or the renewables would not have been such a good substitute, the oil price would have jumped up by a discrete moment of the breakthrough. A sufficient condition for this not to occur is that $(b/(b-\Delta))^{\sigma} > (r+h)/h = 3.5$.

4.3. Sensitivity of Outcomes

Figure 2 plots the expected present value of oil profits at time zero, $V^B(S_0) = K(b-\Delta, h)S_0^{1-1/\varepsilon}$, and the speed at which oil is extracted, L, against the hazard rate h, both for a potential cost reduction in renewables Δ of 80 and 40. The highest feasible level of expected oil profits is 35.36, which occurs if there is no chance of a breakthrough. The expected present value of oil profits is lower for higher hazard rates and for larger potential cost reductions in renewables.

Figure 1

Impact of Threat of Breakthrough Renewables on Oil Extraction and Prices

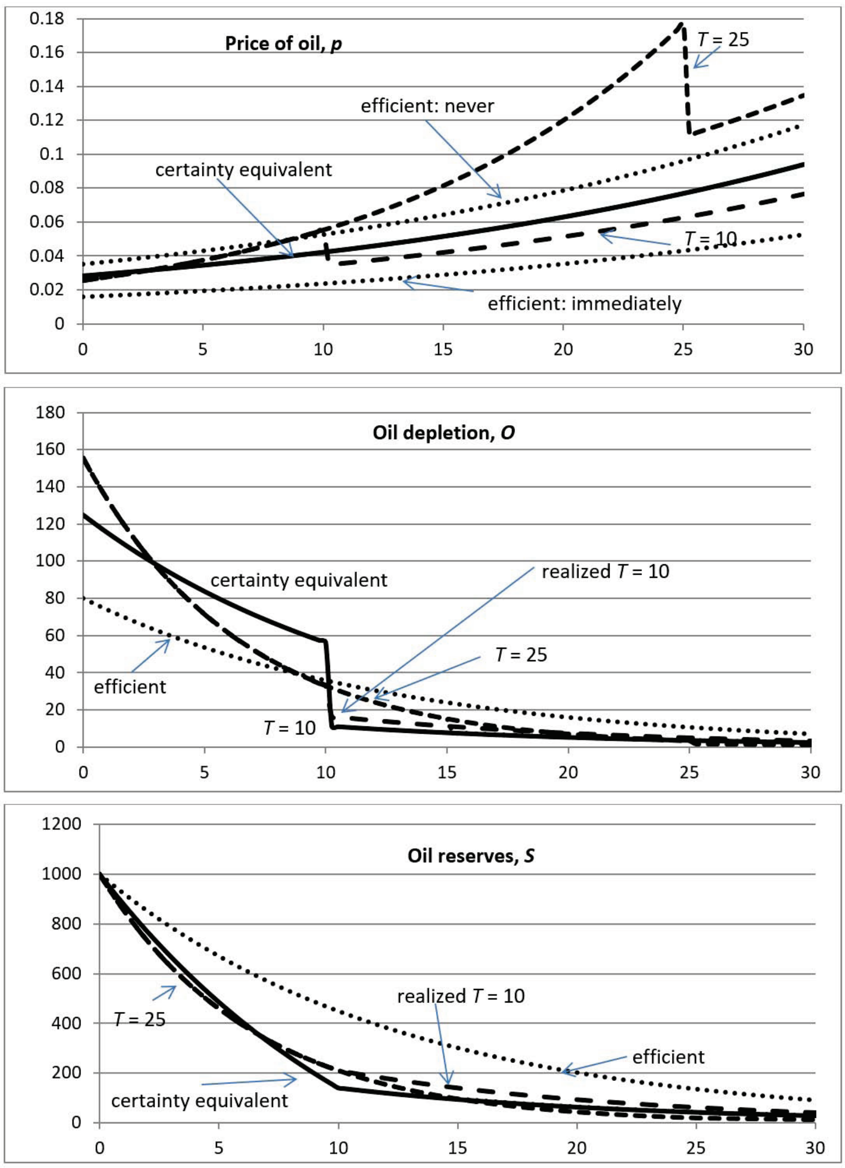
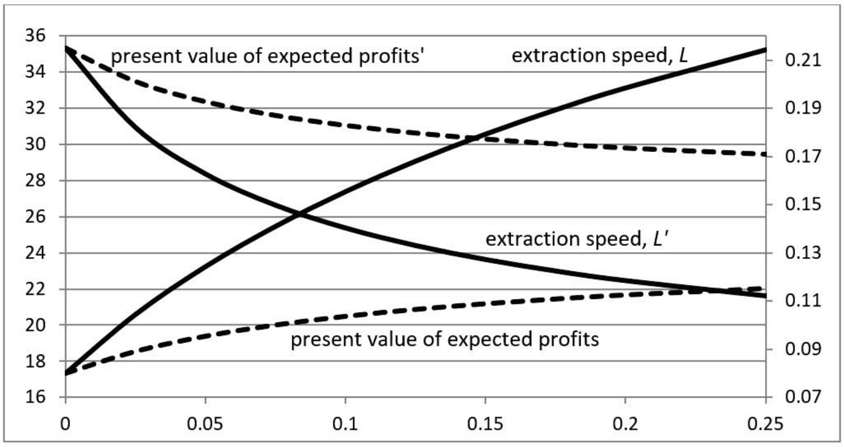


Figure 2
Bigger Risk of Bigger Breakthrough Cuts Oil Profits and Boosts Extraction Speed



Note: Hazard rate h on horizontal axis, speed of oil extraction L on right vertical axis and present value of expected profits on left vertical axis. Dashed indicates $\Delta = 40$ instead of 80.

Effectively, a more serious threat of being put out of business by a revolution in breakthrough technology damages prospects for oil producers. As a result, when the threat of a breakthrough and the size of the breakthrough become more substantial, oil producers start extracting oil and more rapid rates before renewables come to market and oil prices fall.

5. Exploration Investment and the Hold-up Problem

The final stage of solving the problem stated in section 2 is to solve for the optimal level of exploration investment I. Using the oil exploration investment schedule (3) and the expression for the value function at time zero, $V^B(\Theta(I)) = K(b-\Delta,h)\Theta(I)^{1-1/\varepsilon}$, we find that this requires setting the marginal return on marginal exploration investment to its cost:

$$(1-1/\varepsilon)K(b-\Delta,h)\Theta(I)^{-1/\varepsilon}\Theta'(I) = q. \tag{19}$$

Total differentiation of (19) gives $q[[\Theta'(I)/\varepsilon\Theta(I)] - \Theta''(I)/\Theta'(I)]dI = q(K_b d(b-\Delta) + K_h dh)/K - dq$, so that optimal exploration investment declines with its cost q, the breakthrough hazard h and the size of the reduction in the cost of renewables after the breakthrough Δ:

$$I = I(b-\Delta,h,q), \quad I_{b-\Delta} > 0, I_h, I_q < 0. \tag{20}$$

If exploration investment is subsidized at the rate θ, the optimality condition (19) becomes $(1-1/\varepsilon)K(b-\Delta,h)\Theta(I)^{-1/\varepsilon}\Theta'(I)=q-\theta$. In the efficient outcome with never a breakthrough or a breakthrough from the outset, one has $(1-1/\varepsilon)K(b-\Delta,0)\Theta(I)^{-1/\varepsilon}\Theta'(I)=q$ with $K(b-\Delta,0)=(\varepsilon r/\Upsilon b^{\sigma})^{-1/\varepsilon}$. Hence, it follows that the optimal exploration investment subsidy which corrects for the investment inefficiency increases in the breakthrough hazard and the potential cost advantage of breakthrough renewables:

$$\begin{aligned}\theta &= \left[(\varepsilon r/\Upsilon b^{\sigma})^{-1/\varepsilon}-K(b-\Delta,h)\right](1-1/\varepsilon)\Theta(I)^{-1/\varepsilon}\Theta'(I)\\ &\equiv \theta(\Delta,h)>0, \quad \theta_{\Delta}>0,\theta_h>0.\end{aligned} \tag{21}$$

Equations (20) and (21) give rise to the following proposition.

Proposition 2 The inefficiencies induced by the uncertain timing of a breakthrough in renewables are exacerbated by a drop in exploration investment, especially if the risk of a better carbon-free substitute and the potential cost reduction are higher. These inefficiencies can be eliminated by subsidizing exploration investment at the rate (21).

This proposition is an illustration of the hold-up problem (e.g., Rogerson, 1992; Holmström and Roberts, 1998). One way to overcome this is vertical integration, which may be feasible if the government can nationalize the oil firm. There may also be contractual solutions. Here an appropriate exploration investment subsidy gets rid of the inefficiency. As oil producers are typically in different jurisdictions to oil users, such a subsidy is unlikely to be implemented.

Finally, note that if the market knows with certainty that a breakthrough technology will arrive at a given known future date, the optimal exploration investment subsidy is zero as there is no point of a subsidy if the breakthrough will definitely happen at a future date and cannot be brought forward by a subsidy.

6. Climate Policy and the Green Paradox

It is easy to see that, again under the assumptions of iso-elastic demand and zero oil extraction costs, a *constant* ad valorem carbon tax which is equivalent to a fall in oil demand (lower Υ) or a *constant* subsidy to renewables use (lower b throughout) do not affect the paths of oil extraction and oil reserves. A rising path of *ad valorem* carbon taxes or a constant *specific* carbon tax will affect the rate of oil depletion, but we will abstract from these for it is difficult to muster political support for these policies. Policy makers find the carrot easier than the stick, so they focus at subsidizing green R&D instead of taxing carbon emissions. Subsidizing green R&D is meant to bring forward the introduction of carbon-free substitutes for fossil fuel, so we specify $h = H(\varpi), H' > 0, H$«0, where ϖ is the subsidy for green R&D. The subsidy thus increases the hazard rate h and cuts the expected time of the breakthrough, $1/h$. We see from (20) and Proposition 2 (abstracting from exploration investment subsidies) that subsidizing green R&D makes oil more obsolete and thus depresses exploration investment I and thus curbs the total amount of recoverable oil reserves, S_0, and the total amount of carbon that can be emitted into the atmosphere. We also know from Proposition 1 that a higher hazard rate slows down the speed of oil extraction and thus the speed at which carbon is emitted into the atmosphere (see also figure 2). Subsidizing green R&D to speed up the development of carbon-free substitutes thus leads to a Green Paradox in the short run. However, as initial exploration investment and initial fossil fuel reserves are curbed, the total amount of carbon that can be emitted into the atmosphere is curbed, the Green Paradox can in principle be reversed in the long run.

To see this a little more precisely, define the present value of global warming damages by:

$$G \equiv \int_0^\infty D(E_0 + \Theta(I(b-\Delta, H(\varpi))) - S(t))e^{-\rho t}\,dt, \qquad D' > 0, \quad D'' > 0, \tag{22}$$

where $E_0 > 0$ denotes the initial stock of carbon in the atmosphere, $\rho > 0$ the social rate of discount and $D(.)$ denotes the damages from atmospheric carbon (as a proxy for global warming). This formulation supposes, for simplicity, that all carbon that is emitted into the atmosphere stays there forever. Total carbon in the biosphere is thus $E_0 + S_0$ and $E \equiv E_0 + S_0 - S$ thereof is in the atmosphere and contributes to global warming. Global warming damages depend on total carbon emissions and are described by the convex function $D(E) = D(E_0 + \Theta(I) - S)$. As oil in measured in Giga tons of carbon, we then have:

$$\frac{\partial G}{\partial \varpi} = \underbrace{-\int_0^\infty \frac{\partial S(t)}{\partial h} H'(\varpi) D'(E(t)) e^{-\rho t}\,dt}_{+} + \underbrace{\Theta' I_h H' \tau}_{-}, \tag{23}$$

where $\tau \equiv \int_0^\infty D'(E(t))e^{-\rho t}dt > 0$ defines the social cost of carbon (the present value of marginal damages caused by emitting one extra ton of carbon today). The first term on the right-hand side of (23) indicates that a green R&D subsidy speeds up oil extraction and thus exacerbates damages, which is the usual (weak) Green Paradox, and the second term shows that the subsidy discourages exploration of oil fields and thus limits the total amount of carbon emitted into the atmosphere. If the latter effect dominates, the Green Paradox is reversed and a subsidy for investment in green breakthrough R&D exacerbates damages from global warming. More generally, this occurs if the price elasticity of supply of fossil fuel reserves is high and that of demand for fossil fuel is low (van der Ploeg, 2016). In the original analysis of the Green Paradox initial reserves were given and supply completely inelastic in which case a renewable energy subsidy also produces a Green Paradox and higher damages from global warming.

7. Conclusion

We have used a tractable model of a resource-owning monopolist with iso-elastic demand and zero variable oil extraction costs to gain more insights into the Green Paradox. The anticipation of the arrival of a carbon-free substitute at an uncertain moment of time in the future induces oil well owners to pump oil more quickly and to push down the oil price. Since this leads to more rapid emissions of the *given* amount carbon in the crust of the earth, global warming is exacerbated which is the (weak) Green Paradox. As soon as the carbon-free substitute has arrived, the oil depletion rate jumps down. If the new carbon-free fuel is not a very good substitute and the cost reduction is not too substantial, the oil price jumps up by a discrete amount at that moment. If the breakthrough is substantial enough and a good enough substitute for oil, the oil price jumps down. From then on the oil depletion rate declines at the Hotelling rate, albeit starting out from a lower level of reserves than would be the case if there was no anticipation of renewables being introduced, and the oil price rises at the market rate of interest. Interestingly, if the carbon-free substitute was introduced from the outset with certainty, oil extraction would be unaffected.

An uncertain introduction date for the carbon-free substitute depresses oil exploration investment and thus more oil is left in the crust of the earth which is due to hold-up problem. Hence, the total amount of carbon emitted into the atmosphere is reduced, albeit that what is emitted is emitted more rapidly. This can easily overturn the Green Paradox in the sense that the present discounted value of global warming damages can increase. The exploration investment inefficiency can be corrected for with an appropriate subsidy, which increases in the chance of the breakthrough occurring and the cost reduction arising from the breakthrough technology.

The consequences of cheaper renewables are thus not necessarily as bleak as suggested by proponents of the Green Paradox. Future work will benefit from a better grasp of regime shifts, whether they relate to arrival of carbon-free substitutes or potential tipping points and climate disasters.[2] It is also of interest to put the breakthrough approach in a strategic setting.[3]

8. Appendix: Derivation of the HJB Equation (10)

Since the probability of a regime shift in an infinitesimally small time period Δt is $h\Delta t$, the Principle of Optimality from the perspective of time zero can be written as follows:

$$e^{-rt}V^B(S(t)) = \max_{O^B}\Big[\int_t^{t+\Delta t} e^{-rs}p\big(O^B(s)\big)O^B(s)ds$$
$$+(1-h\Delta t)e^{-r(t+\Delta t)}V^B(S(t+\Delta t))+h\Delta t e^{-r(t+\Delta t)}V^A(S(t+\Delta t))\Big] \quad (24)$$

(suppressing the arguments b, Δ and h in the value function $V^B(.)$). Multiplying both sides by e^{rt}, rearranging and dividing by Δt, we rewrite (24) as:

$$\max_{O^B}\Bigg[\frac{\int_t^{t+\Delta t}e^{-r(s-t)}p(O^B(s))O^B(s)ds}{\Delta t} - he^{-r\Delta t}V^B(S(t+\Delta t))$$
$$+he^{-r\Delta t}V^A(S(t+\Delta t))\frac{\big(e^{-r\Delta t}-1\big)V^B(S(t+\Delta t))}{\Delta t}$$
$$+\frac{V^B(S(t+\Delta t))-V^B(S(t))}{\Delta t}\Bigg]=0. \quad (25)$$

Evaluating the integral in (25) for infinitesimally small Δt and taking the limit as $\Delta t \to 0$ whilst using l'Hôpital's Rule for $\lim_{\Delta t\to 0}\frac{\exp(-r\Delta t)-1}{\Delta t} = -r$, we get:

$$\max_{O^B}\big[p\big(O^B(t)\big)O^B(t) - \dot{V}^B(S(t)) - hV^B(S(t))$$
$$+hV^A(S(t)) - rV^B(S(t))\big] = 0. \quad (26)$$

Substituting $\dot{V}^B = V_S^B\dot{S}$and using (2), rearranging and dropping the time index, we get (10).

2 If positive feedback effects in the carbon cycle are strong enough, runaway global warming may result (Winter, 2014). However, this will lead to an irreversible doomsday scenario.

3 Hoel (1978) and Gerlagh and Liski (2011) analyze the strategic pricing policies of an oil producer faced with a substitute coming to market. Jaakkola (2012) analyzes in more detail the strategic dynamic interactions with oil importers developing substitutes. Jaakkola and van der Ploeg (2017) perform primarily a numerical analysis of an international dynamic game in carbon pricing, on the one hand, and investments in breakthrough technology, on the other hand.

References

Benchekroun, H., Gaudet, G., and van Long, N. (2006), Temporary Natural Resource Cartels, Journal of Environmental Economics and Management 52, 663–674.

Bohn, H., and Deacon, R. (2000), Ownership Risk, Investment, and the Use of Natural Resources, American Economic Review 90, 526–549.

Cherian, J.A., and Perotti, E. (2001), Option Pricing and Foreign Investment under Political Risk, Journal of International Economics 55, 359–377.

Clarke, H.R., and Reed, W.J. (1994), Consumption/Pollution Tradeoffs in an Environment Vulnerable to Pollution-Related Catastrophic Collapse, Journal of Economic Dynamics and Control 18, 991–1010.

Cropper, M.L. (1976), Regulating Activities with Catastrophic Environmental Effects, Journal of Environmental Economics and Management 3, 1–15.

Dasgupta, P., and Heal, G.M. (1974), The Optimal Depletion of Exhaustible Resources, Review of Economic Studies 41, 3–28.

Dasgupta, P., and Stiglitz, J. (1981), Resource Depletion under Technological Uncertainty, Econometrica 49, 85–104.

Daubanes, J., and Laserre, P. (2012), Non-Renewable Resource Supply: Substitution Effect, Compensation Effect and All That, mimeo., ETH, Zurich.

Gaudet, G., and Laserre, P. (1988), On Comparing Monopoly and Competition in Exhaustible Resource Exploitation, Journal of Environmental Economics and Management 15, 412–418.

Gerlagh, R. (2011), Too Much Oil, CESifo Economic Studies 57, 79–102.

Gerlagh, R., and Liski, M. (2011), Strategic Resource Dependence, Journal of Economic Theory 146, 699–727.

Grafton, R.Q., Kompas, T., and van Long, N. (2010), Biofuels and the Green Paradox, Working Paper 2960, CESifo, Munich.

Grafton, R.Q., Kompas, T., and van Long, N. (2012), Substitution between Bio-Fuels and Fossil Fuels: Is there a Green Paradox?, Journal of Environmental Economics and Management 64, 328–341.

Heal, G.M. (1984), Interactions between Economy and Climate: A Framework for Policy Design Uncertainty, in: Kerry Smith, V., and Witte, A. (eds.), Advances in Applied Micro-Economics, JAI Press, Greenwich, Connecticut, 151–168.

Hoel, M. (1978), Resource Extraction, Substitute Production, and Monopoly, Journal of Economic Theory 19, 28–37.

Hoel, M. (2010), Is there a Green Paradox?, Working Paper 3168, CESifo, Munich.

Holmström. B., and Roberts, J. (1998), The Boundaries of the Firm Revisited, Journal of Economic Perspectives, 12, 73–94.

Hotelling, H. (1931), The Economics of Exhaustible Resources, Journal of Political Economy 39, 137–175.

Jaakkola, N. (2012), Strategic Oil Supply and Development of Substitutes, mimeo., University of Oxford.

Jaakkola, N., and van der Ploeg, F. (2017), Non-Cooperative and Cooperative Climate Policies with Anticipated Breakthrough Technology, OxCarre Research Paper 190, University of Oxford.

Kamien, M.I., and Schwartz, N.L. (1971), Optimal Maintenance and Sale Age for a Machine Subject to Failure, Management Science 17, 495–504.

Laurent-Luchetti, J., and Santuguni, M. (2012), Ownership Risk and the Use of Common-Pool Natural Resources, Journal of Environmental Economics and Management 63, 242–259.

van Long, N. (1975), Resource Extraction under the Uncertainty about possible Nationalization, Journal of Economic Theory 10, 42–53.

van Long, N., and Sinn, H.W. (1985), Surprise Price Shifts, Tax Changes and the Supply Behavior of Resource Extracting Firms, Australian Economic Papers 24, 278–289.

Naevdal, E. (2006), Dynamic Optimization in the Presence of Threshold Effects when the Location of the Threshold is Uncertain – With an Application to a possible Disintegration of the Western Antarctic Ice Sheet, Journal of Economic Dynamics and Control 30, 1131–1158.

Pittel, K., van der Ploeg, F., and Withagen, C. (2014), Climate Policy and Non-Renewable Resources – The Green Paradox and Beyond, MIT Press, Cambridge, Mass.

van der Ploeg, F. (2016), Second-Best Carbon Taxation in the Global Economy: The Green Paradox and Carbon Leakage Revisited, Journal of Environmental Economics and Management 78, 85–105.

van der Ploeg, F. (2017), Political Economy of Dynamic Resource Wars, Journal of Environmental Economics and Management, forthcoming.

van der Ploeg, F., and Withagen, C. (2012a), Is there Really a Green Paradox?, Journal of Environmental Economics and Management 64, 342–363.

van der Ploeg, F., and Withagen, C. (2012b), Too Much Coal, Too Little Oil, Journal of Public Economics 96, 62–77.

van der Ploeg, F., and Withagen,C. (2015), Global Warming and the Green Paradox: A Review of Adverse Effects of Climate Policy, Review of Environmental Economics and Policy 9, 285–303.

Polaski, S., de Zeeuw, A.J., and Wagener, F. (2011), Optimal Management with Potential Regime Shifts, Journal of Environmental Economics and Management 62, 229–240.

Rogerson, W. (1992), Contractual Solutions to the Hold-Up Problem, Review of Economic Studies 59, 774–794.

Sinn, H.W. (2008a), Public Policies Against Global Warming, International Tax and Public Finance 15, 360–394.

Sinn, H.W. (2008b), Das Grüne Paradoxon: Warum mann das Angebot bei der Klimapolitik nicht vergessen darf, Perspektiven der Wirtschaftspolitik 9, 109–142.

Sinn, H.W. (2008c), Das Grüne Paradoxon – Plädoyer für eine illusionsfrei Klimapolitik, Econ, Berlin.

Stiglitz, J.E. (1976), Monopoly and the Rate of Extraction of Exhaustible Resources, American Economic Review 66, 655–661.

Tsur, Y., and Zemel, A. (1996), Accounting for Global Warming Risks: Resource Management under Event Uncertainty, Journal of Economic Dynamics and Control 20, 1289–1305.

van der Werf, E., and Maria, C. Di (2012), Imperfect Environmental Policy and Polluting Emissions: The Green Paradox and Beyond, International Review of Environmental and Resource Economics 6, 153–194.

Winter, R. A. (2014), Innovation and the Dynamics of Global Warming, Journal of Environmental Economics and Management 68, 124–140.

de Zeeuw, A.J., and Zemel, A. (2012), Regime Shifts and Uncertainty in Pollution Control, Journal of Economic Dynamics and Control 36, 939–950.

From the Linear Economy to the Circular Economy: A Basic Model

Peter Birch Sørensen*

Received 11 August 2017; in revised form 01 October 2017; accepted 04 October 2017

This paper sets up a Ramsey model with natural resources to study the optimal recycling of polluting raw materials. Under plausible conditions it is optimal for the economy to go through an initial *linear* phase with no recycling followed by a *circular* phase where a fraction of materials is recycled to alleviate growing natural resource scarcity and environmental degradation. In the presence of a Pigouvian tax on nonrecycled materials a competitive market economy will ensure the optimal degree of recycling.

Keywords: circular economy, linear economy, optimal recycling, Hotelling rule, Pigouvian taxation

JEL classification: Q 53, Q 58, H 21

1. Introduction: The Concept of a Circular Economy

"Reuse, recycle, reduce, rethink!" With this slogan an advisory group of business leaders recently urged the Danish government to move from the current "linear economy" to a "circular economy" (Advisory Board for Circular Economy, 2017). According to this vision the present linear economy is characterized by a "buy-and-throw-away" mentality involving excessive exploitation of natural resources and accumulation of polluting waste products: increasingly scarce raw materials are being extracted from the environment and returned to it as harmful waste as they are put through the "linear" process of production and consumption. By contrast, a circular economy seeks to minimize the use of raw materials per unit of output and to recycle waste products as much as possible in order to reuse them as inputs in production.

The concept of the circular economy is becoming increasingly popular among environmentalists and policymakers and in parts of the business community. The idea has been pushed for some time by think tanks such as the Ellen Macarthur Foundation (2012), and it has featured in the last two Five

* University of Copenhagen, Øster Farimagsgade 5, Building 26, 1353 Copenhagen K, Denmark (pbs@econ.ku.dk). I wish to thank Ronnie Schöb for valuable comments on an earlier version of this paper. All remaining shortcomings are my own responsibility.

FinanzArchiv 74 (2017), 71–87
ISSN 0015-2218
doi: 10.1628/001522118X15097191506475

Year Plans of the Chinese government (Zhijun and Nailing, 2007). The European Commission (2015) has recently proposed an EU action plan for the circular economy, and many governments around the world are currently considering policies to promote recycling and more efficient waste treatment.

To economists trained in public economics or environmental economics this hype about the circular economy may seem somewhat exaggerated. For one thing, the idea of promoting recycling is hardly new. For example, in his famous paper on the economics of the coming "spaceship earth," Kenneth Boulding wrote: "Man must find his place in a cyclical ecological system which is capable of continuous reproduction of material form even though it cannot escape having inputs of energy" (Boulding, 1966, pp. 7–8). Boulding also anticipated the concept of the linear economy with his image of the "cowboy economy" where "… the success of the economy is measured by the amount of the throughput from the 'factors of production,' a part of which … is extracted from the reservoirs of raw materials and another part of which is output into the reservoirs of pollution." (Boulding, 1966, p. 8).

For another thing, economists will be skeptical of the idea that the economic system should maximize the amount of output per unit of natural resource input and the degree of recycling. In standard welfare economics and environmental economics the goal is to ensure an efficient use of *all* economic resources, including man-made goods as well as those given to us by nature. After all, a basic tenet of environmental economics is that the optimal level of pollution is generally larger than zero. As William Baumol (1977) pointed out, recycling of the byproducts of production and consumption requires the use of resources that at some point may generate more harm to the environment than the damage prevented through recycling.

Yet the present paper will show that there is a rational core to the proposition that the government should promote the transition from a *linear* economy with little or no recycling to a *circular* economy where a part of the materials used in production is recouped and recycled as inputs. To illustrate this, I will set up a simple model of an economy where production of final goods uses an exhaustible natural resource and (human and physical) capital as inputs and where the use of raw materials generates pollution, which can be mitigated by investing part of the capital stock in a recycling process. If the economy starts out with a good quality of the environment and a sufficiently large reserve stock of the natural resource, it will be optimal for it to go through an initial linear phase with no recycling of materials, but at some point a growing scarcity of natural resources relative to man-made capital and a deteriorating quality of the environment makes it optimal to enter a circular phase with positive recycling. However, in a laissez-faire economy the initial linear phase will involve excessive use of raw materials and the transition to the circular phase will not take place at the appropriate time. Hence government intervention in the form of a Pigou tax on nonrecycled materials is needed to steer the economy to the first-best transition path with the optimal level and timing of recycling.

Earlier writings on recycling such as Smith (1972), Schultze (1974), Lusky (1975, 1976), Hoel (1978), Di Vita (2001, 2007), and Pittel et al. (2010) have had little focus on explaining the transition from the linear to the circular economy and the design of public policy to ensure the optimal timing of this transition. The present paper seeks to fill this gap.

The paper adds to a relatively small environmental economics literature on recycling. An early contribution was made by Smith (1972), who focused on the reuse of household waste. Schultze (1974) illustrated how the recycling of raw materials could ameliorate the exhaustion of nonrenewable resources, and Lusky (1975, 1976) studied the allocation of household time between work in the labor market and recycling activity, showing how the optimal amount of recycling might be secured through a tax on consumption. The more recent papers by Di Vita (2001, 2007) investigate how endogenous technical change driven by R&D may affect the recycling of waste and thereby consumer welfare, and Pittel et al. (2010) set up a Ramsey-type model of exogenous growth with recycling of waste to study how the optimal level of recycling may be implemented through government subsidies. Like the present paper, the article by Andersen (2007) makes the point that the policy problems discussed within the circular economy paradigm can be tackled via the classical Pigouvian policy instruments emphasized in conventional environmental economics.

In contrast to the present paper, the contributions mentioned above did not focus on explaining the transition from a linear to a circular economy. The closest predecessor to the present study is the paper by Hoel (1978), who analyzed the optimal path of economic development and the role of recycling when natural-resource extraction harms the environment. However, Hoel's study was a microeconomic partial-equilibrium analysis, and in his simple model resource extraction and recycling will never take place simultaneously, whereas the present macroeconomic general-equilibrium analysis finds that the two activities can go on at the same time.

Section 2 sets up the model, which is used in section 3 to derive the first-best allocation of resources. Section 4 describes the first-best transition from a linear to a circular economy, and section 5 analyzes the resource allocation and recycling activity generated by a competitive market economy. Section 6 explains how a laissez-faire market economy will fail to attain the optimal volume and timing of recycling and how this failure can be corrected through Pigouvian taxation. The main conclusions are summarized in section 7.

2. The Model

We consider an economy inhabited by a representative family dynasty with an infinite horizon. In each period the family derives utility $u(C)$ from consumption of final goods (C) and utility $v(E)$ from the quality of the environment (E). At time zero the present value of the family's lifetime utility U is

$$U = \int_0^{\infty} [u(C) + v(E)]e^{-\rho t} dt, \quad u' > 0, \quad u'' < 0, \quad v' > 0, \quad v'' < 0, \tag{1}$$

where $\rho > 0$ is the constant rate of time preference, and the variables C and E are understood to be functions of time t. The total output of final goods (Y) may be used for consumption or for investment (I):

$$Y = C + I. \tag{2}$$

The output of final goods is given by the linearly homogeneous production function

$$Y = F\left(K^Y, M\right), \quad F_K > 0, \quad F_{KK} < 0, \quad F_M > 0, \quad F_{MM} < 0, \tag{3}$$

where the subscripts indicate first and second partial derivatives. The variable K^Y is the stock of capital used in final-goods production, and M is the input of a flow of raw materials. A part of these materials may be recycled by investing a capital stock K^R in the recycling process. The flow of recycled materials is given by the following recycling technology:

$$\begin{gathered} R = g(K^R/M)M, \quad g(0) = 0, \quad g' > 0, \quad g'' < 0, \\ \lim_{K^R/M \to \infty} g(K^R/M) = 1. \end{gathered} \tag{4}$$

According to the last assumption in (4) a complete recycling of all materials ($g = 1$) would require an infinitely high capital intensity of the recycling process and is therefore infeasible due to the Second Law of Thermodynamics discussed by Georgescu-Roegen (1971). The assumption $g(0) = 0$ reflects that no recycling is possible if no capital is invested in recycling equipment.

Raw materials may be extracted at zero cost from a stock of an exhaustible natural resource. When there is recycling, the flow of new materials extracted from the ground each period is $M - R > 0$. Abstracting from new discoveries, the reserve stock of the natural resource (S) therefore evolves as

$$\dot{S} = -(M - R), \tag{5}$$

where a dot above a variable indicates its derivative with respect to time. The total stock of man-made capital (K) is

$$K = K^Y + K^R. \tag{6}$$

We may think of K as a composite of physical and human capital where optimizing behavior ensures that investment in the two forms of capital yields the same marginal return. Ignoring depreciation, the change in the capital stock over time is

$$\dot{K} = I. \tag{7}$$

The throughput of raw materials in the production process generates polluting waste products, so the quality of the environment deteriorates by an amount γ for each unit of raw material that is not recycled. The ability of the environment to assimilate waste and regenerate itself is proportional to the existing stock of environmental goods (proxied by E), with a proportionality factor δ. Hence the change in environmental quality over time is

$$\dot{E} = \delta E - \gamma(M - R), \quad \rho > \delta > 0, \quad \gamma > 0. \tag{8}$$

The assumption $\rho > \delta$ ensures that the shadow value of environmental quality is finite (cf. equation (20) below).

3. The First-Best Allocation

A utilitarian social planner will maximize the lifetime utility function (1) subject to the constraints implied by (2) through (8), given the predetermined initial values of K, S, and E. The current-value Hamiltonian for this optimal control problem can be written as

$$\begin{aligned} H = {} & u(C) + v(E) + \mu \overbrace{[F(K - K^R, M) - C]}^{\dot{K}} \\ & + \lambda \overbrace{[g(K^R/M) - 1]M}^{\dot{S}} + \eta \overbrace{\{\delta E - \gamma[1 - g(K^R/M)]M\}}^{\dot{E}} \end{aligned} \tag{9}$$

where μ, λ, and η are the current shadow values of the state variables K, S, and E, respectively, and the control variables are C, K^R, and M. The first-order conditions for the solution to the social planning problem are found to be

$$u'(C) = \mu, \tag{10}$$

$$mF_M = \frac{\lambda + \gamma\eta}{\mu}, \quad m \equiv \frac{1}{1 - (1 - \varepsilon)g}, \quad \varepsilon \equiv \frac{dg/g}{d(K^R/M)/(K^R/M)}, \tag{11}$$

$$K^R = 0 \text{ if } g'(0)\left(\frac{\lambda + \gamma\eta}{\mu}\right) \leq F_K, \tag{12a}$$

$$K^R > 0 \text{ and } g'(K^R/M)\left(\frac{\lambda + \gamma\eta}{\mu}\right) = F_K \text{ if } g'(0)\left(\frac{\lambda + \gamma\eta}{\mu}\right) > F_K, \tag{12b}$$

$$\dot{\mu} = (\rho - F_K)\mu, \tag{13}$$

$$\dot{\lambda} = \rho\lambda, \tag{14}$$

$$\dot{\eta} = (\rho - \delta)\eta - v'(E). \tag{15}$$

Equation (10) states that the marginal utility of consumption must equal the marginal welfare gain from investment. The fraction $(\lambda + \gamma\eta)/\mu$ appearing in (11) and (12) is the marginal social cost of using an additional unit of nonrecycled raw material in production. It is measured in units of the final good (since we are dividing by the marginal utility of consumption, μ) and consists of the marginal cost of depleting the natural-resource stock, captured by the shadow price λ/μ, plus the marginal welfare cost $\gamma\eta/\mu$ of the damage to the environment when an extra unit of nonrecycled materials is put through the production process. The variable m in (11) is a *recycling multiplier* reflecting that a unit of materials can be used more than once when there is recycling. Each time an extra unit of materials enters the production process, a fraction $(1-\varepsilon)g$ of it can be used again, so an initial unit increase of materials input results in a total increase of $m \equiv 1/[1-(1-\varepsilon)g]$ units.[1] The presence of the dampening elasticity ε in the expression for m reflects that adding an extra unit of materials to the recycling process while keeping the recycling equipment K^R constant reduces the effectiveness of the process, thereby reducing the fraction of materials that can be recycled. Note that diminishing returns in the recycling process imply that the elasticity ε defined in (11) is smaller than 1.[2]

With these observations in mind, we see that (11) is a condition for optimal use of materials, stating that the marginal productivity of materials should equal the marginal social cost of their use, taking accout of the degree of recycling. The optimal degree of recycling is determined by (12a) and (12b), where the term $g'(0)(\lambda + \gamma\eta)/\mu$ is the marginal social gain from investing a unit of capital in recycling, starting from a level of zero investment. This gain reflects the alleviation of natural-resource scarcity and the improvement of environmental quality resulting from initiating recycling. The right-hand side of (12a) and (12b) is the marginal social opportunity cost of reallocating capital from final-goods production to recycling, given by the marginal productivity of capital in final-goods production. Thus (12a) says that if the marginal social gain from recycling is smaller than its marginal opportunity cost, society should not invest in recycling. But if $g'(0)(\lambda + \gamma\eta)/\mu > F_K$, so

1 To verify this, note that $m = 1 + (1-\varepsilon)g + [(1-\varepsilon)g]^2 + [(1-\varepsilon)g]^3 + \cdots = 1/[1-(1-\varepsilon)g]$.

2 The recycling process specified in (4) can be thought of as resulting from a linearly homogeneous *recycling function* $R = R(K^R, M) = g(K^R/M)M$ where $g(K^R/M) \equiv R(K^R/M, 1)$. With diminishing returns to each of the inputs in the recycling function $R(K^R, M)$, the function $g(K^R/M)$ will also display diminishing returns to the capital intensity K^R/M.

that some amount of recycling is worthwhile, (12b) says that investment in recycling should be carried to the point where its marginal social benefit equals its marginal social opportunity cost.

We can boil down the conditions for a first-best allocation into a *wealth accumulation rule* determining how much wealth society should transfer from the present to the future and a *portfolio composition rule* indicating how society should allocate its wealth between man-made capital and natural capital. The wealth accumulation rule in the present model is the familiar Keynes–Ramsey rule for an optimal intertemporal allocation of consumption that is implied by (10) and (13):

$$\frac{\dot{C}}{C} = \frac{1}{\sigma}(F_K - \rho), \quad \sigma \equiv -\frac{u''C}{u'} > 0. \tag{16}$$

The portfolio composition rule can be found by differentiating (11) with respect to time and inserting (11) plus (13) through (15) into the resulting expression to obtain

$$F_K = \frac{\dot{F}_M}{F_M} + \left(\frac{\gamma\eta}{\lambda+\gamma\eta}\right)\left(\delta + \frac{v'(E)}{\eta}\right) + \frac{\dot{m}}{m}. \tag{17}$$

The left-hand side of (17) is the marginal social rate of return on investment in man-made capital, given by its marginal productivity. In optimum this must equal the marginal social rate of return on investment in natural capital appearing on the right-hand side of (17). The investment in natural capital takes the form of postponing the extraction of an extra unit of materials from "today" until "tomorrow." A part of the gain from doing so consists in the rise of the marginal productivity of materials as they become scarcer over time. This is captured by the first term on the right-hand side of (17). The second term reflects that postponing extraction implies a lower current use of materials, which generates an environmental gain, partly because the lower current emission of waste products increases the future assimilative capacity of the environment (captured by the parameter δ), and partly because the postponement of emissions directly benefits consumers by delaying the damage to the environment (reflected in the term v'/η). We see that the environmental gain carries a heavier weight the greater the importance of improving environmental quality relative to the importance of alleviating natural-resource scarcity, i.e., the larger the fraction $\gamma\eta/(\lambda+\gamma\eta)$. Finally, there is a gain from postponement of extraction to the extent that the *materials multiplier* m increases over time so that materials can be used more effectively in the future. This is captured by the third term on the right-hand side of (17). From the definition of m stated in (11) it follows that if the elasticity ε is roughly constant, we have $\frac{\dot{m}}{m} \approx \frac{(1-\varepsilon)\dot{g}}{1-(1-\varepsilon)g}$, so that (17) may be written as

$$F_K = \frac{\dot{F_M}}{F_M} + \left(\frac{\gamma\eta}{\lambda+\gamma\eta}\right)\left(\delta + \frac{v'(E)}{\eta}\right) + \frac{(1-\varepsilon)\dot{g}}{1-(1-\varepsilon)g}. \tag{18}$$

Recalling that $g < 1$ and $\varepsilon < 1$ because of diminishing returns to recycling, we see from (18) that an increase over time in the recycling rate g increases the marginal gain from postponing the extraction and use of materials, which is intuitive.

4. From the Linear to the Circular Economy

If the economy starts out at an early stage of economic development, it is likely that an optimal development path will involve an initial *linear* stage with no recycling and a deteriorating environment followed by a *circular* stage with positive recycling that reduces the pressure on the environment and slows down the depletion of the natural-resource stock.

To see this, note that (14) and (15) imply

$$\lambda(t) = \lambda(0)e^{\rho t}, \tag{19}$$

$$\eta(t) = \int_t^\infty v'(E(z))e^{-(\rho-\delta)(z-t)}dz. \tag{20}$$

According to (19) the shadow value of an extra unit of the natural resource rises steadily over time at the rate ρ as the resource gets scarcer. Equation (20) states that the shadow value of a unit improvement in environmental quality equals the present value of the future marginal utilities of environmental quality.[3]

Now suppose the economy starts out at an early stage of economic development where the reserve stock of the natural resource is large, the quality of the environment is good, and the stock of man-made capital is relatively low. With abundant natural resources, a well-preserved environment, and a relatively low level of material consumption due to a low capital stock, the marginal social cost $(\lambda+\gamma\eta)/\mu$ of using a unit of nonrecycled raw material will be low, since λ and η will be small whereas μ (the marginal utility of consumption) will be large. At the same time the marginal productivity of capital in final-goods production will be high due to its scarcity. In these circumstances the marginal social gain $g'(0)(\lambda+\gamma\eta)/\mu$ from investing in recycling will most likely be lower than the marginal opportunity cost F_K of doing so. According to (12a)

3 Note that since $\rho > \delta$ by assumption, the integral in (20) is finite. The presence of the parameter δ in the effective discount rate $\rho-\delta$ reflects that an improvement in current environmental quality increases the future ability of the environment to absorb waste, thereby increasing the future quality of the environment.

the economy should therefore start out in a linear phase with no recycling. During this phase, where $g = 0$, the optimality conditions (11) and (18) simplify to

$$F_M = \frac{\lambda + \gamma\eta}{\mu}, \tag{21}$$

$$F_K = \frac{\dot{F_M}}{F_M} + \left(\frac{\gamma\eta}{\lambda + \gamma\eta}\right)\left(\delta + \frac{v'(E)}{\eta}\right), \tag{22}$$

and with $R = 0$ it follows from (8) that the quality of the environment will evolve as

$$\dot{E} = \delta E - \gamma M. \tag{23}$$

When the marginal social cost of materials use is low, the optimality condition (21) will encourage a large input of materials in final-goods production. In the absence of recycling it is therefore likely that the pollution from materials use (γM) will exceed the absorption capacity of the environment (δE), causing the environment to deteriorate. Since the marginal utility of environmental quality increases as the quality goes down, it follows from (20) that the fall in environmental quality will drive up its shadow value η over time. According to (19) the shadow value λ of natural resource reserves will likewise increase systematically with time. Moreover, as long as man-made capital is relatively scarce, its marginal product is likely to exceed the rate of time preference, inducing positive savings and capital accumulation that will cause consumption to rise (cf. (16)) and drive down the marginal utility of consumption μ over time. At the same time the accumulation of capital will gradually reduce its marginal productivity.

Thus the linear economy is likely to be characterized by falling values of μ and F_K and rising values of η and λ as capital and pollution accumulate and the natural-resource stock diminishes. With the passing of time the economy will therefore reach a point where $g'(0)\left(\frac{\lambda+\gamma\eta}{\mu}\right) = F_K$. Beyond this point it becomes optimal to move from the linear phase to a circular phase with a positive level of recycling determined by the arbitrage condition (12b), which ensures identical marginal social returns to investment in recycling and investment in final-goods production. The transition from the linear economy with $R = 0$ to the circular economy with $R > 0$ alleviates the pressure on the environment as the evolution of environmental quality becomes governed by (8) rather than (23).

5. Resource Allocation in the Market Economy

Let us now compare the resource allocation generated by competitive markets with the socially optimal allocation described above. Consider a representative competitive mining firm owning a natural-resource stock S from which it extracts a flow of new raw materials N per period. Extraction is costless, and raw materials can be sold at the real market price p. In each period the mining firm can therefore pay out the following (time-varying) net dividend D^M to its owners:

$$D^M = pN. \tag{24}$$

The market value V_t^M of the mining firm at time t is the present value of its future dividend payouts, which is

$$V_t^M = \int_t^\infty D_z^M e^{-\int_t^z r_q dq} dz, \tag{25}$$

where r is the real market interest rate. The mining firm draws up a plan for the future levels of extraction that will maximize its market value (25) at time t subject to the stock-flow constraint $\dot{S} = -N$ and the predetermined initial reserve stock S_t. The first-order conditions for the solution to this problem yield the classical Hotelling rule stating that the equilibrium natural-resource price rises at the rate of interest:

$$r = \frac{\dot{p}}{p}. \tag{26}$$

The mining firm sells the extracted raw materials to the representative competitive firm in the final-goods industry, and the price of materials adjusts to ensure that supply equals demand, so that

$$N = M - R. \tag{27}$$

The final-goods firm uses the production technology (3) and the recycling technology (4) (when recycling is profitable). The government may choose to levy a unit tax at the (time-varying) rate τ on materials that are not recycled. Using the final good as numeraire, the real dividend D^Y paid out by the final-goods firm after deduction for investment expenditure may therefore be written as

$$\begin{aligned} D^Y &= Y - (p+\tau)(M-R) - I \\ &= F(K-K^R, M) - (p+\tau)[1-g(K^R/M)]M - I. \end{aligned} \tag{28}$$

By analogy to (25), the market value V^Y of the final-goods firm is

$$V_t^Y = \int_t^\infty D_z^Y e^{-\int_t^z r_q dq} dz. \tag{29}$$

Given (28) and its initial total stock of capital, the final-goods firm chooses K^R, M, and I with the purpose of maximizing (29) subject to the stock-flow constraint $\dot{K} = I$. The first-order conditions for the solution to this problem imply that

$$F_K = r, \tag{30}$$

$$mF_M = p + \tau, \tag{31}$$

$$K^R = 0 \text{ if } g'(0)(p+\tau) \leq F_K, \tag{32a}$$

$$g'(K^R/M)(p+\tau) = F_K \text{ if } g'(0)(p+\tau) > F_K. \tag{32b}$$

Equation (30) is the standard condition for profit maximization, that the marginal productivity of capital must equal the real rate of interest. Equation (31) says that materials are used until their marginal productivity equals their tax-inclusive price, allowing for the multiplier effect of recycling captured by the variable m. According to (32a), no capital is invested in recycling unless the resulting saving on materials expenses exceeds the marginal revenue from investing capital in final-goods production. In the early stage of development where natural resources are abundant and man-made capital is scarce, the materials price p will be low and the marginal productivity of capital in final-goods production will be high, so (32a) suggests that the market economy will go through an initial linear phase with no recycling. However, (26) implies that the materials price will rise over time, and as capital accumulates its marginal productivity will fall. At some point recycling therefore becomes profitable, and the market economy will enter the circular phase where the profit-maximizing level of recycling is determined by the arbitrage condition (32b), which requires identical marginal returns to investment in recycling and investment in final-goods production.

The household finances its consumption by the net dividends received from firms and by a government lump-sum transfer B financed by the revenue from the tax on nonrecycled materials. Hence

$$C = D^M + D^Y + B, \quad B = \tau(M - R). \tag{33}$$

Note that D^M and D^Y are dividend payouts *minus* any new capital that households inject in firms, so (33) allows for financial savings. The total household wealth V is

$$V \equiv V^M + V^Y. \tag{34}$$

From the expressions for V^M and V^Y in (25) and (29) it follows that total wealth evolves as

$$\dot{V} \equiv \dot{V^M} + \dot{V^Y} = r(V^M + V^Y) - D^M - D^Y = rV - (D^M + D^Y). \tag{35}$$

Combining (33) and (35), we obtain the dynamic household budget constraint:

$$\dot{V} = rV + B - C. \tag{36}$$

The household maximizes the present value of its lifetime utility (1) subject to the budget constraint (36) and the initial stock of wealth, taking the government transfer B as given. The first-order conditions for the solution to this problem yield the standard Keynes–Ramsey rule,

$$\frac{\dot{C}}{C} = \frac{1}{\sigma}(r-\rho), \quad \sigma \equiv -\frac{u''C}{u'} > 0. \tag{37}$$

When the condition for profit maximization $r = F_K$ is inserted, (37) takes the same form as the wealth accumulation rule (16) for the planned economy. The market economy will therefore accumulate wealth at the optimal rate provided the marginal product of capital $F_K(K^Y, M)$ is at its first-best level at each point in time. For this to be the case, resource allocation in the market economy must also obey the portfolio composition rule (17). The next section shows how this may be achieved.

6. Securing the Optimal Transition from the Linear to the Circular Economy

Differentiating (31) with respect to time and inserting (26), (30), and (31) into the resulting equation, we obtain the following expression characterizing the portfolio composition in the market economy:

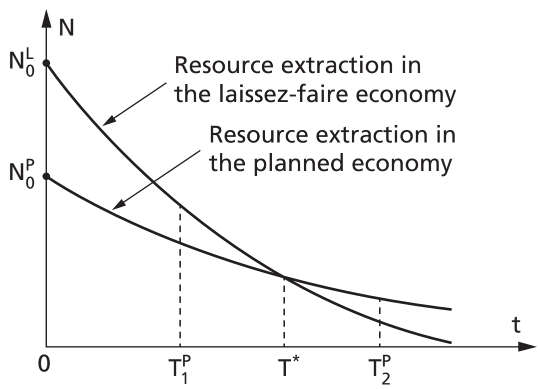
$$F_K = \frac{\dot{F_M}}{F_M} + \frac{\dot{m}}{m} + \frac{r\tau - \dot{\tau}}{P}, \quad P \equiv p + \tau. \tag{38}$$

Comparing (17) with (38), we see that, in a laissez-faire economy where $\tau = \dot{\tau} = 0$, the marginal private gain from postponing resource extraction given by the right-hand side of (38) will tend to be lower than the marginal social rate of return, which includes the environmental gain from slower extraction. In the initial linear phase of the laissez-faire economy, natural-resource extraction will therefore tend to be too rapid relative to the first-best pace of extraction. Intuitively one would also expect the transition to the circular phase to occur too late in the laissez-faire economy. However, this cannot be taken for granted, since the more intensive use of raw materials in the linear laissez-faire economy also means that the scarcity of natural resources increases faster over time.

The situation is illustrated in figure 1, where the flatter curve starting at the initial extraction level N_0^P shows the time path of materials extraction in the planned economy, and the steeper curve starting at the higher extraction level N_0^L depicts the evolution of extraction in the laissez-faire economy. Since the total area under each curve must add up to the same initial reserve stock S_0, the curve for the laissez-faire economy must cut through the curve for the planned economy from above at some point in time, denoted by T^* in figure 1. Now suppose it is optimal for the planned economy to move from the linear to the circular stage at time T_1^P. At that time, where the recycling multiplier m is still 1 but just about to become larger than 1, it follows from (11) and (12) that

$$\overbrace{g'(0)F_M(K,M)}^{\substack{\text{Marginal return to investment in recycling} \\ \text{(starting from zero recycling)}}} = \overbrace{F_K(K,M)}^{\substack{\text{Marginal return to investment} \\ \text{in final-goods production}}} . \tag{39}$$

Figure 1



In the laissez-faire economy, where $\tau = 0$, (31) and (32) likewise imply that the transition to the circular economy will take place at the time when the condition (39) is met. However, at time T_1^P the laissez-faire economy is seen to involve a larger materials input and is therefore likely to have a higher materials intensity M/K than the planned economy, implying a lower marginal productivity of materials and a higher marginal productivity of capital. In the laissez-faire economy the left-hand side of (39) will then be smaller than the right-hand side at time T_1^P, so the transition to the circular phase will not take place until some later time when the materials intensity has fallen sufficiently to satisfy the equality in (39). In this example the laissez-faire economy will thus move too slowly to the circular phase.

But suppose the initial marginal return to investment in recycling, $g'(0)$, is very low, so that it is not optimal for the planned economy to become circular until time T_2^P in figure 1. At that time the laissez-faire economy has a lower materials use and therefore most likely a lower materials intensity than the planned economy, implying (by simple reversal of the reasoning above) that it must have moved from the linear to the circular phase at some earlier time. Without imposing further restrictions on the model, we therefore cannot say whether the transition from the linear to the circular phase in the laissez-faire economy happens too early or too late.

What we *can* say is that the transition will take place at the "wrong" time and that the levels of materials use and recycling at any given point in time will be distorted compared to the first-best levels. These market failures may be corrected by imposing a Pigouvian tax on nonrecycled materials at a rate equal to the present value of the marginal environmental cost of materials use. Specifically, this Pigou tax must be levied at the following rate, where η and λ are the shadow values of the environment and of the natural resource stock prevailing along the economy's first-best time path, and where mec_z is the marginal external cost of using a unit of nonrecycled materials in some future period z, measured as a fraction of its tax-inclusive price P_z:

$$\tau_t = \int_t^{\infty} mec_z P_z e^{-\int_t^z r_q dq} dz, \quad mec_z \equiv \left(\frac{\gamma\eta}{\lambda+\gamma\eta}\right)\left(\delta + \frac{v'(E)}{\eta}\right). \tag{40}$$

To see that this tax rate does indeed guarantee optimality, note that (40) implies

$$\dot{\tau} = r\tau - mec \cdot P = r\tau - \left(\frac{\gamma\eta}{\lambda+\gamma\eta}\right)\left(\delta + \frac{v'(E)}{\eta}\right)P. \tag{41}$$

When (41) is inserted in (38), the resulting portfolio composition rule for the market economy becomes identical to the corresponding portfolio composition rule (17) for the planned economy, both in the linear phase with $m = 1$ and $\dot{m} = 0$ and in the circular phase with $m > 1$ and $\dot{m} \neq 0$. At any point in time the values of K and M in the market economy will then be at their first-best levels, and profit-maximizing behavior will therefore ensure that the transition from the linear to the circular economy takes place at the right time determined by (39).

In his influential study of the Green paradox, Sinn (2008) pointed out that an environmental tax on the use of a polluting exhaustible raw material may actually backfire if the present value of the tax rate increases over time, since resource owners will then have an incentive to accelerate the extraction of the resource, thereby accelerating the accumulation of pollution in the environment. The optimal Pigouvian tax determined by (40) is not vulnerable to such a Green paradox, since we see from (41) that the tax rate will grow at a rate below the rate of interest, so that its present value will fall over time.

On the other hand, from (40) and (41) we cannot exclude the possibility that the Pigou tax should start out from a high level and be gradually lowered as the relative price of raw material increases over time. As Sinn (op. cit.) pointed out, there may be serious political-economy obstacles to such a time profile of environmental taxation.

7. Conclusions

Our simple Ramsey model with natural resources that can be recycled has generated the following insights.

First, the proponents of the circular-economy paradigm are right in claiming that the economy should at some point move from a linear phase with no recycling to a circular phase where a part of the polluting materials used in production is recycled. The rationale for moving to the circular economy is a growing scarcity of natural resources relative to man-made capital combined with a deterioration of environmental quality as a result of their use.

Second, as natural resources become scarcer, the transition to a circular economy will occur even in a laissez-faire economy, but it will happen at the wrong time, and the volume of recycling will be distorted due to lacking internalization of the environmental cost of materials use.

Third, this market failure can be eliminated through a Pigouvian tax on nonrecycled materials that reflects their marginal environmental costs. When such a tax is levied, there is no need for further intervention, as profit-maximizing behavior will then secure the appropriate level and timing of recycling.

Thus the analysis suggests that the subsidy schemes and other forms of regulation (like mandatory sorting of waste) that have been implemented in many countries with the aim of promoting recycling may be poor substitutes for environmental taxes designed and calibrated according to time-honored Pigouvian principles. It seems that the case for other forms of regulation must rest mainly on political-economy barriers to Pigouvian taxation and/or a lack of information or administrative capacity to implement Pigou taxes at the correct level.

Our simple model could be extended in numerous ways. A fruitful topic for future research might be to include pollution from waste generated in the process of consumption and the possibility of sorting and recycling of household waste. In such a setting the optimal policy is likely to include a tax on nonrecycled household waste in addition to a tax on nonrecycled materials used by firms.

8. Appendix

8.1. Derivation of Equation (17)

Rearrangement of (11) yields

$$\mu m F_M = \lambda + \gamma \eta. \tag{42}$$

Differentiating both sides of (42) with respect to time, we get

$$\mu(\dot{m}F_M + m\dot{F_M}) + mF_M\dot{\mu} = \dot{\lambda} + \gamma\dot{\eta}$$
$$\Leftrightarrow \quad \mu m F_M\left(\frac{\dot{m}}{m} + \frac{\dot{F_M}}{F_M} + \frac{\dot{\mu}}{\mu}\right) = \dot{\lambda} + \gamma\dot{\eta}. \quad (43)$$

Dividing through by $\mu m F_M$ in (43) and inserting (42), we find

$$\frac{\dot{m}}{m} + \frac{\dot{F_M}}{F_M} + \frac{\dot{\mu}}{\mu} = \frac{\dot{\lambda} + \gamma\dot{\eta}}{\lambda + \gamma\eta}. \quad (44)$$

Using the first-order conditions (13), (14), and (15) to eliminate $\dot{\mu}/\mu$, $\dot{\lambda}$, and $\dot{\eta}$ from (44), we obtain

$$\frac{\dot{m}}{m} + \frac{\dot{F_M}}{F_M} + \rho - F_K = \frac{\rho(\lambda+\gamma\eta) - \gamma[\eta\delta + v'(E)]}{\lambda+\gamma\eta} \quad \Leftrightarrow$$
$$\frac{\dot{m}}{m} + \frac{\dot{F_M}}{F_M} - F_K = -\left(\frac{\gamma\eta}{\lambda+\gamma\eta}\right)\left(\delta + \frac{v'(E)}{\eta}\right). \quad (45)$$

A simple rearrangement of (45) now yields (17).

8.2. Derivation of Equation (38)

Differentiation of both sides of the first-order condition (31) with respect to time gives

$$\dot{m}F_M + m\dot{F_M} = \dot{p} + \dot{\tau} \quad \Leftrightarrow \quad mF_M\left(\frac{\dot{m}}{m} + \frac{\dot{F_M}}{F_M}\right) = \dot{p} + \dot{\tau}. \quad (46)$$

According to (31) we have $mF_M = p + \tau$, which may be inserted in (46) to give

$$\frac{\dot{m}}{m} + \frac{\dot{F_M}}{F_M} = \frac{\dot{p} + \dot{\tau}}{p + \tau}. \quad (47)$$

From the Hotelling rule (26) we know that $\dot{p} = rp$, and according to (30) value maximization by the final-goods firm implies $r = F_K$. Using these results, we can rewrite (47) as

$$\frac{\dot{m}}{m} + \frac{\dot{F_M}}{F_M} - F_K = \frac{rp + \dot{\tau}}{p+\tau} - r\left(\frac{p+\tau}{p+\tau}\right) \quad \Leftrightarrow \quad F_K = \frac{\dot{m}}{m} + \frac{\dot{F_M}}{F_M} + \frac{r\tau - \dot{\tau}}{p+\tau}, \quad (48)$$

which is identical to (38).

References

Advisory Board for Circular Economy (2017), Anbefalinger til Regeringen, Copenhagen, http://mfvm.dk/fileadmin/user_upload/MFVM/Miljoe/Cirkulaer_oekonomi/Advisory_Board_for_cirkulaer_oekonomi_Rapport.pdf (Access Date: 2017-07-19).

Baumol, W. (1977), On Recycling as a Moot Environmental Issue, Journal of Environmental Economics and Management 4, 83–87.

Boulding, K. (1966), The Economics of the Coming Spaceship Earth, in: Jarrett, H. (Ed.), Environmental Quality in a Growing Economy, Resources for the Future/Johns Hopkins University Press, Baltimore, MD, 3–14.

Di Vita, G. (2001), Technological Change, Growth and Waste Recycling, Energy Economics 23, 549–567.

Di Vita, G. (2007), Exhaustible Resources and Secondary Materials: A Macroeconomic Analysis, Ecological Economics 63, 138–148.

Ellen Macarthur Foundation (2012), Towards the Circular Economy, Vol. 1: Economic and Business Rationale for a Circular Economy, Cowes: Ellen Macarthur Foundation.

European Commission (2015), Closing the Loop – An EU Action Plan for the Circular Economy, COM (2015) 614 Final, https://ec.europa.eu/transparency/regdoc/rep/1/2015/EN/1-2015-614-EN-F1-1.pdf (Access Date: 2017-07-19).

Georgescu-Roegen, N. (1971), The Entropy Law and the Economic Process, Harvard University Press, Cambridge, MA.

Hoel, M. (1978), Resource Extraction and Recycling with Environmental Costs, Journal of Environmental Economics and Management 5, 220–235.

Lusky, R. (1975), Optimal Taxation Policies for Conservation and Recycling, Journal of Economic Theory 11, 315–328.

Lusky, R. (1976), A Model of Recycling and Pollution Control, Canadian Journal of Economics 9, 91–101.

Pittel, K., Amigues, J.-P., and Kuhn, T. (2010), Recycling under a Material Balance Constraint, Resource and Energy Economics 32, 379–394.

Schultze, W. (1973), The Optimal Use of Non-Renewable Resources: The Theory of Extraction, Journal of Environmental Economics and Management 1, 53–73.

Sinn, H.-W. (2008), Public Policies Against Global Warming: A Supply Side Approach, International Tax and Public Finance 15, 360–394.

Smith, V. (1972), Dynamics of Waste Accumulation: Disposal Versus Recycling, Quarterly Journal of Economics 86, 600–616.

Zhijun, F., and Nailing, Y. (2007), Putting a Circular Economy into Practice in China, Sustainability Science 2, 95–101.

Social Insurance for Long-Term Care with Deductible and Linear Contributions

Justina Klimaviciute and Pierre Pestieau*

Received 02 February 2017; in revised form 31 July 2017; accepted 29 September 2017

With the rapid increase in long-term care (LTC) needs, the negligible role of the market, and the declining role of informal family care, one might expect that the government would take a more proactive role in the support of dependent elderly, particularly those who cannot, whatever the reason, count on assistance from their family. The purpose of this paper is to analyze the possibility of designing a sustainable public LTC scheme that would meet a widespread concern, that about going bankrupt and being unable to bequeath any saving to one's children.

Keywords: long-term care, deductible theorem, capped spending, optimal taxation

JEL classification: H 21, I 13, J 14

1. Introduction

Due to the aging process, the rise in long-term care needs constitutes a major challenge of the coming decades. Long-term care (LTC) concerns individuals who are no longer able to carry out basic daily activities such as eating, washing, dressing, etc. Nowadays, the number of persons in need of LTC is substantial. According to Frank (2012), in 2010 nearly 10 million Americans required ongoing help through LTC. This number is expected to grow to 15 million by 2020. Similarly, in Europe the number of persons in need of LTC is expected to grow from 27 million in 2013 to 35 million by year 2060 (see European Commission, 2015).

* Klimaviciute: Université de Liège, Liège, Belgium (justina.klimaviciute@ulg.ac.be); Pestieau: Université de Liège, Liège, Belgium (p.pestieau@ulg.ac.be). This paper is part of the special issue of FinanzArchiv/Public Finance Analysis commemorating Hans-Werner Sinn's 70th birthday. We thank the Editor Alfons Weichenrieder and two anonymous referees for their helpful comments and suggestions. We are also grateful to the participants of the 16th Journées Louis-André Gérard-Varet, the APET 2017 conference and seminars in Pisa, Strasbourg, and Paris Nanterre for useful remarks and questions. Financial support from the Belgian Science Policy Office (BELSPO) research project CRESUS is gratefully acknowledged. P. Pestieau gratefully acknowledges financial support from the Chaire "Marché des risques et création de valeur" of the FDR/SCOR.

FinanzArchiv 74 (2017), 88–108
ISSN 0015-2218
doi: 10.1628/001522118X15084133837809

The expected rise in the number of persons in need of LTC raises the question of the provision of care. As stressed by Norton (2000), about two-thirds of LTC is generally provided by informal caregivers (mainly the family, i.e., spouses, daughters, and stepdaughters). Recent figures in Frank (2012) show that about 80 % of dependent individuals in the U.S. receive informal care from relatives and friends. The remaining LTC is provided formally, that is, through services that are paid for on the market. Formal care can be provided either at the dependent's home, or in an institution (care center or nursing home). Whereas LTC services do not require high skills, they are nonetheless extremely expensive. Those large costs raise the question of the funding of formal LTC. And that question will become increasingly important in the future, when it is expected that the role of informal LTC provision will decrease. According to the 2015 Aging Report (European Commission, 2015), one can foresee at the same time an increase in the needs for LTC and "a shift from informal care towards formal care-giving as typical caregivers get more involved in the labor market and the new family structures may imply less support to the older generations" (European Commission, 2015, p. 147). The implication of this is that financial risks associated with meeting LTC needs will grow and therefore the development of mechanisms for absorbing these risks will gain in importance.

Given that each person has a large probability of entering a nursing home when becoming old and given the large costs of these institutions, one would expect that private LTC insurance markets would expand, in order to insure individuals. However, although markets for private LTC insurance exist, these remain thin in most countries. According to Brown and Finkelstein (2007), only about 9 to 10 % of the population at risk of facing future LTC costs has purchased a private LTC insurance in the U.S. This is the so-called *long-term care insurance puzzle*.[1] For various reasons pertaining both to the demand side (myopia, denial of LTC, crowding out by the family, etc.) and to the supply side of that market (high loading factors, unattractive reimbursement rules, etc.), only a small fraction of the population buys LTC private insurance. One can thus hardly rely only on the development of private LTC insurance markets to fund the cost of LTC.

In the light of the expected decline in informal care, and of the difficulties faced by the market for private LTC insurance, one may expect that the public sector will play a more important role in the provision and funding of LTC. Nowadays, in most advanced economies, the state is involved either in the provision or in the funding of LTC services, but to an extent that varies strongly across countries. However, the involvement of the public sector in LTC is not as comprehensive and generous as it is for the funding of general health ser-

1 Pestieau and Ponthière (2012).

vices. The LTC pillar of the welfare state remains quite thin in comparison with other pillars of the social insurance system.

Recently a number of papers have looked at the design of optimal social insurance for LTC.[2] In most cases, they assume at the outset that the LTC public benefit is flat and thus not related to the severity of the dependence, nor to the amount of contributions. Those papers do not meet one of the concerns of most dependents, which is that they might incur very large costs that would force them to sell all their assets and prevent them from bequeathing any of them. This concern is not met by current LTC practices either. This concern could be dealt with by a system in which individuals' contributions to their LTC costs are capped at a certain amount after which individuals would be fully covered for all further expenditures. Such a system was proposed in the UK by the Dilnot Commission (2011). The Dilnot Commission describes the rationale for this suggestion in terms of the benefits of insurance. While only a fraction of the dependents (in their estimates around a third) would reach the proposed cap of about £35,000, everyone would benefit from knowing that if they ended up in the position of facing these costs, they would be covered, removing the fear and uncertainty of the current system (Dilnot Commission, 2011, p. 32).

We argue that this proposed formula can be justified as an efficient insurance policy, applying Arrow's (1963) theorem on insurance deductibles. This theorem goes as follows: "If an insurance company is willing to offer any insurance policy against loss desired by the buyer at a premium which depends only on the policy's actuarial value, then the policy chosen by a risk-averting buyer will take the form of 100 % coverage above a deductible minimum" (Arrow, 1963). In an earlier paper, Klimaviciute and Pestieau (2017), we show that optimal social LTC insurance indeed features a deductible as long as there are loading costs. In that paper, we study a nonlinear policy allowing for the deductibles to differ between the individual types and the states of nature. In the present paper, we want to explore a more restricted policy in which the government is constrained to use linear instruments and the same deductible for all types and in both dependence states of nature. We consider thus a social insurance scheme that consists of a linear payroll tax and 100 % coverage of LTC risks above a deductible. Another feature of this paper is that besides the heterogeneity in income we consider the reasonable hypothesis that there is a negative correlation between the income levels and the probability of dependence.

Before proceeding, a comment is in order. Both Dilnot Commission (2011) and our analysis are looking for the optimal design of a social insurance that would meet the concern of most individuals, namely avoiding losing all one's

2 See, e.g., Cremer et al. (2016), Cremer and Roeder (2013), Pestieau and Sato (2008).

assets in case of a too long and too severe state of dependence. Such a scheme does not exist in the real world. Yet through a number of social assistance programs many countries offer some kind of protection against those "catastrophic" risks. The best example of this is Medicaid in the U.S., which provides LTC support to the poor and to the middle class elderly who incur a long and costly period of disability. Unfortunately, these programs are not as generous and universal as required by those risks, which are expected to grow rapidly in the next decades.

The rest of the paper is organized as follows. Section 2 presents the model. Section 3 discusses optimal public policy in the absence of private LTC insurance, and section 4 looks at the case when private insurance is available. Section 5 concludes. Some additional analysis and more technical material are provided in the appendix (section 6).

2. The Model

We consider a two-period model with a society consisting of N types of individuals. Individuals differ in their first-period income y_i $(i = 1,\ldots,N)$[3] and in their probabilities of becoming dependent in the second period. In the first period, individuals choose how to allocate their disposable income between their first-period consumption c_i and their savings s_i $(i = 1,\ldots,N)$ for the second period. In the second period, individuals face the risk of becoming dependent. With probability π_{1i} $(i = 1,\ldots,N)$, they experience a low severity level of dependence, in which case they have LTC needs (expressed in terms of costs incurred) L_1; with probability π_{2i} $(i = 1,\ldots,N)$, they face a heavy dependence with LTC needs $L_2 > L_1$; and with probability $1-\pi_{1i}-\pi_{2i}$, they remain healthy. We assume that the risk of dependence is negatively correlated with individual income, i.e., $\pi_{1j} > \pi_{1k}$ and $\pi_{2j} > \pi_{2k}$ for all j,k for which $y_j < y_k$. We first assume that there is no market for private LTC insurance (reflecting the fact that, as mentioned in the introduction, private LTC insurance market is in most countries very small or nonexistent), but later on, in section 4, we also consider the case where private insurance purchases are possible.

We consider a government that introduces a public policy consisting of a linear income tax of rate τ used to finance social LTC insurance and a demogrant A provided in the first period (in section 6.1 we also discuss the case without a demogrant). Most of our analysis focuses on the case of a utilitarian government, but in section 3 we also look at the case of a Rawlsian social

3 For simplicity, we do not explicitly model individual choices of labor supply and consider individual income as exogenously given.

welfare function, which allows us to derive some deeper insights into the influence of redistributional concerns. We discuss the Rawlsian case further in section 4. We allow for inefficiency in tax collection by assuming that a tax rate τ is associated with a quadratic cost $\frac{\gamma\tau^2}{2}$, with $\gamma > 0$. We also assume that insurance provision is not costless for the government, i.e., the government faces loading costs $\lambda > 0$ that reflect, for instance, the associated administrative expenses. Following Arrow's (1963) theorem of the deductible, we consider a social LTC insurance scheme in which individuals have to pay for their LTC needs themselves up to a certain amount D, above which the costs are fully covered by the government. Note, however, that if LTC costs in some state of nature are lower than D, the government provides no insurance in that state and the individuals simply pay the entirety of their costs. We will assume that D is always lower than the costs in the state of heavy dependence (L_2), but will consider the possibility that it is higher than the costs in the state of light dependence (L_1).[4]

Denoting by $c_i^{D_1}$, $c_i^{D_2}$, and c_i^I the second-period individual wealth levels (net of LTC costs) in respectively the light-dependence, the heavy-dependence, and the healthy states, the expected utility of an individual i $(i = 1,\ldots,N)$ can be written as follows:[5]

$$U_i = u(c_i) + \pi_{1i} u\left(c_i^{D_1}\right) + \pi_{2i} u\left(c_i^{D_2}\right) + (1 - \pi_{1i} - \pi_{2i}) u(c_i^I),$$

$$\text{where} \qquad c_i = y_i(1-\tau) + A - s_i,$$

$$c_i^{D_1} = \begin{cases} s_i - D & \text{if } D \leq L_1, \\ s_i - L_1 & \text{if } D > L_1, \end{cases}$$

$$c_i^{D_2} = s_i - D \text{ and } c_i^I = s_i.$$

It should be noted that we do not model individuals' bequests explicitly, but rather focus on their total second-period wealth. An alternative could be to add a joy-of-giving to the utility function, but this would not change the essence of the analysis. We therefore concentrate on individuals' total wealth, assuming implicitly that they decide how to allocate this wealth between their old-age consumption and bequests left to their children. As long as bequests are considered as normal goods, wealthier individuals will leave higher bequests. In other words, individuals want to smooth both their consumption and their bequests across the states of nature.

4 In this paper we assume that dependence occurs in the whole second period of life and that it is measured in monetary units. In a more realistic model, dependence could occur at any age and last as long as life. In that case an insurance with deductible would cover all LTC expenses beyond a given length of dependence. For this, see Drèze et al. (2016).

5 For simplicity, we assume that individuals have the same utility functions in both periods and in all states of nature. Another way would be to assume state-dependent preferences, but this makes the problem much more complicated.

The individual choice of savings is made so as to satisfy the following first-order condition (FOC):

$$u'(c_i) = \pi_{1i} u'\left(c_i^{D_1}\right) + \pi_{2i} u'\left(c_i^{D_2}\right) + (1 - \pi_{1i} - \pi_{2i}) u'(c_i^I). \tag{1}$$

Note that, in the presence of a negative correlation between individual income and the risk of dependence, the comparison of savings chosen by different individual types is generally ambiguous and depends on the differences in y and on dependence probabilities. If, for instance, there are only small differences in y but large differences in dependence probabilities between the types, it is possible that poorer individuals will save more than the richer ones because they have a higher risk of experiencing the states of nature with losses (LTC costs). It seems, nevertheless, that such a situation is less likely to occur and that it is more reasonable to expect differences in y to be larger than differences in dependence probabilities. In what follows, we therefore assume this more reasonable scenario and consider that savings of richer individuals are higher than those of poorer ones.

In order to focus on the influence of redistributional concerns, we also make an assumption that the loading costs λ are not too large, so that, from the pure point of view of insurance provision, insuring individuals against LTC costs (i.e., proposing $D < L_2$) is desirable. More specifically, we assume that at the point $D = L_2$, we have

$$\lambda < \frac{\sum n_i \left[u'\left(c_i^{D_2}\right) - u'(c_i)\right]}{\sum n_i u'(c_i)}, \tag{2}$$

where n_i is the share of type i individuals in the society ($\sum n_i = 1$).[6]

3. Optimal Linear Policy without Private Insurance

We now turn to the derivation of the optimal public policy, and we first study the case of a utilitarian government.

6 This condition is derived from the fourth term in equation (7) or the third term in equation (21).

3.1. Utilitarian Case

The Lagrangian of the government's problem can be written as follows:[7]

$$\begin{aligned}\mathcal{L} = \sum n_i [u(y_i(1-\tau)+A-s_i)+\pi_{1i}u(s_i-D)\\ +\pi_{2i}u(s_i-D)+(1-\pi_{1i}-\pi_{2i})u(s_i)]\\ +\mu\sum n_i\Big[\Big(1-\frac{\gamma\tau}{2}\Big)\tau y_i - A-(1+\lambda)\pi_{1i}(L_1-D)\\ -(1+\lambda)\pi_{2i}(L_2-D)\Big],\end{aligned} \tag{3}$$

where μ is the Lagrange multiplier associated with the government's budget constraint. Note that (3) applies as long as $D \leq L_1$ holds. If $D > L_1$, the term $\pi_{1i}u(s_i-D)$ becomes $\pi_{1i}u(s_i-L_1)$ and the term $(1+\lambda)\pi_{1i}(L_1-D)$ disappears.

Using the envelope theorem, the FOCs for the policy variables can be written in the following way:

$$\frac{\partial\mathcal{L}}{\partial\tau} = -\sum n_i u'(c_i)y_i + \mu\sum n_i y_i(1-\gamma\tau) = 0, \tag{4}$$

$$\frac{\partial\mathcal{L}}{\partial A} = \sum n_i u'(c_i) - \mu = 0, \tag{5}$$

$$\begin{aligned}\frac{\partial\mathcal{L}}{\partial D} = -\sum n_i\pi_{1i}u'\big(c_i^{D_1}\big) - \sum n_i\pi_{2i}u'\big(c_i^{D_2}\big)\\ +\mu\sum n_i(1+\lambda)\pi_{1i} + \mu\sum n_i(1+\lambda)\pi_{2i} = 0.\end{aligned} \tag{6}$$

Note that for $D > L_1$, the first and third terms disappear from equation (6).

We can then define the following compensated FOCs:

$$\frac{\partial\mathcal{L}^c}{\partial\tau} = \frac{\partial\mathcal{L}}{\partial\tau} + \frac{\partial\mathcal{L}}{\partial A}\frac{dA}{d\tau} = 0$$

and

$$\frac{\partial\mathcal{L}^c}{\partial D} = \frac{\partial\mathcal{L}}{\partial D} + \frac{\partial\mathcal{L}}{\partial A}\frac{dA}{dD} = 0$$

with $\frac{dA}{d\tau} = (1-\gamma\tau)\bar{y}$ and $\frac{dA}{dD} = (1+\lambda)\bar{\pi}_1 + (1+\lambda)\bar{\pi}_2$ derived from the budget constraint, where $\bar{y} = \sum n_i y_i$, $\bar{\pi}_1 = \sum n_i\pi_{1i}$, and $\bar{\pi}_2 = \sum n_i\pi_{2i}$.

After some manipulations, the compensated FOC for τ can be written as

$$\frac{\partial\mathcal{L}^c}{\partial\tau} = -\mathrm{cov}\big[u'(c), y\big] - \gamma\tau\bar{y}\sum n_i u'(c_i) = 0,$$

7 We focus on the policy including a demogrant. For comparison, the utilitarian case without a demogrant is provided in section 6.1.

where $\mathrm{cov}[u'(c), y] = \sum n_i u'(c_i) y_i - \sum n_i u'(c_i) \bar{y}$. This gives

$$\tau = \frac{-\mathrm{cov}[u'(c), y]}{\gamma \bar{y} \sum n_i u'(c_i)} > 0.$$

The optimal tax rate thus exhibits the usual trade-off between efficiency (the denominator) and redistribution (the numerator, which is positive, since $\mathrm{cov}[u'(c), y]$ is negative).

Similarly, the compensated FOC for D can be written as

$$\begin{aligned}\frac{\partial \mathcal{L}^c}{\partial D} &= -\mathrm{cov}\left[u'\left(c^{D_1}\right), \pi_1\right] - \mathrm{cov}\left[u'\left(c^{D_2}\right), \pi_2\right] \\ &\quad + \bar{\pi}_1 \sum n_i \left[(1+\lambda) u'(c_i) - u'\left(c_i^{D_1}\right)\right] \\ &\quad + \bar{\pi}_2 \sum n_i \left[(1+\lambda) u'(c_i) - u'\left(c_i^{D_2}\right)\right] = 0, \end{aligned} \tag{7}$$

where $\mathrm{cov}\left[u'\left(c^{D_j}\right), \pi_j\right] = \sum n_i \pi_{ji} u'\left(c_i^{D_j}\right) - \sum n_i u'\left(c_i^{D_j}\right) \bar{\pi}_j$, with $j = 1, 2$.

The compensated derivative $\frac{\partial \mathcal{L}^c}{\partial D}$ has four terms (note again that for $D > L_1$, the first and third terms will disappear). The last two terms reflect purely the motive of insurance and would be present even if all individuals were identical. The first two terms, on the other hand, reflect the motive of redistribution. Given the assumption that differences in y are sufficiently large compared to differences in dependence probabilities, so that richer individuals save more than poorer ones, we see that the two covariances are positive and thus the first two terms call for a lower deductible. Indeed, since those who are worse off (i.e., the poor) have a higher probability of becoming dependent, transferring resources to the dependence states of nature reinforces redistribution.

It is instructive to study $\frac{\partial \mathcal{L}^c}{\partial D}$ by evaluating it at $D = 0$ (which means full insurance provided by the government).[8] It can first be noted that if $\lambda = 0$, the last two terms of $\frac{\partial \mathcal{L}^c}{\partial D}$ are then equal to zero, which means that, because of the negative first two terms, the compensated derivative is negative, implying that it is optimal to have $D < 0$. Thus, if there are no loading costs, the possibility to use insurance for redistribution calls for providing more than full insurance (whereas in the case of identical probabilities, with the covariance terms being equal to zero, full insurance would be optimal under $\lambda = 0$). On the other hand, if $\lambda > 0$, the last two terms are positive at $D = 0$, which makes the sign of the whole derivative ambiguous. Indeed, since insurance is costly, it might be no longer optimal to provide more than full, or even just full, insurance. Note,

8 Note that in this case, the assumption about the relative size of differences in y and dependence probabilities is not needed: in the presence of full insurance, wealth levels in the three second-period states of nature are equalized, and differences in dependence probabilities thus play no role in the individual saving decisions. Richer individuals therefore always save more than poorer ones.

however, that, differently from the case of identical probabilities (where we have less than full insurance as soon as $\lambda > 0$), full insurance is not necessarily excluded under heterogeneous probabilities and might still be optimal if the loading costs are not too large compared to the redistributional concerns.

To gain a deeper insight into how the optimal deductible is influenced by redistributional concerns, we will now look at the solution obtained under a Rawlsian social welfare function implying the maximization of the least well-off individual's welfare.

3.2. Rawlsian Case

The least well-off individual in the considered society is the one having the lowest income and the highest probability of dependence. Let us assume that this individual is of type $i = N$ and, for simplicity, that $y_N = 0$. Let us also focus on the case of $D \leq L_1$ to allow for D being smaller than or equal to zero. The Lagrangian of the government's problem can thus be written as follows:

$$\begin{aligned}\mathcal{L} = & u(A - s_N) + \pi_{1N} u(s_N - D) + \pi_{2N} u(s_N - D) \\ & + (1 - \pi_{1N} - \pi_{2N}) u(s_N) \\ & + \phi \left[\left(1 - \frac{\gamma\tau}{2}\right) \tau \bar{y} - A - (1+\lambda)\bar{\pi}_1 (L_1 - D) - (1+\lambda)\bar{\pi}_2 (L_2 - D) \right],\end{aligned} \tag{8}$$

where ϕ is the Lagrange multiplier associated with the government's budget constraint, and $\bar{y}$, $\bar{\pi}_1$, and $\bar{\pi}_2$ are the average values of y, π_1, and π_2 as defined before.

The FOCs for the policy variables can now be written in the following way:

$$\frac{\partial \mathcal{L}}{\partial \tau} = \phi \bar{y}(1 - \gamma\tau) = 0, \tag{9}$$

$$\frac{\partial \mathcal{L}}{\partial A} = u'(c_N) - \phi = 0, \tag{10}$$

$$\frac{\partial \mathcal{L}}{\partial D} = -\pi_{1N} u'\left(c_N^{D_1}\right) - \pi_{2N} u'\left(c_N^{D_2}\right) + \phi(1+\lambda)\bar{\pi}_1 + \phi(1+\lambda)\bar{\pi}_2 = 0. \tag{11}$$

From (9) we have that the optimal tax rate is simply $\tau = \frac{1}{\gamma}$. As far as the optimal deductible is concerned, combining (11) with (10), using (1), and noting that for $D \leq L_1$ we have $u'\left(c_N^{D_1}\right) = u'\left(c_N^{D_2}\right)$, we obtain the following FOC:

$$\begin{aligned}& (1+\lambda)(\bar{\pi}_1 + \bar{\pi}_2)(1 - \pi_{1N} - \pi_{2N}) u'\left(c_N^I\right) \\ & - (\pi_{1N} + \pi_{2N}) u'\left(c_N^{D_1}\right)[1 - (1+\lambda)(\bar{\pi}_1 + \bar{\pi}_2)] = 0.\end{aligned} \tag{12}$$

It can be easily verified that if $\frac{\pi_{1N}+\pi_{2N}}{\bar{\pi}_1+\bar{\pi}_2} = 1+\lambda$, we have $D = 0$, and if $\frac{\pi_{1N}+\pi_{2N}}{\bar{\pi}_1+\bar{\pi}_2} > (<)\ 1+\lambda$, we have $D < (>)\ 0$. The optimal deductible is thus influenced by the ratio between the sum of the dependence probabilities of the poorest individual (which is more generally the poorest individual's probability of becoming dependent, whatever the severity level) and the sum of the population's average dependence probabilities (which is the population's average probability of becoming dependent, whatever the severity level). If the poorest individual's dependence probability is much higher than the population average, it might be optimal to have a negative deductible even in the presence of loading costs. In section 6.2 we show more generally that the optimal deductible decreases when the ratio $\frac{\pi_{1N}+\pi_{2N}}{\bar{\pi}_1+\bar{\pi}_2}$ goes up. The more likely the poorest individual is to become dependent, compared to the average in the society, the more resources need to be transferred to the dependence states of nature.

The main results of this section can be summarized in the following proposition:

Proposition 1 Consider a setting wherein individuals differ in income and dependence probability and wherein a LTC social insurance consists of a deductible and a linear income tax (with a demogrant). A negative correlation between income and risk makes the case for social insurance stronger and may trigger a departure from Arrow's theorem: a zero or even negative deductible may be optimal despite insurance loading costs. This is particularly clear at a Rawlsian optimum, which implies a negative deductible if the ratio between the worst-off individual's and the average dependence probability is greater than one plus the loading cost.

4. The Case with Private Insurance

So far we have assumed away the possibility for individuals to purchase insurance on the private market. We are now going to introduce this possibility. Rochet (1991) shows, in the context where both private and social insurance have no loading costs, that a utilitarian optimum implies no use of private insurance as long as there is a negative correlation between individual productivity and the probability of loss. He also shows that private insurance is not used when the government's objective is Rawlsian. We are going to explore if these results are valid in our context.

We therefore assume that there is a market for private LTC insurance and that private insurance can cover part of the social insurance deductible, thus reducing the amount of LTC expenses that the individual effectively incurs. More precisely, we denote by α_{1i} $(0 \leq \alpha_{1i} \leq 1)$ the fraction of the social insurance deductible to be covered in the state of light dependence, and by α_{2i} $(0 \leq \alpha_{2i} \leq 1)$ the fraction to be covered in that of heavy dependence $(i = 1, \ldots, N)$. Note that private LTC insurance is possible only when the social insurance deductible is strictly positive (i.e., there is a loss in the dependence states of nature); otherwise, no private insurance is provided. We also assume that private insurers face the same loading costs (λ) as the government.

The timing we consider is the following. First, the government announces its policy consisting of a linear income tax of rate τ, a demogrant A, and social LTC insurance with a deductible D. Given this policy, individuals then choose their savings s_i and their private insurance coverage characterized by fractions α_{1i} and α_{2i} of the social insurance deductible. Reasoning backwards, we will first discuss individual choices and then we will look at the government's policy.

4.1. Individual Choices

The expected utility of an individual i can be written as follows:

$$U_i = u(c_i) + \pi_{1i} u\left(c_i^{D_1}\right) + \pi_{2i} u\left(c_i^{D_2}\right) + (1-\pi_{1i}-\pi_{2i}) u(c_i^I),$$

where $c_i = y_i(1-\tau) + A - P_i - s_i$,

$$c_i^{D_1} = \begin{cases} s_i - (1-\alpha_{1i})D & \text{if } 0 < D \leq L_1, \\ s_i - (1-\alpha_{1i})L_1 & \text{if } D > L_1, \end{cases}$$

$c_i^{D_2} = s_i - (1-\alpha_{2i})D$, $c_i^I = s_i$, and P_i is the private insurance premium given by[9]

$$P_i = (1+\lambda)[\pi_{1i}\alpha_{1i} + \pi_{2i}\alpha_{2i}]D \tag{13}$$

if $0 < D \leq L_1$, or by

$$P_i = (1+\lambda)[\pi_{1i}\alpha_{1i}L_1 + \pi_{2i}\alpha_{2i}D] \tag{14}$$

if $D > L_1$. The FOC for s_i can be written as in (1), whereas the FOCs for α_{1i} and α_{2i} are respectively

$$-u'(c_i)(1+\lambda) + u'\left(c_i^{D_1}\right) \leq 0 \tag{15}$$

and

$$-u'(c_i)(1+\lambda) + u'\left(c_i^{D_2}\right) \leq 0. \tag{16}$$

Assuming interior solutions and combining (15) and (16), we have $u'\left(c_i^{D_1}\right) = u'\left(c_i^{D_2}\right)$, which implies $(1-\alpha_{1i})D = (1-\alpha_{2i})D$ (or $(1-\alpha_{1i})L_1 = (1-\alpha_{2i})D$).

We can define $M_i \equiv (1-\alpha_{1i})D = (1-\alpha_{2i})D$ (or $M_i \equiv (1-\alpha_{1i})L_1 = (1-\alpha_{2i})D$), M_i being the true deductible that an individual i has to pay. We can then rewrite the individual problem in terms of M_i as follows:

$$\max_{s_i, M_i}\left[U_i = u(c_i) + \pi_{1i} u\left(c_i^{D_1}\right) + \pi_{2i} u\left(c_i^{D_2}\right) + (1-\pi_{1i}-\pi_{2i}) u(c_i^I)\right],$$

9 We assume that the private insurers know the individual risk probability. This is quite a standard assumption, made, for instance, in the economics of annuities.

where $c_i = y_i(1-\tau) + A - P_i - s_i$, $c_i^{D_1} = c_i^{D_2} = s_i - M_i$, $c_i^I = s_i$, and $P_i = (1+\lambda)\pi_{1i}(D - M_i) + (1+\lambda)\pi_{2i}(D - M_i)$ if $D \leq L_1$, but $P_i = (1+\lambda)\pi_{1i}(L_1 - M_i) + (1+\lambda)\pi_{2i}(D - M_i)$ if $D > L_1$. The FOC for s_i again can be written in the same way as in (1), while the FOC for M_i can be written as

$$u'(c_i)[(1+\lambda)\pi_{1i} + (1+\lambda)\pi_{2i}] - \pi_{1i}u'\left(c_i^{D_1}\right) - \pi_{2i}u'\left(c_i^{D_2}\right) = 0. \tag{17}$$

Evaluating the left-hand side of (17) at $M_i = 0$, it can be easily verified that, as long as $\lambda > 0$, the optimal level of M_i is always greater than zero. In other words, as long as there are loading costs, private insurance always features a strictly positive deductible (individuals purchase less than full insurance).

For further analysis, it is useful to explore how the optimal level of M_i differs between individual types, and in particular how it depends on the two individual characteristics: income and dependence probabilities. In section 6.3, we show that the way in which M_i is influenced by these two variables depends on the absolute risk aversion (ARA) exhibited by the utility function. As far as income is concerned, we show that M_i is increasing in y_i under decreasing absolute risk aversion (DARA), decreasing in y_i under increasing absolute risk aversion (IARA), and constant in y_i under constant absolute risk aversion (CARA) preferences.[10] To see the intuition of this result, recall that a higher deductible means less insurance. Since under DARA (under IARA) wealthier people are less (more) risk-averse, they require less (more) insurance. On the other hand, we find that M_i is increasing in dependence probability under CARA and IARA preferences, while the effect is ambiguous under DARA. To understand this result, first note that an increase in dependence probability raises the price of insurance. There is then a substitution effect that pushes for buying less insurance (i.e., increasing the deductible). However, there is also a wealth effect in the sense that an increase in the price of insurance makes the individual poorer. In the case of IARA, this translates into the individual becoming less risk-averse, which, like the substitution effect, pushes for a higher deductible. The deductible thus clearly increases under IARA. In contrast, under DARA the wealth effect pushes in the opposite direction to the substitution effect, since poorer individuals are more risk-averse in that case and thus require lower deductibles. The overall effect is thus ambiguous. Finally, under CARA, the wealth effect plays no role and the deductible increases only due to the substitution effect. The results on the effect of income and dependence probability on M_i are summarized in table 1.

Let us now discuss what conclusions can be drawn about the differences in M_i between individual types. Under CARA, M_i does not depend on income

10 DARA (IARA, CARA) means that absolute risk aversion decreases (increases, remains constant) when wealth increases. For more details, see section 6.3.

Table 1
Effect of Income and Dependence Probability on M_i

ARA	$\frac{\partial M_i}{\partial y_i}$	$\frac{\partial M_i}{\partial \pi_{1i}}$ (or $\frac{\partial M_i}{\partial \pi_{2i}}$)
DARA	>0	$\lesseqgtr 0$
IARA	<0	>0
CARA	$=0$	>0

but increases with dependence probability, which, taking into account the negative correlation between income and dependence probabilities, implies that poorer (and thus higher-probability) individuals will clearly choose higher deductibles than richer ones. Under IARA, M_i also increases with dependence probability and, in addition to this, decreases with increasing income, which again makes it clear that the deductible will be higher for poorer individuals. On the other hand, this is not necessarily the case under DARA. First, under DARA, M_i increases with income, which pushes for poorer individuals having lower deductibles. Second, the effect of dependence probability is ambiguous. If it is negative, i.e., if M_i decreases with increasing dependence probability, then poorer individuals will indeed have lower deductibles than richer ones. If it is positive, i.e., if M_i increases with dependence probability, then the total effect is not clear. No clear-cut comparison can therefore be made in the case of DARA.

4.2. Public Policy

We can now turn to public policy. Let us first consider the utilitarian case discussed in section 3.1 but in the presence of the above-described private insurance market. Using the envelope theorem, it can be verified that the FOCs of the social planner's problem are the same as in the case without private insurance (equations (4)–(6)). The compensated FOC for D is thus also the same as equation (7). Let us now analyze this equation, given the presence of private insurance.

When there is little social insurance (D is high), all individuals buy private insurance (assuming that everyone can afford it) and we have $(1+\lambda)u'(c_i) - u'\left(c_i^{D_1}\right) = 0$ and $(1+\lambda)u'(c_i) - u'\left(c_i^{D_2}\right) = 0$ for all i. However, this level of D is not optimal, since the compensated derivative is then negative due to the covariance terms. When we decrease the level of D, there will be a point where some individuals, those with the highest optimal M, will stop buying private insurance. In the cases of CARA and IARA, these will be the poorest individuals, while under DARA, that is not necessarily the case. Other individuals, those with lower levels of optimal M, will continue insuring themselves on the private market. For these individuals we will thus still have $(1+\lambda)u'(c_i)-u'\left(c_i^{D_1}\right)=0$ and $(1+\lambda)u'(c_i)-u'\left(c_i^{D_2}\right)=0$, whereas for those who stop buying private insurance we will now have $(1+\lambda)u'(c_i)-u'\left(c_i^{D_1}\right)>0$ and $(1+\lambda)u'(c_i)-u'\left(c_i^{D_2}\right)>0$. The last two terms of (7) will thus be positive, and this might be the optimal solution if the covariance terms are not too large. On the other hand, it might be optimal to reduce D even more, so that all individuals stop buying private insurance. Thus, we might have the result of no use of private insurance, as in Rochet (1991), but a situation where some individuals insure themselves privately cannot be ruled out either. It is, however, clear that the social optimum implies a nonpurchase of private insurance at least by some individuals in the society (the poorest ones in the cases of CARA and IARA). These individuals get more social insurance than they would purchase on the private market.

The reason why our conclusions differ from those of Rochet (1991) is that we consider a setting where insurance (both social and private) involves loading costs. Indeed, if we assumed, as Rochet (1991), that both social and private insurance were actuarially fair, we would also have a conclusion of no private insurance. To see this, let us suppose for a moment that $\lambda = 0$. In that case, the optimal level of M for all individuals is zero (i.e., full insurance). Thus, if the government provides less than full insurance (i.e., $D > 0$), all the individuals insure themselves privately to reach full insurance. However, this is not optimal from the social point of view, since the last two terms on the left-hand side of (7) are then zero and the first two are negative. Even full social insurance (i.e., $D = 0$) is not optimal, since the last two terms then remain zero as well. D therefore has to be reduced even more and becomes negative, i.e., more than full insurance is provided. It is clear that then there is no private insurance.[11]

On the other hand, when $\lambda > 0$, it might be optimal to have a strictly positive social insurance deductible (i.e., less than full insurance). Moreover, in that case, different individual types require different levels of insurance. Thus, if social insurance is less than full, this might be sufficient for some individual types but insufficient for others, who would then insure themselves on the private market. Note, however, that the case for private insurance becomes weaker if private insurers have higher loading costs than the government.

Let us now look at the Rawlsian case discussed in section 3.2. Again, using the envelope theorem, it can be verified that the FOCs of the social planner's problem have the same form as in the case without private insurance (equa-

11 Note that in Rochet (1991) insurance is not allowed to be more than full, and his result is thus that full public insurance is optimal, which also implies no private insurance.

tions (9)–(11)). Combining (11) with (10), we have the following FOC for D:

$$-\pi_{1N}u'\left(c_N^{D_1}\right)-\pi_{2N}u'\left(c_N^{D_2}\right)+u'(c_N)(1+\lambda)(\bar{\pi}_1+\bar{\pi}_2)=0. \tag{18}$$

We know from section 3.2 that if $\frac{\pi_{1N}+\pi_{2N}}{\bar{\pi}_1+\bar{\pi}_2}=(>)\,1+\lambda$, we have $D=(<)\,0$. In these cases, it is clear that there will be no private insurance. On the other hand, if $\frac{\pi_{1N}+\pi_{2N}}{\bar{\pi}_1+\bar{\pi}_2}<1+\lambda$, then $D>0$ and private insurance might occur. As long as $D\geq M_N$, equation (17) holds for type N and, using it in (18), we have

$$u'(c_N)(1+\lambda)(\bar{\pi}_1+\bar{\pi}_2-\pi_{1N}-\pi_{2N})<0.$$

This means that it is optimal to have $D<M_N$. Thus, the worst-off individual will clearly not purchase private insurance. On the other hand, as in the utilitarian case, some other individuals might still find it desirable to insure themselves on the private market.

Proposition 2 summarizes the main insights of this section.

Proposition 2 Introducing the possibility of private insurance with the same loading cost as the social insurance and keeping the setting of Proposition 1, it can be shown that under the utilitarian optimum at least some individuals will not purchase private insurance, but a situation where some other individuals insure themselves privately cannot be excluded. Whether these individuals belong to the top or to the bottom of the income distribution depends on the absolute risk aversion. With the Rawlsian objective, the worst-off individual never purchases private insurance.

5. Conclusion

In this paper we have looked at the design of a social insurance for LTC that consists of a linear payroll tax (with a demogrant) and a deductible. We were thus following Arrow's (1963) proposal that an efficient way of providing insurance when there are loading costs is to let the insurees pay all the costs below a given deductible and reimburse them for any expenses above that deductible. We were in particular interested in exploring how the design of such policy is affected by a reasonable assumption that income and the probability of dependence are negatively correlated. In the first part of the paper, we assumed that there was no market for private LTC insurance, whereas we introduced that possibility in the second part.

We show that the presence of a negative correlation between income and dependence probability makes the case for social insurance stronger and might trigger a departure from Arrow's theorem in the sense that, due to redistributional concerns, a zero or even a negative deductible might be optimal despite the presence of loading costs. The influence of redistributional concerns is particularly clearly seen in the case of a Rawlsian social welfare function. In that case, a negative deductible becomes optimal as soon as the ratio between the worst-off individual's and the population's average probability of dependence becomes greater than one plus the loading cost.

The introduction of private LTC insurance allows us to compare our results with those of Rochet (1991), who shows, in a context without loading costs, that a negative correlation between individual productivity and the probability of loss implies no use of private insurance. We find that this result does not necessarily hold in our setting involving loading costs. In particular, with a utilitarian social welfare function, we find that the social optimum implies nonpurchase of private insurance at least by some individuals in the society (these are the poorest individuals under CARA and IARA preferences, but not necessarily under DARA), but a situation where some other individuals insure themselves privately cannot be ruled out as long as the optimal social insurance is less than full. With a Rawlsian social welfare function, private insurance is clearly not purchased by the least well-off individual, while it might be purchased by some other ones (but also only if social insurance is less than full).

6. Appendix

6.1. Utilitarian Case without a Demogrant

Here we consider a more restrictive version of the utilitarian case presented in section 3, namely, a policy in which the government is not able to use a demogrant. The government's problem appears as in section 3.1 except that we now set $A = 0$. The FOCs for τ and D also appear as in (4) and (6). We now define the following compensated FOC:

$$\frac{\partial \mathcal{L}^c}{\partial \tau} = \frac{\partial \mathcal{L}}{\partial \tau} + \frac{\partial \mathcal{L}}{\partial D}\frac{dD}{d\tau} = 0,$$

where $\frac{dD}{d\tau} = \frac{(\gamma\tau - 1)\bar{y}}{(1+\lambda)\bar{\pi}_1 + (1+\lambda)\bar{\pi}_2}$ is derived from the budget constraint.

After some manipulations, this FOC can be written as

$$\begin{aligned}\frac{\partial \mathcal{L}^c}{\partial \tau} = &-[(1+\lambda)\bar{\pi}_1 + (1+\lambda)\bar{\pi}_2]\text{cov}\left[u'(c), y\right] \\ &+ \bar{y}\text{cov}\left[u'\left(c^{D_1}\right), \pi_1\right] + \bar{y}\text{cov}\left[u'\left(c^{D_2}\right), \pi_2\right] \\ &+ \bar{y}\bar{\pi}_1 \sum n_i\left[u'\left(c_i^{D_1}\right) - (1+\lambda)u'(c_i)\right] \\ &+ \bar{y}\bar{\pi}_2 \sum n_i\left[u'\left(c_i^{D_2}\right) - (1+\lambda)u'(c_i)\right] \\ &- \gamma\tau\bar{y}\left[\sum n_i\pi_{1i}u'\left(c_i^{D_1}\right) + \sum n_i\pi_{2i}u'\left(c_i^{D_2}\right)\right] = 0. \end{aligned} \tag{19}$$

Note that for $D > L_1$, equation (19) is missing the second and fourth terms, as well as the term $(1+\lambda)\bar{\pi}_1$ in the first and the term $\sum n_i \pi_{1i} u'\left(c_i^{D_1}\right)$ in the last brackets.

We can then express the optimal tax rate as

$$\tau = \frac{\begin{array}{c} -[(1+\lambda)\bar{\pi}_1 + (1+\lambda)\bar{\pi}_2]\mathrm{cov}\left[u'(c), y\right] \\ +\bar{y}\mathrm{cov}\left[u'\left(c^{D_1}\right), \pi_1\right] + \bar{y}\mathrm{cov}\left[u'\left(c^{D_2}\right), \pi_2\right] \end{array}}{\gamma \bar{y}\left[\sum n_i \pi_{1i} u'\left(c_i^{D_1}\right) + \sum n_i \pi_{2i} u'\left(c_i^{D_2}\right)\right]}$$

$$+ \frac{\begin{array}{c} \bar{y}\bar{\pi}_1 \sum n_i \left[u'\left(c_i^{D_1}\right) - (1+\lambda)u'(c_i)\right] \\ +\bar{y}\bar{\pi}_2 \sum n_i \left[u'\left(c_i^{D_2}\right) - (1+\lambda)u'(c_i)\right] \end{array}}{\gamma \bar{y}\left[\sum n_i \pi_{1i} u'\left(c_i^{D_1}\right) + \sum n_i \pi_{2i} u'\left(c_i^{D_2}\right)\right]} \tag{20}$$

with the above-mentioned terms disappearing for $D > L_1$.

The denominator of (20) is again the efficiency term, which is positive. The numerator, however, unlike in the case with a demogrant, now takes into account not only the motive of redistribution in the first period (the first term, which pushes for a higher tax rate), but also the motives of insurance (the last two terms) and of redistribution in the second period achieved through insurance provision (the second and third terms). As discussed in section 3, the two covariances entering the second and third terms are positive and call for increasing insurance coverage (i.e., lowering the deductible), which also means increasing the tax rate so that this coverage can be financed.

To gain somewhat more insight, we can look at the compensated FOC $\frac{\partial \mathcal{L}^c}{\partial \tau}$ evaluated at $\tau = 0$. From the budget constraint, $\tau = 0$ obviously implies that no insurance coverage is provided, which in other words means that D is equal to L_2. Noting that we are now in the case $D > L_1$, and recalling the assumption (2), we can write

$$\begin{aligned} \frac{\partial \mathcal{L}^c}{\partial \tau}\Big|_{\tau=0} = &-(1+\lambda)\bar{\pi}_2 \mathrm{cov}\left[u'(c), y\right] + \bar{y}\mathrm{cov}\left[u'\left(c^{D_2}\right), \pi_2\right] \\ &+ \bar{y}\bar{\pi}_2 \sum n_i \left[u'\left(c_i^{D_2}\right) - (1+\lambda)u'(c_i)\right] > 0. \end{aligned} \tag{21}$$

Equation (21) tells us that the optimal tax rate is $\tau > 0$, which also implies that the optimal deductible D is lower than L_2, i.e., it is desirable to provide social LTC insurance. We can also note that the heterogeneity of individuals makes the case for social insurance stronger. Indeed, even if the assumption (2) is not satisfied and the third term of (21) is negative, it could still be possible to have $D < L_2$ if the covariance terms are large enough. In other words, even if providing LTC insurance is inefficient from a pure insurance point of view, there may still be a case for social insurance due to redistributional concerns. Note also that in this case, without a demogrant, social insurance may be justified even in the absence of a negative correlation between income and dependence probabilities (i.e., with the second covariance equal to zero).[12] To some extent insurance now also plays the role of a demogrant, since taxes are collected proportionally to income but insurance provision is the same to everyone. Introducing a positive tax and using the proceeds to finance social insurance thus enhances redistribution, as reflected by the first term of (21).

6.2. Optimal Deductible in the Rawlsian Case

We are now going to show that the optimal deductible in the Rawlsian case decreases when the probability ratio $\frac{\pi_{1N}+\pi_{2N}}{\bar{\pi}_1+\bar{\pi}_2}$ goes up. To do this, let us first note that the ratio $\frac{\pi_{1N}+\pi_{2N}}{\bar{\pi}_1+\bar{\pi}_2}$ can increase when π_{1N} and/or π_{2N} increases (and the increase in $\bar{\pi}_1+\bar{\pi}_2$ is sufficiently small) or when π_{1N} and π_{2N} remain the same but the probabilities of other individuals decrease, implying a decrease in $\bar{\pi}_1+\bar{\pi}_2$. We look at these two cases.

For the first case, we assume for simplicity that π_{1N} increases while π_{2N} and the sum $\bar{\pi}_1+\bar{\pi}_2$ remain the same (i.e., we assume that the probabilities of some other individuals decrease in such a way that $\bar{\pi}_1+\bar{\pi}_2$ remains unchanged). We therefore need to verify how the optimal deductible changes due to the increase in π_{1N}. From (12) we obtain

$$\frac{\partial D}{\partial \pi_{1N}} = \frac{-(1+\lambda)(\bar{\pi}_1+\bar{\pi}_2)u'(c_N^I)+u'\left(c_N^{D_1}\right)[(1+\lambda)(\bar{\pi}_1+\bar{\pi}_2)-1]}{-SOC_D} < 0, \tag{22}$$

where $SOC_D < 0$ is the second-order condition for D and $(1+\lambda)(\bar{\pi}_1+\bar{\pi}_2)-1<0$ from the FOC (12).

Turning to the case when π_{1N} and π_{2N} do not change but $\bar{\pi}_1+\bar{\pi}_2$ decreases, we get

$$-\frac{\partial D}{\partial(\bar{\pi}_1+\bar{\pi}_2)} = \frac{-\begin{bmatrix} (1+\lambda)(1-\pi_{1N}-\pi_{2N})u'(c_N^I) \\ +(\pi_{1N}+\pi_{2N})u'(c_N^{D_1})(1+\lambda) \end{bmatrix}}{-SOC_D} < 0. \tag{23}$$

6.3. Comparative Statics in the Individual Problem with Private Insurance

In this subsection, we derive the comparative statics of individual savings s_i and the effectively faced deductible M_i (chosen simultaneously) with respect

12 This is not true in the case with a demogrant. Indeed, if the assumption (2) does not hold and there is no correlation between income and dependence probabilities, then, evaluating (7) at $D = L_2$, we find that decreasing D is never optimal.

to income y_i and the probability of dependence π_{1i} (the case of π_{2i} is analogous).

Fully differentiating (17) and (1) with respect to y_i, we get respectively

$$\begin{aligned}&(1+\lambda)(\pi_{1i}+\pi_{2i})u''(c_i)(1-\tau)\\&-\frac{\partial s_i}{\partial y_i}\left[(1+\lambda)(\pi_{1i}+\pi_{2i})u''(c_i)+\pi_{1i}u''(c_i^{D_1})+\pi_{2i}u''(c_i^{D_2})\right]\\&+\frac{\partial M_i}{\partial y_i}\left[(1+\lambda)^2(\pi_{1i}+\pi_{2i})^2u''(c_i)+\pi_{1i}u''(c_i^{D_1})+\pi_{2i}u''(c_i^{D_2})\right]=0\end{aligned}\tag{24}$$

and

$$\begin{aligned}&\frac{\partial s_i}{\partial y_i}\left[u''(c_i)+\pi_{1i}u''(c_i^{D_1})+\pi_{2i}u''(c_i^{D_2})+(1-\pi_{1i}-\pi_{2i})u''(c_i^{I})\right]\\&-\frac{\partial M_i}{\partial y_i}\left[(1+\lambda)(\pi_{1i}+\pi_{2i})u''(c_i)+\pi_{1i}u''(c_i^{D_1})+\pi_{2i}u''(c_i^{D_2})\right]\\&-u''(c_i)(1-\tau)=0.\end{aligned}\tag{25}$$

For ease of exposition, let us define the following:

$$\begin{aligned}&[1]\equiv\left[(1+\lambda)(\pi_{1i}+\pi_{2i})u''(c_i)+\pi_{1i}u''(c_i^{D_1})+\pi_{2i}u''(c_i^{D_2})\right]<0,\\&[2]\equiv\left[(1+\lambda)^2(\pi_{1i}+\pi_{2i})^2u''(c_i)+\pi_{1i}u''(c_i^{D_1})+\pi_{2i}u''(c_i^{D_2})\right]<0,\\&[3]\equiv\left[u''(c_i)+\pi_{1i}u''(c_i^{D_1})+\pi_{2i}u''(c_i^{D_2})+(1-\pi_{1i}-\pi_{2i})u''(c_i^{I})\right]<0.\end{aligned}$$

Solving the system of equations (24) and (25) for $\frac{\partial M_i}{\partial y_i}$ and $\frac{\partial s_i}{\partial y_i}$, we obtain

$$\frac{\partial s_i}{\partial y_i}=\frac{u''(c_i)(1-\tau)\left[[2]-[1]\cdot(1+\lambda)(\pi_{1i}+\pi_{2i})\right]}{[3]\cdot[2]-[1]^2}>0$$

and

$$\frac{\partial M_i}{\partial y_i}=\frac{-u''(c_i)(1-\tau)\cdot[4]}{[3]\cdot[2]-[1]^2},$$

where

$$\begin{aligned}[4]\equiv\Big[&(1+\lambda)(\pi_{1i}+\pi_{2i})(1-\pi_{1i}-\pi_{2i})u''(c_i^{I})\\&-(1-(1+\lambda)(\pi_{1i}+\pi_{2i}))\left(\pi_{1i}u''(c_i^{D_1})+\pi_{2i}u''(c_i^{D_2})\right)\Big].\end{aligned}$$

It can be verified that $[2]-[1]\cdot(1+\lambda)(\pi_{1i}+\pi_{2i})<0$ and $[3]\cdot[2]-[1]^2>0$, from which the sign of $\frac{\partial s_i}{\partial y_i}$ then follows immediately. On the other hand, the sign of $\frac{\partial M_i}{\partial y_i}$ depends on the sign of [4]. The sign of [4] is, however, ambiguous in the general case and depends on the absolute risk aversion (ARA) exhibited by the utility function. In particular, we are now going to show that $[4]>0$ under decreasing absolute risk aversion (DARA), $[4]<0$ under increasing absolute risk aversion (IARA), and $[4]=0$ under constant absolute risk aversion (CARA).

To see this, let us first note that DARA (IARA, CARA) means that

$$ARA(c) = \frac{-u''(c)}{u'(c)} < (>,=)\ ARA(d) = \frac{-u''(d)}{u'(d)} \quad \text{for } c > d,$$

where $\frac{-u''(x)}{u'(x)}$ is the Arrow–Pratt measure of absolute risk aversion at wealth x. Thus, noting that with $M_i > 0$ we have $c_i^I > c_i^{D_1}$, under DARA (IARA, CARA) preferences we can write

$$\frac{-u''(c_i^I)}{u'(c_i^I)} < (>,=)\ \frac{-u''(c_i^{D_1})}{u'(c_i^{D_1})}$$
$$\Longleftrightarrow$$
$$u''(c_i^I) > (<,=)\ \frac{u''(c_i^{D_1})}{u'(c_i^{D_1})}\ u'(c_i^I).$$

We can then multiply both sides by $(1+\lambda)(\pi_{1i}+\pi_{2i})(1-\pi_{1i}-\pi_{2i})$ and subtract from both sides $(1-(1+\lambda)(\pi_{1i}+\pi_{2i}))\big(\pi_{1i}u''(c_i^{D_1})+\pi_{2i}u''(c_i^{D_2})\big)$, which gives

$$\begin{aligned} &(1+\lambda)(\pi_{1i}+\pi_{2i})(1-\pi_{1i}-\pi_{2i})u''(c_i^I) \\ &-(1-(1+\lambda)(\pi_{1i}+\pi_{2i}))\big(\pi_{1i}u''(c_i^{D_1})+\pi_{2i}u''(c_i^{D_2})\big) \\ &> (<,=)\ \frac{u''(c_i^{D_1})}{u'(c_i^{D_1})}\big[u'(c_i^I)(1+\lambda)(\pi_{1i}+\pi_{2i})(1-\pi_{1i}-\pi_{2i}) \\ &\qquad -(1-(1+\lambda)(\pi_{1i}+\pi_{2i}))\big(\pi_{1i}u'(c_i^{D_1})+\pi_{2i}u'(c_i^{D_2})\big)\big] = 0, \end{aligned} \tag{26}$$

where we have used the fact that $c_i^{D_1} = c_i^{D_2}$ and that the expression in the last bracket is equal to zero (this follows from combining (17) with (1)).

The left-hand side of the inequality (26) is exactly the definition of [4]; we therefore indeed have, under DARA (IARA, CARA), that $[4] > (<,=)\ 0$. Coming back to $\frac{\partial M_i}{\partial y_i}$, we can thus conclude that $\frac{\partial M_i}{\partial y_i} > (<,=)\ 0$ with DARA (IARA, CARA) preferences.

Fully differentiating (17) and (1) with respect to π_{1i}, we get respectively

$$\frac{\partial M_i}{\partial \pi_{1i}}\cdot[2] - \frac{\partial s_i}{\partial \pi_{1i}}\cdot[1] - (1+\lambda)^2(\pi_{1i}+\pi_{2i})u''(c_i)(D-M_i) = 0 \tag{27}$$

and

$$\frac{\partial s_i}{\partial \pi_{1i}}\cdot[3] - \frac{\partial M_i}{\partial \pi_{1i}}\cdot[1] + (1+\lambda)u''(c_i)(D-M_i) + u'(c_i^{D_1}) - u'(c_i^I) = 0. \tag{28}$$

Solving the system of equations (27) and (28) for $\frac{\partial M_i}{\partial \pi_{1i}}$ and $\frac{\partial s_i}{\partial \pi_{1i}}$, we obtain

$$\begin{aligned} \frac{\partial s_i}{\partial \pi_{1i}} &= \frac{[2]\cdot\big[u'(c_i^I) - u'(c_i^{D_1})\big]}{[3]\cdot[2]-[1]^2} + \frac{(1+\lambda)u''(c_i)(D-M_i)\cdot[2]\cdot[4]}{[1]\cdot\big[[3]\cdot[2]-[1]^2\big]} \\ &> 0 (\lesseqgtr 0) \end{aligned} \tag{29}$$

under CARA and IARA (under DARA), and

$$\frac{\partial M_i}{\partial \pi_{1i}} = \frac{[1]\cdot\left[u'(c_i^I)-u'(c_i^{D_1})\right]}{[3]\cdot[2]-[1]^2} + \frac{(1+\lambda)u''(c_i)(D-M_i)\cdot[4]}{[3]\cdot[2]-[1]^2} > 0\,(\lesseqgtr 0)$$

under CARA and IARA (under DARA).

References

Arrow, K. (1963), Uncertainty and the Welfare Economics of Medical Care, American Economic Review 53, 941–973.

Brown, J. and Finkelstein, A. (2007), Why is the Market for Long-Term Care Insurance So Small?, Journal of Public Economics 91, 1967–1991.

Cremer, H., Pestieau, P., and Roeder K. (2016), Social Long-Term Care Insurance with Two-Sided Altruism, Research in Economics 70, 101–109.

Cremer, H. and Roeder, K. (2013), Long-Term Care Policy, Myopia and Redistribution, Journal of Public Economics 108, 33–43.

Dilnot Commission (2011), Fairer Care Funding: The Report of the Commission on Funding of Care and Support, London: Commission on Funding of Care and Support.

Drèze, J., Pestieau, P., and Schokkaert, E. (2016), Arrow's Theorem of the Deductible and Long-Term Care Insurance, Economics Letters 148, 103–105.

European Commission (2015), The 2015 Ageing Report. Economic and Budgetary Projections for the 28 EU Member States (2013–2060), European Economy 3.

Frank, R. G. (2012), Long-term Care Financing in the United States: Sources and Institutions, Applied Economic Perspectives and Policy 34, 333–345.

Klimaviciute, J. and P. Pestieau (2017), Long-Term Care Social Insurance: How to Avoid Big Losses?, International Tax and Public Finance, forthcoming.

Norton, E. (2000), Long-Term Care, in: Cuyler, A. and J. Newhouse (Eds.), Handbook of Health Economics, Volume 1b, Chapter 17, North-Holland, Amsterdam, 955–994.

Pestieau, P. and Ponthière, G. (2012), Long-Term Care Insurance Puzzle, in: Costa-Font, J. and C. Courbage (Eds.), Financing Long-Term Care in Europe, Palgrave Macmillan, London, 41–52.

Pestieau, P. and Sato, M. (2008), Long-Term Care: the State, the Market and the Family, Economica 75, 435–454.

Rochet, J.-C. (1991), Incentives, Redistribution and Social Insurance, Geneva Papers on Risk and Insurance Theory 16, 143–165.

Ties between Health Policy, Early Health Problems, and Lifetime Earnings

Manuel Flores and Barbara Wolfe*

Received 31 May 2017; in revised form 18 December 2017; accepted 19 December 2017

Extant literature indicates that early-life health affects later labor market outcomes such as earnings and work effort. We examine whether this holds for multiple dimensions of health and regardless of a country's health care system. We ask whether mental and physical health problems and poor general health by age 15 have similar or different influences on lifetime earnings. We then ask whether the health care system influenced the estimated effects of early health problems on lifetime earnings. We expect that early health problems reduce earnings and that the most generous system is tied to the least negative long-term effects.

Keywords: early-life health, lifetime earnings, health care system

JEL classification: D 10, H 51, I 14, J 2

1. Introduction

Disparities in health and their effects on well-being, including labor market earnings, are a topic of considerable interest across the world. Some of these effects may start early in life, as extant literature has demonstrated that early-life health has long-lasting effects on labor market outcomes later in life, such as earnings and work effort (Almond and Currie, 2011). But does this hold for multiple dimensions of health and does this hold regardless of the health care system in a country? In this paper, we attempt to shed light on these questions. In particular, we start with the question of whether three separate dimensions of health (mental health problems, physical health problems, and poor general health) by age 15 have similar or different influences on lifetime earnings; second, we ask if they are sizeable. Our main focus is to ask whether the estimated effects of early health problems on lifetime earnings are influenced by the health care system in which the child lived. Our expectation is

* Flores: Organisation for Economic Co-operation and Development, 2, rue André-Pascal 75775 Paris CEDEX 16, France (manuel.flores@oecd.org); Wolfe: University of Wisconsin-Madison, 1180 Observatory Drive, Madison WI 53706, USA (Bwolfe@wisc.edu). Responsibility for the contents lies with the authors and does not necessarily represent the views of the OECD or the Governments of OECD member countries.

FinanzArchiv 74 (2018), 109–130 doi: 10.1628/001522118X15156739491360
ISSN 0015-2218

that earnings are reduced by early health problems, especially mental health problems, and that the system does make a difference, with the most generous system tied to the smallest long-term negative effects. However, given the limited treatment for mental health problems at the time our survey participants were young, our expectation is that policy differences will not play as great a role in influencing earnings outcomes tied to early mental health problems.

2. Prior Literature

Much of the work tying early-life health problems to labor market outcomes of adults focuses on a single age later in life and typically reports a positive association between childhood healthiness and labor market outcomes such as employment, earnings, and occupation (Case et al., 2005; Case and Paxson, 2008; Flores and Kalwij, 2014). A few studies use a comparison among siblings to account for unobserved family and neighborhood effects; usually their results are robust to a fixed effects (FE) approach (Black et al., 2007; Fletcher, 2014; Delaney and Smith, 2012; Smith, 2009a). Other studies have used very specific exogenous events around birth such as pandemics and famines, and found that the experience of such health shocks reduced earnings in mid-adulthood (Almond, 2006), but only to a limited extent in early adulthood (Chen and Zhou, 2007), or at the end of working life (Nelson, 2010). Flores et al. (2015) use retrospective and prospective data from the Survey of Health Aging and Retirement in Europe (SHARE) to investigate the association of childhood health and socioeconomic status (SES; education, employment, and income) with labor market outcomes over men's and women's entire life cycle in thirteen European countries. They find a long-term association between childhood health and both average annual earnings and lifetime earnings.

Previous studies have illustrated the multidimensional nature of health (Kalwij and Vermeulen, 2008) and have shown that both childhood-specific diseases and childhood health summary variables contain useful information about adult health (Smith, 2009b). Goodman et al. (2011) use prospectively collected data from the British National Child Development Study (NCDS) to assess the long-term effects of childhood psychological and physical health problems on economic outcomes at ages 23, 33, 42, and 50 years. They find that childhood psychological problems are associated with about 15 % lower hourly wages from early adulthood into middle age. For family income, they also show that the associations with psychological conditions are substantially larger than the ones estimated for suffering from physical health conditions during childhood. Smith (2009a) uses a subsample of U.S. siblings from the Panel Survey of Income Dynamics (PSID) aged 25–47 in 1999 to estimate the associations of childhood self-reported health (SRH) status on an individual's initial level of annual earnings at age 25 and its average growth between age 25 and age as of 1999. He finds that about 50 % of the overall effect of poor general health was present by age 25, while the remaining 50 % is the consequence of differential individual income growth after age 25.[1] The oldest individuals in his sample are still relatively young from a perspective of lifetime earnings (namely, 47 years old). A follow-up study using data from 2005 (Smith and Smith, 2010) suggests about a 30-percentage point difference in log family income in 2005 for those with psychological problems as a child, with little difference in estimates using siblings and FE or OLS (consistent with Fletcher and others as noted above).

A major issue with studying the long-term influence of early health conditions on lifetime earnings and other outcomes is the data requirements. As pointed out by Smith (2009b), there are limited options for creating such data: (1) collect data prospectively such as existing cohort studies that begin at birth or close to birth, but here the wait is very long; (2) link survey and administrative data, but this is especially problematic on the health side; or (3) collect information retrospectively, especially early health history. With the last option, the issue is accuracy including a fear of coloring bias. Smith explores the question of accuracy in his 2009 *Demography* article (2009b). Smith collected retrospective data from respondents to the U.S.-based Health and Retirement Study, HRS, and from the Panel Study of Income Dynamics (PSID). The HRS is a panel survey of those aged 50 and over at the initial survey. Smith asked whether respondents had had any of a long list of childhood illnesses before age 17. For those who replied yes, he then employed a "set of markers used in the Calendar Life History (CLH) in order to gather additional early health data. The specific markers included house moves, marital events of parents, and date of entry into different levels of schooling before age 17" (Smith, 2009b, p. 391). Smith then used the best past data on prevalence he could find to validate the retrospective data from those in the HRS and those aged 50-plus in the PSID. He found the prevalence rate for all three categories of disease he created to be very close, providing one type of evidence on the accuracy of recall data. He also conducted a retest seven years later of the same HRS sample and found high rates of consistency (95 % for hypertension, 93 % for diabetes, 96 % for cancer, and so on). The self-rated health scale had slightly less robust responses in the retest; a comparison of 1998 responses to 2005 or 2007 found that 60 % gave the same response, 33 % were one point apart on the five-point scale, while 6 % were 2 points apart. Of the cases that differed,

1 Interestingly, when unobserved family effects are controlled for (using within-siblings estimates), the estimate of childhood SRH on post-age 25 individual income growth is substantially larger, which Smith attributes to a diminishing role of measurement error due to reporting bias in childhood SRH.

about the same percentage showed improvements as declines. Additional exploration that made use of the incidence of new health issues over the period between waves did *not* find a link between the direction of change in the responses and onset of illnesses, *neither* between the onset of later life minor and major illnesses and changes in self-reported childhood health (Smith, 2009a). Smith summarizes his research on this by saying that "recalled information of health conditions during childhood appears to be a quite useful tool that can be readily added to important demographic and health surveys" (Smith, 2009b, p. 401). We agree and are pleased that on the basis of this work by Smith, similar questions were added to SHARE, which we are able to use in the research reported below.

3. Data

Based on the extant literature, we explore the influence that separate dimensions of early-life health such as the experience of a significant mental, physical, or general health problem have on future earnings over the entire working life. Our measures of health problems are retrospective, consistent with the Smith approach, while our measure of earnings, although mainly retrospective, is also based on answers to an ongoing panel data survey.

We use individual-level data from the first three waves of SHARE, a multidisciplinary and representative cross-national panel of the European population aged 50-plus. Waves 1 (2004/05) and 2 (2006/07) include information on sociodemographic background characteristics, current health, and socioeconomic status, as well as expectations of retirement age.[2] Most of our data are from the third wave, SHARELIFE (2008/09), which is a retrospective survey conducted in 13 European countries as part of the SHARE project. It contains retrospective information on the early life circumstances and work careers of about 75 % of the individuals who participated in Waves 1 or 2. Additionally, about 78 % of the individuals who participated in SHARELIFE are also included in Wave 4. Our analysis focuses on men. We do not study women since their labor force participation and earnings may differ for reasons that go beyond health, including maternity, childcare, etc. Descriptive statistics on our sample are in table 1 and include, amongst others, all the right-hand side variables we employ below.

We use our respondent's country of childhood and a four-way system to characterize the health care systems they lived in as children based on descrip-

2 Currently, more waves of SHARE data are available, but only the first three waves contain information on (net) wages. However, we also use Wave 4 data to update and replace missing values for expectations with regard to retirement age for individuals who participated in Waves 1 to 3.

Table 1

Sample Means of EXPLANATORY VARIABLES for Europe and Health Care System During Respondent's Childhood[a]

	Europe	Full coverage	Considerable cost sharing	Less than full coverage	Socialist
Childhood health variables (0–15 years)[b]					
Emotional problems, epilepsy or depression 2+ weeks (0–1) (*Mental*)	0.015	0.024	0.018	0.009	0.014
Poor self-reported health (0–1) (*Poor*)	0.019	0.018	0.027	0.014	0.016
Physical health index (worst 2 percentiles) (0–1) (*Physical*)	0.020	0.022	0.025	0.017	0.016
Other childhood health and SES variables					
Severe headaches or migraines when 0–15 years (0–1)	0.028	0.048	0.036	0.018	0.018
Rooms per person when 10 years old[c]	0.730	0.857	0.885	0.630	0.526
Number of books at home when 10 years old (1–5)[d]	2.101	2.740	2.193	1.713	2.165
Main breadwinner's occupation (in ISCO-88 skill levels) when 10 years old (0–4)[e]	1.971	2.139	1.966	1.883	2.012
Number of facilities at home when 10 years old (0–5)[f]	2.037	3.185	2.228	1.584	1.539
Born in an urban area (0–1)	0.370	0.452	0.334	0.400	0.289
No usual source of care when 0–15 years (0–1)	0.052	0.037	0.030	0.069	0.073
Other control variables					
Age (in years)	62.72	62.73	62.40	62.78	63.20
Mediating factors					
Education level ISCED 0–1[g]	0.244	0.201	0.148	0.383	0.162
Education level ISCED 2[g]	0.168	0.096	0.134	0.214	0.207
Education level ISCED 3–4[g]	0.358	0.390	0.444	0.231	0.445
Education level ISCED 5–6[g]	0.229	0.314	0.274	0.172	0.186
Years in full-time education	12.61	14.53	13.92	10.26	13.44
Ever married (0–1)	0.940	0.926	0.942	0.939	0.951
Years married since first marriage (Fraction)	0.903	0.857	0.895	0.922	0.920
1+ periods of ill health in adulthood (0–1)	0.195	0.186	0.154	0.204	0.265
2+ periods of ill health in adulthood (0–1)	0.053	0.049	0.039	0.055	0.080
Retired due to own ill health (0–1)[g]	0.124	0.155	0.127	0.109	0.122
N	9,199	1,472	2,907	3,355	1,465

Note: [a] This table depicts the final sample of all men. The first three rows show childhood mental, physical, and self-reported health.
[b] Differences across coverage systems are statistically significant at 5 % level of significance for Mental and Poor, but not for Physical.
[c] The number of rooms includes bedrooms, but exclude kitchen, bathrooms, and hallways.
[d] 1 = none or very few (0–10 books), 2 = enough to fill one shelf (11–25 books), 3 = enough to fill one bookcase (26–100 books), 4 = enough to fill two bookcases (101–200 books), 5 = enough to fill two or more bookcases (> 200 books).
[e] 0 = armed forces or no main breadwinner, 1 = first skill level, 2 = second skill level, 3 = third skill level, 4 = fourth skill level.
[f] Includes fixed bath, cold and hot running water supply, inside toilet, and central heating.
[g] Overall sample size for "Education level" and "Retired due to own ill health" is 9,142 and 5,525 observations, respectively.

tions in the U.S. Social Security Administration's Office of Policy (2002). These four groupings are: full coverage; considerable use of co-payments; limited coverage; and Socialist (full coverage but limited care). Table 2 shows the country of childhood of our respondents by these four groups. Countries with health care systems with full coverage in our data include Sweden, Denmark, and United Kingdom. Those with considerable cost sharing include Austria, Belgium, (West) Germany, Finland, France[3], Norway, and Switzerland. The countries with less than full coverage include Greece, Italy, Netherlands, Portugal, and Spain, and finally the countries in our sample with full coverage but limited care (Socialist at the time our sample members were children) include Czechoslovakia, Hungary, Poland, East Germany, and Russia. Because France has attempted to move to a Beveridge system, albeit unsuccessfully, and has characteristics that overlap with such a system that would reduce cost-sharing, we also do a sensitivity test excluding those who grew up in France. In addition, in this sensitivity test we exclude those who grew up in Greece, since as of 1983 there was an expansion of coverage that would change their group (see WHO, 1996).

The sample sizes, with a brief description of the health care system by country, are presented in table 2. Descriptive statistics of all the right-hand-side variables by health care system are given in table 1.

3.1. Measures of Childhood Health

In this research, we use retrospective data on general, physical, and mental health that refer to the period before an individual attained 16 years of age. This categorization is fairly typical in studies of health status: physical health is based on reports of experiencing a set of illnesses that are primarily physical in nature; mental health is based on a set of severe mental illnesses; and general health is based on a more subjective overall assessment of general health (cf. Smith, 2009b). The general measure is commonly used in economic analysis including studies on the income gradient in health (see, for example, Case et al., 2002). We create one measure of general health, one of physical health, and one of mental health. Our general health measure is based on the commonly used self-reported five-point scale, with *excellent* being the highest and *poor* the lowest category of health. From this we select those who report poor health as our *general health* measure. For physical health we ran a polychoric Principal Component Analysis (PCA) using count variables for respiratory

3 Although France has regulated fees that determine insurance reimbursement, providers often charge more (excess charging). Patients pay for outpatient care and then are reimbursed. The average amount coverage paid in the late 1990s was 75 % for doctor visits but 90.2 % for hospital care (Sandier et al., 2004).

Table 2

Distribution of RESPONDENTS in Final Sample by COUNTRY OF ORIGIN with Description of HEALTH CARE COVERAGE

	Full coverage	Considerable cost sharing	Less than full coverage	Socialist	Total	% pop. covered 1960–61 (OECD)	Description for SSA 2002 social security programs, Europe 2002
Czechoslovakia	0	0	0	637	637	100	1956 cash sickness benefit; 1994 public health coverage
Finland	0	11	0	0	11	55	Reimbursed partly, so cost sharing since 1963.
Hungary	0	0	0	6	6	-	Public health service
Austria	0	301	0	0	301	78	Cost sharing. Current law since 1955.
Sweden	670	0	0	0	670	100	Some co-pays since 1962.
Netherlands	0	0	830	0	830	71	No co-pays for those included 1964. Higher income separate
Spain	0	0	670	0	670	54	Few co-pays since 1994, but similar before. Co-pay drugs.
Italy	0	0	988	0	988	87	Few co-pays since 1978, but similar since 1943.
France	0	676	0	0	676	76	Reimbursed partly with co-pays since 1945; changes in 1967.
Denmark	796	0	0	0	796	95	Few co-pays since 1971 with limited choice of provider.
Greece	0	0	854	0	854	44	Few co-pays since 1951 and 1984 except on drugs.
Switzerland	0	440	0	0	440	74	Sizeable co-pays since 1996.
Russia	0	0	0	23	23	-	No co-pays since 1991. OOP for specialists.
Belgium	0	979	0	0	979	59	Co-pays since 1994, but similar since 1894.
Norway	0	3	0	0	3	100	Co-pays since 1997, but similar since 1907. Ceiling on OOP.
Poland	0	0	0	609	609	-	No copays as of 1974, but added in 1999.
Portugal	0	0	13	0	13	19	No co-pays since 1988/90, but similar since 1935.
United Kingdom	6	0	0	0	6	100	Co-pays drugs since 1977, but similar since 1911.
West Germany	0	497	0	0	497	-	Co-pays with exceptions.
East Germany	0	0	0	190	190	-	From 1949 to 1989; co-pays since 1989.
Total	1472	2907	3355	1465	9199		

Note: US Social Security Administration, Office of Policy (2002), Social Security Programs Throughout the World: Europe.

problems (asthma, other respiratory problems and allergies); infectious diseases (polio, severe diarrhea, meningitis/encephalitis, appendicitis, and other infectious diseases); cardiovascular diseases (diabetes or high blood sugar and heart troubles); disorders of the sense organs (chronic ear problems, speech impairment, and difficulty in seeing even with eyeglasses); and other serious health conditions an individual suffered before age 16. Using the first principal component, we create a dummy variable indicating whether an individual is in the bottom or worst two percentiles. This is our index of *childhood physical health*. Our measure of *mental health* is based on responses to questions of whether the individuals experienced emotional, nervous, or psychiatric problems, or epilepsy fits or seizures, or symptoms of depression that lasted at least 2 weeks before age 16.[4]

The frequency of these conditions and the other explanatory variables that we include in our empirical analysis are reported in table 1. In particular, we include various indicators of childhood socioeconomic status (SES) to proxy for parental cultural background (the variable: *number of books at home when 10 years old*); parental occupation (the variable: *main breadwinner's occupation when 10 years old*, in ISCO-88 skill levels); and the household's financial status (the variables: *rooms per person* and *number of facilities at home when 10 years old*). Regarding the parents' financial status dimension, Cavapozzi et al. (2011) show that our variable number of rooms per person in the household is strongly and positively correlated (0.82) with the OECD average disposable income of households with children aged 0–17 and thus serves as "a sound indicator of parental financial status during childhood years" (Cavapozzi et al., 2011, p. 32).[5]

The percentage of the population with each of the three health measures is similar. The three measures of health are largely independent. The correlations between any two of them overall and for each group of countries by health care system are quite low (see table 7). It is worth remembering that one of our main interests in this paper is to compare the relative effects/associations of multiple childhood health measures. As far as we know, there is no exogenous variation that can be used to identify the separate causal effects of childhood physical, mental, and general health. Nevertheless, from a policy perspective (e.g., for the design of prevention policies), it is of crucial importance to gain insight into which of them is most harmful in terms of, for example, the lifetime earnings that an individual accumulates over his working life.

4 We also have information on severe headaches or migraines. Since it is not clear whether to call these physical or mental health, we instead control for whether an individual reports she had severe headaches or migraines when aged 0–15 years in our regression estimates.

5 The facilities variable includes fixed bath, cold and hot running water supply, inside toilet, and central heating and is meant to serve as an additional proxy for the household's financial situation.

3.2. Outcome Variable: Lifetime Earnings

We create a measure of lifetime earnings or compounded labor income as described in Alessie et al. (2013) and Flores et al. (2015). Briefly, this measure uses the first monthly wage on each job, the last monthly wage on the main job, as well as the current wage from Waves 1–3.[6] All countries' monetary values are converted to 2006 Euros following the procedure explained in Trevisan et al. (2011). We use a compound real interest rate, $r = 2\%$ (Haider and Solon, 2006). For future labor income (from W3 interview to retirement, R) we use weighted survival probabilities from country- and sex-specific 2009 period life tables from Eurostat.[7] More accurately, we use $L_{0t} = \sum_{\tau=1}^{t}(1+r)^{t-\tau}E_\tau$ to estimate the compounded labor income if the individual is retired at age t – where E_τ are annual earnings from employment at age τ – and $L_{1t} = \sum_{\tau=t+1}^{R}(1+r)^{t-\tau}E_\tau$ to estimate future income, where we assume that future real annual earnings remain constant ($E_\tau = E_t, \tau = t+1,\ldots,R$). Table 3 below shows sample means and medians of lifetime earnings for Europe and the different health care systems when using our final estimation sample (columns 1) and when restricting the sample to individuals with positive working years (columns 2). The overall mean lifetime earnings for our sample is 1,263,112 Euro (€). The mean is highest in the countries with considerable cost sharing, followed in order by those in countries with full coverage, less than full coverage, and finally Socialist. We note, though, that the calculations are based on country of childhood rather than current country of residence. For the majority of our sample (over 90 %), these two locations are the same.

Table 3

Sample Means and Medians of LIFETIME EARNINGS (LTE) for European Men by Health Care Systems During Childhood

	Europe		Full coverage		Considerable cost sharing		Less than full coverage		Socialist	
	Final sample (1)	LTE > 0 (2)	Final sample (1)	LTE>0 (2)	Final sample (1)	LTE > 0 (2)	Final sample (1)	LTE > 0 (2)	Final sample (1)	LTE > 0 (2)
Sample mean	1,263,112	1,278,680	1,265,740	1,269,189	1,598,515	1,609,032	1,197,398	1,223,658	745,423	754,174
Sample median	939,636	948,932	1,044,144	1,046,160	1,249,256	1,255,415	882,067	892,441	485,292	489,632
N	9,199	9,087	1,472	1,468	2,907	2,888	3,355	3,283	1,465	1,448

Note: The table shows sample median and mean values for lifetime earnings, as well as sample sizes (N) for Europe and by health care system. All amounts are discounted and in purchasing power parity (PPP)-adjusted German Euros of 2006.

6 The retrospective data on wages are taken from the retrospective SHARE Job Episodes Panel Data (Brugiavini et al., 2013.)

7 We use within-period survival probabilities, i.e., between age t and $t+1$, and allow these to vary across country, gender, and age. We assume the survival probabilities remain constant after 2009.

4. Results

When addressing our first question on the tie between our three health dimensions and lifetime earnings, we find strong negative ties between all three early health dimensions and lifetime earnings (see table 4). In terms of particular health problems, we find that poor general health and mental health have stronger negative ties to lifetime earnings than physical health. The strong tie to mental health, which we hypothesized above, implies that a male who suffers from such problems during childhood is estimated to earn up to 202,606–243,094 Euros less income during his working life compared to those who do not suffer from such issues. Somewhat unexpectedly, the point estimates of poor general health suggest even stronger ties with estimates of 217,835–256,420 lower lifetime earnings than those men who do not report poor general health during ages 0–15.[8] These values are about 5 % greater than those for mental health conditions. For childhood physical health, the estimates suggest a range of 132,069–167,832 Euros less, on average, than those without physical health problems, a considerably smaller penalty than for those who experienced a mental health condition or overall poor health. The estimates for physical health are not statistically significant at standard levels. The penalties we estimated for general poor health and mental health problems are approximately one and a half times as large as those for physical health. Thus, it appears that the type of early health problem is relevant for estimating the expected earnings penalty due to an early health condition. The big difference appears to be the smaller expected influence of physical health problems compared to poor general health and mental health.[9]

The strong negative ties between lifetime earnings and both childhood mental health problems and childhood general poor health are consistent with prior research (Goodman et al., 2011; Smith, 2009a) and show that these conditions are likely to lead to lower earnings, possibly due to lower productivity. The lower productivity might be tied to continuing health effects of early health problems, leading to poorer health as an adult. The links could also be tied to early health problems leading to fewer years of schooling, poorer performance in school, or alternatively might represent discrimination, with poten-

8 These results are from regressions that either include only the single health measure or include all three health measures. They also include severe headaches (with a negative but insignificant coefficient), various measures of childhood SES, and whether the individual was born in an urban area, in addition to dummy variables for birth year and country at the time of the SHARELIFE interview.

9 In the case of physical health, we conducted a sensitivity analysis and included a larger share of individuals in our physical health index, namely those who were at least one standard deviation below the mean (approximately 17 % of the sample). We found that estimating with a larger share of individuals with less severe physical health problems results in smaller negative effects on lifetime earnings (see appendix table 8).

Table 4
Estimates of EARLY HEALTH PROBLEMS on the LIFETIME EARNINGS of Men in Europe

	(1)	(2)	(3)	(4)
Mental	−202606**	−243094***		
	(80053)	(79305)		
Physical	−132069		−167832	
	(128940)		(126901)	
Poor	−217835***			−256420***
	(63974)			(60952)
R-squared	0.075	0.074	0.074	0.074
Observations	9199	9199	9199	9199

Note: OLS estimates for mental health problems, physical health problems, and poor self-reported health prior to age 16 obtained from estimating linear models on the lifetime earnings (see table 3) of men. All models include all other variables listed in table 1, except age and the mediating factors, plus country dummies and birth-year dummies. Robust standard errors in parentheses. Significance levels: *** $p < 0.01$ ** $p < 0.05$ * $p < 0.10$.

tially negative effects also in the marriage market (Smith, 2009a). In the next analysis we explore the likelihood of several factors that might account for the general pattern of lower earnings of those with each of the three early health problems.

Education, marital status, and health in adulthood can mediate the associations between childhood health and labor market earnings. Table 5 analyzes the associations of general, physical, and mental health problems during childhood with education (years in full-time education and four broad categories of education), marital status (ever being married and duration of marriage), and health through the life cycle (having 1+ or 2+ periods of ill health during adulthood, and retired due to own ill health). Regarding education, only for poor general health do we find evidence of a significant negative association with years in full-time education (column 1). The coefficient suggests a year less education for those with poor health before age 16. Consistent with this finding, the estimates suggest that those with poor health are more likely to have achieved ISCED levels 0–1 (column 2a) and 2 (column 2b) and less likely to have achieved ISCED levels 3–4 (column 2c) and 5–6 (column 2d).[10] There is no association between either of our other two health indicators and education. Thus on the basis of this analysis, with the exception of poor gen-

10 We use four levels of education defined from the 1997 International Standard Classification of Education (UNESCO, 1997): no education or primary education (ISCED 0–1), lower secondary education (ISCED 2), upper secondary and postsecondary nontertiary education (ISCED 3–4), and tertiary education (ISCED 5–6).

eral health, we do not have evidence that the mediating factor behind lower earnings is less schooling.

For marital status, we find that men who experienced mental or general health problems as children are less likely to have ever been married with the largest association found among those with mental health problems, who are slightly more than 9 % less likely to have ever married than those without mental health problems (column 3). Among those with early mental health problems or early poor health who do marry, the results show that their duration of marriage, as captured by years married since first year of marriage, is also below those of their peers without these health problems (column 4).

Finally, regarding life cycle health, individuals with mental, physical, or general health problems are all more likely to experience at least one episode of ill health during adulthood (column 5). In addition, those with mental health problems or poor general health during childhood are significantly more likely, at a 5 % significance level, to experience two or more episodes of ill health (column 6). The strongest tie is between poor general health as a child and two or more episodes of ill health as an adult. These patterns are all consistent with the pattern of earnings penalties we report above. These results suggest that a major mediating factor in the tie of early poor health to lower lifetime earnings is health issues as an adult, which issues are likely to reduce productivity either on the intensive or extensive margin (lower hourly productivity or fewer days worked). Interestingly, we do not find a statistically significant link between early health problems and an increased probability of retirement due to own ill health, although the coefficients are positive (column 7).

5. Policy Analysis

Does a country's health care system influence the penalty of early health problems? The primary focus of this paper is to ask whether the health care system of a country has an influence on the long-term effects of early childhood health problems. We use the four-way dichotomy to capture the essence of alternative programs: full coverage, considerable cost sharing, less than full coverage, and Socialist. The countries of birth are included and their categorizations are described in table 2.

To explore our question, in table 6 we run a regression in which our aggregate measure of lifetime earnings is the dependent variable and each of our early health indicators are included along with interactions with the four types of health care systems. In addition to the childhood variables in table 1, we include dummy variables for health care systems, year of birth, and current country of residence. As for table 4 (and table 5), we control for the current

Table 5

Marginal Effects of EARLY HEALTH PROBLEMS on EDUCATION, MARTIAL STATUS, and HEALTH DURING ADULTHOOD for Men in Europe[a]

	Full-time education (1)	Educational attainment (2a)	(2b)	(2c)	(2d)	Ever married (3)	Years married since year of marriage (4)	1+ ill health periods (5)	2+ ill health periods (6)	Retired due to own ill health (7)
	Years	ISCED 0–1	ISCED 2	ISCED 3–4	ISCED 5–6	(0–1)	Fraction	(0–1)	(0–1)	(0–1)
Mental	−0.097 (0.359)	0.030 (0.026)	0.007 (0.006)	−0.009 (0.008)	−0.028 (0.024)	−0.094*** (0.030)	−0.145*** (0.034)	0.094** (0.039)	0.062** (0.027)	0.004 (0.036)
Physical	−0.003 (0.312)	−0.023 (0.020)	−0.005 (0.005)	0.007 (0.006)	0.021 (0.019)	−0.026 (0.020)	−0.034 (0.024)	0.097*** (0.034)	0.016 (0.019)	0.054 (0.038)
Poor	−0.988*** (0.340)	0.047** (0.022)	0.011** (0.005)	−0.014** (0.007)	−0.044** (0.021)	−0.057** (0.024)	−0.057** (0.027)	0.104*** (0.035)	0.091*** (0.026)	0.053 (0.035)
Severe headaches or migraines	0.011 (0.275)	−0.020 (0.018)	−0.005 (0.004)	0.006 (0.005)	0.019 (0.017)	0.003 (0.014)	−0.008 (0.019)	0.026 (0.026)	0.002 (0.014)	0.056* (0.029)
Rooms (an increase of one room)	0.871*** (0.135)	−0.070*** (0.009)	−0.016*** (0.002)	0.021*** (0.003)	0.065*** (0.009)	−0.005 (0.006)	−0.003 (0.008)	−0.022* (0.011)	0.001 (0.006)	−0.017 (0.013)
Number of books at home when 10 years old	0.869*** (0.051)	−0.065*** (0.003)	−0.015*** (0.001)	0.019*** (0.001)	0.061*** (0.003)	0.006** (0.003)	0.004 (0.003)	0.002 (0.005)	0.001 (0.002)	−0.012** (0.005)
Main breadwinner's occupation when 10 (in ISCO−88 skill levels)	0.510*** (0.067)	−0.032*** (0.004)	−0.007*** (0.001)	0.010*** (0.001)	0.030*** (0.004)	−0.005 (0.003)	−0.005 (0.004)	−0.010* (0.005)	−0.005* (0.003)	−0.001 (0.006)
Number of facilities at home when 10 years old	0.277*** (0.038)	−0.022*** (0.002)	−0.005*** (0.001)	0.007*** (0.001)	0.021*** (0.002)	−0.000 (0.002)	−0.001 (0.002)	−0.006* (0.003)	−0.005** (0.002)	−0.006* (0.003)
Born in urban area	0.492*** (0.103)	−0.052*** (0.007)	−0.012*** (0.002)	0.015*** (0.002)	0.048*** (0.006)	−0.006 (0.005)	−0.015** (0.006)	−0.006 (0.009)	0.009* (0.005)	−0.019** (0.010)
Had no usual source of care when between 0–15 years old	−0.622** (0.246)	0.002 (0.015)	0.000 (0.003)	−0.000 (0.004)	−0.002 (0.014)	0.013 (0.011)	0.006 (0.011)	0.003 (0.019)	−0.002 (0.010)	0.008 (0.020)
R-squared	0.312						0.037			
Pseudo R-squared		0.145	[b]	[b]	[b]	0.036		0.034	0.056	0.072
Observations	9199	9142	[b]	[b]	[b]	9199	9199	9199	9199	5525

Note: [a] All models include all other variables listed in Table 1, except age plus country dummies and birth-year dummies. Models 1 and 4 show OLS estimates. Model 2 in columns (2a)-(2d) shows marginal effects from an ordered probit model and Models 3–7 show marginal effects from probit models. Robust standard errors in parentheses. Significance levels: *** $p < 0.01$ ** $p < 0.05$ * $p < 0.10$.
[b] Model 2 Pseudo R-squared and observations reported in column for model 2a.

Table 6

Estimates of Interaction Terms Between EARLY HEALTH PROBLEMS and Indicators of HEALTH CARE SYSTEMS on LIFETIME EARNINGS of Men in Europe

Mental × Full coverage	−265823**
	(118764)
Mental × Considerable cost sharing	−163892
	(173446)
Mental × Less than full coverage	−163885
	(122353)
Mental × Socialist	−296989***
	(108709)
Poor × Full coverage	−192
	(151048)
Poor × Considerable cost sharing	−183690
	(116369)
Poor × Less than full coverage	−428427***
	(107970)
Poor × Socialist	−214922**
	(100461)
Physical × Full coverage	−411359***
	(122544)
Physical × Considerable cost sharing	87180
	(302594)
Physical × Less than full coverage	−348379**
	(138345)
Physical × Socialist	49546
	(93610)
Had no usual source of care when between 0–15 years old	32220
	(76284)
Severe headaches or migraines	−17371
	(86424)
Rooms (an increase of one room)	35588
	(43075)
Number of books at home when 10 years old	23914
	(15551)
Main breadwinner's occupation when 10 (in ISCO-88 skill levels)	34705*
	(19389)
Number of facilities at home when 10 years old	44878***
	(11483)
Born in urban area	42494
	(37472)
Considerable cost sharing	−259675*
	(152662)
Less than full coverage	−680078***
	(180729)
Socialist	−532525***
	(158910)
Constant	1287569***
	(188076)
R-squared	0.076
Observations	9199

Note: OLS estimates from linear models on the lifetime earnings of men. Regression also includes year of birth and country dummy variables. Health care systems are dummy variables with Full Coverage as the reference system. Robust standard errors in parentheses. Significance levels: *** $p < 0.01$ ** $p < 0.05$ * $p < 0.10$.

country of residency using a set of dummy variables in these estimates in order to take into account basic differences in the economy, which are likely to also influence earnings. We acknowledge that there may be country-specific time varying factors that we are not able to take into account.

Our findings for the influence of a country's health care system on long-term earnings effects of an early health problem are clearest for those with early poor health. And in this case, the results match our hypothesis. Those with early poor health who lived in a country with less than full coverage face the greatest lifetime earnings penalty followed by those who grew up in a Socialist country with limited health care. In sharp contrast, those who grew up in a country with full coverage did not experience any penalty in terms of lifetime earnings (those with poor early health who grew up in a country with considerable cost sharing have a negative coefficient but it is not statistically significant). So, our results are consistent with the idea that more comprehensive health care coverage with little if any cost sharing would lead to more effective care for children, perhaps because parents were in a position to pursue care for their child with poor health without regard for direct costs of that care. Excluding those living in France and Greece, we have fully consistent results for those with early poor health (see table 9).

Our predictions for the tie between health care policy and the penalties for those with early mental health problems are far less precise, presumably because at the time most of these men were boys, the health care system did not really cover mental health problems or did so only for those institutionalized. Nevertheless, we noted in our results above that there were large average penalties in terms of labor market earnings for those with early mental health problems. Here our results suggest that those who grew up in Socialist countries with likely limited, if any, access to care for mental health problems indeed were the most penalized in terms of lifetime earnings on average. The next most heavily penalized were those in countries with full coverage, a somewhat surprising finding. Those with mental health problems as a child in the other systems (considerable cost sharing and less than full coverage) also show similar negative earnings penalties but the coefficients are not statistically significant. Still, the full set of results suggests that those with early mental health problems were penalized heavily in terms of lifetime earnings under all of these health care systems. This is not surprising given the very limited care for mental health problems under all four health systems at the time when these boys were growing up.[11]

Turning to early physical health problems, we noted earlier that on average those with an early physical health problem did not experience a statisti-

11 Again, our results excluding France and Greece in table 9 are fully consistent with these results.

cally significant penalty in lifetime earnings, as captured by the considerably smaller coefficients on early physical health problems than for mental health or poor early health problems. Still, we explored if there were differences by health care system, though the reader should view these with some skepticism. We find that for men who experienced physical health problems as a boy, the penalty is largest if they grew up in a system with full coverage followed by those who grew up in a system with less than full coverage. We find this first result surprising. Our explanation for this is that in the countries with full coverage, children were more likely to survive (the mortality rate for these countries is lower in the 1950s, 1960s, and 1970s across all ages of children [infants, 1–4, 5–9 and 10–14], suggesting that those who survived were more seriously ill on average than those in other countries included in our analysis [based on data from the Human Mortality Database]). We also note that there is no suggestion of any penalty for those with physical health problems as a boy on lifetime earnings for those who grew up under a Socialist system or a system with considerable cost sharing. Our thought on this is that our index may be too heterogeneous such that only some of the underlying diseases might be expected to have long-run implications for the individual. Our sensitivity estimates excluding France and Greece again show that those who grew up in a system of full coverage do worst, but these results suggest that those in a system with considerable cost sharing also do significantly worse in terms of lifetime earnings (table 9).

We conducted one other sensitivity test in which we interact only the experience of early poor health and of early physical health by our four categories of health care systems. We do this because at the time most of these men were less than age 16, there was little care available for mental health problems. These results, reported in table 10, are consistent with those discussed above.

6. Conclusions

In this paper, we explore the effect of three dimensions of early-life health problems on lifetime earnings. We use individual-level data from the first three waves of SHARE, a multidisciplinary and representative cross-national panel of the European population aged 50-plus. Waves 1 (2004/05) and 2 (2006/07) include information on sociodemographic background characteristics, current health, and socioeconomic status. Most of our data are from the third wave, SHARELIFE (2008/09), which is a retrospective survey conducted in 13 European countries as part of the SHARE project. We use the country of childhood for those in our sample and a four-way system to characterize their country's health care systems. These include full coverage, considerable use of copayments, limited coverage, and Socialist (full coverage but limited care).

Countries with health care systems with full coverage in our data are Sweden, Denmark, and United Kingdom. Those with considerable cost sharing include Austria, Belgium, (West) Germany, Finland, France, Norway, and Switzerland. Countries with less than full coverage include Greece, Italy, Netherlands, Portugal, and Spain, and those with full coverage but limited care (Socialist at the time our sample members were children), include Czechoslovakia, Hungary, East Germany, Russia, and Poland. We use this four-way classification to capture what we expected would be important determinants of early health care use—fullness of coverage and cost sharing requirements.

In terms of particular health problems, we find that the early experience of poor general health and mental health have stronger negative ties to lifetime earnings than physical health. Our results for health care policy find that for those who experienced poor health as a child, those growing up under Socialist health care systems, and those in systems characterized by less than full coverage experience the greatest earnings losses as an adult. In contrast, those with poor early health who grew up in a system characterized by full coverage do not experience any earnings loss. For those growing up with an early mental health problem, those living under a Socialist policy again experience the greatest loss of earnings over the life cycle. In this case, however, all those who had an early mental health problem experience sizeable penalties in terms of lifetime earnings.

Our current hypothesis tied to the pattern of penalties for those with early mental health conditions is that at the time these males were young there was little available care except for those with severe mental health problems, who were likely institutionalized. Our guess is that under-diagnosis or lack of treatment lies behind the results. We hope that in the future, as data become available, researchers will study the pattern for boys who grew up in the 1990s or later, when diagnosis and care for mental health conditions began to be more available.

Our policy conclusions are that early health conditions matter, and matter differentially, with poor general health and mental health mattering more to lifetime earnings than early physical health problems. The costs to these individuals in terms of earnings losses are large and deserving of resources. Second, regardless of type of health condition, those who grew up with full coverage tended to experience smaller lifetime earnings losses, with the exception of those with physical health problems who were more likely to survive if they grew up in a country providing full coverage. Health policy would seem to have long-run consequences.

7. Appendix

Table 7

Correlations Between EARLY HEALTH PROBLEMS in Europe by Indicators of Health Care Systems

Europe	Mental	Physical	Poor
Mental	1.00		
Physical	0.07	1.00	
Poor	0.13	0.12	1.00
Full coverage			
Mental	1.00		
Physical	0.13	1.00	
Poor	0.11	0.12	1.00
Considerable cost sharing			
Mental	1.00		
Physical	0.06	1.00	
Poor	0.17	0.18	1.00
Less than full coverage			
Mental	1.00		
Physical	0.04	1.00	
Poor	0.15	0.08	1.00
Socialist			
Mental	1.00		
Physical	0.03	1.00	
Poor	0.03	0.03	1.00

Table 8

Estimates of EARLY HEALTH PROBLEMS on LIFETIME EARNINGS of Men in Europe[a]

	(1)	(2)
Mental	−208452***	
	(79714)	
Physical[b]	−12199	−27056
	(42416)	(41894)
Poor	−228843***	
	(63115)	
R-squared	0.075	0.074
Observations	9199	9199

Note: [a] See note to table 4 for details on the specification, standard errors, and levels of statistical significance.
[b] A larger fraction of respondents are considered: all those who scored at least one standard deviation below the mean, which includes about 17 % of all respondents.

Table 9

Estimates of Interaction Terms Between EARLY HEALTH PROBLEMS and Indicators of HEALTH CARE SYSTEM on LIFETIME EARNINGS of Men in Europe

Mental × Full coverage	−266226**
	(118987)
Mental × Considerable cost sharing	−111035
	(131978)
Mental × Less than full coverage	−174948
	(129941)
Mental × Socialist	−314404***
	(109242)
Poor × Full coverage	11238
	(150230)
Poor × Considerable cost sharing	−35508
	(103246)
Poor × Less than full coverage	−429045***
	(105151)
Poor × Socialist	−235691**
	(97325)
Physical × Full coverage	−419154***
	(119127)
Physical × Considerable cost sharing	−216207**
	(92914)
Physical × Less than full coverage	−6863
	(155161)
Physical × Socialist	62860
	(95625)
Had no usual source of care when between 0–15 years old	27990
	(63786)
Severe headaches or migraines	−38054
	(56451)
Rooms (an increase of one room)	70580*
	(40013)
Number of books at home when 10 years old	42996***
	(13857)
Main breadwinner's occupation when 10 (in ISCO-88 skill levels)	11370
	(18957)
Number of facilities at home when 10 years old	30360***
	(9382)
Born in urban area	51185
	(35352)
Considerable cost sharing	−267618*
	(158289)
Less than full coverage	−661187***
	(168166)
Socialist	−535380***
	(162955)
Constant	1216676***
	(175391)
R-squared	0.115
Observations	7669

Note: Sensitivity analysis excluding respondents who spent most of their childhood in France or Greece. OLS estimates from linear models on the lifetime earnings of men. Regression also includes year of birth and country dummy variables. Health cares systems are dummy variables with Full Coverage as the reference system. Robust standard errors in parentheses. Significance levels: *** $p < 0.01$ ** $p < 0.05$ * $p < 0.10$.

Table 10

Estimates of Interaction Terms Between EARLY HEALTH PROBLEMS and Indicators of HEALTH CARE SYSTEM on LIFETIME EARNINGS of Men in Europe

Poor × Full coverage	−6240
	(150159)
Poor × Considerable cost sharing	−177422
	(114809)
Poor × Less than full coverage	−422919***
	(107045)
Poor × Socialist	−217252**
	(101287)
Physical × Full coverage	−418046***
	(121570)
Physical × Considerable cost sharing	88486
	(302242)
Physical × Less than full coverage	−347386**
	(138184)
Physical × Socialist	46979
	(94860)
Mental	−210405***
	(79512)
Had no usual source of care when between 0–15 years old	32227
	(76272)
Severe headaches or migraines	−18587
	(86264)
Rooms (an increase of one room)	35133
	(43053)
Number of books at home when 10 years old	23816
	(15549)
Main breadwinner's occupation when 10 (in ISCO-88 skill levels)	34721*
	(19383)
Number of facilities at home when 10 years old	44914***
	(11478)
Born in urban area	42507
	(37459)
Considerable cost sharing	−256588*
	(152193)
Less than full coverage	−677168***
	(180167)
Socialist	−531752***
	(158508)
Constant	1285105***
	(187820)
R-squared	0.076
Observations	9199

Note: Sensitivity analysis not interacting mental health problems. OLS estimates from linear models on the lifetime earnings of men. Regression also includes year of birth and country dummy variables. Health cares systems are dummy variables with Full Coverage as the reference system. Robust standard errors in parentheses. Significance levels: *** $p < 0.01$ ** $p < 0.05$ * $p < 0.10$.

References

Alessie, R., Angelini, V., and van Santen, P. (2013), Pension Wealth and Household Savings in Europe: Evidence from SHARELIFE, European Economic Review 63, 308–328.

Almond, D. (2006), Is the 1918 Influenza Pandemic Over? Long-Term Effects of in Utero Influenza Exposure in the Post-1940 U.S. Population, Journal of Political Economy 114, 672–712.

Almond, D., and Currie, J. (2011), Human Capital Development before Age Five, in: Ashenfelter, O., and Card, D. (Eds.), Handbook of Labor Economics, Vol. 4B, Elsevier, Amsterdam, 1315–1486.

Black, S. E., Devereux, P. J., and Salvanes, K. G. (2007), From the Cradle to the Labor Market? The Effect of Birth Weight on Adult Outcomes, Quarterly Journal of Economics 122, 409–439.

Brugiavini, A., Cavapozzi, D., Pasini, G., and Trevisan, E. (2013), Working Life Histories from SHARELIFE: A Retrospective Panel, SHARE Working Paper Series No. 11–13.

Case, A., D. Lubotsky, and C. Paxson (2002), Economic Status and Health in Childhood: The Origins of the Gradient, American Economic Review 92, 1308–1334.

Case, A., Fertig, A., and Paxson, C. (2005), The Lasting Impact of Childhood Health and Circumstance, Journal of Health Economics 24, 365–389.

Case, A., and Paxson, C. (2008), Stature and Status: Height, Ability, and Labor Market Outcomes, Journal of Political Economy 116, 499–532.

Cavapozzi, D., Garrouste, C., and Paccagnella, O. (2011), Childhood, Schooling and Income Inequality, in: Börsch-Supan, A., Brandt, M., Hank, K., and Schröder, M. (Eds.), The Individual and the Welfare State. Life Histories in Europe, Springer, Heidelberg, 31–43.

Chen, Y., and Zhou, L.-A. (2007), The Long-Term Health and Economic Consequences of the 1959–1961 Famine in China, Journal of Health Economics 26, 659–681.

Delaney, L., and Smith, J. P. (2012), Childhood Health: Trends and Consequences over the Life-Course, Future of Children 22, 43–63.

Fletcher, J. M. (2014), The Effects of Childhood ADHD on Adult Labor Market Outcomes, Health Economics 23, 159–181.

Flores, M., García-Gómez, P., and Kalwij, A. (2015), Early Life Circumstances and Life Cycle Labor Market Outcomes, Tinbergen Institute Discussion Paper No. 2015-094/V.

Flores, M., and Kalwij, A. (2014), The Associations Between Early Life Circumstances and Later Life Health and Employment in Europe, Empirical Economics 47, 1251–1282.

Goodman, A., Joyce, R., and Smith, J.P. (2011), The Long Shadow Cast by Childhood Physical and Mental Problems on Adult Life, Proceedings of the National Academy of Sciences 108, 6032–6037.

Haider, S., and Solon, G. (2006), Life-Cycle Variation in the Association Between Current and Lifetime Earnings, American Economic Review 96, 1308–1320.

Human Mortality Database, University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany). Available at www.mortality.org or www.humanmortality.de (Access Date: 2017-11-15).

Kalwij, A., and Vermeulen, F. (2008), Health and Labor Force Participation of Older People in Europe: What Do Objective Health Indicators Add to the Analysis?, Health Economics 17, 619–638.

Nelson, R.E. (2010), Testing the Fetal Origins Hypothesis in a Developing Country: Evidence from the 1918 Influenza Pandemic, Health Economics 19, 1181–1192.

Sandier S, Paris V, and Polton D. (2004), Health Care Systems in Transition: France, WHO Regional Office for Europe on Behalf of the European Observatory on Health Systems and Policies, Copenhagen.

Smith, J. P. (2009a), The Impact of Childhood Health on Adult Labor Market Outcomes, Review of Economics and Statistics 91, 478–489.

Smith, J. P. (2009b), Reconstructing Childhood Health Histories, Demography 46, 387–403.

Smith, J. P., and Smith, G. C. (2010), Long-Term Economic Costs of Psychological Problems During Childhood, Social Science and Medicine 71, 110–115.

UNESCO (1997), International Standard Classification of Education: ISCED1997, Paris: UNESCO.

US Social Security Administration, Office of Policy (2002), Social Security Programs Throughout the World: Europe.

World Health Organization (1996), Health Care Systems in Transition: Greece, WHO Regional Office for Europe, Copenhagen, http://www.euro.who.int/__data/assets/pdf_file/0020/120278/E72454.pdf (Access Date: 2017-11-15).

Real Firms in Tax Systems

Joel Slemrod and William C. Boning*

Received 17 August 2017; in revised form 03 November 2017; accepted 09 December 2017

Economic analysis of taxation often assumes a homogeneous, usually perfectly competitive production sector in which individual firms are immaterial. This paper discusses some recent developments bringing key characteristics of real firms into the analysis of tax systems, which include enforcement rules and remittance regimes alongside tax rates and bases. Introducing more realistic firms into the analysis of tax systems has enabled progress in understanding the role of information in tax administration, the tradeoff between production efficiency and minimizing the administrative costs of tax collection, the consequences of remittance responsibility, and the fundamental role of firm heterogeneity in tax incidence.

Keywords: tax systems, heterogeneous firms, tax incidence

JEL classification: H 26, H 21, H 22

1. Introduction

Firms are the workhorses of modern tax systems. Firms remit between 85 and 90 % of all tax revenue in most OECD countries and in India (OECD, 2017; Slemrod and Velayudhan, 2017). Tax authorities rely on firms to provide information about other businesses, employees, owners, and customers. The treatment of firms is thus an important ingredient in the analysis of tax systems, which consist of tax rates and bases and the accompanying remittance and enforcement rules. Tax systems analysis should not treat firms as immaterial, exemplified by the homogeneous, constant-returns-to-scale, perfectly competitive production sector in the seminal model of Diamond and Mirrlees (1971), but instead should capture the diversity of real-world firms that underpins both the technology of tax administration and the tradeoffs tax systems face between production efficiency and obtaining tax revenue.

* Slemrod: University of Michigan, 611 Tappan Avenue, Ann Arbor 48109, USA (jslemrod@umich.edu). Boning: University of Michigan, 611 Tappan Avenue, Ann Arbor 48109, USA (wcboning@umich.edu). This essay is based on a keynote address given by Slemrod at the Workshop on Public Economics of the 2017 RIDGE May Forum in Montevideo, Uruguay on May 26, 2017.

FinanzArchiv 74 (2018), 131–143 doi: 10.1628/fa-2018-0002
ISSN 0015-2218

Incomplete information is fundamental to many of the challenges tax administrations face, and real firms vary widely both in their ability to conceal information from tax authorities and in their ability to provide information about other firms and people to tax authorities. From insular family firms to platform companies with a near-omniscient view of a market, real firms are far from homogeneous information environments. Informal firms, concealing even their existence from tax authorities, are rarely if ever large, although it is far from obvious how best to measure the relevant dimension of firm size. Large corporations employ tax advisors who guide precisely how to characterize information in disclosures to tax authorities to manage their tax liability. Perfect-information models cannot address many important tax-systems concerns, among them detecting income and evasion and the use of firms as wage-reporting agents.

Efficient administrative technologies–the tax bases and remittance and enforcement rules that yield the greatest tax revenue at the lowest administrative cost–often come at a cost in production efficiency, contra Diamond and Mirrlees (1971), precisely because of certain differences between real firms. Real firms differ in both productivity and evasion technology, with the result that changes in remittance rules can alter tax incidence and production efficiency. It is often costlier to obtain information about tax liability and tax payments from smaller firms, leading to, for example, VAT exemption thresholds. In what follows, we address several issues that arise when considering the role of real firms in tax systems. We begin by defining the key terms comprising the essay's title.

1.1. Tax Systems

A tax system is a set of rules, regulations, and procedures with three aspects. It defines what events or states of the world trigger tax liability, and what rate of tax applies; these are *tax bases and rates*. It specifies who or what entity must remit that tax and when: these are *remittance rules*. It details procedures for ensuring compliance, including audit coverage and third-party information reporting requirements as well as the consequences of not remitting legal liability: these are *enforcement rules*. Until recently, tax economists have focused overwhelmingly on the first aspect-the positive and normative aspects of the choice of tax bases and rates. Although this paper focuses on remittance rules and enforcement rules, these rules do interact with tax rates and bases; for example, the impact of tax rates on behavior depends on the efficacy of the enforcement regime.

Although an analysis of real firms in tax systems raises many issues that standard analysis largely ignores, it shares several features with standard analysis. Firms are (merely) instrumental in normative analysis: individuals are the ultimate concern of welfare analysis, and are affected by taxation in their roles as consumers, workers, entrepreneurs, capital providers, and so on. A natural starting point would be to keep a representative consumer and a focus on efficiency, although this approach cannot address some material issues. In particular, analysis with a representative individual sets aside any distributional effects of firm taxation, notably including the popular notion that conflates taxing larger businesses with assigning the incidence of taxation to wealthier people.

1.2. Real Firms

In several ways that matter for tax systems, particularly but not solely remittance and enforcement regimes, actual firms depart from the idealized characterization that has been the standard treatment in, especially, optimal taxation. Some real firms are larger than others, of course, and the technology of tax administration differs greatly between large and small firms, as does the technology of tax avoidance and evasion. Avoidance and evasion technologies also differ across sectors; for example, the presence of intangible assets provides greater opportunities for avoidance.

The word "real" also connotes behavioral responses to taxation that materially alter production, such as firms changing their volume of investment, input choice or the location of production. Firm behavioral responses to taxation are both broader and more interrelated than this: most real firms (legally) avoid taxes, and many also seek to (illegally) evade taxes, through a combination of decisions that may or may not directly affect production. Moreover, firms can respond to taxation using means that are not applicable to individuals. Real firms can break up, or merge, in ways that individuals cannot. They can proliferate to obfuscate true ownership and control by creating "shell" corporations.

2. Real Firms Shape the Availability of Information in Tax Systems

If costlessly provided with perfect information, tax authorities would no longer need to devote considerable resources to the difficult task of ascertaining tax liability and could instead concentrate on enforcing the remittance of tax payments. Conversely, raising government revenue with no information whatsoever would require *capricious* assignment of tax liability. The efficiency costs of capricious taxation are hard to quantify, and evaluating the equity implications would require incorporating differences between people and would not be straightforward given the lack of an accepted framework for addressing horizontal inequity. The social benefits of informed tax administration remain unclear, if likely substantial given the expenditures made to obtain information and facilitate taxation on bases that are not capricious.

Employers pervasively report information about employee wages and salaries to tax authorities, as well as information about dividends, interest, share sales and real estate sales. Information reporting is also built into most invoice-credit value-added tax systems, when credits for purchases from other firms are allowed only if accompanied by information on the seller–confirming that they remitted tax on their sales–which can in principle be checked against the VAT returns of those sellers; this "self-enforcing" aspect of VATs does not apply to the final sale in the value-added chain. Recent evidence corroborates the importance of the VAT's information reporting regime. For example, Pomeranz (2015) shows that, upon mailing increased-audit-threat letters to Chilean firms, the increase in VAT receipts (and therefore the inferred level of previous evasion) induced by the letters is concentrated at the level of sales from firms to final consumers, for which there is no paper trail. Consumers do serve as information reporters in the São Paulo VAT regime Naritomi (2016) studies. Business income tax systems struggle to monitor business expenses, which could in principle be addressed by requiring that businesses report the identity of their customers and input providers, as they do under a VAT, and having the tax authority link and monitor these reports.

2.1. Informal Firms and the Self-Employed

Modern tax systems rely heavily not on all types of firms, but specifically on medium-sized and large firms, because efficiently collecting tax from small firms and the self-employed is ubiquitously problematic. Theory provides several possible explanations, but not a clear way to adjudicate between them. Tax systems struggle to implement third-party information reporting of non-employee income, as information reporting often benefits from opposition between the interests of two parties, which is absent for self-employment income. Small businesses face fewer potential arm's-length or employee whistleblowers, an argument formalized by Kopczuk and Slemrod (2006) and Kleven et al. (2016). Small businesses are also less reliant on the financial system, and Gordon and Li (2009) suggest that this provides them with less reason to keep records that the tax authority could use in an audit.

A mountain of individual- and firm-level evidence using multiple methodologies, surveyed in Slemrod (2017), documents a strong association between self-employment and noncompliance and between self-employment and the "flexibility" of reported taxable income locally to kinks and notches in tax schedules. Kleven (2014) plots for over 80 countries the fraction of workers who are self-employed against the tax/GDP ratio, and documents a strong negative relationship: countries that have more self-employed collect less tax. Although no causal inferences can be drawn from such a graph, it seems clear that the availability of third-party information on employee income provided by employers plays a key role in tax compliance and in explaining a country's overall tax take. Consistent with this conclusion, Jensen (2016) shows that, as countries develop, their employment structures shift away from self-employment, and then exemption thresholds for income tax liability fall, a pattern that is consistent with tax authorities setting the threshold at a level that justifies enforcement costs.

2.2. Family Firms

Family firms provide a fascinating special case of several of the phenomena already discussed. Bertrand and Schoar (2006) note the pervasiveness of family firms around the world, and analyze cross-country data to inform why they are so prevalent without considering the possible role of taxation. Kopczuk and Slemrod (2011) sketch a model of the taxation of family firms, stressing that in some developing countries the weakness of legal institutions encourages the formation of family firms, whose bonds provide an informal means to discourage employee theft and misbehavior. While family ties are beneficial replacements for weak legal institutions, these bonds have a social cost because they increase the opacity of firms, making tax enforcement more difficult. The same threat of family ostracism that constrains theft also inhibits the kind of whistleblowing that aids tax enforcement. This calls into question whether family firms should receive the tax preferences they often do, perhaps because it is difficult to determine and tax the labor income attributable to family members, and raises the issue of whether optimal enforcement policy ought to take into account whether a business is a family firm.

One piece of evidence suggests that, in at least one setting, family firms are relatively less tax-aggressive, which is not the same as tax noncompliant. Using data from S&P 1500 firms in the period 1996–2000, Chen et al. (2010) show that family firms exhibit lower tax aggressiveness than their non-family counterparts, as measured by their having higher effective tax rates and lower book-tax differences. This could be due to family owners being willing to forgo tax benefits to avoid the non-tax cost that might arise from minority shareholders' concern with family rent-seeking masked by tax avoidance activities. Alternatively, family owners may be more concerned with the potential penalty and reputation damage from an IRS audit than non-family firms. Not explored is whether this finding applies to tax evasion in addition to aggressive tax avoidance. Nor is it clear that this finding for the largest family-owned firms applies to the much larger population of small family firms, where the opportunities family structure provides for evasion are larger.

2.3. The Platform Economy

The rise of the platform economy, in which businesses provide web-based platforms to facilitate transactions between individual buyers and sellers, has intriguing implications for the tax system. Major platforms intermediate borrowing goods, loaning money, performing tasks, and selling places to sleep and car rides. Tax rules are beginning to adapt to the growing numbers of people using platforms not only as customers but as small-scale sellers, effectively functioning as tiny firms, by exempting such activities below a threshold amount from taxation or reporting rules. For example, the U.K. has a Rent a Room Scheme that permits £7,500 per year of rental revenue tax-free, and in the U.S. there is a 15-day threshold below which rental income from a property need not be reported as business income, reducing the information required.

What role should platform providers play in tax systems? Their role as transaction intermediaries equips them to remit taxes on large numbers of transactions at low cost. Platform providers are likewise well-situated to provide information reports to tax authorities. The novel relationships between platform providers and the buyers and sellers using the platform also pose challenges for the rules in existing tax systems. As of April 2017, the online retailer Amazon remits sales taxes on orders shipped to all 45 U.S. states that have a state sales tax and to Washington, D.C.; before these agreements the consumer was responsible for remitting the "use tax" levied at the same rate as sales tax on out-of-state purchases, for which compliance was universally assumed to be abysmally low. Airbnb remits the hotel and occupancy taxes due on the short-term rentals and accommodations it provides in 275 jurisdictions in the United States and France; absent these agreements, the property owner was responsible for remitting these taxes, and it was widely suspected that compliance was minimal. Wilking (2016) finds that the change in remittance responsibility increased tax-inclusive rental prices, suggesting that consumers bear a greater share of the tax burden when the remittance obligation is shifted to a party with fewer evasion opportunities. A primary tax difference under the U.S. income tax system between classifying drivers for the ride-sharing service Uber as employees or as independent contractors (as is the current interpretation of their relationship to Uber) is that only in the former case would Uber be responsible for withholding (i.e., remitting) an approximation of the income tax liability the driver incurs. The growth of the platform economy thus both enables new remittance and enforcement regimes and poses challenges to existing regimes.

3. The Tradeoff Between Production Efficiency and Tax Administration Cost Minimization

The venerable result that even a distortionary tax should not interfere with production efficiency, due to Diamond and Mirrlees (1971), holds only under certain assumptions about how real firms and tax systems interact. For example, Diamond and Mirrlees (1971) note that if some production sectors, in their example agriculture, are "untaxable", taxation should take place at the boundary between the rest of the production sector and the untaxable sector, rather than at the boundary between the untaxable sector and consumption. Diamond and Mirrlees assume a constant-returns-to-scale setting from which real firms are entirely absent. The result that production efficiency should be preserved fails, for example, when there are fixed per-firm administrative costs, as Dharmapala et al. (2011) show. Best et al. (2015) show theoretically that, in the presence of evasion, the optimal tax base sacrifices some production efficiency in order to curtail evasion levels. Exploiting a size-based policy notch in Pakistan, they estimate that the switch from a profit tax to a turnover tax reduces evasion levels by up to 60 to 70 % of corporate income. It should be possible to derive more general conditions on the nature of administrative costs and the structure of production that clarify when production efficiency should and should not be disturbed.

3.1. Firm Size

A growing but incomplete literature studies the role firm size plays in tax systems. To the extent that firm size predicts compliance behavior, optimal tax systems may well treat firms differently based on their size, even though this may distort the structure of production. Several aspects of real-world tax systems provide firms with incentives to change their size; for example, the cascading nature of a gross receipts tax without exemptions for business-to-business sales provides incentives to vertically integrate (while a pure VAT, for example, does not). In this case, tax liability depends on where economic activity takes place relative to the border between firms.

One particular example has received substantial attention in the literature: given fixed per-firm costs of tax enforcement and in particular of conducting an audit, it may make sense to exempt small firms from the tax net altogether, even in light of the incentives this gives firms to stay small and the distortion induced by taxing some firms in an industry and not others. Dharmapala et al. (2011) lay out the logic of business income tax administration with an exemption threshold. Keen and Mintz (2004) develop a simple rule characterizing the optimal threshold (when firms' sizes are fixed) for a VAT in terms of a trade-off between tax revenues and collection costs and then consider the implications for the optimal threshold of the production inefficiencies implied by the differential treatment of those above and below the threshold. Liu and Lockwood (2016) study behavior at one such threshold in the United Kingdom, finding suggestive evidence that firms bunch below the VAT threshold by underreporting sales.

Taxes can be collected at lower cost when the tax authority can make use of information generated (and reported) by arms-length transactions between firms and between firms and employees. Firm boundaries alter the nature and extent of these information flows, and so firm size may produce externalities for purposes of tax administration. All this suggests that, in the presence of taxes, the equilibrium distribution of firm size need not be optimal, contrary to the suggestion of Coase (1937). Sole proprietorships and family and other small businesses are difficult for the tax authority to penetrate, for example, and replacing them with larger firms could be desirable on tax administration grounds even if it results in production inefficiency, contrary to the classic result of Diamond and Mirrlees (1971). Larger firms are not a panacea for tax collection, however, because while they are easier to detect and monitor, they are also better able to take advantage of returns to scale in tax avoidance. Multinational firms in particular have access to a variety of tools they can use both to avoid tax and to reduce the transparency of their operations to tax authorities. Further research could more precisely trace out the shape of the relationship between firm size and administrative efficiency.

4. Real, Heterogeneous Firms, Remittance Responsibility, and Tax Incidence

As already mentioned, firms are the linchpin of modern tax systems' remittance and collection regimes, remitting over 85 percent of taxes in most countries for which data are available. This fact may seem irrelevant given the common notion that who or what entity remits taxes does not matter. Although it is commonly asserted in as irrefutable truth in undergraduate public-finance textbooks, the assertion that remittance is irrelevant is certainly just folk wisdom, in the sense that it is rarely posed formally, laying out the assumptions required and addressing what happens when these conditions do not hold. This remittance-is-irrelevant folk wisdom does not accord well to a world of real firms.

There are several ways to show that the remittance-is-irrelevant folk wisdom does not hold in general. Its truth in fact rests on severe assumptions, including that evasion opportunities do not depend on who remits. Empirical studies confirm that, in contrast, the remittance regime can matter. Brockmeyer and Hernandez (2017) show that doubling the rate of withholding by credit-card companies in Costa Rica increased sales tax collection from those subject to withholding by 33 percent, even as the information reported to the tax authority did not change. Kopczuk et al. (2016) present empirical evidence that the identity of the remitting party in the U.S. diesel fuel market affects both the revenue collected and the pass-through of taxes; retail diesel prices are higher, and a larger fraction of diesel taxes are passed through to retail prices, in states where the point of collection is at the distributor or prime supplier level rather than at the retail level, suggesting that this collection regime reduces evasion and alters the incidence of the tax.

The prevalence of particular remittance patterns is also at odds with the folk wisdom. As stated above, firms remit most tax. Among consumption taxes, value-added taxes are far more common than retail sales taxes, although the substantive difference between the two is in which firms must remit–the remittance regime. Tax systems rarely, if ever, feature individuals remitting taxes in their role as consumers. When consumers do feature, the system is designed so that they remit minimal or negative taxes. The average individual receives a refund, for example, in the U.S. income tax system, and consumers receive lottery tickets for providing information to the tax authority in the VAT system Naritomi (2016) documents in São Paolo.

The remittance-is-irrelevant folk wisdom has extreme implications: not only does it not matter if the buyer or seller remits, it doesn't matter if anyone else remits a given tax liability, as long as there is a sufficiently thick web of connections among firms and people. To the extent that the implications of the extended model are patently false, the model must be flawed. This exercise would be in the spirit of Bernheim and Bagwell (1988), who cast doubt on the dynastic model underlying Ricardian equivalence by showing that carrying the model to its logical extreme implies that everything is neutral– including the irrelevance of all public redistributions, distortionary taxes, and prices.

Surprisingly little research attention has been devoted to compliance by firms in their role as withholders for taxes "on" workers. A recent exception is the randomized experiment Boning et al. (2017) study, in which U.S. employers received an in-person visit or letter intended to increase timely compliance with their income tax withholding and payroll tax obligations. Both treatments increased compliance, and the in-person visit also increased compliance among firms connected to a visited firm by a shared tax preparer. In the U.S., there is a special, and especially stringent, penalty regime for employer remittances of payroll and income taxes, which are called trust fund taxes because legally the firm holds the employee's money in trust until it makes a federal tax deposit. Noncompliance can trigger a 100 % "trust fund recovery penalty" that pierces the corporate veil, and can be levied on any person who has the duty to perform and the power to direct the collecting, accounting, and payment of trust fund taxes, including but not restricted to officers or an employee of a corporation as well as a corporate director or shareholder. Whether this qualitatively different penalty feature has a qualitatively different deterrent effect than standard penalties for, e.g., corporate income tax noncompliance, is not known. In principle, reports from a firm's employees could be matched to the reports of the withholding employers to monitor their payments. This is, however, likely less effective than using employer reports to verify payments received by workers. Some employees may have income below the threshold at which filing an income tax return becomes mandatory, and thus may not provide information about their employer. More generally, discrepancies between a firm's total report and the sum of its employees' reports is a weaker sign of noncompliance than disagreement between one employee's self-report and the firm's report about that employee.

Finally, workers and firms, especially small firms, may in some situations collude to facilitate evasion, as was explored theoretically by Yaniv (1993). Best (2014) finds that firms in Pakistan aggregate the preferences of workers and facilitate tax avoidance (not evasion) by bunching salary offers around kinks in the tax schedule. If and under what circumstances such collusion facilitates evasion in practice is worth exploring. One setting in which this is suspected is firms' reclassifying workers as independent contractors rather than employees. As discussed above, this eliminates the firm's responsibility to withhold and remit tax. This makes it more costly to enforce the worker's compliance because remittance responsibility is not irrelevant, and lower compliance can reduce the firm's labor costs.

How should an optimal tax system strike a balance between keeping firms formal so that they provide tax revenue and information reports and at the same time maximizing the value of the taxes and information reports provided? Again, a model of the social benefits of information reporting is needed.

4.1. Heterogeneous Firms and Tax Incidence

The differences between real firms fundamentally change the analysis of tax incidence, especially the analysis of the incidence of tax evasion. For example, Suárez Serrato and Zidar (2016) find that the incidence of state taxes diverges from the predictions of standard open-economy models when firms' location-specific productivities are allowed to differ, and so some firms' location decisions are inframarginal. In general, many real firms earn economic profits, and taking that possibility into consideration alters incidence calculations.

If one took the standard model of commodity taxation literally, or at least naively, one might think that consumers are responsible for remitting the tax due. Then, if retailers charge the same price for a given good to everyone, those consumers who successfully evade the tax due end up relatively better off. The amount of aggregate evasion, if it affects purchasing decisions, would increase demand and the consumer price depending on the relative supply and demand elasticities, so that non-evaders may bear an additional burden due to others' evasion.

However, in practice consumption taxes are almost always remitted by businesses, and as discussed above real businesses differ in their ability to evade taxes. As Kopczuk et al. (2016) demonstrate in the case of diesel taxes, tax evasion opportunities differ depending on which firms in the supply chain are required to remit the tax. Moving remittance responsibility to a relatively small number of upstream producers can make it easier to monitor tax compliance.

An analogy can be drawn between settings in which firms differ in production technologies and settings in which firms differ in evasion technologies. Tax incentives to use one production technology instead of another parallel the incentives remittance responsibility provides to evade taxes. In either case, production distortions result. A substantial extension to any incidence framework is needed to account for differential evasion technology across firms.

5. Conclusion

Although modern tax systems rely overwhelmingly on firms to collect revenue, tax economists have only recently begun to explore the implications of this empirical reality. Several interesting directions of research in tax systems stem from replacing the representative firm with real firms. The costs tax authorities incur to obtain information and the benefits that information provides both stem from the varied information environments real firms provide. Differences between firms make administrative policies that vary across firms according to their size cost-effective. Some firms are better positioned to evade than others, and both administrative policy and evasion technology can provide a competitive advantage to firms that need not be the most productive. Remittance responsibility has consequences when firms are realistic, and differences between firms can be essential for incidence.

More work is needed on several fronts. A solid theory of the social benefit of information to the tax authority would help to quantify the efficiency and equity improvements that result from better information. Theory can also help to explain which differences between firms are material and why, though empirical work will be needed to distinguish between the competing explanations. More realistic theories of firms will enable researchers to better evaluate the optimal design of firm-specific policies, like VAT thresholds, that tax authorities have adopted as a matter of practical necessity. As firms are central to the remittance systems of OECD countries and of India (Slemrod and Velayudhan, 2017), while countries may differ in their ability to raise revenue, what revenue they do raise is mostly obtained via firms. Future research can help to illuminate the relationships between real firms and differences in fiscal capacity. Empirical work can also broaden our understanding of which features of real firms are essential when considering information reporting, remittance responsibility, and incidence. Fortunately, administrative microdata on firms are becoming more widely available for research purposes, and these data enable researchers to study real firms in ever greater detail.

References

Bernheim, B. D., and Bagwell, K. (1988), Is Everything Neutral?, Journal of Political Economy 96, 308–338.

Bertrand, M., and Schoar, A. (2006), The Role of Family in Family Firms, Journal of Economic Perspectives 20, 73–96.

Best, M. C. (2014), Salary Misreporting and the Role of Firms in Workers' Responses to Taxes: Evidence from Pakistan, London School of Economics, mimeo.

Best, M. C., Brockmeyer, A., Kleven, H. J., Spinnewijn, J., and Waseem, M. (2015), Production versus Revenue Efficiency with Limited Tax Capacity: Theory and Evidence from Pakistan, Journal of Political Economy 123, 1311–1355.

Boning, W. C., Guyton, J., Hodge, R., Slemrod, J., and Troiano, U. (2017), Heard It Through the Grapevine: Direct and Network Effects of a Tax Enforcement Field Experiment, University of Michigan, mimeo.

Brockmeyer, A., and Hernandez, M. (2017), Taxation, Information, and Withholding: Evidence from Costa Rica, World Bank, mimeo.

Chen, S., Chen, X., Cheng, Q., and Shevlin, T. (2010), Are Family Firms More Tax Aggressive than Non-Family Firms?, Journal of Financial Economics 95, 41–61.

Coase, R. (1937), The Nature of the Firm, Economica 4, 386–405.

Dharmapala, D., Slemrod, J., and Wilson, J.D. (2011), Tax Policy and the Missing Middle: Optimal Tax Remittance with Firm-Level Administrative Costs, Journal of Public Economics 95, 1036–1047.

Diamond, P. A., and Mirrlees, J.A. (1971), Optimal Taxation and Public Production I: Production Efficiency, American Economic Review 61, 8–27.

Gordon, R., and Li, W. (2009), Tax Structures in Developing Countries: Many Puzzles and a Possible Explanation, Journal of Public Economics 93, 855–866.

Jensen, A. (2016), Employment Structure and the Rise of the Modern Tax System, London School of Economics, mimeo.

Keen, M., and Mintz, J. (2004), The Optimal Threshold for a Value-Added Tax, Journal of Public Economics 88, 559–576.

Kleven, H. J. (2014), How Can Scandinavians Tax So Much?, Journal of Economic Perspectives 28, 77–98.

Kleven, H. J., Kreiner, C.T., and Saez, E. (2016), Why Can Modern Governments Tax So Much? An Agency Model of Firms as Fiscal Intermediaries, Economica 83, 219–246.

Kopczuk, W., Marion, J., Muehlegger, E., and Slemrod J. (2016), Does Tax-Collection Invariance Hold? Evasion and the Pass-through of State Diesel Taxes, American Economic Journal: Economic Policy 8, 251–286.

Kopczuk, W., and Slemrod, J. (2006), Putting Firms into Optimal Tax Theory, American Economic Review 96, 130–134.

Kopczuk, W., and Slemrod, J. (2011), Taxation and Family Firms, University of Michigan, mimeo.

Liu, L., and Lockwood, B. (2016), VAT Notches, Voluntary Registration and Bunching: Theory and UK Evidence, Oxford University Centre for Business Taxation, mimeo.

Naritomi, J. (2016), Consumers as Tax Auditors, London School of Economics, mimeo.

OECD (2017), Legal Tax Liability, Legal Remittance Responsibility & Tax Incidence: Three Dimensions of Business Taxation, Committee on Fiscal Affairs Centre for Tax Policy and Administration, mimeo.

Pomeranz, D. (2015), No Taxation without Information: Deterrence and Self-Enforcement in the Value Added Tax, American Economic Review 105, 2539–2569.

Slemrod, J. (2017), Tax Compliance and Enforcement, University of Michigan, mimeo.

Slemrod, J., and Velayudhan, T. (2017), Do Firms Remit At Least 85 % of Tax Everywhere? New Evidence from India, University of Michigan, mimeo.

Suárez Serrato, J. C., and Zidar, O. (2016), Who Benefits from State Corporate Tax Cuts? A Local Labor Markets Approach with Heterogeneous Firms, American Economic Review, 106, 2582–2624.

Wilking, E. (2016), Hotel Tax Incidence with Heterogeneous Firm Evasion: Evidence from Airbnb Remittance Agreements, University of Michigan, mimeo.

Yaniv, G. (1993), Collaborated Employer-Employee Tax Evasion, Public Finance/Finances Publiques 47, 312–321.

Toward Understanding the Relationship of Temporal Changes in Demographic Structure to Changes in U.S. Poverty

Robert Haveman and Kathryn Wilson*

Received 02 September 2017; in revised form 26 September 2017; accepted 04 October 2017

This paper attempts to quantify how changes in demographic trends have affected the poverty rate in the United States since the start of the "War on Poverty" in the 1960's. The analysis uses both the official Census poverty definition and a supplemental poverty measure that better captures both the resources available to families and their expenditure needs. Using regression estimates to construct a counterfactual, our results reveal that, while some demographic change increase poverty and others decrease poverty, the net effect of the changes in the demographic structure of the U.S. population was to reduce both of these two measures of poverty.

Keywords: poverty, demographic change, labor market change

JEL classification: I 32, J 11

1. Introduction

In 1964, President Johnson declared a "War on Poverty". Many things have changed since the 1960's, including the demographics of the population, the structure of the labor market, and changes in public policy, all of which affect the poverty rate. In this paper, we attempt to quantify the effect of changes in the demographic structure of the population on changes in two measures of poverty over the 1968–2012 period—the Official Poverty Rate (OPM) and the Supplemental Poverty Rate (SPM).

In this paper, we use micro-data from the Panel Study of Income Dynamics to explore the effects of the changing pattern of the nation's demographic structure on both the OPM and the SPM. Our results reveal that, while some demographic changes increase poverty and others decrease poverty, the net effect of the changes in the demographic structure of the U.S. population reduces both of these measures of poverty.

* Haveman: University of Wisconsin-Madison 1225 Observatory Drive Madison, WI 53706-1211, USA (haveman@lafollette.wisc.edu); Wilson: Kent State University, PO Box 5190, Kent, OH 44242, USA (kwilson3@kent.edu).

FinanzArchiv 74 (2018), 144–157
ISSN 0015-2218
doi: 10.1628/001522118X15123891774237

Understanding the evolution of poverty in the nation is a crucial issue both from intellectual and policy perspectives. Many studies have speculated as to the causes of the rises and falls of the nation's poverty rate, and our findings provide quantitative evidence of the role of one of the most important of these causes. Policymakers will find these results helpful in that they suggest anti-poverty policy interventions to offset these demographic changes. For example, measures to reverse the growth of single parent families could offset the negative effect of this change on the poverty rate.

In section 2, we present the prior literature that precedes this study. In section 3, we define the OPM and SPM poverty measures. In section 4, we examine descriptive statistics about the changes over the 1968–2012 period in both demographic variables and the poverty rate for each demographic characteristic. Section 5 presents regression results for these poverty rates. Section 6 presents a counterfactual used to calculate the changes in the poverty rates associated with changes in the demographic structure. Section 7 concludes.

2. Prior Literature

At the fiftieth anniversary of President Johnson's declaration, several reviews and assessments have been made of the War on Poverty (for example, Bailey and Danzinger, 2013; President's Council of Economic Advisors, 2014; Haveman, et al., 2014). Many of these studies have sought to understand the difference in poverty trends using both the OPM and a version of the SPM.

Fox et al. (2015) compares poverty using a supplemental poverty measure, which includes taxes and in-kind government transfers, to poverty measured by the official poverty measure. Examining trends in poverty over from 1967 to 2012 using both measures, they find the trends in poverty with the supplemental measure have been more favorable than the trends using the official poverty measure; they interpret this as suggesting that public policy has been more effective in reducing poverty than official poverty measures capture. In particular, they find government programs play a particularly large and growing role in reducing childhood poverty. They do not explore the effect of demographic changes on either of these poverty measures.

Wimer et al. (2016) is an extension of this work, using an anchored supplemental poverty measure to look at historical trends in poverty since 1967. Although the official poverty measure has remained relatively flat over the decades, Wimer et al. show that the poverty rate with the post-tax/post-transfer anchored supplemental poverty measure has fallen by more than 40 percent during the past fifty years. They conclude that "government policies, not market incomes, are driving the declines observed over time" (p. 1207).

This paper takes a different approach in answering the same question as Fox et al. (2015) and Wimer et al. (2016). By incorporating regression analysis and presenting counterfactuals for two measures of poverty in 1968 and 2012, we are able to estimate the effect demographic changes on two U.S. poverty rates. The analysis uncovers both the changes in demographic characteristics and the changes in how these characteristics are associated with poverty since the start of the war on poverty.

3. Poverty Measures

As we have indicated above, our analysis uses two measures of the poverty rate, the Official Poverty Rate and the Supplemental Poverty Rate. The Official Poverty Rate (OPM) is the most well-known of the poverty rates that we use. The U.S. Census Bureau determines family poverty status by comparing pre-tax cash income with a threshold that is set at three times the cost of a minimum food diet in 1963, updated annually for inflation using the Consumer Price Index, and adjusted for family size, composition, and age of householder.[1] Hence, the OPM is insensitive to the fact that non-food expenditures may increase more than proportionally. While the official poverty rate is relatively easy to measure, it faces criticism for not accurately reflecting the income and needs of families. For example, since the measure is pre-tax cash income it does not include many in-kind and tax-based transfer programs such as food stamps and the Earned Income Tax Credit. Similarly, the needs of a family do not reflect the changing composition of what families buy, such as the increase in child care and work costs as more mothers are in the labor market.

The Supplemental Poverty Rate (SPM) differs in several ways from the OPM. The SPM begins with family pre-tax cash income but also takes into account the dollar value of in-kind benefit programs (e.g. the Food Stamp program) and benefits conveyed through the tax system (e.g. the Earned Income Tax Credit) in the resource measure. The SPM also deducts estimates of FICA taxes (to support Social Security and Medicare), work-related expenses, net federal income tax and net state income tax (including the federal and state Earned Income Tax Credit) and out-of-pocket health-care costs. The SPM poverty thresholds are based on expenditures on food, housing, and clothing (rather than just food) and are adjusted over time as the composition of expenditures changes; hence the SPM is a quasi-relative poverty measure.

1 "Family" is defined by the official poverty measure as persons living together who are related by birth, marriage, or adoption. The thresholds do not vary geographically. This description is drawn from the web site of the Institute for Research on Poverty: http://www.irp.wisc.edu/faqs/faq2.htm

Differences in housing costs by type of housing (own home with no mortgage, own home with mortgage, and rent) and an improved equivalence scale are also used to determine the thresholds for different types of families.[2] For many, the SPM provides a more reliable national poverty measure than does the OPM.

The following table shows the primary differences between the OPM and the SPM.

Table 1

Poverty Measure Concepts: OPM and SPM

Concept	Official poverty measure (OPM)	Supplemental Poverty Measure (SPM)
Household Unit definition	Conventional definition: Families and unrelated individuals	Broadened definition: All related individuals who live at the same address, including any cohabiters and their relatives and foster children
Resource measure	Before-tax cash income	Cash income *plus* noncash transfers (such as food stamps and housing subsidies) and refundable tax credits *minus* income and payroll taxes, medical out-of-pocket expenses, and work expenses (includes childcare expenses)
Threshold level for base two-adult/two-child unit	Three times the cost of a minimum food diet (from the Department of Agriculture), updated by the U.S. Consumer Price Index	33rd percentile of expenditures on food, clothing, shelter, and utilities (from recent Bureau of Labor Statistics surveys) multiplied by 1.2
Threshold adjustments	Implicit equivalence scale that varies by family size, composition, and age of the family head	Explicit equivalence scale that varies by unit size and composition, but not by age of unit head; also, adjustments for differences in housing costs by (1) housing status (e.g. owner with a mortgage) and (2) geographic area

Sources: Short (2012). DeNavas-Walt, Proctor, and Smith (2012). See also http://www.census.gov/newsroom/releases/archives/income_wealth/cb12-172.html.

2 The SPM is also often adjusted for differences in housing costs between areas, but our measure does not contain these cost adjustments.

4. Descriptive Statistics on Demographic Variables and Poverty Rates

There are a variety of ways that changes in the demographic composition of the population (age, race, gender, education, marital status etc.) could have affected the OPM and SPM poverty rates over the 1968–2012 period. Some demographic changes would be associated with lower expected poverty, such as the increase in female labor force participation and the increases in educational attainment. Other demographic changes would be associated with higher expected poverty, such as the increase in single-parent households and the aging of the population.

The demographic composition of the population in 1968 and 2012 is shown in Table 2. The data are tabulated from individual observations included in the Panel Study of Income Dynamics supplemented with tax data calculated using the National Bureau of Economic Research TAXSIM tax simulation program.[3] The table highlights the statistically significant changes in the demographic composition of the population. Most variables have changed in statistically significant ways. Some of these changes are things that would be expected to increase poverty (e.g. the increase in female headed families, and divorced and single individuals) while other factors would be expected to reduce poverty (e.g. the decrease in large families and the increase in educational attainment). The largest changes are 1) the shift in the age composition of the population with relatively fewer children and relatively more older people, 2) a substantial increase in the prevalence of female headed families, divorced individuals and singles, 3) a large increase in racial minorities (especially Hispanics), 4) a large increase in the prevalence of households with no workers (a part of this is accounted for by the increase in older households many of which are retirees)[4], 5) a large decrease in the prevalence of large families, 6) a very large increase in the prevalence of those with a college or advanced degree and the associated decrease in the prevalence of those with little education, and 7) the large decrease in the prevalence of households with children less than 18 years living at home.

In Table 3,we show the changes in the OPM and the SPM poverty rates for the nation and for each of the demographic categories. The changes from 1968 to 2012 are quite different between the two measures; while the OPM remains about the same over the period, the SPM decreases by nearly 4 percentage points. The inclusion of the value of in-kind transfers in the SPM largely ac-

3 The Panel Study of Income Dynamics (PSID) is a U.S. longitudinal data set that began in 1968. The sample includes 15,937 observations in 1968 and 17,403 observations in 2012. PSID sample weights are used in all analysis to make the sample nationally representative.

4 When the sample is limited to those 18-65, the percent of households with no workers is 6.26 percent in 1968 and 11.7 percent in 2013.

Table 2
Changes in Demographic Characteristics 1968 to 2012

	% of Population In 1968	% of Population In 2012	Percentage Point Difference	t-statistic
Age				
5 and under	12.0 %	6.9 %	**−5.1 %**	**−15.84**
6 to 17	27.3 %	15.2 %	**−12.2 %**	**−27.34**
18 to 39	29.6 %	29.5 %	−0.1 %	−0.17
40 to 64	23.6 %	33.2 %	**9.6 %**	**19.53**
65 to 79	6.3 %	10.8 %	**4.5 %**	**14.69**
80 plus	1.1 %	4.5 %	**3.3 %**	**18.66**
Race				
Non-Hispanic White	82.3 %	69.5 %	**−12.8 %**	**−27.72**
Black	12.5 %	14.3 %	**1.8 %**	**4.90**
Hispanic	3.7 %	12.8 %	**9.1 %**	**30.87**
Other	1.1 %	3.3 %	**2.1 %**	**13.34**
Gender				
Female	51.2 %	51.1 %	−0.1 %	−0.16
Family Structure				
Female-Headed Household	13.7 %	24.0 %	**10.3 %**	**24.29**
Head's Marital Status				
Married	83.6 %	59.6 %	**−23.9 %**	**−50.54**
Widow	6.7 %	5.7 %	**−0.9 %**	**−3.52**
Single	3.3 %	17.6 %	**14.3 %**	**44.49**
Divorced	6.4 %	17.0 %	**10.6 %**	**30.65**
Home Ownership				
Owns home	65.5 %	63.8 %	**−1.7 %**	**−3.30**
Number of Workers in Household				
Zero workers	10.3 %	18.9 %	**8.5 %**	**22.35**
One worker	52.0 %	42.2 %	**−9.8 %**	**−17.92**
Two workers	37.7 %	38.9 %	**1.2 %**	**2.30**
Number of Children Under 18 in Household				
Zero	25.9 %	52.3 %	**26.4 %**	**51.44**
One	15.3 %	15.9 %	0.5 %	1.35
Two	19.1 %	17.9 %	**−1.2 %**	**−2.88**
Three	15.7 %	8.3 %	**−7.4 %**	**−20.69**
Four or more	24.0 %	5.6 %	**−18.4 %**	**−48.24**
Education Level of Head or Wife, whichever Higher				
High School Dropout	33.6 %	8.3 %	**−25.3 %**	**−59.01**
High School Graduate	36.3 %	25.3 %	**−11.0 %**	**−21.74**
Some College	15.3 %	26.1 %	**10.7 %**	**24.51**
College Graduate	9.4 %	19.8 %	**10.4 %**	**27.38**
Advanced Degree	5.3 %	20.4 %	**15.1 %**	**42.69**
Region				
Northcentral	30.2 %	26.1 %	**−4.1 %**	**−8.27**
Northeast	24.3 %	17.3 %	**−7.0 %**	**−15.78**
South	29.5 %	33.5 %	**4.0 %**	**7.92**
West	16.0 %	22.5 %	**6.5 %**	**15.02**

Table 3
Changes in Poverty Rates by Demographic Variables

	Census Poverty rate				SPM Poverty Rate			
	1968	2012	Change	t-Stat	1968	2012	Change	t-Stat
Overall	11.1 %	11.1 %	−0.1 %	−0.24	**13.6 %**	**10.1 %**	**−3.6 %**	**−10.04**
Age								
5 and under	**13.1 %**	**16.4 %**	**3.3 %**	**3.17**	**15.7 %**	**11.8 %**	**−3.8 %**	**−3.80**
6 to 17	*13.3 %*	*14.7 %*	*1.4 %*	*1.91*	**15.6 %**	**10.7 %**	**−4.9 %**	**−7.04**
18 to 39	**7.5 %**	**12.8 %**	**5.3 %**	**8.96**	**10.7 %**	**12.2 %**	**1.5 %**	**2.40**
40 to 64	8.4 %	8.6 %	0.1 %	0.20	**10.4 %**	**8.5 %**	**−2.0 %**	**−2.79**
65 to 79	**21.1 %**	**5.6 %**	**−15.5 %**	**−8.65**	**23.3 %**	**6.6 %**	**−16.7 %**	**−8.95**
80 plus	**34.6 %**	**10.7 %**	**−23.9 %**	**−5.08**	**33.6 %**	**11.5 %**	**−22.1 %**	**−4.71**
Race								
Non-Hispanic White	7.2 %	6.8 %	−0.4 %	−0.97	**9.0 %**	**6.5 %**	**−2.5 %**	**−6.16**
Black	**35.7 %**	**24.7 %**	**−11.0 %**	**−13.80**	**41.3 %**	**21.0 %**	**−20.3 %**	**−25.78**
Hispanic	18.3 %	18.9 %	0.6 %	0.34	**24.7 %**	**17.1 %**	**−7.6 %**	**−3.97**
Other	*7.5 %*	*11.6 %*	*4.1 %*	*1.89*	10.2 %	11.3 %	1.1 %	0.49
Gender								
Male	10.3 %	10.0 %	−0.3 %	−0.54	**12.9 %**	**9.3 %**	**−3.6 %**	**−7.21**
Female	12.0 %	12.1 %	0.1 %	0.18	**14.4 %**	**10.8 %**	**−3.5 %**	**−7.00**
Family Structure								
Fem.-Head Household	**29.1 %**	**24.2 %**	**−4.9 %**	**−5.17**	**31.4 %**	**22.1 %**	**−9.3 %**	**−9.72**
Not Fem.-Head Hhld.	**8.3 %**	**6.9 %**	**−1.4 %**	**−4.03**	**10.8 %**	**6.3 %**	**−4.5 %**	**−12.72**
Head's Marital Status								
Married	**8.0 %**	**4.4 %**	**−3.6 %**	**−10.78**	**10.6 %**	**3.9 %**	**−6.7 %**	**−19.23**
Widow	**18.0 %**	**25.4 %**	**7.4 %**	**4.67**	20.3 %	23.0 %	2.7 %	1.63
Single	**28.3 %**	**13.2 %**	**−15.1 %**	**−8.24**	**28.7 %**	**13.9 %**	**−14.8 %**	**−7.98**
Divorced	**30.8 %**	**18.7 %**	**−12.0 %**	**−9.94**	**33.7 %**	**17.0 %**	**−16.7 %**	**−13.74**
Home Ownership								
Owns home	**6.9 %**	**4.1 %**	**−2.8 %**	**−7.90**	**7.7 %**	**3.7 %**	**−4.0 %**	**−11.28**
Does not own home	**19.2 %**	**23.3 %**	**4.1 %**	**6.35**	**24.9 %**	**21.3 %**	**−3.6 %**	**−5.38**
Number of Workers in Household								
Zero workers	**41.5 %**	**26.2 %**	**−15.2 %**	**−11.11**	**43.0 %**	**26.1 %**	**−16.8 %**	**−12.26**
One worker	**10.8 %**	**12.8 %**	**1.9 %**	**3.89**	**14.0 %**	**11.2 %**	**−2.8 %**	**−5.43**
Two workers	**3.3 %**	**1.9 %**	**−1.4 %**	**−4.79**	**5.1 %**	**1.1 %**	**−4.1 %**	**−12.38**
Number of Children Under 18 in Household								
Zero	**10.4 %**	**8.8 %**	**−1.6 %**	**−2.46**	**12.0 %**	**9.6 %**	**−2.4 %**	**−3.44**
One	**7.1 %**	**10.2 %**	**3.1 %**	**4.04**	10.6 %	10.7 %	0.1 %	0.11
Two	**7.0 %**	**9.3 %**	**2.2 %**	**3.24**	**9.1 %**	**7.2 %**	**−1.9 %**	**−2.64**
Three	**9.9 %**	**16.8 %**	**6.9 %**	**6.83**	**11.7 %**	**9.6 %**	**−2.1 %**	**−2.29**
Four or more	**18.6 %**	**31.7 %**	**13.1 %**	**9.72**	22.2 %	22.1 %	−0.1 %	−0.04
Education Level of Head or Wife, whichever Higher								
High School Dropout	**25.2 %**	**40.6 %**	**15.4 %**	**12.01**	**29.6 %**	**37.0 %**	**7.4 %**	**5.83**
High School Graduate	**5.4 %**	**16.3 %**	**10.9 %**	**17.55**	**7.9 %**	**14.4 %**	**6.5 %**	**10.38**
Some College	**3.9 %**	**9.6 %**	**5.7 %**	**9.23**	**4.7 %**	**9.0 %**	**4.3 %**	**6.72**
College Graduate	**0.9 %**	**3.1 %**	**2.2 %**	**5.15**	**0.8 %**	**2.9 %**	**2.1 %**	**5.03**
Advanced Degree	**0.2 %**	**2.1 %**	**2.0 %**	**6.06**	**0.5 %**	**2.0 %**	**1.5 %**	**3.73**
Region								
Northcentral	**6.8 %**	**11.1 %**	**4.3 %**	**6.85**	*8.4 %*	*9.5 %*	*1.1 %*	*1.77*
Northeast	**4.9 %**	**7.7 %**	**2.8 %**	**3.95**	7.8 %	6.9 %	−1.0 %	−1.26
South	**21.4 %**	**13.2 %**	**−8.2 %**	**−13.08**	**24.1 %**	**12.1 %**	**−12.0 %**	**−18.94**
West	10.1 %	10.0 %	−0.1 %	−0.07	**13.0 %**	**9.7 %**	**−3.2 %**	**−3.74**

counts for this difference. The patterns shown in Table 3 give insight into how changes from 1968 to 2012 in the poverty rate by demographic characteristic contribute to the overall change in the two poverty rates. For example, if the returns to education have increased over this period, we would expect the poverty rate to go up for those who are high school dropouts relative to those with higher education levels.

These changes mean that even if the demographic characteristics of the population were unchanged from 1968 to 2012, we would expect to see a different poverty rate in 2012 than 1968 because the poverty rate associated with each demographic characteristic has changed over this period. The most striking patterns are, for both poverty rates: 1) a large decrease in poverty for older people, 2) a large decrease for Blacks, 3) a sizable decrease for female headed families, the divorced and the widowed, 4) a decrease for households with one or more workers, 5) a sizable increase (decrease) for those with low education (higher education), and 6) decrease for those living in the South. These changes suggest that it is not just the changing demographics, but also the changes in the relationship between poverty and these demographics that must be considered. The regression analysis and counterfactuals in the next two sections provide a framework for examining this.

5. Poverty Regressions

Table 3 fails to isolate the effect of any given demographic characteristic on poverty. For example, there is a lower poverty rate in 2012 for households with zero workers than there was in 1968. We would expect the opposite given the transition from AFDC to EITC. The reason, though, is in part due to the large change in the age distribution over this time period with a rapid growth of retirees who are just above the poverty line.

We have undertaken a series of probit regressions revealing the correlates of the OPM and the SPM poverty rates. For both 1968 and 2012 sample we estimate the following probit regression:

$$Poverty_i = a + \beta X_i + \epsilon_i \tag{1}$$

where $Poverty_i$ is a dummy variable that equals 1 if the individual is in poverty and 0 otherwise, X_i is a vector of demographic variables, and ϵ_i is an individual-specific error term. We note that the regression estimates are reduced form rather than structural. However, the coefficient estimates indicate the extent to which different demographic factors are related to poverty and changes in these patterns over time. By comparing the coefficients from the probit (β) from estimating Equation (1) separately on the 1968 and the 2012 samples, we can report the extent to which a particular demographic variable has increased or decreased its relationship to poverty rate over time.

Table 4a presents the results of the probit regression for the OPM and Table 4b presents the results for the SPM. The first four columns show the marginal effect[5] (and standard error) of changes in the demographic variables, while the last two columns show the change in the coefficient estimate and a t-statistic indicating whether the coefficient estimate is statistically significantly different in 2012 than 1968. While many of the demographic variables are statistically significant in both years for the two measures, we highlight where the coefficient estimates have changed over time in a statistically significant way. For both older and younger people, Blacks, and living in the South, the differences in the marginal effect over time are consistently statistically significant across both measures of poverty; the marginal effects decrease over time for these three characteristics across both measures of poverty. The changes in these coefficient estimates suggests that while changes in demographic variables are important, they are also occurring in the context of changing economic and policy considerations.

6. Counterfactuals

The changes in the demographic characteristics (Table 2), combined with the changes in the reduced form regressions (Table 4), indicate how the changing distribution of demographic characteristics are correlated with poverty. This information provides the framework for the decomposition methodology and counterfactuals we use. The counterfactuals are calculated using the demographic characteristics (X from Equation (1)) in 1968 and 2012 with the coefficients (β from Equation (1)) from the 1968 and 2012 probit regressions. These counterfactuals indicate the extent to which changes in demographic characteristics are associated with changes in poverty.

Consider first the Census poverty measure (OPM) in Panel A of Table 5. Using 1968 demographics (X_{1968}) and 1968 probit coefficient estimates (β_{1968}), the 1968 census poverty rate is predicted to have been 11.2 % (cell a). However, with the 2012 demographics (X_{2012}), the predicted poverty rate using 1968 probit coefficient estimates falls to 10.1 % (cell b). The demographic changes are associated with a 1.1 percentage point (9.8 percent) reduction in poverty. If, instead, the 2012 probit coefficients are used (β_{2012}), the difference in results are even greater. Comparing cell c and cell d, the changing demographics are associated with a 5.1 percentage point (31 percent) reduction in poverty.

5 The marginal effects reported show the change in the poverty rate for a change in a category from the base value.

Table 4a

Marginal Effects from Probit on Census Poverty in 1968 and 2012

	1968		2012		Difference (2012–1968)	t-Statistic
	Marginal Effect	Stand. Error	Marginal Effect	Stand. Error		
Age (40 to 64 is omitted category)						
5 and under	0.0023	0.0056	**0.0187**	**0.0078**	0.0164	1.709
6 to 17	**−0.0062**	**0.0029**	−0.0004	0.0041	0.0058	1.153
18 to 39	−0.0083	0.0044	0.0052	0.0045	**0.0135**	**2.144**
65 to 79	0.0035	0.0082	**−0.0287**	**0.0045**	**−0.0322**	**−3.440**
80 plus	**0.0710**	**0.0314**	**−0.0244**	**0.0053**	**−0.0954**	**−2.999**
Race (White is omitted category)						
Black	**0.0655**	**0.0136**	**0.0142**	**0.0067**	**−0.0513**	**−3.384**
Hispanic	0.0446	0.0273	**0.0217**	**0.0104**	−0.0229	−0.784
Other	0.0090	0.0172	0.0259	0.0191	0.0170	0.661
Gender and Household Composition						
Individual is Female	0.0012	0.0024	**0.0067**	**0.0034**	0.0055	1.323
Household head is fem.	0.0102	0.0127	0.0034	0.0065	−0.0068	−0.474
Head's Marital Status (Married is omitted category)						
Widow	0.0046	0.0136	−0.0065	0.0101	−0.0111	−0.658
Single	0.0292	0.0221	**0.0548**	**0.0125**	0.0255	1.007
Divorced	0.0105	0.0145	**0.0237**	**0.0100**	0.0131	0.746
Home Ownership						
Owns home	**−0.0390**	**0.0090**	**−0.0596**	**0.0076**	−0.0206	−1.755
Number of Workers in Household (Two or more is the omitted category)						
Zero workers	**0.3387**	**0.0457**	**0.3064**	**0.0318**	−0.0324	−0.581
One worker	**0.0525**	**0.0090**	**0.0672**	**0.0103**	0.0147	1.070
Number of Children Under 18 in Household (Two is omitted category)						
Zero	**−0.0184**	**0.0073**	**−0.0260**	**0.0077**	−0.0076	−0.714
One	−0.0051	0.0097	−0.0096	0.0071	−0.0045	−0.375
Three	0.0132	0.0138	**0.0595**	**0.0194**	0.0464	1.945
Four or more	**0.0480**	**0.0158**	**0.1007**	**0.0334**	0.0527	1.429
Educ. Level of Head or Wife, whichever Higher (HS Dropout is omitted)						
High School Graduate	**−0.0456**	**0.0106**	**−0.0690**	**0.0163**	−0.0234	−1.204
Some College	**−0.0128**	**0.0096**	**−0.0271**	**0.0071**	−0.0143	−1.194
College Graduate	**−0.0319**	**0.0091**	**−0.0222**	**0.0071**	0.0097	0.844
Advanced Degree	**−0.0255**	**0.0124**	−0.0014	0.0098	0.0241	1.525
Region (North Central is omitted category)						
Northeast	**−0.0287**	**0.0072**	**−0.0162**	**0.0059**	0.0125	1.353
South	**0.0659**	**0.0126**	0.0026	0.0059	**−0.0633**	**−4.571**
West	0.0255	0.0150	**−0.0159**	**0.0060**	**−0.0414**	**−2.558**

Table 4b

Marginal Effects from Probit on SPM Poverty in 1968 and 2012

	1968		2012		Difference (2012–1968)	t-Statistic
	Marginal Effect	Stand. Error	Marginal Effect	Stand. Error		
Age (40 to 64 is omitted category)						
5 and under	−0.0038	0.0074	0.0140	0.0073	0.0177	1.712
6 to 17	**−0.0113**	**0.0048**	−0.0007	0.0039	0.0106	1.721
18 to 39	−0.0045	0.0064	0.0077	0.0044	0.0122	1.578
65 to 79	0.0120	0.0126	**−0.0250**	**0.0041**	**−0.0370**	**−2.798**
80 plus	**0.0817**	**0.0363**	**−0.0236**	**0.0043**	**−0.1053**	**−2.877**
Race (White is omitted category)						
Black	**0.0994**	**0.0197**	0.0070	0.0056	**−0.0924**	**−4.510**
Hispanic	**0.0666**	**0.0326**	**0.0219**	**0.0102**	−0.0447	−1.309
Other	0.0072	0.0224	0.0255	0.0186	0.0183	0.629
Gender and Household Composition						
Individual is Female	−0.0001	0.0034	0.0032	0.0033	0.0034	0.723
Household head is fem.	0.0288	0.0223	0.0057	0.0062	−0.0231	−0.997
Head's Marital Status (Married is omitted category)						
Widow	−0.0190	0.0149	−0.0096	0.0080	0.0093	0.553
Single	0.0085	0.0227	**0.0315**	**0.0100**	0.0230	0.928
Divorced	−0.0127	0.0157	0.0086	0.0077	0.0212	1.211
Home Ownership						
Owns home	**−0.0818**	**0.0128**	**−0.0595**	**0.0075**	0.0223	1.503
Number of Workers in Household (Two workers is omitted category)						
Zero workers	**0.3306**	**0.0439**	**0.3235**	**0.0325**	−0.0071	−0.130
One worker	**0.0708**	**0.0115**	**0.0728**	**0.0105**	0.0020	0.128
Number of Children Under 18 in Household (Two is omitted category)						
Zero	**−0.0255**	**0.0101**	−0.0017	0.0069	0.0238	1.940
One	0.0074	0.0153	0.0101	0.0091	0.0027	0.153
Three	0.0098	0.0163	0.0160	0.0135	0.0062	0.291
Four or more	**0.0616**	**0.0199**	**0.0474**	**0.0237**	−0.0142	−0.459
Educ. Level of Head or Wife, whichever Higher (HS Dropout is omitted)						
High School Graduate	**−0.0677**	**0.0137**	**−0.0608**	**0.0151**	0.0069	0.339
Some College	**−0.0281**	**0.0122**	**−0.0177**	**0.0061**	0.0104	0.763
College Graduate	**−0.0541**	**0.0106**	**−0.0219**	**0.0062**	**0.0322**	**2.630**
Advanced Degree	−0.0111	0.0281	−0.0005	0.0091	0.0105	0.357
Region (North Central is omitted category)						
Northeast	**−0.0306**	**0.0111**	**−0.0129**	**0.0056**	0.0178	1.424
South	**0.0783**	**0.0147**	0.0050	0.0055	**−0.0733**	**−4.665**
West	**0.0424**	**0.0202**	−0.0084	0.0059	**−0.0508**	**−2.412**

Table 5
Predicted Poverty Rate Based on Probit Coefficient Estimates and Demographics: Census Poverty Measure and SPM

Panel A: Official Census Poverty Definition	1968 Probit Coefficients β_{1968}	2012 Probit Coefficients β_{2012}
1968 Demographics X_{1968}	11.2 % [a]	16.2 % [c]
2012 Demographics X_{2012}	10.1 % [b]	11.1 % [d]

Panel B: Supplemental Poverty Definition	1968 Probit Coefficients β_{1968}	2012 Probit Coefficients β_{2012}
1968 Demographics X_{1968}	13.7 % [A]	12.1 % [C]
2012 Demographics X_{2012}	11.1 % [B]	10.1 % [D]

Panel B shows a similar exercise for the SPM measure. Using the 2012 probit coefficients, the change in demographic characteristics (cell C to cell D) reduced the poverty rate from 12.1 % to 10.1 %, a reduction of 2.0 percentage points (16.5 percent). Using the 1968 probit coefficients (cell A to cell B) the change is similar, at 2.6 percentage points (19.0 percent).[6]

The counterfactuals show what would happen if all of the demographics were changed to reflect 1968 or 2012 demographics, but it is also of interest to disaggregate this for the different sets of demographic characteristics. For example, which demographic changes are associated with increases in poverty and which are associated with decreases in poverty? Since the predicted probability from the probit regression is a non-linear function of the demographic characteristics, it is not possible to isolate the effect of individual characteristics within the counterfactuals since the effects vary depending on the values of the other demographic variables. However, as an approximation, we use the marginal effects from the regression and the changes in the demographic variables between 1968 and 2012 to get a sense of the direction and relative importance of categories of demographic variables. Most of the demographic changes between 1968 and 2012 had the effect of increasing both OPM and SPM poverty, with large increases associated with the changes in marital status and number of households with no workers. The changes in racial compo-

6 It is interesting to note that while the demographic changes lead to a reduction in poverty for both the OPM and SPM, using the coefficient estimates from 2012 increase poverty relative to the 1968 coefficients for the OPM but decrease poverty for the SPM. This suggests that changes in the other factors, such as the labor market and public policy, have changed in a way that differentially affects the narrower pre-tax/transfer cash income measure of the OPM compared to the broader after-tax/transfer income measure of the SPM. An understanding of these differences are beyond the scope of this paper, but warrant further investigation.

sition, female-headed families, and home ownership are also associated with smaller increases in poverty. However, these increases are outweighed by a large decrease in poverty associated with the educational changes and households having fewer children, with the net result being the counterfactual result of reduced poverty from demographic changes.

In summary, for both poverty measures, the changes in demographic characteristics results in a lower poverty rate. For the OPM, the magnitude of the reduction is larger using the 2012 probit coefficients than the 1968 coefficients, while for the SPM poverty measure the result is similar regardless of which year's coefficient estimates are used. However, for all of the counterfactual measures the net results of the demographic changes is a reduction in poverty. This net reduction is driven by the educational changes and having fewer children in the household, poverty-reducing demographic changes that outweigh the other demographic changes, particularly in marital status and households with no workers, that work to increase poverty.

7. Conclusion

The analysis in this paper has examined two different measures of poverty in 1968 and 2012 in an attempt to quantify the effect of changing demographics on the poverty rate in the United States. There have been substantial changes in demographic patterns over this time period, particularly related to the age of the population, family structure, racial diversity, labor market attachment, and education levels. In addition, the poverty rate associated with different demographic changes has changed over time, both in the descriptive statistics and the regression analysis. Some of these changes, such as the increased educational attainment, would be expected to be associated with a lower poverty rate while others, like the increase in single-parent families, would be expected to be associated with a higher poverty rate. The analysis uses a counterfactual calculated using the probit regressions and the demographic changes to show that the net effect of the changes in demographics reduces poverty across both poverty measures.

References

Bailey, M.J., and Danzinger, S. (2013), Legacies of the War on Poverty, Russell Sage Foundation, New York, NY.

Council of Economic Advisors (2014), Economic Report of the President, U.S. Government Printing Office, Washington, D.C.

DeNavas-Walt, C., Proctor, B., and Smith, J. (2012), Income, Poverty, and Health Insurance Coverage in the United States: 2011, U.S. Census Bureau, Washington, D.C.

Fox, L., Garfinkel, I., Kaushall, N. Waldfogel, J., and Wimer, C. (2015), Waging War on Poverty: Poverty Trends Using a Historical Supplemental Poverty Measure, Journal of Policy Analysis and Management 34, 567–592.

Haveman, R., Blank, R., Moffitt, R., Smeeding, T., and Wallace, G. (2014), The War on Poverty: Measurement, Trends, and Policy, Journal of Policy Analysis and Management 34, 593–638.

Short, K. (2012), The Research Supplemental Poverty Measure: 2012, U.S. Census Bureau, Washington, D.C.

Wimer, C., Fox, L., Garfinkel, I., Kaushal, N., and Waldfogel, J. (2016), Progress on Poverty? New Estimates of Historical Trends Using an Anchored Supplemental Poverty Measure, Demography 53, 1207–1218.

Migration-Induced Redistribution with and without Migrants' Voting

Assaf Razin and Efraim Sadka*

Received 17 May 2017; in revised form 25 September 2017; accepted 09 October 2017

We are motivated by the unique migration experience of Israel, of a supply-side shock triggering skilled immigration and the concurrent decline in welfare-state redistribution. This paper develops a model that can provide an explanation for the mechanism through which a supply-side shock, triggering high-skill migration, can also reshape the political-economy balance and the redistributive policies. The paper highlights the differences in the political-economy-based redistribution policies between the cases in which migrants participate in the electoral system and the case in which they do not. When migrants are allowed to vote, and take advantage of this right, then, all income groups gain (in their net income), except the low-skilled immigrants, who lose. However, when migrants are not allowed to vote, or choose not to participate in elections, all income groups gain, except the skilled migrants who lose.

Keywords: immigration episode as a "natural experiment", majority voting, progressivity of the welfare state, gainers and losers

JEL classification: F 22, H 24, H 55

1. Introduction

Following the collapse of the Soviet Union some three decades ago, large numbers of immigrants (about 20 % of the Israeli population at the time) went to Israel. Relative to the native-born Israelis, these immigrants were poor in wealth, but rich in skills.[1]

In history, immigrants often shift the balance of politics among ethnic groups, economic classes, and age groups, so that they could generate political backlash. In Israel, however, the political backlash has been moderate, whereas the change in the political balance has been substantial. Israel's Law of Return grants returnees immediate citizenship and consequently vot-

* Razin: Tel Aviv University, Tel Aviv 69978, Israel (razin@tauex.tau.ac.il); Sadka: Tel Aviv University, Tel Aviv 69978, Israel (sadka@post.tau.ac.il). We wish to thank a referee and the editor for useful comments. Tslil Aloni and Eyal Ben-David provided competent research assistance, and the Sapir Center for Development, Tel-Aviv University, provided financial support.

1 See Razin (2017).

FinanzArchiv 74 (2017), 158–172
ISSN 0015-2218
doi: 10.1628/001522117X15105745052709

ing rights. Immigrants' voting is key to understanding the political-economy mechanism that determines income distribution and redistribution (see Razin, Sadka, and Swagel, 2002a,b). An early study by Avner (1975) found that the voter turnout rate of new immigrants was markedly lower than that of the established population. This would mean that immigrants did not fully exercise their voting rights and therefore did not influence the political-economy equilibrium in Israel as much as the established population.[2] However, a later study conducted by Arian and Shamir (2002) about voter turnout patterns of new immigrants to Israel in the 2001 elections reverses the earlier finding. The new immigrants in this study were predominantly from the former Soviet Union (FSU). Arian and Shamir find no marked difference in voter turnout rates between the new immigrants and the established population.

Migration differs from the movement of other factors of production (such as capital) in one fundamental way. Migrants become part of the society of the receiving country, including its evolving culture and politics.[3] A highly developed social welfare system in the receiving country may greatly complicate matters, as emphasized by Razin, Sadka, and Swagel (2002b).[4] While high-skilled and therefore high-wage migrants may be net contributors to the fiscal system, low-skilled migrants are likely to be net recipients, thereby imposing an indirect tax on the taxpayers of the receiving country. A sizeable wave of migrants may shift the balance of politics among ethnic groups, economic classes, or age groups, and reshape the distribution of wealth and disposable income. That is, immigrants could influence the size of the welfare state directly through the electoral system, and indirectly through their effect on market-based inequality.

Figure 1 depicts the standard Gini coefficients of the distribution of gross income and disposable income (the two upper graphs). The bottom graph, which is the difference between them, measures the degree of redistribution; a higher graph indicates redistribution that is more intensive.

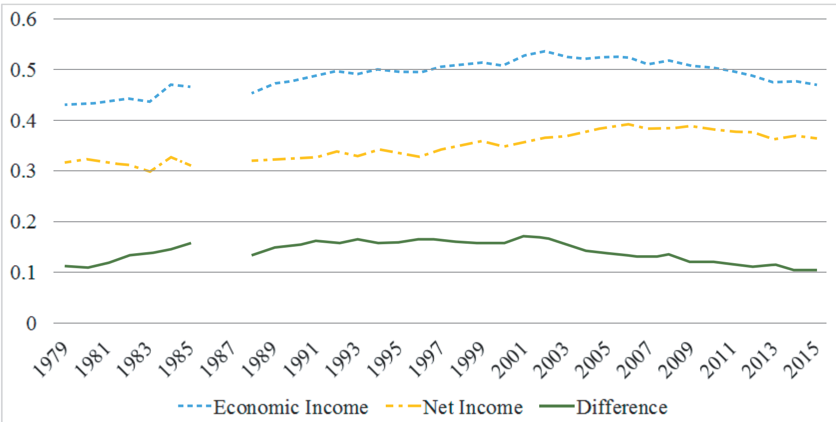
Figure 1 demonstrates a marked decline in economic income inequality, starting at the beginning of the present century, and a noticeable decline in redistribution, resulting in a rather moderate rise in net income inequality. We suggest that these trends are driven by the influx of immigrants from the FSU in the preceding decade. Thanks to high-skill migration the rising middle class lowered economic income inequality, but it reoriented the income redistribution policies.

2 Messina (2007) and Bird (2011) report a similar low voter turnout pattern for migrants for Western Europe.

3 The Swiss playwright Max Frisch put it poetically: "We asked for workers. We got people".

4 A related issue is the implications of aging population for the size of the welfare state; see Razin, Sadka, and Suwankiri (2011).

Figure 1
Gini Coefficients: Gross Economic Income, Net Income, and Redistribution – 1979–2015***



Notes: * Redistribution is measured by the difference between economic and net (disposable) income standard, Gini coefficients. ** The break in the data is in the source. Source: Dahan (2017).

The literature has addressed several issues in the political economics of immigration. For instance, Gradstein and Schiff (2006) deal with redistribution between the majority native-born and the minority immigrants. Mayda, Peri, and Steingress (2015) study empirically how immigrants shape political party voting in the U.S. The novelty of our paper is in analyzing simultaneously how immigration affects the nationwide income redistribution, and how redistribution affects the volume and the skill mix of immigration. Specifically, the paper aims at developing formally a political-economy mechanism that may explain the aforementioned conflicting effects on income inequality and the skill mix of immigration driven by an immigration supply shock. We develop an analytical model in which immigrants' voting is key for the explanation of the migration cum redistribution trends.

The organization of the paper is as follows. Section 2 describes the model, and section 3 presents the political-economy equilibrium. In section 4, we discuss the redistribution with and without migrants' voting. Section 5 provides concluding remarks.

2. The Model

The basic ingredients of the model are as follows.

2.1. Human-Capital Investment

There are just two types of workers: *skilled* (with symbol S) and *unskilled* (with symbol U). The wage per unit of labor of a skilled worker is w, whereas that of an unskilled worker is ρw, where $\rho < 1$. All native-born (with symbol N) are initially unskilled. However, a native-born can acquire education at some cost (c) and become skilled. Individuals differ from one another through their cost of education: there is a continuum of native-born individuals, distinguished only by their cost of education. For notational simplicity, we normalize the number of native-born individuals to one. An individual is identified by her cost of education, so that an individual with a cost of c is termed a c-individual. We assume for simplicity that the cost of education is uniformly distributed over the interval $[0, \overline{c}]$.

All native-born individuals are endowed with E units of a composite good, the single good in this economy[5]. All individual inelastically supply one unit of labor. If a c-individual acquires education and becomes skilled, her income[6] (denoted by I_S^N) is

$$I_S^N(c) = (1-t)w + b + (E-c)(1+r), \tag{1}$$

where t is a flat wage tax rate, b is a uniform (lump sum) per capita social benefit, and r is the interest rate – the return to capital. If a c-individual decides not to acquire education and remain unskilled, her income (denoted by I_U^N) is

$$I_U^N = (1-t)\rho w + b + E(1+r). \tag{2}$$

(Note that $I_S^N(c)$ depends on c, whereas I_U^N does not.)

Thus, there is a cutoff level of cost, c^*, so that all c-individuals with $c \le c^*$ will choose to become skilled, and all the others (with $c \ge c^*$) will remain unskilled. This c^* is defined by

$$(1-t)w + b + \left(E - c^*\right)(1+r) = (1-t)\rho w + b + E(1+r).$$

The variable c^* is solved for from the equality between the return to education and its cost. A c^*-individual is just indifferent between acquiring education (and thereby becoming skilled) and staying unskilled. Upon further rearrangement, c^* is expressed by

$$c^* = \frac{(1-t)(1-\rho)w}{1+r}. \tag{3}$$

Note that c^* may well exceed E, which means that those c-individuals with c below but close to c^* (which is endogenous) actually *borrow* in order

5 To simplify the analysis, we assume that E and c are uncorrelated. A possible extension of the model is to assume some distribution of E, which is negatively correlated with c, so that more capable individuals (with low c) have possession of larger endowments (higher E).

6 Note that this specification assumes that capital does not depreciate at all.

to acquire education. Naturally, the payoff due to the higher wage will more than offset the borrowing cost. For those individuals $E-c$ is negative.

Also, note that we are employing a static framework within which all economic and political processes occur simultaneously with no time dimension.[7] For instance, we do not distinguish between the time when the education is acquired and the time when the earnings occur. Similarly, capital earns its return r at the same time it is employed.

The number of c-individuals with $c \leq c^*$ is the number of native-born skilled individuals. Denoting this number by n_S, it follows that

$$n_S = \frac{c^*}{\overline{c}}. \tag{4}$$

Then, the number of native-born unskilled individuals, n_U, is given by

$$n_U = 1 - n_S. \tag{5}$$

The aggregate investment in human capital (education), denoted by H, and is then given by

$$H = \int_0^{c^*} c \cdot \frac{1}{\overline{c}} dc = \frac{(c^*)^2}{2\overline{c}}. \tag{6}$$

Therefore, the aggregate stock of physical capital, K, is equal to

$$K = E - H. \tag{7}$$

There are also two types of migrants: the skilled, who can earn a wage w in the host country, and the unskilled, who earn a wage ρw in the host country. None of them has any initial endowment. The migrants come to the host country after they have already made and implemented the decision whether to acquire or not acquire education.[8] Thus, it is exogenously determined who is skilled and who is unskilled. In other words, the economy benefits from the skilled migrants because it does not have to pay the cost of investment.

2.2. Income Groups

The income of skilled and unskilled migrants, respectively, is

$$I_S^M = (1-t)w + b \tag{8}$$

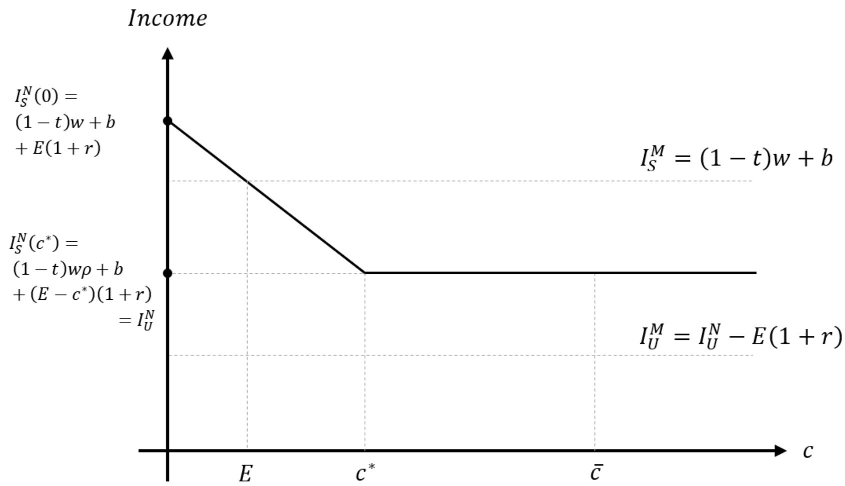
and

$$I_U^M = (1-t)\rho w + b. \tag{9}$$

7 Such a framework is akin to a steady state in a dynamic model with rational expectations.

8 For simplicity we assume that migrants come with no initial endowment and no debt from abroad. That is, their E is zero.

Figure 2
Income Groups and Cost of Education



The income of the native-born as a function of c is depicted in Figure 2. Note that $I_S^N(c)$ declines in a straight line until it reaches c^*, where

$$I_S^N\left(c^*\right) = (1-t)w + b + \left(E - c^*\right)(1+r)$$
$$= (1-t)\rho w + b + E(1+r) = I_U^N.$$

The labor income of the unskilled native-born and the unskilled migrants is the same, but the total income of an unskilled migrant, which is $(1-t)\rho w + b$, is definitely below the income of an unskilled native-born, the difference being the capital income enjoyed by the unskilled native-born, namely $E(1+r)$. The total income of a skilled migrant is definitely higher than the total income of the unskilled migrant, because of the higher wage earned by the skilled, whereas neither has any other income. The income of the skilled migrants exceeds the income of the skilled native-born with $c > E$, but falls short of the income of the skilled native-born with $c < E$.

The income of a skilled migrant is $I_S^M = (1-t)w + b$, whereas the income of a skilled c-individual is $(1-t)w + b + (E-c)(1+r)$. Therefore, as long as $E - c$ is positive (i.e., the c-individual does not borrow in order to invest in human capital), then $I_S^N(c) > I_S^M$. However, if $E - c < 0$ (i.e., the individual borrows in order to invest in human capital), then the income of the skilled migrant (I_S^M) is greater than the income of the skilled native-born (I_S^N). In sum, we have the following ranking of incomes:

$$I_U^M < I_U^N = I_S^N\left(c = c^*\right) < I_S^N(c > E) < I_S^N(c = E) = I_S^M < I_S^N(c < E).$$

2.3. Supply of Immigrants

Recall that the country has an unrestricted migration policy. We envisage an economy that allows any migrant to come. Thus, the decision whether to immigrate or not rests solely with the migrant. Each potential migrant has some reservation income, so that she will migrate if and only if she will be accorded a higher income in the destination country.

Due to various factors (skill, family ties, age, etc.), this reservation income is not the same for all, but there is rather a continuum of reservation incomes. Distinguishing between the two skill groups, we then assume that there is an upward-sloping supply function for each skill group, depending on the income accorded to immigrants in the destination country. Denoting the number of skilled migrants by m_S, the supply function of skilled migrants is given by an isoelastic function:

$$m_S = B_S\left(I_S^M\right)^{\sigma_S}, \tag{10}$$

where B_S and σ_S are positive parameters. Similarly, the supply function of unskilled migrants is given by

$$m_U = B_U\left(I_U^M\right)^{\sigma_U}, \tag{11}$$

where m_U is the number of unskilled migrants and B_U and σ_U are positive parameters.

2.4. Production and Factor Prices

We employ a Cobb–Douglas production function

$$Y = AK^{\alpha}L^{1-\alpha}, \quad A > 0, \quad 0 < \alpha < 1, \tag{12}$$

where Y is the gross domestic product, A is a total factor productivity (TFP) parameter, and α is the capital-share parameter (with $1-\alpha$ the labor-share parameter). L is the total labor supply in efficiency units and is given by

$$L = n_S + \rho n_U + m_S + \rho m_U. \tag{13}$$

The competitive wage per efficiency unit of labor (w) and the competitive interest rate (r) are given by the marginal productivity conditions

$$w = (1-\alpha)A\left(\frac{K}{L}\right)^{\alpha} \tag{14}$$

and

$$r = \alpha A\left(\frac{K}{L}\right)^{1-\alpha}. \tag{15}$$

We assume that capital is immobile across countries. This is meant to say that there is some immobile, nontradable factor, such as land or housing, whose returns are determined in the confines of the domestic economy, and are affected by immigration.

2.5. The Redistribution System

We employ a simple system of redistribution. Wages are taxed at a flat rate of t. The revenues are redistributed by a uniform per capita transfer b.

We assume that the migrants qualify for all the benefits of the welfare state, and they are subject to the state taxes. Therefore, the government budget constraint is as follows:

$$twL = b(1 + m_S + m_U), \tag{16}$$

assuming that the government has no other revenue needs than for redistribution.[9] Note that it follows from equation (16) that t and b must be of the same sign. A positive wage tax (t) allows the government to accord a positive transfer (b) to all. A subsidy to wages (namely, a negative t) requires the government to impose a lump-sum tax (namely, a negative b) on all. When t and b are positive, the tax-transfer system is progressive. When they are negative, the system is regressive.

3. Equilibrium

With unrestricted migration, the flows of migrants m_S and m_U are determined by the migrants themselves according to their reservation incomes (embedded in the supply functions (10) and (11)) and the incomes available to them in the host country. There are therefore only two policy variables – the tax rate t and the social benefit b. However, as the government is constrained by a balanced budget (the condition (16)), it follows that there is essentially only one policy variable; once t is chosen, all the other economic variables are determined in equilibrium, including the tax revenue (twL), the numbers of migrants (m_S and m_U), and b. Or, alternatively, once b is chosen, all the other economic variables are determined in equilibrium.

Choosing t as the single policy variable, we note that there remain 15 endogenous variables:

$$w,\ b,\ r,\ c^*,\ I_S^M,\ I_U^M,\ n_S,\ n_U,\ I_U^N,\ m_S,\ m_U,\ H,\ K,\ Y,\ L.$$

There are also 15 equations in the model, (2)–(16), which are solved for the endogenous variables. In addition, the income of the skilled native-born, which depends on their education cost, is given by the function defined in (1).

9 One may wonder why there is no tax on the initial endowment E, which could be taken to be nondistortive. However, in a dynamic setting, which we have preferred to transform into a static framework, E represents accumulated savings, and taxing it would be distortive. Furthermore, because all native-born possess the same initial endowment, taxing it in our static model would not distribute income across native-born income groups, but rather would amount to transferring income from the native-born to the migrants.

4. Redistribution with and without Migrants' Voting

As explained in the introduction, we aim at studying the effect of migration on the progressivity of the welfare state, and the resulting distribution of disposable income. This depends on the skill composition of migrants and the extent of their integration in the political system, that is, whether or not they participate in the electoral process. We consider the extreme cases: case (a) when migrants do not participate in elections, and case (b) when they do so fully. For each of these two cases, we also study how a skilled migration shock affects the political-economy equilibrium and the ensuing functional and size distribution of income. For this, we resort to numerical simulations.

The policy variable is chosen by a natural (and plausible) version of majority voting, as described below.

Case (a): Migrants do not vote. In this case, the political equilibrium is rather straightforward. Note that if a c_0-individual would like to raise t, then all c-individuals with $c \geq c_0$ (whether skilled or unskilled) would certainly support such a move. This means that the distribution of the voters over the most preferred t is single-peaked. Hence, the t that will be chosen in equilibrium is the median voter's most preferred t.

Note that the story of the immigration to Israel from the former Soviet Union, described in the introduction, is characterized by the immigrants being on average more skilled than the native-born. To focus on this feature we considered the case where

$$\frac{\overline{c}}{2} > c^*,$$

that is, the median native born is unskilled.

Then the median voter is also an unskilled native-born (for ρ sufficiently large, this will indeed be the case). Then the equilibrium t will be at the (endogenously determined) Laffer point. The equilibrium is described in row (a)(1) in table 1, and in figure 2(a,b).

Now suppose that there is a skilled migration supply shock. In order to generate a marked structural change in the political-economy equilibrium, we specifically let B_s rise exogenously from 1.2 to 8.2, whereas B_u is kept unchanged. Note that as immigrants do not vote, the identity of the median voter does not change. As expected, the wage per efficiency unit falls, and the interest rate rises. The policy becomes more progressive. Both t and b rise. Note that the skilled-migration shock is strong, and the number of skilled migrants (m_s) rises sharply even though their income (I_s^M) falls. The fall in their income stems from an increase in the tax (t), which is somewhat offset by the rise in the transfer (b).

In fact, the median voter, who is an unskilled native-born and as such a net beneficiary of the welfare state, encourages an inflow of skilled migrants in order to exploit these net fiscal contributors to the welfare state. Indeed, the tax rate is raised, and more importantly, the social benefit (b) rises significantly. Interestingly, all skilled and unskilled native-born are better off as a result of this supply-side shock of skilled migrants. Note that the native-born unskilled benefit mainly both because the interest rate (r) rises (and they save all of their initial endowment), and because the transfer (b) is more generous.

Table 1(a) and figure 2 suggest also that the average income of the native-born skilled ($\overline{I}_s^N$) rises. It is worth pointing out that *all* skilled native-born (regardless of their cost of education, c) are better off. By revealed preferences, the income of every skilled native-born is at least as high as that of an unskilled native-born, because a skilled person could have chosen to stay unskilled.

Case (b): Migrants vote. Suppose now that migrants do vote. Formally, everything takes place at one point in time, as the model is static. That is, migration, education, and voting decisions, and the resulting factor incomes, are all made simultaneously with the voting decisions, so that voting decisions are made while taking into account the effects of the voting outcome on immigration and all other variables, and vice versa.[10]

Due to the lack of the single peakedness property, we assume a two-stage voting system. First, the majority determines whether the system is progressive or regressive. Second, the largest subgroup determines the parameter values of t and b.

Upon observation, we can see from equations (2) and (9) that the direct effect of the tax-transfer policy on the incomes of the unskilled native-born and the unskilled migrants is the same, and works through the net wage income $(1-t)\rho w + b$. For the unskilled migrant this is the only effect of the tax-transfer system. However, for unskilled native-born, there is also an indirect effect through capital income $E(1+r)$ (note that r depends on t). However, our calculations indicate that this indirect effect is of second-order magnitude compared to the direct effect.

Similarly, upon observation of equations (1) and (2), we can see that the direct effect of the tax-transfer policy on the incomes of the skilled native-born and the skilled migrants is the same and works through the net wage income $(1-t)w + b$. Here again, there is also an indirect effect on the income of the skilled native-born (but not on the income of the skilled migrants) through the capital income $(E-c)(1+r)$. Again, our calculations suggest that the indirect effect is of second-order magnitude.

10 We are essentially assuming perfect foresight. In a dynamic model, it is important to specify the sequencing of decisions. In our static model, the simultaneous determination of all variables may be viewed as a steady state of a dynamic setup.

Thus, all unskilled (both native-born and migrants) are affected by the tax-transfer policy mainly through $(1-t)\rho w+b$, whereas all skilled (both native-born and migrants) are affected mainly through $(1-t)w+b$. It is therefore natural that all the unskilled, whose wage is only ρw, would prefer to tax wage income and take advantage of all the skilled, whose wage, w, is higher. Thus, the most preferred policy of the unskilled entails a positive tax and a positive transfer. Therefore, if the unskilled (both native-born and migrants) constitute a majority, then the political-economy equilibrium tax and transfer will be positive – a progressive tax-transfer system. However, due to the indirect effect, which applies only to the unskilled native-born, the most preferred tax and transfer policy is not necessarily the same for the unskilled native-born and the unskilled migrants. Therefore, the tax-transfer policy chosen is the policy most preferred by the larger of the two subgroups (the unskilled native-born or the unskilled migrants), because the smaller subgroup will naturally support the larger subgroup.[11]

Similarly, the skilled (both native-born, and migrants whose wage is higher than that of the unskilled) would opt to grant a subsidy to the wage, financed by a lump-sum tax. That is, they opt for negative t and b – a regressive tax-transfer policy. In this case too, there is also an indirect effect that applies only to the skilled native-born. Thus, the most preferred tax-transfer policy is not the same for the two subgroups of skilled native-born and skilled migrants. In this case too, we postulate that the political-economy tax-transfer policy is the most preferred policy of the larger subgroup.

Note that indirect effect of the tax-transfer policy, which works through the capital income $(E-c)(1+r)$, is not the same for all members of the skilled native-born subgroup (because it depends on c). In this case, we assume that the median voter within this group prevails.

As before, we start with $B_s = 1.2$, and parameter values that entail the unskilled (both native-born and migrants) as a majority: $x_U + m_U > x_S + m_S$. This is described in row (b)(1) of table 1, and in figure 2(b). As predicted, the political-economy tax-transfer policy is progressive: t and b are positive. Also, the unskilled native-born form a majority of the unskilled: $x_U > m_U$.

We then contemplate a skilled migration supply shock, that is, we keep all other parameter values constant and increase the value of B_S from 1.2 to 8.2 (as in case (a)). The results are described in figure 3(a) and in row (b)(2) of table 1. This supply-side shock triggers a wave of skilled migration. The results are shown in the second row of table 1. The number of migrants (m_S) rose sharply. As a result, the skilled constitute now the majority: $x_S + m_S > x_U + m_U$. Also, the skilled migrants form the larger of the two skilled subgroups (i.e., $m_S > x_S$), and their most preferred tax transfer

11 Note that we implicitly exclude bargaining between the two subgroups.

Table 1

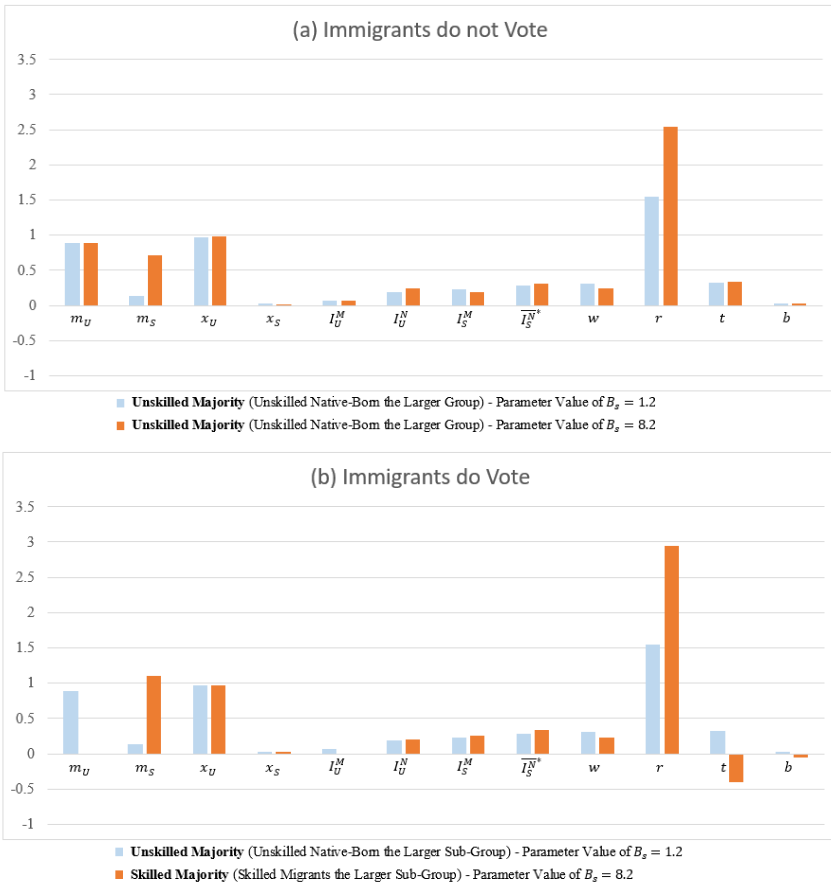
The Effect of a Supply Shock of Skilled Migration: (a) Immigrants Do Not Vote, (b) Immigrants Do Vote

	m_U	m_S	x_U	x_S	I_U^M	I_U^N	I_S^M	$\overline{I_S^N}$ *	w	r	t	b
Immigrants do not Vote												
Case (a)												
(1) Unskilled Majority (Unskilled Native-Born the Larger Group); Parameter Value $B_s = 1.2$	**0.8909**	**0.1380**	**0.9660**	**0.0339**	**0.0632**	**0.194**	**0.236**	**0.281**	**0.312**	**1.553**	**0.3234**	**0.0252**
(2) Unskilled Majority (Unskilled Native-Born the Larger Group); Parameter Value $B_s = 8.2$	**0.8917**	**0.7138**	**0.9811**	**0.0188**	**0.0633**	**0.244**	**0.196**	**0.311**	**0.245**	**2.537**	**0.3382**	**0.0341**
Immigrants Vote												
Case(b)												
(1) Unskilled Majority (Unskilled Native-Born the Larger Subgroup); Parameter Value $B_s = 1.2$	**0.8909**	**0.1380**	**0.9660**	**0.0339**	**0.0632**	**0.194**	**0.236**	**0.281**	**0.312**	**1.553**	**0.3234**	**0.0252**
(2) Skilled Majority (Skilled Migrants the Larger Subgroup); Parameter Value $B_s = 8.2$	**0**	**1.1059**	**0.9666**	**0.0333**	**0**	**0.202**	**0.262**	**0.334**	**0.228**	**2.940**	**-0.4058**	**-0.0577**

Note: In both case (a)(1) and case (b)(1) the unskilled native-born is the decisive voter; in case (b)(2) the skilled migrant is the decisive voter; in case (a)(2) the unskilled native-born is the decisive voter. Since the income of the native skilled population is not constant but a linear function of an individual's c, we report this group's average income. Other (common) parameter values: $B_U = 56$, $\rho = 0.18$, $\overline{c} = 2$, $E = 0.05$, $\alpha = 0.33$, $\sigma_S = \sigma_U = 1.5$, $A = 1$.

now becomes the political-equilibrium tax-transfer policy. As predicted, the political-economy tax-transfer policy becomes now regressive: t and b are negative. Furthermore, the politically dominant subgroup of skilled migrants drives out all unskilled migrants ($m_U = 0$), by according them zero income ($I_U^M = 0$). As skilled labor is assumed a perfect substitute for unskilled labor, the group of skilled migrants have no need for the unskilled migrants, who pose a fiscal burden, and therefore the former drive the number and income of the latter to zero. It is noteworthy that the unskilled native-born were initially the politically dominant subgroup and dictated their most preferred progressive tax transfer. Following the supply-side shock of skilled migration, the unskilled native-born lose their dominance to the skilled migrants, who are now dictating their most preferred regressive tax-transfer policy. Nevertheless, the unskilled native-born are better off, because the return to their capital income (namely, r) rises.

Figure 3
The Effect of a Supply Shock of Skilled Migration



The comparison between the two cases is insightful. When not given the right to vote, the supply-side shock of skilled migration (case (a)) renders the fiscal system more progressive. By contrast, when the migrants have the right to vote (which they fully exercise), they cause the fiscal system to be regressive. Notably, when they are not allowed to vote, the skilled migrants lose and all other income groups gain. When they are allowed to vote, it is the unskilled migrants who lose, and all other income groups gain.

Note that among the model's parameters, there are two crucial ones: ρ and E. The former determines the income gap between skilled and unskilled (both native-born and immigrants). The second determines the income gap between native born and immigrants. If the skilled–unskilled productivity parameter ρ rises, the income gap between skilled and unskilled labor shrinks. If the native-born endowment parameter E rises, the income gap between immigrants and native-born rises. As long as ρ deviates significantly below one (so that there is a marked premium to investment in education), and as long as E is sizable (so that the native-born are in general richer than immigrants are), our qualitative results are likely to hold.

The model helps explain what is shown figure 1: a moderate rise in net income inequality after 2000, which is a combination of declining market income inequality and an offsetting fall in income redistribution. The influx of high-skilled immigrants can explain both: a rising middle class and a rebalanced political-economy equilibrium.

5. Concluding Remarks

This paper develops a model that can explain the mechanism through which a supply-side shock of skilled immigration substantially alters political-economy-based policies. In particular, we show that when migrants do not vote, the fiscal system becomes more progressive. When they do vote, the fiscal system becomes less progressive. In both cases, the native-born gain in net income.

The paper assumes a static model. The dynamics of the interactions between immigration and income redistribution are left for future research.

References

Arian, A., and Shamir, M. (2002), Abstaining and Voting in 2001, in: Arian, A., and Shamir, M. (Eds.), The Elections in Israel–2001, Israel Democracy Institute, Tel Aviv, 150–175.

Avner, U. (1975), Voter Participation in the 1973 Election, in: Arian, A. (Ed.), The Elections in Israel–1973, Academic Press, Tel Aviv, 203–218.

Bird, K. (2011), Voter Turnout among Immigrants and Visible Minorities in Comparative Perspective, in: Bird, K., Saalfeld, T., and Wuest, A. (Eds.), The Political Representation of Immigrants and Minorities: Voters, Parties, and Parliaments in Liberal Democracies, Routledge/ECPR, London, 51–76.

Dahan, M. (2007), Why Has the Labor-Force Participation Rate of Israel Men Fallen?, Israel Economic Review 5, 95–128.

Dahan, M. (2017), Income Inequality in Israel: A Distinctive Evolution, CESifo Working Paper No. 6542.

Gradstein, M., and Schiff, M. (2006), The Political Economy of Social Exclusion with Implications for Immigration Policy, Journal of Population Economics 19, 327–344.

Mayda, A.-M., Peri, G., and Steingress, W. (2015), Immigration to the US: A Problem to Republicans or Democrats, CEPR Discussion Paper 110001.

Messina, A. M. (2007), The Logics and Politics of Post-WWII Migration to Western Europe, Cambridge University Press, New York.

Razin, A. (2017), Israel and the World Economy: Power of Globalization, MIT Press, Cambridge, MA.

Razin, A., and Sadka, E. (1993), The Economy of Modern Israel: Malaise and Promise, University of Chicago Press, Chicago.

Razin, A., and Sadka, E. (2014), Migration States and Welfare States: Why Is America Different from Europe?, Palgrave-Macmillan, London.

Razin, A., Sadka, E., and Swagel, P. (2002a), The Aging Population and the Size of the Welfare State, Journal of Political Economy 110, 900–918.

Razin, A., Sadka, E., and Swagel, P. (2002b), Tax Burden and Migration: A Political Economy Theory and Evidence, Journal of Public Economics 85 167–190.

Razin, A., Sadka, E., and Suwankiri, B. (2011), Migration and the Welfare State: Political-Economy Policy Formation, MIT Press, Cambridge, MA.