Essays in Macroeconomics and Public Finance

by

Christian P. A. Gillitzer

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Economics)
in the University of Michigan
2013

Doctoral Committee:

Professor Joel B. Slemrod, Chair
Professor James R. Hines, Jr.
Assistant Professor Martin C. Schmalz
Professor Matthew D. Shapiro

To my parents.

Acknowledgements

Contents

Chapter 3: Evidence on Unclaimed Charitable Contributions from the Introduction of Third-Party Information Reporting in Denmark   118

List of Tables

List of Figures

Abstract

This dissertation contains three chapters at the intersection of macroeconomics and public finance.

The first chapter demonstrates that deep recessions can stimulate investment in state fiscal capacity. Large negative income shocks endanger the revenue-raising capability of existing narrow tax bases, particularly when the ability to borrow is limited, making an increase in fiscal capacity desirable relative to its implementation cost. An increase in fiscal capacity enables a given amount of revenue to be raised by taxing a wider range of economic activity at lower tax rates, which reduces the efficiency cost of taxation. Evidence from U.S. state governments during the Great Depression supports the model's predictions: governments in states experiencing larger than average negative income shocks were significantly more likely to adopt a retail sales tax (and income taxes) than were governments in states experiencing smaller than average income shocks, and state governments entering the Great Depression with a high level of debt were more likely to adopt new tax bases than those with low levels of debt.

The second chapter proposes a model of consumption commitments—costly adjustment of spending for some goods—that can be easily incorporated into an otherwise standard representative agent DSGE model. The model explains several features of aggregate consumption data: (i) excess smoothness and excess sensitivity; (ii) hump-shaped dynamics; (iii) attenuated response to transitory real interest rate changes; and (iv) some aspects of the equity-premium puzzle. The model provides a microfoundation for reference dependent consumption.

The third chapter, co-authored with Peer Skov, uses a reform in Denmark affecting reporting of charitable tax deductions to shed light on taxpayer behavior. We find that the introduction of information reporting and pre-population of charitable tax deductions in 2008 coincided with a doubling in the number of deductions claimed, and attribute this change to incomplete claiming of eligible deductions under the prior self-reporting regime. We estimate the per-year average amount of forgone tax benefits to be small, but find that many taxpayers repeatedly failed to claim eligible charitable tax deductions under the self-reporting regime.

Chapter 1

What Determines Investment in Fiscal Capacity? The Role of Macroeconomic Income Shocks, Indebtedness and Sovereign Risk

1.1 Introduction

Much of the modern public finance literature takes the set of tax bases as exogenously fixed, and studies the optimal tax rate to levy on those bases. But a defining feature of economic development is growth and compositional change in the set of tax bases used to raise revenue. In developed economies, modern consumption and income tax bases have largely replaced comparatively inefficient trade, seignorage, licensing, and property taxes (except at the local level for property taxes). These tax base changes—improvements in state fiscal capacity—have reduced the efficiency cost of raising revenue and facilitated the growth of government in developed economies; limited ability to raise revenue at tolerable efficiency cost is widely seen as a barrier to economic growth in the developing world today (Besley and Persson, 2009).

Understanding the determinants of improvement in state fiscal capacity is an active area of research. Historians and economists have noted the coincidence of external wars and the upgrading of fiscal capacity, captured in Tilly's (1975, p. 42) famous words "War made the state, and the state made war." A glance at U.S. history is revealing: the first U.S. income tax was proposed during the War of 1812 (although the war ended before the tax was instituted); income taxes were imposed on a small number of taxpayers during the U.S. Civil War; the modern income tax was introduced during World War I, following the ratification of the Sixteenth Amendment to the U.S. Constitution; and during World War II withholding for wage and salary income was introduced, strengthening the ability of tax administrators to enforce the income tax code.

In a seminal body of research, Besley and Persson (2009; 2010; 2013) and Besley, Ilzetzki and Persson (2013) present a framework in which political frictions limit the willingness of governments to invest in fiscal capacity: politicians undervalue the future benefits of higher fiscal capacity because of political turnover, and a lack of social cohesion makes incumbent politicians unwilling to invest in fiscal institutions that can be used by future governments to redistribute money to disfavored groups. External wars act as a stimulant to investment in fiscal capacity in their framework because, in the language of Besley and Persson (2009, p. 1218), military spending is "an archetypal public good representing broadly common interests for citizens." Their framework assumes non-distortionary lump-sum taxes, with fiscal upgrading corresponding to investment in compliance infrastructure that limits evasion and avoidance behavior. In reality, however, upgrading of fiscal institutions includes both improvement in compliance infrastructure and the adoption of new tax bases, both of which lower the efficiency cost of raising revenue.

This paper pursues a complementary explanation for upgrading in fiscal capacity, studying the role of macroeconomic income shocks. A decline in income, particularly in the presence of a high level of indebtedness, endangers the existing tax infrastructure and makes the benefit of an increase in fiscal capacity high relative to its fixed implementation cost. Macroeconomic fluctuations are transitory but, when the fixed cost incurred to upgrade fiscal capacity is large,

improvements in fiscal capacity can be enduring.

This intuition is formalized in a model in which a benevolent government provides a public good by taxing either a narrow share of private consumption goods at a high tax rate or a broad set of goods at a low tax rate. The broader is the tax base—corresponding to a higher level of fiscal capacity—the lower is the marginal efficiency cost of raising tax revenue, because a broader tax base permits the same amount of revenue to be raised at a lower tax rate. In the model, revenue collections vary one-for-one with income (given an unchanged tax rate and tax base) but, matching an empirical regularity, government spending does not; in bad times, a government unable to accumulate debt must immediately cover its revenue shortfall by either raising the tax rate on its existing tax base or undertaking a base broadening reform: sufficiently large income shocks make an increase in fiscal capacity optimal. A government with the ability to borrow can accumulate debt when income is low and postpone the decision on how to meet a revenue shortfall, until net debt repayment is required. Net debt repayment ordinarily occurs when income is high but, with high levels of indebtedness, may occur in bad times, because sovereign risk makes lenders unwilling to permit net lending. Net debt repayment adds to the revenue requirement, raises the required tax rate, increases the distortionary cost of taxation, and thus increases the benefit of improving fiscal capacity. An increase in fiscal capacity may also reduce the cost of borrowing for a heavily indebted government: an increase in fiscal capacity is typically more beneficial in the repayment than default state of the world (because debt repayment makes the revenue requirement larger in the repayment state) which reduces default incentives.

The model's implications are tested by examining the behavior of U.S. state governments during the Great Depression. The size of income shocks during the 1930s was historically unprecedented, creating strong incentives for governments to broaden their tax bases: real per capita GDP for the U.S. as a whole fell by 29 percent between 1929 and 1933. The empirical analysis makes use of the substantial cross-state variation in the magnitude of income shocks experienced: South Dakota experienced a fall in real per capita personal income of 56 percent between 1929 and 1933, while at the other extreme Virginia experienced

a decline of "only" 12 percent over the same period. Across state governments there was also significant heterogeneity in debt levels on the eve of the Great Depression: several states had almost no debt, while Arkansas had a debt-to-revenue ratio of 4.4 in 1929 and defaulted on some of its debt obligations during the Great Depression.

The Great Depression had a profound impact on U.S. state government tax structure. Before the Great Depression no state had a retail sales tax, and few states had individual income or corporate income taxes. But by 1938 22 states had a retail sales tax, and many had also adopted income taxes. States adopting a retail sales tax in the 1930s raised on average about one-fifth of total tax revenue from the retail sales tax by 1942, and the importance of the sales tax as a source of revenue has grown since then. Income taxes introduced during the 1930s were a non-trivial, but less significant, source of revenue.

The cross-sectional pattern of tax base adoption is consistent with the model: states suffering larger than average declines in income were significantly more likely to introduce a retail sales tax than those experiencing smaller than average falls in income. The estimated effect of income shocks is both economically significant and precisely estimated: each 10 percentage point fall in per capita real income is estimated to have increased the probability that a state government adopted a retail sales tax during the 1930s by about 0.18 (the average state experienced a fall in real per capita income of about 30 percent). There was also a tendency for states entering the Great Depression with high debt-to-income ratios to adopt a retail sales tax.

Strikingly, however, state governments entering the Great Depression with little debt tended to not increase deficit spending when incomes fell sharply in the early 1930s. Further, the most indebted states at the onset of the Great Depression had begun reducing the real value of debts before incomes recovered their pre-Depression level. Some of the unwillingness of state governments to engage in deficit spending appears to be due to binding institutional constraints: 20 states had constitutional balanced budget requirements in 1929, while others had constitutional or procedural restrictions placing long-term debt issue in the hands of voters, rather than legislatures (Rodriguez-Tejedo and Wallis, 2010; Ratchford, 1938). But

even for states in which the effective power to issue debt resided in the state government legislature, the magnitude of deficit spending was small compared to the size of income shocks experienced. Some of this behavior appears to be due to a deterioration in borrowing conditions: only a minority of states maintained their pre-Depression Aaa credit rating.

The remainder of the paper proceeds as follows: Section 1.2 lays out and discusses a formal model endogenizing the upgrading of fiscal capacity; Section 1.3 contains the empirical analysis, which uses the behavior of U.S. state governments during the Great Depression to investigate the relationship between macroeconomic income shocks, indebtedness and fiscal capacity; and some concluding thoughts are offered in Section 1.4.

1.2   Model

1.2.1   Overview of the Model

The model assumes a government that raises revenue via a distortionary tax to provide a public good. It is inspired by Yitzhaki (1979) and Wilson (1989), but differs in a number of important ways. Households receive a time-varying income endowment and consume a continuum of private goods, for which the government chooses both the tax rate and the breadth of the tax base (the set of taxed commodities).[1] A broader tax base corresponds to a higher level of fiscal capacity, and these two terms are used interchangeably. Cobb-Douglas utility is assumed because with these preferences a uniform rate on all taxed goods is optimal, permitting the analysis to use a single tax rate and thus sidestep the issue of differentiated tax rates among goods, which is not a central issue in this context.

An increase in tax base breadth lowers the excess burden of taxation, because there are fewer untaxed goods for taxpayers to substitute toward, but raises marginal administrative cost, which is assumed to be smoothly increasing in tax base breadth. As an example of the relationship between tax base breadth and marginal administrative cost, expanding the

---

[1] Slemrod and Kopczuk (2002) analyze the optimal setting of tax base breadth in an extended version of this model with taxpayers of differing ability levels and a social preference for redistribution.

sales tax base to include services is widely believed to raise administrative cost because service transactions are generally more costly to observe, and therefore to tax, than are goods transactions.[2] At an optimum, the reduction in the excess burden of taxation from a marginal increase in tax base breadth is equal to the marginal increase in administrative costs. Because excess burden is convex in the tax rate levied on a given tax base, the optimal breadth of the tax base is increasing in the government's revenue requirement.

The model is written in terms of a commodity tax base but, by letting the set of untaxed goods be income tax deductible items, the model can be interpreted as a description of the income tax base as well. Much of the discussion that follows emphasizes the sales tax interpretation of the model because income tax changes were a less important development for U.S. state government finances during the 1930s than was adoption of the sales tax.

Two variants of the model are solved: a special case in which the government must run a per-period balanced budget (20 U.S. states had such restrictions in 1929), and a general version in which tax revenue shortfalls can be smoothed via sovereign borrowing (U.S. state governments are sovereign with respect to their debts). To model the interdependence between income shocks, fiscal capacity, and sovereign risk, the model of optimal tax base breadth sketched out above is incorporated into an Eaton and Gersovitz (1981) type model of sovereign debt accumulation. Modeling the interaction between fiscal capacity and sovereign risk is a key contribution of this paper: few models of sovereign borrowing explicitly include a government sector, and, to the best of my knowledge, none have endogenized tax base breadth.

In the Eaton and Gersovitz (1981) model, borrowers are willing to lend to sovereign governments because default imposes reputation costs on borrowers that preclude access to international financial markets; governments are willing to maintain their repayment obligations for as long as the expected present value costs of financial autarky exceed the cost of repayment. While the threat of financial autarky is sufficient at a theoretical level to support sovereign borrowing, plausible calibrations indicate that the increase in consumption

---

[2]Note that there may be cases where administrative cost is not smoothly increasing in tax base breadth; excluding a small number of commodities from the tax base is likely to raise collection costs because extra resources are required to police avoidance and evasion opportunities provided by exemptions.

volatility under financial autarky imposes small costs and can support only small amounts of debt in equilibrium (see Aguiar and Gopinath, 2006, and Lucas, 1987).[3] The reputation cost model is now widely used in the modern quantitative sovereign debt literature but, in order to yield a plausible quantitative description of debt dynamics, these models generally assume there is also an output cost experienced upon default (see Arellano, 2008, and Aguiar and Amador, forthcoming).

The following subsection begins the formal description of the model.

1.2.2 Model Setup

1.2.2.1 Privately Consumed Goods

There is a representative consumer who has Cobb-Douglas period utility for privately consumed goods given by

$$u_t = \int_0^1 \alpha_i log\,(c_{i,t})\,di, \tag{1.2.1}$$

where the sum of the parameters α_i has been normalized to unity: $\int_0^1 \alpha_i di = 1$. At an optimum for the consumer, the parameter α_i is equal to the consumer's expenditure share on good $i \in [0,1]$. The representative consumer receives an exogenous income endowment A_t, and spends all their income each period. The assumption of an endowment economy means that there is no labor/leisure trade-off, and that leisure is not included in the set of goods $i \in [0,1]$. The income endowment A_t is assumed to evolve according to the autoregressive process $A_t = \rho A_{t-1} + (1-\rho)\bar{A} + \varepsilon_t$, where ρ is the autocorrelation parameter, $\bar{A} = 1$ is the steady-state income level, and $\varepsilon_t \sim iid\,(0,\sigma_\varepsilon^2)$. As a consequence of the representative agent assumption, there are no differences in income or time preference across households that would give rise to borrowing or lending. However, the government may be able to borrow

---

[3]The Eaton-Gersovitz equilibrium relies on governments being unable to save at the world interest rate, and thereby accumulate surpluses. Bulow and Rogoff (1989b) show that if a government can save at the world interest rate, then the reputation cost model can sustain no debt in equilibrium: rather than repay debt, a sovereign government would be better of by defaulting at its point of maximum indebtedness and using the funds that would have gone to debt repayment to self-insure against fluctuations in national income.

outside the jurisdiction on behalf of households. The set of goods $i \in [0, I_t]$ are subject to a uniform tax rate, and the remaining set of goods $i \in (I_t, 1]$ are not taxed. The larger is the index of taxed goods, I_t, the broader is the tax base because a wider set of commodities is subject to tax. The expenditure share of taxed goods is $b(I_t) \equiv \int_0^{I_t} \alpha_i di$ and, because $b(I_t)$ is a monotonic transformation of I_t, the planner can equivalently set the breadth of the tax base by choosing $b(I_t)$ or I_t. Normalizing the exogenous pre-tax price of all goods to unity, households face the price $p_{i,t} = \frac{1}{1-\tau_t}$ for goods $i \in [0, I_t]$, and $p_{i,t} = 1$ for goods $i \in (I_t, 1]$. The utility-maximizing choice of privately consumed goods is:

$$
c_{i,t} = \begin{cases} (1-\tau)\alpha_i A_t & for \ i \leq I_t \\ \\ \alpha_i A_t & for \ i > I_t, \end{cases}
$$

(1.2.2)

implying period indirect utility for privately-consumed goods equal to

$$
\widetilde{v}(A_t, \tau_t, b_t) = \gamma + log A_t + b_t log(1-\tau_t),
$$

(1.2.3)

where $\gamma \equiv \int_0^1 \alpha_i log(\alpha_i) di$ is a constant. *Ceteris paribus*, utility from privately consumed goods is increasing in the income endowment A_t, decreasing in the tax rate τ_t, and decreasing in the share of goods subject to tax b_t.

The excess burden of taxation is the cost to the representative consumer of paying taxes on a narrow tax base relative to a world in which they remit the same tax liability via a lump-sum tax.[4] Measured in units of utility, the excess burden of taxation for a tax policy that collects tax revenue $R = \tau_t b_t A_t$ is equal to:

$$
\begin{aligned} EB(A_t, \tau_t, b_t) &\equiv \widetilde{v}(A_t - R, 0, b_t) - \widetilde{v}(A_t, \tau_t, b_t) \\ &= log(A_t - R) - [log(A_t) + b_t log(1-\tau_t)]. \end{aligned}
$$

(1.2.4)

In what follows, this excess burden of taxation will sometimes be equivalently referred as

---

[4]In this model, a comprehensive tax base is equivalent to a lump-sum tax, but is assumed to be prohibitively expensive to administer.

efficiency cost, or deadweight loss. Next, consider the reduction in the excess burden of taxation due to a revenue-neutral marginal increase in tax base breadth. This can be found by differentiation of Equation (1.2.4) with respect to b_t, subject to the requirement that tax-revenue collected is unchanged: $R = \tau_t b_t A_t$. The decline in excess burden is:

$$
\begin{aligned}
\frac{\partial EB}{\partial b_t} &= -log\left(1 - \tau_t\right) - \frac{\tau_t}{1 - \tau_t} \\
&\simeq -\frac{\tau_t^2}{2},
\end{aligned}
\tag{1.2.5}
$$

where the approximate equality in Equation (1.2.5) follows from taking a second-order Taylor series approximation around $\tau = 0$. The decline in excess burden due to a marginal increase in tax base breadth is approximately proportional to the tax rate squared; as discussed in detail later, this convexity plays a crucial role in explaining the occurrence of tax base changes.[5]

Reflecting the fact that the marginal excess burden of a tax rate increase is decreasing in tax base breadth, Slemrod and Kopczuk (2002, p. 108) show that the compensated tax base elasticity (TBE) with respect to the net-of-tax price of goods in this setting is equal to the expenditure-weighted share of goods not subject to tax:

$$
TBE_t \equiv \frac{\partial \left[b_t A_t\right]^c}{\partial \left(1 - \tau_t\right)} \frac{\left(1 - \tau_t\right)}{b_t A_t} = \left(1 - b_t\right).
\tag{1.2.6}
$$

Because the parameter b_t controls both the share of goods subject to tax and the tax base elasticity, the model cannot, in general, match both these statistics for a real-world commodity tax base. The direct dependence of the TBE on tax base breadth can be relaxed by assuming more general constant elasticity of substitution preferences (see Slemrod and Kopczuk 2002, fn. 11), but the loss in tractability exceeds any gain in realism for the application here. In Section 1.2.5, where the model is calibrated and solved quantitatively, b_t is chosen to match the aggregate TBE, not the empirically observed tax base breadth.

---

[5]Note that the increase in excess burden due to a marginal increase in the tax *rate* is approximately $\frac{1}{2}\tau_t b_t$, near $\tau = 0$.

1.2.2.2 Public Good Provision and Administrative Costs

Revenue raised by taxation of the privately consumed goods is used to fund provision of a public good, and net repayment of debt obligations incurred. The government's period budget constraint is given by:

$$G_t = \tau_t b_t A_t - \xi\left(b_t, b_{t+1}\right) + q_t D_{t+1} - D_t, \tag{1.2.7}$$

where G_t is spending on the public good, $\tau_t b_t A_t$ is revenue raised from proportional commodity taxes at rate τ_t on a tax base with breadth b_t at an income level A_t, D_{t+1} is the face value of a one-period discount bond repayable at time $t+1$, and q_t is the price of that bond today. The administrative cost function $\xi\left(b_t, b_{t+1}\right)$ depends on both the current level of fiscal capacity, and next period's level of fiscal capacity:

$$\xi\left(b_t, b_{t+1}\right) \;=\; \xi_f\left(b_t\right) + \xi_F\left(b_t, b_{t+1}\right), \tag{1.2.8}$$

where $\xi_f\left(b_t\right)$ is the per-period fixed cost to administer a tax base with breadth b_t, and $\xi_F\left(b_t, b_{t+1}\right)$ is the fixed setup cost incurred in undertaking a tax base broadening reform that expands the expenditure share of commodities subject to tax from b_t to b_{t+1}. There is no fixed setup cost associated with a tax base narrowing reform. The per-period administrative cost $\xi_f\left(b_t\right)$ is assumed to be convex in tax base breadth, reflecting the fact that purchases of some commodities are more difficult to observe, and therefore more costly to include in the tax base. Although in practice it may sometimes be less costly to administer a tax base with few exceptions than a tax base with many exceptions, the assumption that $\xi_f\left(b_t\right)$ is convex in tax base breadth captures the underlying tradeoff faced by tax administrators: an increase in tax base breadth reduces the excess burden of taxation, but raises collection costs.

The fixed setup cost incurred when fiscal capacity is *upgraded* includes all the expenses incurred by the tax administration, such as training employees and purchasing equipment, in readying operations to collect revenue on a new tax base. The infrequency with which

we observe fundamental changes in tax base breadth—such as the introduction, or repeal, of an income or sales tax base—suggests that the one-time fixed cost incurred in changing tax base breadth is large; were this cost small, we would expect tax bases introduced when revenue needs are unusually high to be repealed when revenue needs return to normal levels, and the flow cost of administering a broad tax base exceeds the benefit afforded via reduced efficiency cost of raising revenue. An increase in tax base breadth chosen this period becomes operational in the next period, but the fixed setup cost is incurred today; in contrast, the per-period fixed administrative cost $\xi_f(b_t)$ is incurred only in periods in which the tax base is operational.

Utility is additively separable between private and public good consumption, with period semi-indirect utility for the representative household from consumption of private and public goods given by:

$$v\left(A_t, \tau_t, b_t, b_{t+1}, G_t\right) = \widetilde{v}\left(A_t, \tau_t, b_t\right) + \phi log\left(G_t - \lambda \bar{G}\right), \tag{1.2.9}$$

where ϕ parameterizes the importance of the public good relative to privately consumed goods. Note that utility depends on next period's level of fiscal capacity via the resource cost of any fixed setup cost incurred, which is reflected, holding the tax rate constant, in the level of public good provision (see Equation 1.2.8). Utility for the public good depends on its provision relative to the reference point $\bar{G}$, with the parameter $\lambda > 0$ controlling the weight placed on the reference point.

This specification of public good utility is introduced to model the empirical regularity of counter-cyclical business-cycle variation in government spending as a share of income. When λ is large there is a high degree of curvature in the public good utility function around the reference point, and the representative household prefers greater business-cycle variation in spending on privately-consumed goods than the public good. Hence, the larger is λ, the smaller is the optimal degree of variation in public spending over the business cycle.

The reference level $\bar{G}$ is assumed to be equal to the steady-state (when income A_t is at its

mean) level of public good provision G^{ss}, in which case the marginal utility of government spending, and thus the optimal size of government, in the steady-state is unaffected by the weight placed on the reference level. To see this, note that when $G^{ss} = \bar{G}$, $\phi log\left(G^{ss} - \lambda \bar{G}\right) = \phi log\left(1 - \lambda\right) + \phi log\left(G^{ss}\right)$, and the weight on the reference level λ does not affect marginal utility of the public good. For simplicity, the reference point $\bar{G}$ is assumed to not vary with tax base breadth, even though, in principle, the optimal level of public good provision does depend on the efficiency cost of raising tax revenue, via tax base breadth.[6]

1.2.3 Optimal Commodity Tax Rate

The government enters each period with tax base breadth b_t (determined in the previous period), a level of debt D_t that must be repaid today, and knowing the current period's income endowment A_t. The level of borrowing D_{t+1}, the tax rate τ_t, and next period's tax base breadth b_{t+1} are chosen each period to maximize the discounted value of lifetime utility. The model assumes no frictions affecting the choice of the tax rate τ_t each period, making its choice a static problem conditional on choices for D_{t+1} and b_{t+1}. Taking tax base breadth and net borrowing as given, the planner maximizes welfare for the representative taxpayer (given by Equation 1.2.9), subject to the budget constraint (given by Equation 1.2.7). The resulting first-order condition provides an implicit expression (i.e., conditional on tax base breadth and the amount of debt issued in the current period) for the optimal tax rate:

$$\tau\left(A_t, D_t, D_{t+1}, b_t, b_{t+1}\right) = \frac{\phi A_t + \xi\left(b_t, b_{t+1}\right) + \lambda \bar{G} + D_t - q_t D_{t+1}}{A_t\left(b_t + \phi\right)}. \tag{1.2.10}$$

The optimal tax rate is higher the greater is net debt repayment $D_t - q_t D_{t+1}$, and administrative costs $\xi\left(b_t, b_{t+1}\right)$. Holding net debt repayment and tax base breadth constant, an increase in the income endowment A_t reduces the optimal tax rate, and an increase in the public good preference parameter ϕ increases the steady-state optimal tax rate. The larger

---

[6]Assuming utility for public consumption is globally more concave than utility for private consumption would also lead households to desire less variation in private than public consumption over the business cycle. However, an assumption of globally more concave utility for public than private consumption is difficult to defend in light of the growth in government spending as a share of income over the 20th century.

is λ, the greater is the required increase in the tax rate when income declines to fund the desired level of public good provision. Holding administrative costs constant, a broader tax base requires a lower tax rate to raise a given amount of revenue than a narrow tax base, and thus the optimal tax rate is decreasing in tax base breadth.

Making use of the implicit expression for the optimum tax rate (Equation 1.2.10), let

$$\hat{v}\left(A_t, D_t, D_{t+1}, b_t, b_{t+1}\right) \equiv v\left(A_t, \tau_t^*, b_t, G_t^*\right) \tag{1.2.11}$$

represent utility for the representative taxpayer evaluated at the optimum tax rate (conditional on D_{t+1} and b_{t+1}), where $\tau_t^* = \tau\left(A_t, D_t, D_{t+1}, b_t, b_{t+1}\right)$ and $G_t^* = \tau_t^* b_t A_t - \xi\left(b_t, b_{t+1}\right) + q_t D_{t+1} - D_t$. This substitution re-expresses welfare for the representative household in terms of only two choice variables for the planner at time t, D_{t+1} and b_{t+1}, simplifying the analysis that follows.

1.2.4 Optimal Tax Base Breadth: Balanced Budget Requirement

Before considering government behavior when borrowing can be used to intertemporally smooth revenue shocks, this subsection describes the determination of optimal tax base breadth when there is a binding per-period balanced budget requirement. Faced with a tax revenue shortfall, a government with a balanced budget requirement, unwilling to accommodate a fall in public spending equal to the fall in tax revenue, faces an immediate choice of raising revenue by either levying a higher tax rate on its existing tax base or undertaking a tax base broadening reform. In 1929, 20 U.S. states had a balanced budget requirement, making this special case particularly relevant for the empirical analysis that follows.

With a balanced budget requirement, debt is equal to zero each period ($D_t = D_{t+1} = 0$) and the optimal tax rate (given by Equation 1.2.10) simplifies to $\tau_t^{bb} \equiv \tau\left(A_t, 0, 0, b_t, b_{t+1}\right)$. Noting that tax base breadth at time t is predetermined (chosen in the previous period), Equation (1.2.10) reveals that the optimal tax rate varies inversely with income: as income falls, the value of the tax base ($b_t A_t$) declines and, holding tax base breadth constant, the tax rate must

rise to collect enough revenue to fund administrative costs, which do not vary with income, and to fund provision of the public good, for which demand varies less than one-for-one with income.

A revenue-neutral increase in tax base breadth permits a reduction in the tax rate, which reduces the excess burden of taxation, but a tax base broadening reform incurs higher per-period administrative cost, and a fixed setup cost to implement the increase in tax base breadth. Defining $\hat{v}_t^{bb} \equiv \hat{v}(A_t, 0, 0, b_t, b_{t+1})$, the increase in per-period welfare at time t from having a marginally broader tax base is given by

$$\frac{\partial \hat{v}_t^{bb}}{\partial b_t} = \underbrace{\left[log\left(1 - \tau_t^{bb}\right) + \frac{\tau_t^{bb}}{1 - \tau_t^{bb}} \right]}_{\Lambda_{0,t}} - \underbrace{\frac{\xi_f'(b_t)}{A_t\left(1 - \tau_t^{bb}\right)}}_{\Lambda_{1,t}}. \tag{1.2.12}$$

The term $\Lambda_{0,t}$ measures the reduction in the excess burden of taxation from having a marginally broader tax base (see Equation 1.2.5), and $\Lambda_{1,t}$ is the utility cost of paying for the associated marginal increase in per-period administrative costs. If there were no fixed setup cost incurred to increase tax base breadth the planner would optimally set $\Lambda_{0,t} = \Lambda_{1,t}$ in each period.

The sign of the cross-partial derivative $\partial^2 \hat{v}_t^{bb}/\partial b_t \partial A_t$ is critical: if it is negative then the benefit of having a marginally broader tax base *rises* as income *falls*, and we should expect governments with a balanced budget requirement to be more likely to undertake a tax base broadening reform when income is low than when it is high (and vice versa). In the neighborhood of an optimum for tax base breadth (i.e., assuming $\Lambda_{0,t} = \Lambda_{1,t}$), a marginal increase in income raises the benefit in the current period of having a marginally broader tax base by

$$\left.\frac{\partial^2 \hat{v}_t^{bb}}{\partial b_t \partial A_t}\right|_{\Lambda_0 = \Lambda_1} = \frac{1}{1 - \tau_t^{bb}} \left[-\frac{\partial \tau_t^{bb}}{\partial A_t} log\left(1 - \tau_t^{bb}\right) + \frac{1}{A_t}\left(\tau_t^{bb} + \left(1 - \tau_t^{bb}\right) log\left(1 - \tau_t^{bb}\right)\right) \right]. \tag{1.2.13}$$

Evidence from the U.S. states during the 1930s, discussed in detail in Section 1.3, indicates that state governments experiencing large negative income shocks were significantly more

likely to undertake tax base broadening reforms than states experiencing relatively modest declines in income. This suggests it is important to understand the conditions under which (in the neighborhood of an optimum for tax base breadth) $\partial^2 \hat{v}_t^{bb}/\partial b_t \partial A_t$ is negative. With some manipulation of Equation (1.2.13), it can be shown that a sufficient condition for a negative cross-partial derivative (for a government with a balanced budget requirement) is an elasticity of the optimal tax rate with respect to income less than negative one-half:

$\varepsilon_{\tau^{bb},A} \equiv \frac{\partial \tau^{bb}}{\partial A} \frac{A}{\tau^{bb}} < -\frac{1}{2}.^{7}$

The elasticity of the optimal tax rate with respect to the income endowment $(\varepsilon_{\tau,A})$ depends, among other factors, on the optimal degree of variation in public good provision over the business cycle. The smaller is the variation in public good provision that the representative household desires in response to fluctuations in income the larger, in absolute value, is the elasticity of the tax rate with respect to income. Intuitively, when demand for the public good is insensitive to variation in income, the tax rate must rise sharply when income falls to prevent a revenue shortfall; the sharp rise in the tax rate raises the excess burden of taxation—and so the benefit afforded by a marginal increase in tax base breadth. For a government with a per-period balanced budget requirement, it can be shown that, making use of Equation (1.2.10), for all values of the endowment $\varepsilon_{\tau,A} < -\frac{1}{2}$ provided revenue needs that do not vary with income are sufficiently large: $(\xi_t + \lambda \bar{G}) > \phi A_{max}$.

Figure 1.1 graphs $\Lambda_{0,t}$ and $\Lambda_{1,t}$ as a function of income, assuming that tax base breadth b_t is at an optimum (ignoring the fixed setup component of administrative costs) when income is at its mean level $\bar{A}$. Under the assumption that $(\xi_t + \lambda \bar{G}) > \phi A_{max}$, the curve $\Lambda_{0,t}$ cuts the curve $\Lambda_{1,t}$ from above as a function of income. For values of the income endowment below $\bar{A}$, the vertical distance between $\Lambda_{0,t}$ and $\Lambda_{1,t}$ shown in Figure 1.1 indicates the increase in utility from having a marginally broader tax base in period t. Recall though that tax base breadth in the current period is determined in the previous period, and an increase in tax base breadth requires payment of a fixed setup cost. If income in subsequent periods is expected to be only briefly below $\bar{A}$, a tax base broadening reform is undesirable because the social

---

[7] The tax revenue income elasticity is equal to one plus the tax rate income elasticity: $\varepsilon_{R^{bb},A} = 1 + \varepsilon_{\tau^{bb},A}$.

cost of funding the fixed setup cost exceeds the benefit of having a marginally broader tax base. In contrast, an expected prolonged decline in the income endowment makes it optimal to incur the fixed cost to expand tax base breadth.

The next section relaxes the balanced budget restriction, outlining a model of optimal tax base breadth determination in the presence of sovereign borrowing.

1.2.5   Model of Optimal Tax Base Breadth with Debt

A government that does not face a binding balanced budget requirement chooses $\{D_{t+1}, b_{t+1}\}$ each period to maximize the discounted sum of expected lifetime social welfare. Formally, the maximization problem at time t is:

$$V^o\left(A_t, D_t, b_t\right) = max\left\{V^c\left(A_t, D_t, b_t\right), V^d\left(A_t, b_t\right)\right\}, \tag{1.2.14}$$

where $V^0\left(A_t, D_t, b_t\right)$ is expected discounted lifetime utility for the representative household at time t conditional on the government having access to foreign lenders. It is the maximum of the expected utility provided by two choices available at time t: either maintain contractual obligations and receive expected discounted lifetime utility $V^c\left(A_t, D_t, b_t\right)$, or default on debt obligations and receive expected discounted lifetime utility $V^d\left(A_t, b_t\right)$. Default is assumed to result in full repudiation of debt obligations. The value function conditional on maintaining debt obligations is itself the solution to the following maximization problem:

$$V^c\left(A_t, D_t, b_t\right) = \max_{\{D_{t+1}, b_{t+1}\}}\left\{\hat{v}\left(A_t, D_t, D_{t+1}, b_t, b_{t+1}\right) + \beta E_t V^o\left(A_{t+1}, D_{t+1}, b_{t+1}\right)\right\}. \tag{1.2.15}$$

If the government chooses to default on its debt obligations expected discounted lifetime social welfare is given by:

$$V^d\left(A_t, b_t\right) = \max_{b_{t+1}}\left\{\hat{v}\left(A_t\left(1 - \omega\right), 0, 0, b_t, b_{t+1}\right)\right.$$
$$\left. + \beta E_t\left[\left(1 - \pi\right)V^d\left(A_{t+1}, b_{t+1}\right) + \pi V^o\left(A_{t+1}, 0, b_{t+1}\right)\right]\right\}, \tag{1.2.16}$$

with ω parameterizing the output cost experienced during financial autarky. Upon default the government loses access to credit markets, but regains access with state-independent probability π each period; the government cannot save and must match its revenues and expenditures period-by-period. Financial autarky has a cost to households via an increase in the volatility of some combination of private consumption, public good provision, and the tax rate.

Letting the probability of default next period be given by $\delta\left(A_t, D_{t+1}, b_{t+1}\right)$, perfectly competitive risk-neutral lenders set a price for debt to earn an expected rate-of-return r per period:[8]

$$q\left(A_t, D_{t+1}, b_{t+1}\right) = \frac{1 - \delta\left(A_t, D_{t+1}, b_{t+1}\right)}{1 + r}. \qquad (1.2.17)$$

Lenders require a deeper discount on the face value of debt as the probability they will be repaid falls. If the government enters period t with debt D_t and fiscal capacity b_t, repayment is optimal for the set of income realizations defined by:

$$R\left(D_t, b_t\right) = \left\{ A_t \ : \ V^c\left(A_t, D_t, b_t\right) > V^d\left(A_t, b_t\right) \right\}, \qquad (1.2.18)$$

implying that default is optimal for the set of income realizations defined by the complement set of income realizations: $N\left(D_t, b_t\right) = R^c\left(D_t, b_t\right)$.

Letting $\Gamma_t = \{A_t, D_t, b_t\}$, a recursive competitive equilibrium for this economy consists of (i) policy functions for the tax rate $\tau\left(A_t, D_t, D_{t+1}, b_t, b_{t+1}\right)$, public consumption G_t, and private consumption $c_{i,t}$ for $i \epsilon\left[0, 1\right]$, (ii) the government's chosen debt level $D_{t+1}\left(\Gamma_t\right)$, repayment sets $R\left(D_t, b_t\right)$ and default sets $N\left(D_t, b_t\right)$; (iii) the policy function for next period's level of fiscal capacity $b_{t+1}\left(\Gamma_t\right)$; and (iv) the bond price function $q\left(A_t, D_{t+1}, b_{t+1}\right)$ such that:

1. The policy functions for the government and the representative household satisfy their respective budget constraints each period;

---

[8]The Eaton-Gersovitz model has been extended to allow for risk-averse pricing (see Arellano, 2008).

2. Taking the bond price schedule as given, the policy functions are optimal;

3. The bond price schedule reflects the government's default probability for each history of the state variables Γ_t (i.e., in expectation competitive lenders make zero profit).

Because the value function V^o is neither globally concave nor differentiable, being the maximum of two other value functions, the model, even without the incorporation of endogenous fiscal capacity, yields few analytical insights (see Aguiar and Amador, forthcoming). However, the model can be solved numerically. But before doing so, in the next subsection I analyze a simplified two-period version of the model that permits analytic expressions with a rich set of implications.

1.2.6   Optimal Tax Base Breadth with Debt: Two-Period Model

This section considers a simple two-period version of the general model described above. The income endowment in the second period is expected to be higher than in the first period, creating a desire borrow in the first period. The government enters the first period with no debt, and chooses the level of borrowing and the second period's level of tax base breadth. The model ends in the second period and debt incurred in the first period is due for repayment: if the government defaults on its debt there is a proportional output cost ω. Borrowing is sustained only by the output cost experienced upon default, unlike in the general case where threat of financial autarky also helps sustain borrowing. These simplifications imply the following maximization problem for the planner in the first period:

$$\max_{\{D_{t+1}, b_{t+1}\}} \hat{v}\left(A_t, 0, D_{t+1}, b_t, b_{t+1}\right) + \beta \int_{A_{min}}^{A_{max}} max\left\{\hat{v}^c_{t+1}, \hat{v}^d_{t+1}\right\} dF\left(A_{t+1}|A_t\right), \qquad (1.2.19)$$

where $\hat{v}^c_{t+1} \equiv \hat{v}\left(A_{t+1}, D_{t+1}, 0, b_{t+1}, b_{t+2}\right)$ is utility under repayment, $\hat{v}^d_{t+1} \equiv \hat{v}(A_{t+1}(1-\omega), 0, 0, b_{t+1}, b_{t+2})$ is utility under default, and $F\left(A_{t+1}|A_t\right)$ is the cumulative density function for the income endowment in the second period, conditional on its level in the first period. Because the model ends in the second period, it is assumed that there is no investment in fiscal capacity made in the second period: $b_{t+2} \leq b_{t+1}$.

The model is solved by backward induction, first considering the decision to repay or default on debt obligations in the second period. Because there is no commitment technology binding the government in the final period, the decision to repay or default on debt obligations depends only on which choice gives higher second-period utility: default is optimal if $\hat{v}_{t+1}^d > \hat{v}_{t+1}^c$. With some algebra (see the appendix), it can be shown that default is optimal for the set of endowment income realizations $A_{t+1} \in \left[A_{min}, A_{t+1}^d \right]$, where

$$A_{t+1}^d \equiv \frac{D_{t+1} + \left(1 - \varphi\left(b_{t+1}\right)\right)\left(\xi_{t+1} + \lambda \bar{G}\right)}{b_{t+1}\left(1 - (1 - \omega)\varphi\left(b_{t+1}\right)\right)} \tag{1.2.20}$$

is the threshold income level below which default is optimal, and $\varphi\left(b_{t+1}\right) \equiv (1 - \omega)^{\left(\frac{1 - b_{t+1}}{\phi + b_{t+1}}\right)}$. That default occurs for low income realizations is a general feature of models with non-state-contingent debt (see Aguiar and Amador, forthcoming). Because the cost of repayment rises with indebtedness, but the output cost of default does not, the threshold A_{t+1}^d, and therefore the likelihood of default, is increasing in the amount of debt due for repayment; this is also a general feature of the Eaton-Gersovitz framework (see Arellano, 2008).

With a comprehensive tax base in the second period ($b_{t+1} = 1$), the default threshold given by Equation (1.2.20) simplifies to $A_{t+1}^d = (D_{t+1}/\omega)$: default is optimal whenever debt due for repayment is greater than the output cost of default. The smaller is the output cost of default, the lower is the maximum amount of borrowing that can be sustained in equilibrium.

Because a broader tax base reduces the excess burden of raising revenue in both the repayment and default states, it is not immediately clear how tax base breadth affects default incentives. Critically, but demonstrated in the appendix, under a weak condition on initial tax base breadth, the default threshold A_{t+1}^d is decreasing in the level of fiscal capacity: $\partial A_{t+1}^d / \partial b_{t+1} < 0$. This implies that the default set $\left[A_{min}, A_{t+1}^d \right]$ is shrinking in the level of second-period tax base breadth. To understand this result intuitively, recall first that the reduction in the excess burden of taxation from a marginal increase in tax base breadth is increasing in the tax rate, and thus the revenue requirement (see Section 1.2.2.1). An increase in tax base breadth reduces excess burden in the repayment state by more than in the default state (in

the neighborhood of the default threshold) because—on account of the stock of debt due for repayment—the revenue requirement, and so the tax rate, is greater in the repayment state. In contrast, the change in administrative costs from a marginal increase in tax base breadth reduces resources equally in the repayment and default states.

The probability of default is given by $\delta\left(A_t, D_{t+1}, b_{t+1}\right) = Pr\left(A_{t+1} < A_{t+1}^d|A_t\right) = F\left(A_{t+1}^d|A_t\right)$, and as a consequence of the assumption of competitive risk neutral lenders, the discount bond price in the first period is equal to $q\left(A_t, D_{t+1}, b_{t+1}\right) = \frac{1-F\left(A_{t+1}^d|A_t\right)}{1+r}$ (see Equation 1.2.17). Paying the fixed setup cost in the first period to increase second-period tax base breadth reduces the default threshold $\left(A_{t+1}^d\right)$, and thus the cost of borrowing, because the government is less likely to renege on its debt obligations the higher is the level of fiscal capacity in the second period.

With some manipulation, the first-order condition for optimal second-period tax base breadth (chosen in the first period) can be expressed as follows:

$$\beta\left[\int_{A_{min}}^{A_{t+1}^d}\frac{\partial\hat{v}_{t+1}^d}{\partial b_{t+1}}dF_{t+1} + \int_{A_{t+1}^d}^{A_{max}}\frac{\partial\hat{v}_{t+1}^c}{\partial b_{t+1}}dF_{t+1}\right] = MU\left[\frac{\partial\xi_t}{\partial b_{t+1}} - \frac{\partial q_t}{\partial b_{t+1}}D_{t+1}\right] \qquad (1.2.21)$$

$$= MU\left[\frac{\partial\xi_t}{\partial b_{t+1}} + f_{t+1}\frac{D_{t+1}}{1+r}\frac{\partial A_{t+1}^d}{\partial b_{t+1}}\right], \qquad (1.2.22)$$

where MU is the marginal utility of spending on the public good in period t, $q_t = q(A_t, D_{t+1}, b_{t+1})$, $f_{t+1} = f\left(A_{t+1}^d|A_t\right)$, and $dF_{t+1} = dF\left(A_{t+1}|A_t\right)$. The expression on the left-hand-side of Equation (1.2.21) is the expected benefit in the second period of having a marginally broader tax base; at an optimum, it is balanced against the cost incurred in the first period to increase tax base breadth, shown by the expression on the right-hand-side of Equation (1.2.21). The term $\partial\xi_t/\partial b_{t+1}$ is the marginal administrative cost (in dollars) incurred in the first period to marginally increase second-period tax base breadth; the term $\partial q_t/\partial b_{t+1}$ is the change in the cost of debt (per dollar of repayment next period) with respect to a marginal increase in tax base breadth, which, recalling the discussion above, is decreasing in second-period tax base breadth. Hence, cost incurred in the first-period to expand tax base breadth is partly offset

by a reduction in the cost of borrowing. Equation (1.2.22), which expands the right-hand-side of Equation (1.2.21), indicates that the cost of borrowing falls by more the larger is the decrease in the default threshold with respect to a marginal increase in tax base breadth, $\partial A_{t+1}^d/\partial b_{t+1} < 0$, and the greater is the probability density of output at the default threshold, $f_{t+1} = f\left(A_{t+1}^d|A_t\right)$. Relative to a world in which borrowers can commit to repayment, the presence of sovereign risk raises the marginal benefit of an increase in tax base breadth—via reduced borrowing cost—and thus increases the optimal level of investment in fiscal capacity.

The reduction in *total* debt servicing cost due to a marginal increase in tax base breadth is increasing in the level of indebtedness, for two reasons. First, there is a mechanical effect: the reduction in total borrowing costs is equal to the change in the cost of borrowing multiplied by debt due for repayment, indicated by the D_{t+1} term on the right-hand-side of Equation (1.2.22). Second, the fall in the default threshold—and thus the cost of borrowing—following a marginal increase in tax base breadth is larger the higher is the level of debt: $\partial^2 A_{t+1}^d/\partial b_{t+1}\partial D_{t+1} < 0$ (see the appendix for a derivation).[9] These facts together imply that a marginal increase in tax base breadth reduces total borrowing costs by more the larger is the amount borrowed in the first period.

In summary, from this simple two-period model we have gained these important insights about the interaction of fiscal capacity and sovereign risk: i) for a given level of borrowing, the default probability is decreasing in tax base breadth; ii) the reduction in the default probability due to a marginal increase in tax base breadth is increasing in the level of indebtedness; and iii) the reduction in total debt servicing costs due to a marginal increase in tax base breadth is increasing in the level of indebtedness.

The next section uses numerical techniques to solve the model outlined in Section 1.2.5, allowing us to study the interaction between income shocks, fiscal capacity, and sovereign borrowing in a more general infinite-horizon setting.

---

[9]This result assumes that $\partial f\left(A_{t+1}^d|A_t\right)/\partial A_{t+1} > 0$, which is true if the probability density function f is single-peaked and A_{t+1}^d is at a level of income to the left of the peak in the density function.

1.2.7 Optimal Tax Base Breadth With Debt: Numerical Solution

1.2.7.1 Calibration

Naturally, the model's numerical solution depends on its calibration. The parameter values outlined in this section have been chosen with the goal of describing the optimization problem faced by a typical U.S. state government in the 1930s—the time period and sample of governments that the empirical analysis in Section 1.3 considers. The model is solved using value function iteration, and the solution algorithm is described in the appendix.[10] Matching the typical government's budget cycle, the model is solved at an annual frequency.

For computational simplicity, there are assumed to be only two possible levels of fiscal capacity: $b \in \{b_{low}, b_{high}\}$; two levels of fiscal capacity is sufficient for understanding the circumstances in which a government finds it beneficial to undertake a tax base broadening reform. The calibration of b (tax base breadth) is critical, because it controls the (compensated) tax base elasticity; when it is low, the efficiency cost of raising revenue is high and little debt can be sustained in equilibrium. (Recall that the compensated commodity tax base elasticity with respect to the net-of-tax rate in this model is equal to $1 - b$.) Even putting aside the direct dependence in this model between tax base breadth and the tax base elasticity, calibrating b based on an estimate of the 1930s commodity tax base breadth alone would be misleading, because state governments had other non-commodity tax bases available (e.g., property). The relevant target for b_{low} is a revenue-weighted average tax-base-elasticity for a typical U.S. state government prior to the 1930s wave of tax base broadening reforms, and b_{high} its post-reform value.

In the absence of tax base elasticity estimates for the taxes administered by U.S. state governments in the inter-war years, the appropriate calibration of b is inferred from other work. In their review of the recent literature, Saez et al. (2012) suggest a federal income tax base elasticity with respect to the marginal net-of-tax rate of about 0.1-0.4; Romer and

---

[10]Thanks to Gita Gopinath and Mark Aguiar for making Matlab code available from their (2006) work on sovereign default in emerging markets, which provided a starting point for the code written to solve this model.

Romer (2013) provide the only available estimate for the inter-war period, and estimate the corresponding elasticity to be 0.2. Based on this evidence, the tax base elasticity is conjectured to be a little above this range prior to the 1930s tax base reforms ($b_{low} = 0.5$), and just above the midpoint of this range post reform ($b_{low} = 0.7$).

The per-period administrative cost function is assumed to be quadratic, $\xi'_f(b_t) = f b_t^2$, and the parameter f is calibrated such that, at the mean income level, when $b = b_{low}$ ($b = b_{high}$) the reduction in excess burden from a marginal increase in tax base breadth is greater (slightly less) than the increase in marginal per-period administrative cost. That is, with reference to Section 1.2.4, $\Lambda_{0,t}(b_{low}, \bar{A}) > \Lambda_{1,t}(b_{low}, \bar{A})$, and $\Lambda_{0,t}(b_{high}, \bar{A}) \simeq \Lambda_{1,t}(b_{high}, \bar{A})$. Thus, conditional on having incurred the fixed setup cost to increase tax base breadth, it is optimal with this calibration, for most income realizations, to maintain that broad base; but it is not optimal to incur that fixed setup cost absent a sharp fall in income and/or rise in indebtedness. The assumption that fixed setup cost, rather than per-period marginal administrative cost, is the main factor restricting optimal tax base breadth is consistent with the infrequency with which new tax bases, once implemented, are repealed. Were the opposite true, the cost of administering broad tax bases introduced when revenue needs are temporarily high (e.g., to fund an external war) would be intolerable at normal revenue levels, and we would observe frequent tax base repeal.

Administrative cost is about 1.9 (3.6 percent) of revenue raised when tax base breadth is at a low (high) level. Taking into account that administrative costs do not vary with revenue raised, and that government has growth substantially since World Word II, these assumed per-period administrative cost levels (as a fraction of revenue raised) are broadly in line with recent estimates for OECD countries (OECD, 2011, p. 126-127). In the preferred calibration, the fixed setup cost to increase tax base breadth from $b_{low} = 0.5$ to $b_{high} = 0.7$ consumes resources equal to 1.2 percent of mean income.

Empirically, public good provision varies less than proportionally with income over the business cycle, and the model is calibrated accordingly. Demand for the public good is parameterized such that, were net borrowing infeasible, the government would optimally adjust the

tax rate to offset no less than half the mechanical change in revenue due to an income shock; this assumption corresponds to assuming that $\left(\xi_t + \lambda \bar{G}\right) > \phi A_{max}$. (This is the condition derived in Section 1.2.4, under which a fall in income makes having a marginally broader tax base desirable for a government with a balanced budget requirement.) Government spending is assumed be a little less than 8 percent of mean income, approximately equal to its 1932 level for the unweighted average of U.S. state governments.[11] These two assumptions together pin down the parameters that determine public good provision: ϕ and λ.

The probability each period that a government in financial autarky regains access to financial markets is chosen to be in line with the experience of the U.S. state governments that repudiated their debts during the 1840s default episode. English (1996) documents that all of the U.S. states that repaid their debts in full continued to borrow until the 1860s, but that none of the repudiating states borrowed much before the 1861-1865 U.S. Civil War. Reflecting this, $\pi = 0.05$ is chosen, which implies a mean period in financial autarky of 20 years.

Because increased consumption volatility under financial autarky has only a small welfare cost, the parameterization of the output cost experienced during financial autarky is quantitatively more important than the choice of the per-period redemption probability. (Although the redemption probability does govern the length of time that the output cost is experienced.) The source of any output cost triggered by default is perhaps less clear for U.S. state governments than nation states; for example, a trade embargo against a particular state (often mentioned in the literature as a source of cost experienced upon default) is infeasible because the U.S. Constitution guarantees free trade between the states. During the 1840s wave of U.S. state government defaults, the federal government was unwilling to compensate foreign lenders for debts accumulated by state governments, and the British government refused to use diplomatic or military measures to intervene of behalf of British citizens who held state government debt (English, 1996). Nonetheless, state governments have been able to sustain high levels of indebtedness (measured relative to revenues) at various points in

---

[11]Because the model assumes a representative consumer, there are implicitly no transfers. Although relief programs for the poor, administered or funded by state governments, were important during the Great Depression, transfers to households represented a small share of state government spending in the 1930s.

time. To give a plausible description of empirically observed levels of indebtedness, an output cost of $\omega = 0.02$ (2 percent) in financial autarky is assumed, which is in line with the calibration used by Aguiar and Gopinath (2006). Because the principal goal of solving the model is to qualitatively understand the interaction between income shocks, sovereign risk, and fiscal capacity—not to precisely match empirically observed debt levels—the particular parameterization of the output cost is not crucial.

The rate-of-return required by risk-neutral lenders is chosen to be $r = 0.06$ (6 percent per annum), and the government is assumed to have a subjective rate of time preference β equal to the reciprocal of the gross risk-free interest rate: $\beta = 1/(1+r)$.[12] Lastly, the autocorrelation and shock size parameters of the income endowment process have been chosen to approximately match detrended annual U.S. GDP data over the period 1900-2012: $\rho = 0.9$ and $\sigma_\epsilon = 0.05$.

The numerical solution to the model, using these parameter values, is presented graphically and described in the next section.

1.2.7.2 Discussion of Numerical Results

The shaded area in the lower panel of Figure 1.2 shows combinations of indebtedness (on the x-axis) and income (on the y-axis) for which default is chosen. Consistent with the earlier discussion, default is optimal when indebtedness is high and endowment income is low. The default set is larger for a government with a narrow than a broad tax base, for two reasons: i) debt repayment has a higher utility cost (the excess burden of taxation is higher) for a government with a narrow than a broad tax base, and ii) for some combinations of debt and income, the fixed cost to expand tax base breadth exceeds the punishment incurred by default.

---

[12]It is common in the sovereign debt literature to choose $\beta(1+r) < 1$ (often, substantially less than unity). A low rate of subjective time preference counters a government's incentive to save and outgrow its borrowing constraint, generating high debt levels and frequent default; it also flattens the bond price schedule, increasing the extent of counter-cyclical interest rate variation, helping in open-economy applications to match current account dynamics (see Aguiar and Amador, forthcoming). However, when $\beta(1+r) < 1$, a government never chooses to repay substantial amounts of debt, obscuring the full set of circumstances in which an increase in fiscal capacity is desirable. The open-economy and debt level consequences of choosing a high rate of subjective time preference are not critical in this setting.

The default set is larger still for a hypothetical government permanently constrained to a narrow tax base. The two-period model analyzed in Section 1.2.6 has the same property: for any given level of indebtedness, default is less likely at a high level of fiscal capacity, because there are fewer income realizations for which it is optimal.

The bond price schedules, shown in Figure 1.3, reflect the likelihood of default for each combination of debt, income, and fiscal capacity. A shift in the bond schedule to the right indicates tighter credit conditions, because less money can be borrowed at a given bond price. Consistent with the default sets shown in the lower panel of Figure 1.2, the bond price schedule for a government entering the current period with a low level of fiscal capacity, but the opportunity to upgrade to a high level of fiscal capacity, lies to the right of the bond price schedule for a government that already has a high level of fiscal capacity. A government permanently constrained to a low level of fiscal capacity (shown by the grey dotted line in Figure 1.3) faces less favorable borrowing conditions than a government that has the ability to upgrade to a high level of fiscal capacity.

The shaded area in the upper panel of Figure 1.2 shows the combinations of debt and income for which a government with a narrow tax base finds it desirable to undertake a tax base broadening reform. This region borders the default set for a government with an initially narrow tax base, indicating that, at each income level, a tax base broadening reform is most desirable relative to its implementation cost when indebtedness is close to its maximum level.[13] Near the default boundary, sovereign risk makes lenders unwilling to permit net borrowing; debt repayment occurs to guard against declines in income that would, at an unchanged debt level, result in default. The larger is the magnitude of net debt repayment, the greater is the benefit of a tax base broadening reform, because the tax rate, and thus the excess burden of taxation, is increasing in the revenue requirement.

Recalling that an increase in tax base breadth comes into effect one period ahead, the govern-

---

[13]At sufficiently low levels of the fixed setup cost F, a tax base broadening reform is also be desirable inside the default region, because a reduction in the excess burden of taxation is beneficial in both the repayment and default states. But at this low level of F an increase in tax base breadth would also occur just outside the default region, where the benefit is relatively greater. See the appendix for results with high and low values for F.

ment must guard against adverse income shocks: it chooses to undertake a base broadening reform at levels of indebtedness even before substantial net debt repayments occur and the tax rate must rise sharply; it is also less burdensome to fund the fixed setup cost before substantial net debt repayments are required, because of diminishing marginal utility of (private and public) consumption.

Knowing the government's maximization problem, competitive lenders anticipate the circumstances in which a government with a low level of fiscal capacity chooses to upgrade to a high level of fiscal capacity and offer lending terms that reflect its flexibility in tax base breadth. Because default incentives are decreasing in tax base breadth, lenders recognition of flexibility in fiscal capacity affords benefit even before a tax base broadening reform is undertaken.

These numerical results indicate that the inverse relationship between tax base breadth and borrowing costs derived using a two-period model in Section 1.2.6 carry over to an infinite-horizon setting.

1.2.8   Summary of Model Predictions

This section has presented a formal model in which an optimizing government can vary tax base breadth in response to fluctuations in revenue needs: an increase in tax base breadth reduces the tax rate required to raise a given amount of revenue—and thus reduces the excess burden of taxation—but an increase in tax base breadth incurs higher per-period administrative cost and a fixed setup cost. The model assumes that households prefer less business-cycle variation in public than private consumption, which is consistent with observed variation in public spending; this implies that, with an unchanged tax rate and tax base, a decline in households' income endowment creates a budget deficit.

A government with a balanced budget requirement experiencing a revenue shortfall faces an immediate trade-off between increasing the tax rate on its existing tax base or broadening the set of taxed goods. Raising the tax rate on the existing tax base increases the excess burden

of taxation—which is convex in the tax rate—whereas broadening the tax base permits a lower tax rate—and thus lower excess burden—but incurs higher administrative cost. *For sufficiently large negative income shocks, it is optimal to incur the fixed cost to undertake a tax base broadening reform.*

A government with the ability to borrow can run a budget deficit in response to a negative income shock, and postpone the decision on how to fund its revenue shortfall—either a higher tax rate or an increase in tax base breadth—until debt is due for repayment. Whenever net debt repayment occurs, revenue needs are unusually high, and so is the tax rate; this makes the benefit of a broad tax base high relative to the associated increase in administrative cost. Because having a broad tax base reduces the excess burden of taxation due to debt repayment, an increase in tax base breadth can reduce default incentives, and thus the cost of debt; this interaction between the tax base breadth and borrowing costs reinforces the benefit for a heavily indebted government of undertaking a tax base broadening reform when income is low.

1.3 The U.S. States During the Great Depression: The Role of Tax Base Expansion

1.3.1 Introduction

The model's key predictions are tested by studying the behavior of U.S. state governments during the Great Depression. This time period and group of governments provides an excellent setting to examine the impact of macroeconomic income shocks and indebtedness on fiscal capacity upgrading: the income shocks were large, with the average state experiencing a fall in real per capita personal income of 28.5 percent between 1929 and 1933; there was significant heterogeneity in the size of income shocks experienced across states; and indebtedness at the onset of the Great Depression varied considerably across state governments. Like nation states, U.S. state governments are sovereign with respect to their debts: the Eleventh

Amendment to the U.S. Constitution effectively grants state governments sovereign debt immunity. As English (1996) explains, bond holders may seek payment of debts by turning to state courts, but their experience in Mississippi following the 1840s default suggests that this is not a promising strategy: the Mississippi State Supreme Court upheld the validity of the bondholders' claims but the court could not enforce repayment (English, 1996).[14] However, as discussed in Section 1.2.7.1, the cost of default may be different for a U.S. state government than nation states, because the U.S. Constitution precludes some forms punishment upon default, such as trade embargoes.

Although there is significant heterogeneity, compared to a panel of nation states, U.S. state governments form a relatively homogenous grouping. But there were some differences in fiscal institutions that may have affected upgrading in fiscal capacity during the Great Depression. These differences are discussed next.

1.3.2 Fiscal Institutions

By 1929 20 U.S. state governments had adopted constitutional balanced budget requirements, limiting their ability to use borrowing to smooth income shocks between good and bad times (Rodriguez-Tejedo and Wallis, 2010).[15] For a further 20 states, the effective power to issue debt resided outside the legislature; some of these states had binding constitutional limits on the amount of debt that could be issued, while others required a voter referendum to issue debt.[16] Most of the debt issue restrictions originated from the 1840s: six of the eight states that defaulted in the 1840s adopted procedural restrictions on debt issue by 1851, as did six other non-defaulting states (Wallis, 2005). Many states achieving statehood after this

---

[14]English (1996) discusses barriers to two other means around the Eleventh Amendment. The Supreme Court's ruling in *Hans v. Louisiana* (1890) prevents federal courts ruling in disputes between a citizen and a resident of that state, while the Supreme Court in *Monaco v. Mississippi* (1934) makes U.S. states immune from suits by foreign countries. The Eleventh Amendment does not prevent suits being brought by other U.S. states, or by the U.S. government, but most state government debt has been privately held.

[15]Thanks to Isabel Rodriguez-Tejedo and John Wallis for sharing their data.

[16]See Ratchford (1938) for a classification of states into three categories: i) those requiring a constitutional amendment to issue debt, ii) those requiring voter approval, and iii) those able to borrow without quantitative limit, subject only to legislative approval. Note the caveats to his classification discussed in the accompanying text.

episode adopted debt restrictions as part of their original constitutions. During the 1930s a few states adopted more stringent fiscal institutions: Alabama introduced a constitutional balanced budget requirement in 1933, as did New York in 1938. The only state to default on its debt obligations during the Great Depression, Arkansas, removed the effective power to issue debt from legislators in 1934 and put it in the hands over voters via referendum, and North Carolina, which had entered the Great Depression with a high debt level, but did not default, did the same in 1936 (Ratchford, 1938).[17]

The practical importance of these restrictions had been eroded over time in many states by loose interpretation of the so-called Special Funds Doctrine (Ratchford, 1941): debt issued with earmarked revenue streams does not, in principle, rely on general fund revenues, sidestepping constitutional restrictions on long-term debt issue. Examples are highway revenue obligations tied to gasoline tax revenues, toll bridge revenue bonds, revenue obligations pledged to tax collections from a specific base, and obligations on behalf of state agencies generating their own revenue streams. Even in states where the Special Funds Doctrine was strictly applied, state courts had in some cases permitted significant latitude in the use of short-term debt instruments, such as treasury notes and revenue anticipation warrants.

To the extent they carried some force, balanced budget requirements and debt issue restrictions had the potential to make revenue shortfalls during the Great Depression particularly acute. However, as discussed next, even the group of states without these restrictions on deficit spending did not meaningfully use debt to buffer income shocks experienced during the Great Depression.

---

[17]Despite the extreme level of fiscal stress faced by state governments in the 1930s, Arkansas was the only state government to default on its debts. In 1932 Arkansas had a debt-to-revenue ratio of 6.3, about six times the average debt-to-revenue ratio for state governments in 1932. The Arkansas debt was accumulated principally for highway construction, much of it originally by local road districts. In August 1932 Arkansas defaulted on some district bonds, and by March 1933 had defaulted on all its highway debt. By 1937 almost all highway debt had been refunded into new bonds shifting the repayment burden out to a 40-year horizon, with no new debt maturing before 1943-44. See Ratchford (1941) for a more detailed account.

1.3.3   State Government Borrowing in the 1930s

With a few exceptions, state government indebtedness on the eve of the Great Depression was moderate by historical standards: the median debt-to-revenue ratio in 1929 was 0.67, which is similar to median ratios recorded in 1922, 1912 and 1902, and appreciably lower than the ratio of 1.59 recorded in 1890 (see Table 1.2).[18] As Table 1.2 shows, the distribution of debt-to-revenue ratios across the U.S. states was either similar or more favorable in 1929 compared to earlier years.

Despite mostly modest levels of indebtedness in 1929, there was no general tendency for state governments to run larger budget deficits in the early 1930s than they had prior to the Great Depression. As shown by Figure 1.4, the absence of a sharp increase in state government budget deficits in the early 1930s is evident across states with different fiscal institutions. The median state among those with a constitutional balanced budget requirement remained in approximate budget balance, while those in which effective borrowing power resided in the legislature ran deficits in 1931 and 1932, but of about the same size as they had in 1928 and 1929.[19] As a group, state governments subject to a long-term debt issue restriction moved from approximate budget balance to deficit in 1931 and 1932, but the average magnitude of deficits was small relative to the cross-state average 28.5 percent decline per capita income between 1929 and 1933.

In addition to institutional restrictions potentially affecting some states, a general tightening of credit conditions appears to have been an important factor limiting state governments' ability to run temporary deficits.[20] In 1929 all but two U.S. states received a Aaa rating on their general obligation bonds (North Dakota and South Dakota had the next highest Aa rating), but by 1932 10 states had been downgraded to an Aa rating, and two to an A rating. U.S. real per capita income grew strongly after its 1933 trough, to regain its 1929

---

[18]Debt-to-revenue ratios are reported rather than debt-to-income ratios because income estimates are only available for a few select years before 1929.

[19]Government cost payment (spending) data for capital outlays are unavailable for 1930, and revenue and cost payment data are unavailable from 1933-1936.

[20]Deflation in the early 1930s, which might contemporaneously have been expected to persist for some time, may have made governments unwilling to take on nominal debt. However, the accumulation of debt by the U.S. federal government is not consistent with this being an important deterrent to debt accumulation.

level in 1937, but state government credit ratings continued to deteriorate: by 1937 all but 13 states had received a credit rating downgrade, despite generally declining debt-to-revenue ratios (see Tables 1.2 and 1.3).[21]

Coincident with these credit rating downgrades, the most heavily indebted state governments significantly reduced the real value of their debts between 1932 and 1937 (see Table 1.2 and Figure 1.5), despite real incomes for most states being well below full-employment levels. The model predicts that an optimizing government will only reduce its indebtedness in bad times if there is an increase in the likelihood of default, and lenders are unwilling to permit net lending; the decline in credit ratings for the most indebted state governments suggests this may have been the case. Ratchford (1941) argues that the reduction in real state government debts reflected, in part, the completion of infrastructure projects—principally highway construction—for which most of the pre-Depression debt had been earmarked; but the fall in indebtedness for high debt states, measured by real per capita debt, debt-to-income, or debt-to-revenue ratios, was generally a reflection of rising prices that accompanied recovery in real incomes after 1933 (which boosted nominal state government revenues), rather than a reduction in nominal debt outstanding.

1.3.4 The Great Depression and Upgrading in State Fiscal Capacity

1.3.4.1 Summary of Tax Base Changes

Experiencing large income shocks, and being unwilling or unable to meaningfully increase indebtedness, the Great Depression had a profound and immediate impact on U.S. state government fiscal capacity. Prior to the Great Depression no state had a retail sales tax, only 12 of the 48 U.S. states had an individual income tax, and 10 had a corporate net-

---

[21]Ideally, bond yield data would be used to measure borrowing costs for state governments, but for most states debt on comparable issues was too infrequently traded to provide meaningful prices, if data are even available. Ratings provided in Moody's annual Municipal Government Bond Manual provide a consistent means with which to measure default risk for the U.S. states. States with non-negligible levels of debt were rated in most years, and rating categories remained unchanged over the period of interest. In 1937 Moody's began tracking bond yields for a portfolio of municipal bonds in each of their four highest rating categories. Although not covering uniquely state government debt, the data show persistent differences in borrowing costs across rating categories.

income tax (see Table 1.4). By 1940 22 states had a retail sales tax, another 18 states had adopted an individual income tax, and a further 15 had a corporate net-income tax; all but three of these new tax bases became permanent. Most new tax bases were introduced before real economic activity had returned to pre-Depression levels: Mississippi was the first state the adopt a retail sales tax in 1932, and 19 of the 22 states that adopted a permanent retail sales tax in the 1930s did so by 1935 (see Figure 1.6). No new sales or income tax bases were introduced after 1938 and before 1946. Summing over these three types of tax base, 52 new broad tax bases that ultimately became permanent were in place by 1940 across the 48 U.S. states, compared to only 22 in 1929. In addition, six mostly Northeastern states had levied a retail sales tax for one or two years each during the early to mid-1930s.

Retail sales tax adoption during the 1930s was a more important development in state government fiscal capacity than the adoption of new income tax bases. By 1942, the 22 states that had adopted a retail sales tax during the 1930s raised on average 19 percent of their total tax revenue from the sales tax, compared to an 11 percent income tax revenue share for the 35 states collecting income tax revenue in 1942 (see Figure 1.7).[22] The sales tax was an important revenue source because of its broad base, not high tax rates: of the 22 states with a retail sales tax in 1938, 16 had a 2 percent rate, and the remaining six states had a 3 percent rate (Due and Mikesell, 1995).

In addition to the introduction of retail sales and income tax bases during the 1930s, the relative importance for state governments of existing revenue sources changed substantially, and some other new tax bases were introduced (see Figure 1.7). Having accounted for just over half of all tax revenue for U.S. state governments on average in 1922, the property tax share of total state government tax revenues declined further during the 1930s, from about one-quarter in 1932 to less than one-tenth in 1942. Reflecting in part an increase in automobile use, motor vehicle fuel tax revenue grew to almost one-quarter of total tax revenue by 1942, and increased unemployment compensation taxes accompanied the rise in joblessness.[23] Following the repeal of Prohibition in 1933, all states introduced alcoholic sales

---

[22]Income tax revenue-share data include inheritance taxes.

[23]Unemployment compensation taxes are payroll taxes remitted by employers, and in some states employees,

taxes, but they raised only about 5 percent of total tax revenues in 1942.[24]

State governments continued to adopt retail sales and income tax bases after the 1930s, but at a gradual pace until the 1960s (see Table 1.4). In that decade, 12 additional states adopted a retail sales tax, 7 an individual income tax, and 6 a corporate net-income tax (one of which was later repealed). The 25 new tax bases introduced in the 1960s is substantial, but less than half the number of bases introduced in the 1930s. Due and Mikesell (1995) suggest that the post-World War II adoption of retail sales taxes was due to increased demand for state revenue, in large part to fund new education expenditures. After the 1960s, four more states introduced an individual income tax, and one more a corporate net-income tax.

Because the introduction of the retail sales tax was the most important change in fiscal capacity during the Great Depression, measured by revenue raised, the remainder of the empirical analysis focuses primarily on understanding its adoption. Most of the analysis that follows groups state governments according to whether or not they adopted a sales tax base that ultimately became permanent: the six states that had temporary retail sales taxes in the 1930s maintained these bases for only one or two years each.

1.3.4.2 The Relationship Among Tax Base Adoption, Income, and Indebtedness

States that did and did not adopt a retail sales tax during the 1930s had, on average, a similar level of per capita real spending before and during the 1930s (see Figure 1.8). As a share of income, spending was higher in 1932 and 1937 than before the Great Depression, especially for the states that adopted a retail sales tax; that public spending rose as a share of income during the Great Depression is consistent with the modeling assumption that households prefer public spending to vary less than private consumption over the business cycle. Because borrowing was not meaningfully used to buffer income shocks, and spending differed little between sales tax adopting and non-adopting states, average revenues for the

---

to exclusively fund unemployment compensation benefits.

[24]There was no strong tendency for states without a retail sales tax to rely more heavily on alcoholic beverage taxes: the 22 states with a retail sales tax in 1942 collected on average 4 percent of total tax revenues from alcohol taxes, compared to 6 percent for states without a retail sales tax.

two groups of states were by necessity little different (see Figure 1.9).

While spending and revenues shares were similar across states that did and did not adopt a sales tax, the efficiency cost of raising tax revenues differed markedly. States adopting a retail sales tax in the 1930s experienced, on average, a 7.6 percent larger decline in real per capita personal income between 1929 and 1933 (see Figure 1.10). Grouping states by their magnitude of income decline between 1929 and 1933 reveals a strong relationship between the size of income shocks experienced and a state government's propensity to adopt a retail sales tax during the 1930s (see Table 1.5). States suffering above-average income declines were also more likely to adopt income tax bases during the 1930s (see Table 1.5). The approximately contemporaneous relationship between the magnitude of negative income shocks and retail sales tax base adoption is consistent with the formal model, conditional on governments being unable to accumulate meaningful amounts of debt; this is true whether debt accumulation was restricted by a balanced budget requirement, or because indebtedness was high enough to make lenders were unwilling to permit net borrowing.

Governments with high levels of indebtedness, and unable to accumulate more debt, had a greater incentive to adopt a retail sales tax than governments with little debt, for two reasons: i) they needed additional revenues to fund interest payments, and ii) an increase in fiscal capacity may have reduced default incentives, and thus the cost of borrowing. For the five state government with the highest debt-to-income ratios in 1929, on average 14 percent of revenue was spent servicing debt, compared to only 4 percent for the remaining states. Reflecting their greater revenue needs, the most heavily indebted group of states in 1929 were more likely than low-debt states to adopt a retail sales tax during the 1930s (see Table 1.6). A relationship is also evident between the cost of debt, measured by credit ratings, and the willingness of state governments to adopt a retail sales tax: of the 13 states maintaining a Aaa rating during the 1930s, only three adopted a retail sales tax, while among the five Baa rated states only one did not (see Table 1.7). However, formal regression analysis reported in next sub-section shows that, after controlling for the size of income shocks and debt levels, a relationship between credit ratings and tax base adoption is no longer evident.

This is not surprising in the context of the model presented here, where the level of fiscal capacity, indebtedness, and income fully summarize default probability, implying that there is no additional information contained in credit ratings.

The theoretical model assumes that, given an unchanged tax base and tax rate, tax revenues vary one-for-one with income; the discussion thus far has implicitly assumed such a direct correspondence between changes in income and changes in taxable income. This relationship is, in general, difficult to test: while data on revenue collected by tax base is available, data on taxable value is typically unavailable. The property tax base, which on average accounted for 27 percent of total state government tax collections in 1932, is an important exception. Figure 1.11 indicates an approximately one-for-one relationship between 1929-1933 changes in state incomes (peak-to-trough) and changes in the taxable value of property over the period 1929-1937. (The change in property values is measured over a longer time period than changes in income because assessed property values typically adjust with a lag to changes in the market value of property; the delay is likely to be less than four years, but data unavailability prevent a more timely comparison.)

Reassuringly, given the earlier discussion, states experiencing declines in taxable property values were the most likely to adopt a retail sales tax: of the 28 states experiencing a decline in the real taxable value of property between 1929 and 1937, 18 adopted a retail sales tax during the 1930s, whereas only 4 of the 20 states experiencing a rise in the real taxable value of property adopted a sales tax (see Figure 1.12). Compounding the impact on revenues, states suffering a decline in taxable property values raised an above-average share of revenue from the property tax base: property taxes accounted for 32 percent of total tax revenue in 1932 for states experiencing a fall in assessed property values, compared to 19 percent for states experiencing a rise in assessed property values.

Delayed pass-through of changes in state income to state revenues—particularly for the property tax base—helps explain the timing of tax base changes during the 1930s. Real U.S. per capita GDP fell sharply from 1929-1933, but with the exception of Mississippi, each state government introducing a sales tax during the Great Depression did so coincident with or

after the trough in GDP; Eighteen states introduced a retail sales tax that ultimately became permanent between 1933 and 1935, and five further states had a temporary sales tax base for some of this period (see Figure 1.6).

1.3.4.3   Regression Analysis

Formal regression analysis confirms the findings evident in the bivariate relationships discussed above. In the preferred specification, regression number (1) in Table 1.8, each 10 percent decrease in real income between 1929 and 1933 is estimated to have increased the probability that a state adopted a retail sales tax during the 1930s by 0.18. This relationship is precisely estimated and, given that the average fall in income between 1929 and 1933 was 28.5 percent (0.34 log points), implies that the Great Depression raised the probability of the average state adopting a retail sales tax by about 0.6. A high debt level on the eve of the Great Depression is also estimated to have increased the likelihood of adopting a sales tax: a 1 percentage point increase in the 1929 debt-to-income ratio increased the probability of sales tax base adoption by 0.06. But measuring indebtedness instead by the 1932 debt-to-income ratio, instrumented by the 1929 debt-to-income ratio to sidestep any endogenous response of debt to income shocks, more than halves the estimated effect of indebtedness on the likelihood of sales tax base adoption (see regressions 6 and 7 in Table 1.8). Reflecting the fact that there was limited use of debt to buffer income shocks during the 1930s, cross-sectional differences in income shocks is estimated to be an economically more significant predictor of sales tax base adoption than is differences in indebtedness across state governments. Having an income tax base prior to the Great Depression, and the type of fiscal institutions present in 1929, is not estimated to have significantly affected the likelihood of adopting a sales tax base. Using a Probit rather than OLS estimator for this regression specification changes the magnitude of the estimated effects somewhat, but alters none of the qualitative conclusions (compare regression specifications 1 and 2 in Table 1.8).

As foreshadowed in Section 1.3.4.2, having a less than Aaa credit rating in 1937 is not estimated to have had an independent effect on the likelihood of adopting a sales tax base

during the 1930s. The component of a binary indicator for a non-Aaa credit rating orthogonal to changes in income and to the level of indebtedness during the 1930s (estimated by the residual series for regression 4 in Table 1.8) is insignificant when included in the preferred regression specification (see regression 5 in Table 1.8).

1.4   Conclusion

This paper has established, theoretically and empirically, that deep recessions can be an important stimulant to investment in fiscal capacity. Negative macroeconomic income shocks reduce tax collections, but demand for public spending falls by less, stressing the revenue raising capability of existing tax bases, particularly when the ability to accumulate debt is limited. Raising the tax rate on existing tax bases increases the marginal excess burden of taxation—the marginal welfare loss compared to revenue-equivalent lump-sum taxes— because taxpayers increasingly distort their consumption choices toward untaxed commodities. For sufficiently steep falls in income, it is optimal to incur the fixed cost necessary to upgrade fiscal capacity: this enables taxing a wider range of economic activity at a lower rate, and thus reducing the distortionary cost of taxation. For highly indebted governments, an increase in fiscal capacity can reduce default incentives, because the efficiency cost of raising revenue to repay debt is reduced, and thus can lower borrowing costs. Even though income shocks are transitory, improvement in fiscal institutions can be long-lasting: the fixed cost to upgrade fiscal capacity is sunk, making it optimal to maintain a high level of fiscal capacity even after incomes have recovered.

Evidence from the behavior of U.S. state governments during the Great Depression supports the model's key predictions. At the onset of the Great Depression, none of the state governments levied a retail sales tax, but by 1938 22 state governments had adopted a retail sales tax that ultimately became permanent. Moreover, governments in states experiencing larger than average negative income shocks were significantly more likely to adopt a retail sales tax (and income taxes) than were governments in states experiencing smaller than average

income shocks. As the model predicts, state governments entering the Great Depression with a high debt load were more likely to adopt new tax bases than were those with little debt.

Finally, it is interesting to ponder the impact the recent Great Recession might ultimately have on fiscal institutions in light of the model developed in this paper. To this point, governments in most advanced economies have responded to falls in tax collections by accumulating debt; at some point in the future, substantial net debt repayments will be required, at which time the incentive to invest in increased fiscal capacity will be heightened. Because most developed economies already maintain a broad set of tax instruments, any increase in tax base breadth is likely to involve limiting exemptions and loopholes that provide opportunities for taxpayers to easily substitute from taxed to untaxed commodities; increased enforcement effort, to limit avoidance and evasion behavior, is a complementary way for governments to tax a wider range of economic activity, and thus increase fiscal capacity. Greece—already under substantial fiscal stress, and with a lower level of fiscal capacity than many of its European Union peers—has taken some steps toward increasing fiscal capacity, by seeking to improve tax compliance, under pressure from foreign lenders.

Table 1.1: Quantitative Model: Parameter Values

World interest rate	$r = 0.06$
Rate of time preference	$\beta = \frac{1}{1+r}$
Public good preference parameter	$\phi = 0.037$
Public good curvature parameter	$\lambda = 0.6$
Per-period administrative cost function parameter	$f = 0.006$
Fixed setup cost to upgrade fiscal capacity	$F = 0.012$
Tax base breadth	$b \,\epsilon\, \{0.5, 0.7\}$
AR(1) coefficient on income process	$\rho = 0.9$
Standard deviation of income shocks	$\sigma_\varepsilon = 0.05$
Output cost of default	$\omega = 0.02$
Redemption probability	$\pi = 0.05$

Table 1.2: Debt-to-Revenue Ratios: Number of U.S. States

	1890	1902	1912	1922	1927	1929	1932	1937	1942	1947	1952
0.0-0.5	11	15	22	18	19	20	19	23	29	36	32
0.5-1.0	6	14	6	14	13	15	13	12	12	11	11
1.0-3.0	15	12	18	16	15	11	12	12	7	1	5
> 3.0	13	7	2	0	1	2	4	1	0	0	0
Average	2.62	1.47	0.91	0.80	0.87	0.93	1.09	0.66	0.52	0.27	0.44
Median	1.59	0.83	0.54	0.71	0.71	0.67	0.63	0.51	0.39	0.17	0.30
Federal Govt.	2.78	2.10	1.72	5.70	4.61	4.38	10.13	6.76	4.95	6.71	3.92

Notes: Debt is par value of gross debt less sinking fund assets. Revenue data for 1902 and 1912 are unavailable, and data for 1903 and 1913 has been used instead. Data for three states are missing for 1890. Average values are unweighted. Source: U.S. Department of Commerce (Various Years).

Table 1.3: Moody's General Obligation Bond Ratings
Number of U.S. States by Rating Category

	1922	1927	1929	1932	1937	1942	1947	1952
Aaa	48	44	45	35	13	15	24	22
Aa	0	2	2	10	15	18	16	13
A	0	0	0	2	11	12	7	9
Baa	0	0	0	0	5	2	0	0
Ba	0	0	0	0	1	0	0	0
Unrated	0	2	1	1	3	1	1	4

Notes: Where a state is unrated for the year noted, the state is assigned its rating for the subsequent year. If a rating is unavailable, states with a debt-to-revenue ratio of no more than 0.1 were assigned a Aaa rating. Unrated states noted in the table fall into neither of these categories. The number of assigned ratings for the years shown 1922-1952, respectively, is 2, 0, 3, 3, 6, 7, 11, and 9. According to Moody's, the absence of a rating provides no indication of the credit worthiness of an issuer. Source: Moody's Municipal and Government Manual (1920-1950) .

Table 1.4: Number of U.S. States with Tax Base: By Decade

| | Retail Sales | | Individual Income | | Corporate Income | | |
	Permanent	Temporary	Permanent	Temporary	Permanent	Temporary	Total
1900-09	0	0	0	0	0	0	0
1910-19	0	0	8	0	4	0	12
1920-29	0	0	12	0	10	0	22
1930-39	22	6	28	2	24	1	83
1940-49	27	0	28	2	25	1	83
1950-59	32	0	28	0	28	0	88
1960-69	44	0	35	0	33	1	113
1970-79	44	0	39	0	34	1	118

Notes: Alaska and Hawaii, which achieved statehood in 1959, are excluded. Narrow individual income tax bases are also excluded: New Hampshire and Tennessee have bases taxing only interest and dividend income, and Connecticut a base that taxes only capital gains and dividends. Temporary retail sales taxes in the mid-1930s in Idaho, Kentucky, Maryland, New Jersey, New York and Pennsylvania were in place for only one or two years in each state. Louisiana is classified as permanently introducing the retail sales tax in 1938 because its 1940 repeal lasted only one year. South Dakota and West Virginia had an individual income tax from the 1930s until 1942, and South Dakota had a corporate net-income tax from 1935-1943. Michigan had a corporate net-income tax from 1967-1975. Sources: State retail sales tax data are from Due and Mikesell (1995) and income tax data are from Penniman (1980).

Table 1.5: Income Shocks and Upgrading of Fiscal Capacity

Percent Fall in Real Income: 1929-1933	Number of States	Sales Tax Share 1940	Average No. Broad Bases		
			1929	1940	Δ
< 20	8	0.1	0.8	1.1	0.4
20-30	21	0.3	0.3	1.4	1.0
30-40	15	0.7	0.5	1.9	1.5
40-50	3	1.0	0.7	2.3	1.7
> 50	1	1.0	0.0	3.0	3.0
Federal Govt.			2	2	

Notes: *Real Income* is state per capita personal income deflated by the US real GDP deflator. *Sales Tax Share* is the fraction of states in each category with a retail sales tax in 1940. Mississippi was the first state to introduce a retail sales tax in 1932. A maximum of three broad tax bases is possible: a retail sales tax, an individual income tax, and a corporate net-income tax. Sources: BEA, Due and Mikesell (1995), Penniman (1980), and U.S. Department of Commerce.

Table 1.6: Pre-Depression Debt Levels and Upgrading of Fiscal Capacity

Debt-to-Income Ratio: 1929	Number of States	Average Debt-to-Revenue Ratio: 1929	Sales Tax Share 1940	Average No. Broad Bases		
				1929	1940	Δ
< 0.01	8	0.1	0.3	0.4	1.0	0.6
0.01-0.02	16	0.5	0.4	0.4	1.3	0.9
0.02-0.03	9	1.0	0.4	0.2	1.8	1.6
0.03-0.06	10	1.2	0.5	0.5	1.9	1.4
> 0.06	5	3.2	0.8	1.2	2.6	1.4
Federal Govt.				2	2	

Notes: *Debt-to-Income Ratio* is gross debt less sinking fund assets as a share of state per capita personal income. *Sales Tax Share* is the fraction of states in each category with a retail sales tax in 1940. Mississippi was the first state to introduce a retail sales tax in 1932. A maximum of three broad tax bases is possible: a retail sales tax, an individual income tax, and a corporate net-income tax. Sources: BEA, Due and Mikesell (1995), Penniman (1980), and U.S. Department of Commerce (Various Years).

Table 1.7: Cost of Debt and Upgrading of Fiscal Capacity

Credit Rating: 1937	Number of States	Sales Tax Share 1940	Average No. Broad Bases 1929	1940	Δ
Aaa	13	0.2	0.5	1.2	0.7
Aa	15	0.5	0.1	1.3	1.1
A	11	0.6	0.7	2.1	1.4
Baa	5	0.8	0.8	2.8	2.0
Ba	1	1.0	2.0	3.0	1.0
Unrated	3	0.0	0.0	1.0	1.0
Federal Govt.			2	2	

Notes: The sole Ba rated state, Arkansas, introduced individual and corporate net-income taxes in 1929. *Sales Tax Share* is the fraction of states in each category with a retail sales tax in 1940. Mississippi was the first state to introduce a retail sales tax in 1932. A maximum of three broad tax bases is possible: a retail sales tax, an individual income tax, and a corporate net-income tax. Sources: BEA, Due and Mikesell (1995), Moody's Municipal and Government Manual (1920-1950), Penniman (1980), and U.S. Department of Commerce.

Table 1.8: Upgrading of Fiscal Capacity: Cross-State Regression Analysis

	(1) Sales Tax 1940	(2) Sales Tax 1940	(3) Bases 1940	(4) Down-grade	(5) Sales Tax 1940	(6) Sales Tax 1940	(7) Debt 1932
	OLS	Probit	OLS	OLS	OLS	2SLS	OLS
Inc. Tax 1929	0.065	0.089	0.547*	-0.202	0.013	0.068	-0.002
	(0.173)	(0.218)	(0.312)	(0.249)	(0.152)	(0.163)	(0.013)
Corp. Tax 1929	-0.087	-0.153	0.293	0.080	-0.068	-0.119	0.015
	(0.165)	(0.234)	(0.288)	(0.318)	(0.145)	(0.159)	(0.013)
Debt: 1929	5.604***	9.423***	11.044***	0.065	5.417***		2.591***
	(1.407)	(2.855)	(3.451)	(4.218)	(1.378)		(0.428)
Debt: 1932				1.561		2.163***	
				(3.764)		(0.716)	
Debt: 1937				0.250			
				(4.631)			
ΔInc. 1929-33	-1.827***	-2.962***	-3.332***	-0.194	-1.731***	-1.692***	-0.062**
	(0.382)	(0.825)	(0.850)	(0.626)	(0.439)	(0.357)	(0.030)
ΔInc. 1932-37				0.273			
				(1.049)			
Constitution	0.192	0.255	0.251	0.041	0.171	0.183	0.004
	(0.154)	(0.236)	(0.357)	(0.193)	(0.151)	(0.139)	(0.009)
Referendum	0.269	0.299	0.511	0.164	0.341**	0.264*	0.002
	(0.166)	(0.249)	(0.356)	(0.213)	(0.163)	(0.149)	(0.008)
BBR	0.009	0.019	0.201	-0.033	0.102	0.000	0.004
	(0.151)	(0.173)	(0.314)	(0.188)	(0.165)	(0.140)	(0.010)
Orthog.					0.205		
					(0.150)		
Constant	-0.507***		-0.430	0.406	-0.479**	-0.429***	-0.036*
	(0.171)		(0.484)	(0.339)	(0.183)	(0.156)	(0.019)
Observations	48	48	48	45	45	48	48
R-squared	0.332		0.369	0.185	0.392	0.316	0.892

Notes: *Sales Tax 1940* is a dummy variable taking the value unity if the state had a retail sales tax in 1940. *Bases 1940* is a count of the number of broad tax bases present in 1940, with a maximum of three: a retail sales tax, a personal income tax, and a corporate net-income tax. *Inc. Tax 1929* and *Corp. Tax 1929* are indicators of the presence of those tax bases in 1929. *ΔInc 1929-33* is the change in real state per capita personal income from 1929 to 1933, in log points, with a mean value of -0.34, and *ΔInc 1932-37* is the corresponding change in income between 1932 and 1937, with a mean value of 0.32. *Debt* is the state debt-to-income ratio, with *Debt 1929* having a mean value of 0.03 and a maximum value of 0.18. *Downgrade* is an indicator variable taking the value unity for states that did not have a Aaa credit rating in 1937, with the variable *Orthog.* the residual series from regression specification (4). *Constitution* is a dummy variable equal to unity for states requiring a constitutional amendment to incur debt, *Referendum* is a dummy variable equal to unity for states requiring a referendum to incur debt, and *BBR* is a dummy variable taking the value unity for states with a constitutional balanced budget requirement. Regression specification (7) is the first-stage estimation for 2SLS regression specification (6), with *Debt: 1929* used as an instrument for *Debt: 1932*. Coefficients for regression specification (2) are marginal effects at the mean for each variable, and for a discrete change from 0 to 1 for binary dependent variables. Statistical significance at the 1, 5, and 10 percent levels is denoted by ***, **, and *, respectively. Robust standard errors have been used.

Figure 1.1: Marginal Excess Burden and Marginal Administrative Cost
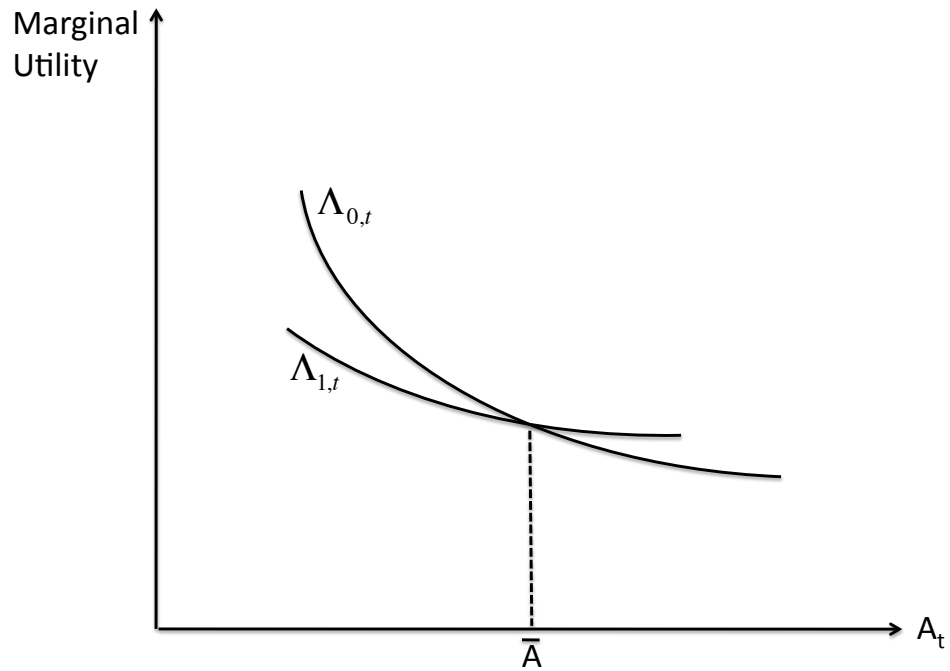
Notes: $\Lambda_{0,t}$ is the reduction in the excess burden of taxation from having a marginally broader tax base, and $\Lambda_{1,t}$ is the social cost of funding the associated increase in per-period administrative costs. The figure is drawn assuming that when income is at its mean level $\bar{A}$ tax base breadth is at an optimum (ignoring the fixed setup cost component of administrative costs): $\Lambda_{0,t} = \Lambda_{1,t}$.

Figure 1.2: Numerical Solution: Upgrading of Fiscal Capacity and Default Sets
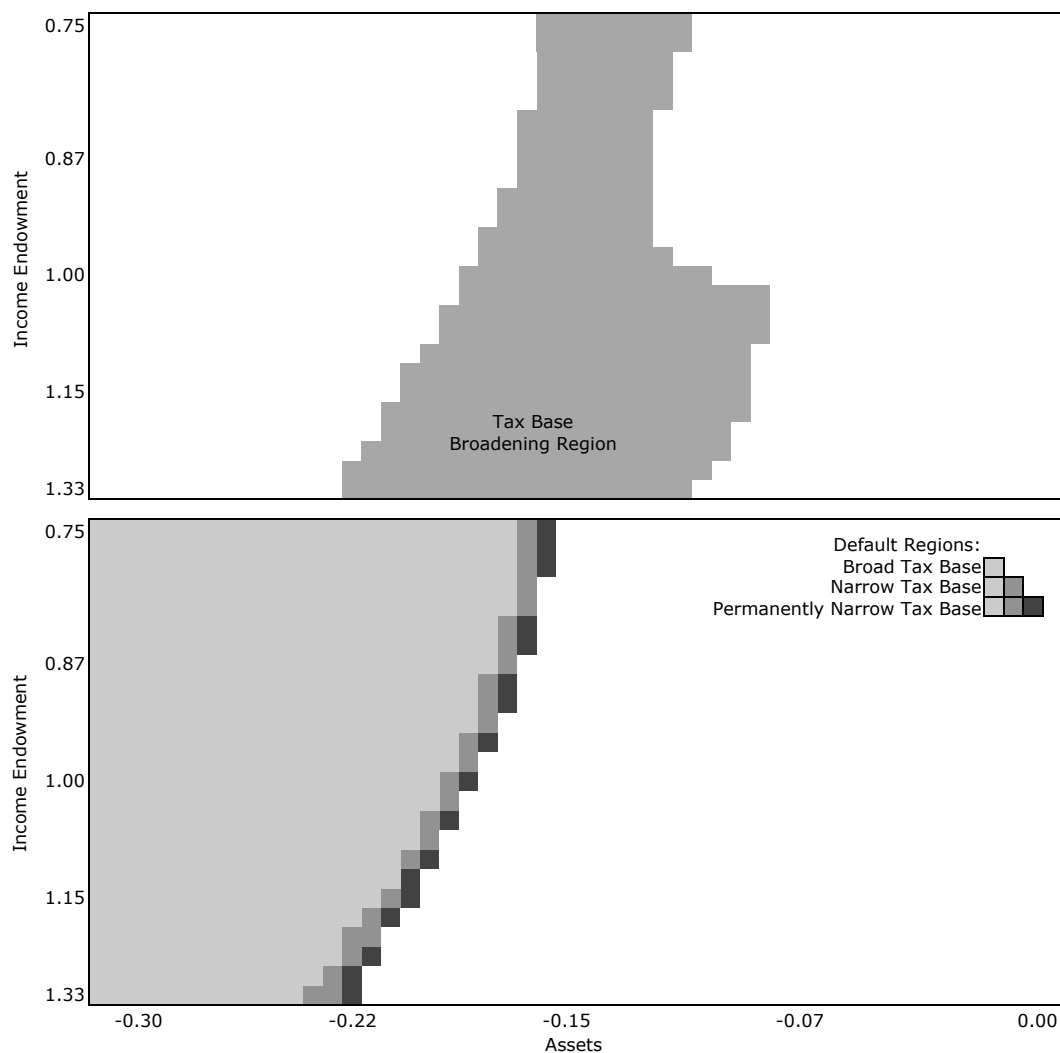
Notes: The shaded region in the upper panel shows combinations of debt (on the x-axis) and income (on the y-axis) for which a government with a low level of fiscal capacity chooses to upgrade to a high level of fiscal capacity. The shaded regions in the lower panel show combinations of debt and income for which a government chooses to default: the lightest grey shaded region shows the default set for a government with a high level of fiscal capacity; the lightest two shaded regions show the default set for a government with a low level of fiscal capacity, but the ability to upgrade to a high level of fiscal capacity; and the combined shaded region shows the default set for a government permanently constrained to a low level of fiscal capacity.

Figure 1.3: Numerical Solution: Bond Price Schedules
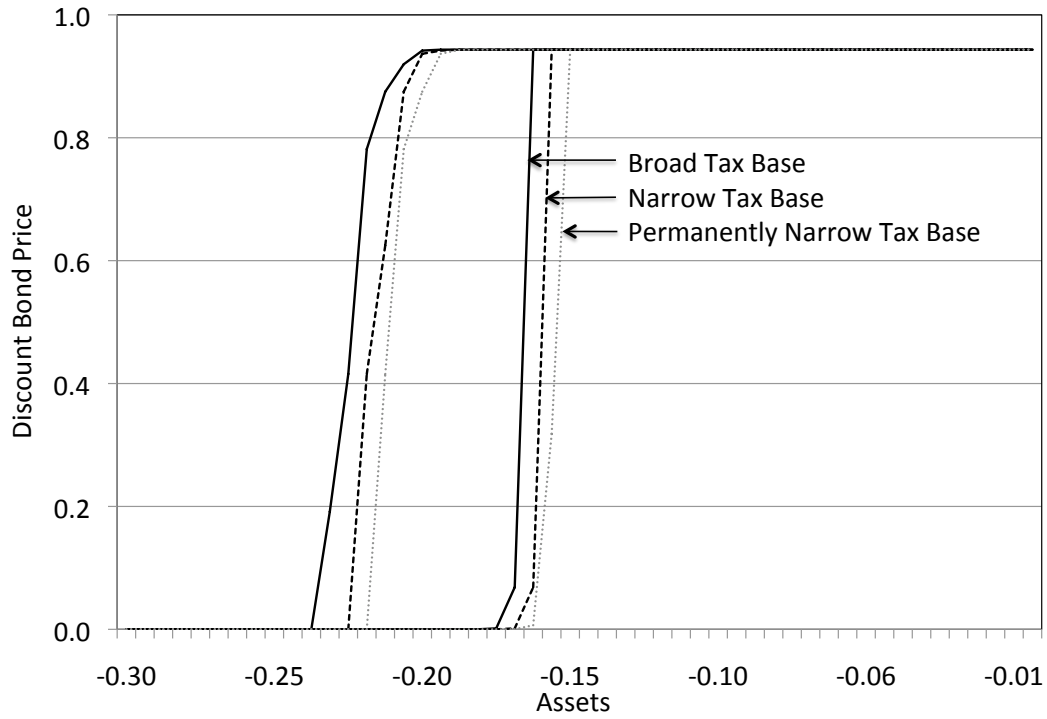
Notes: The price of a discount bond depends on the amount of debt sold, the level of income, and next period's level of fiscal capacity: $q_t = q(A_t, D_{t+1}, b_{t+1})$. The level of debt is shown on the x-axis (recall that mean income in the model is unity) and the discount price of a one-period bond on the y-axis. For each pair of lines, the left-hand line is for income at its maximum level, and the right-hand line for income at its minimum level.

Figure 1.4: Budget Surplus as a Share of Revenue
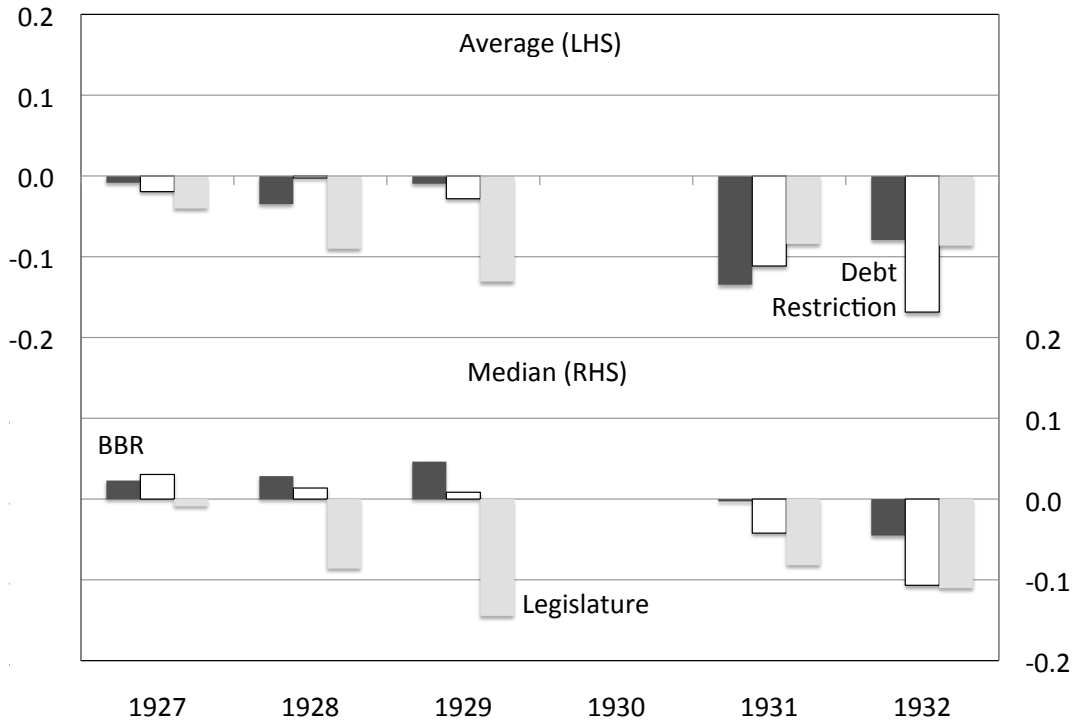
Notes: *BBR* is the group of states that had a balanced budget requirement in 1929, *Debt Restriction* is the group of states that had either a constitutional or procedural (e.g., voter referendum) debt issue restriction, and *Legislature* is the group of states for which the power to issue debt resided in the legislature. The three groups are mutually exclusive: states with a balanced budget requirement are grouped into the *BBR* category whether or not they had constitutional or procedural debt issue restrictions. *Budget Surplus* is total government revenues less total government cost payments. Data are unweighted averages across states. Government cost payment data are not available for 1930. Source: U.S. Department of Commerce (Various Issues), Ratchford (1938) and Rodriguez-Tejedo and Wallis (2010).

Figure 1.5: State Government Debt
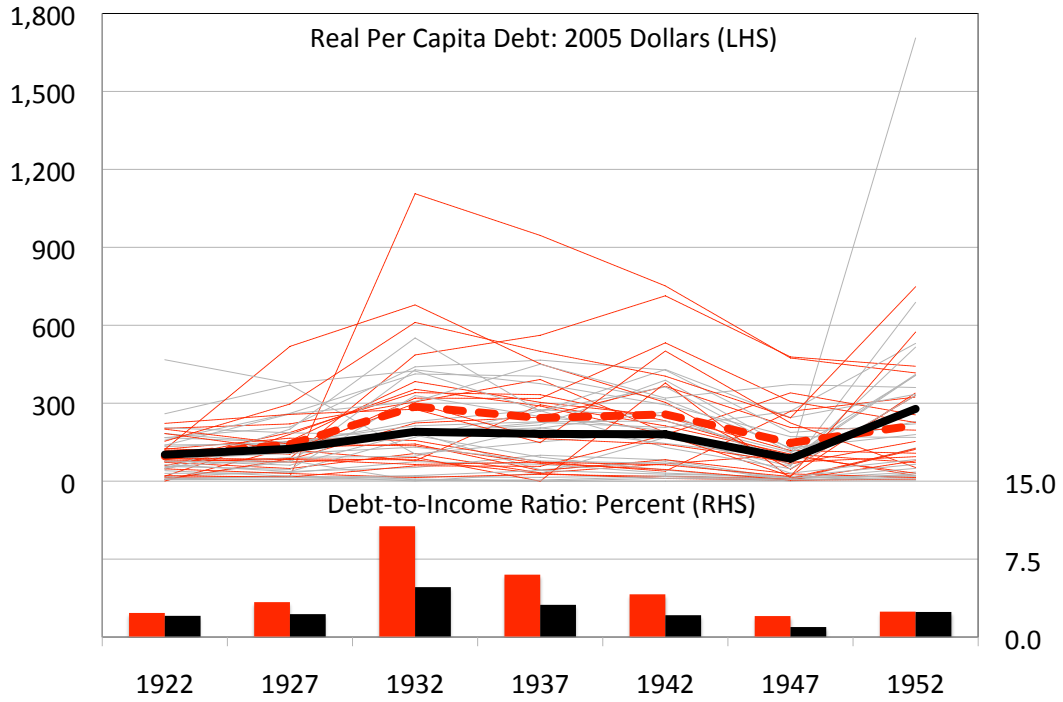
Notes: States with a retail sales tax in 1940 are colored red. *Real Debt* is gross debt less sinking fund assets, deflated by the U.S. GDP deflator. *Income* is state per capita personal income. Sources: BEA, Due and Mikesell (1995), U.S. Department of Commerce (Various Issues).

Figure 1.6: Number of States With Tax Base: By Year for 1930s
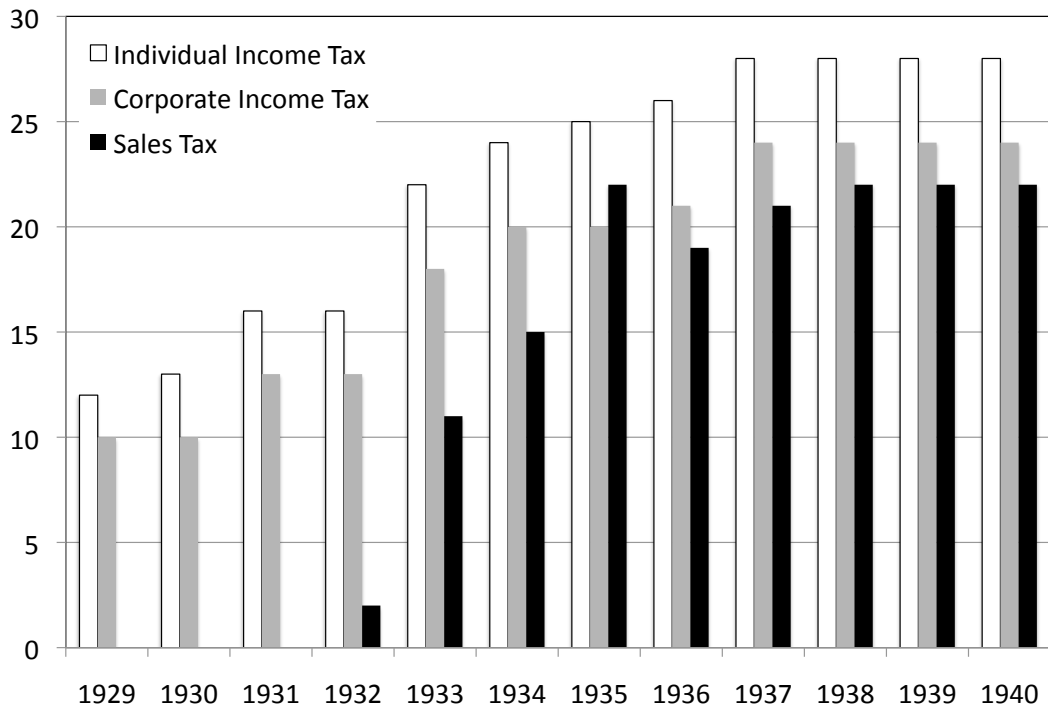
Note: Includes temporary sales tax bases. Sources: Due and Mikesell (1995) and Penniman (1980).

Figure 1.7: Tax Revenue Shares: By Presence of Sales Tax in 1940
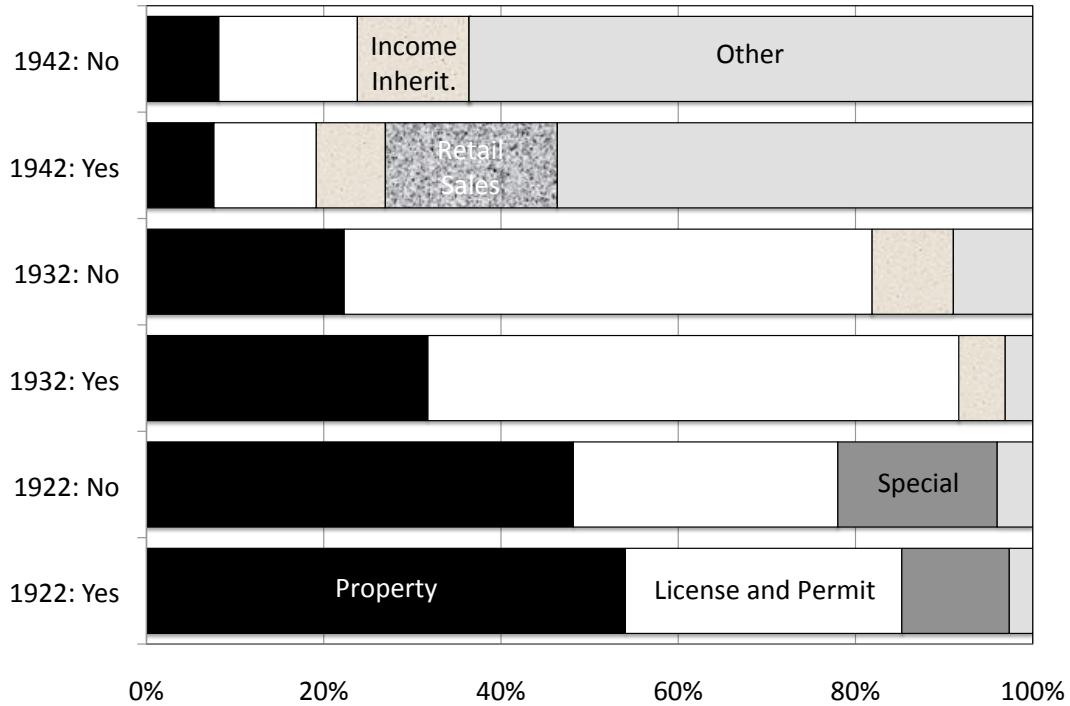
Note: *Yes* indicates the group of states with a retail sales tax in 1940, and *No* indicates the group of states without a retail sales tax in 1940. Data are unweighted averages across states. *Special* taxes in 1922 includes income, inheritance, and other special taxes, which are not reported separately at the state government level for all states. *License and Permit* includes motor vehicle, non-business, and business license taxes. *Retail Sales* taxes in 1942 excludes Arkansas, Michigan, and West Virginia because retail sales and gross receipts tax revenue was not reported separately. *Other* in 1942 includes motor vehicle fuel taxes (44 percent of *Other* tax revenue in 1942 in states with a retail sales tax, and 40 percent in states without a retail sales tax), alcoholic beverage taxes (8/10 percent of *Other*), and unemployment compensation taxes (28/33 percent of *Other*). Source: Due and Mikesell (1995), and U.S. Department of Commerce (Various Issues).

Figure 1.8: State Government Spending
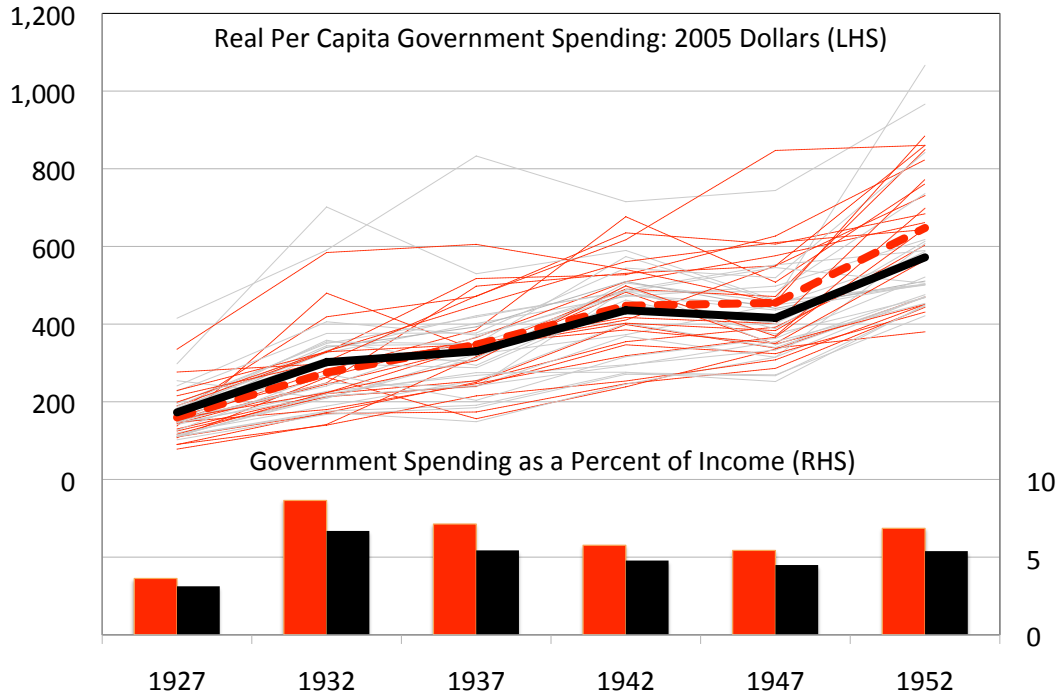
Note: States with a retail sales tax in 1940 are colored red. *Income* is state per capita personal income. Nominal state government spending was converted to real values using the U.S. GDP deflator. Spending data are total state government cost payments for 1922-1937, and total state government expenditures for 1942-1952. Sources: BEA, Due and Mikesell (1995), U.S. Department of Commerce (Various Issues).

Figure 1.9: State Government Revenues
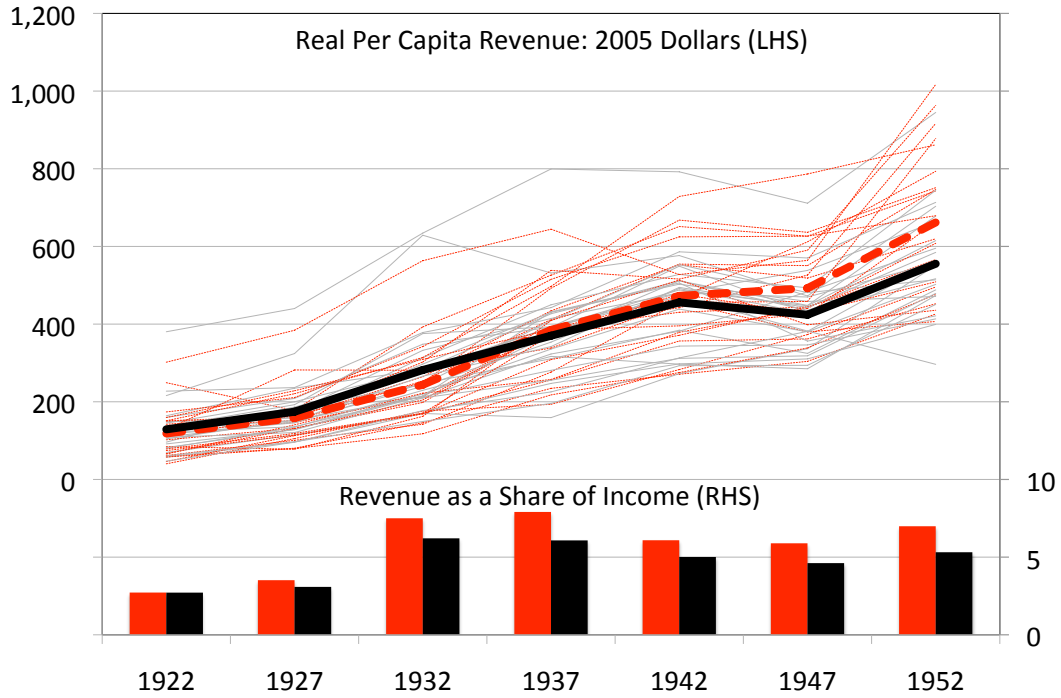
Real Per Capita Revenue: 2005 Dollars (LHS)

Revenue as a Share of Income (RHS)

Note: States with a retail sales tax in 1940 are colored red. *Income* is state per capita personal income. Nominal revenues were converted to real values using the U.S. GDP deflator. Sources: BEA, Due and Mikesell (1995), U.S. Department of Commerce (Various Issues).

Figure 1.10: Retail Sales Tax Adoption: By Size of Income Shock

Real State Per Capita Personal Income: 1929=100
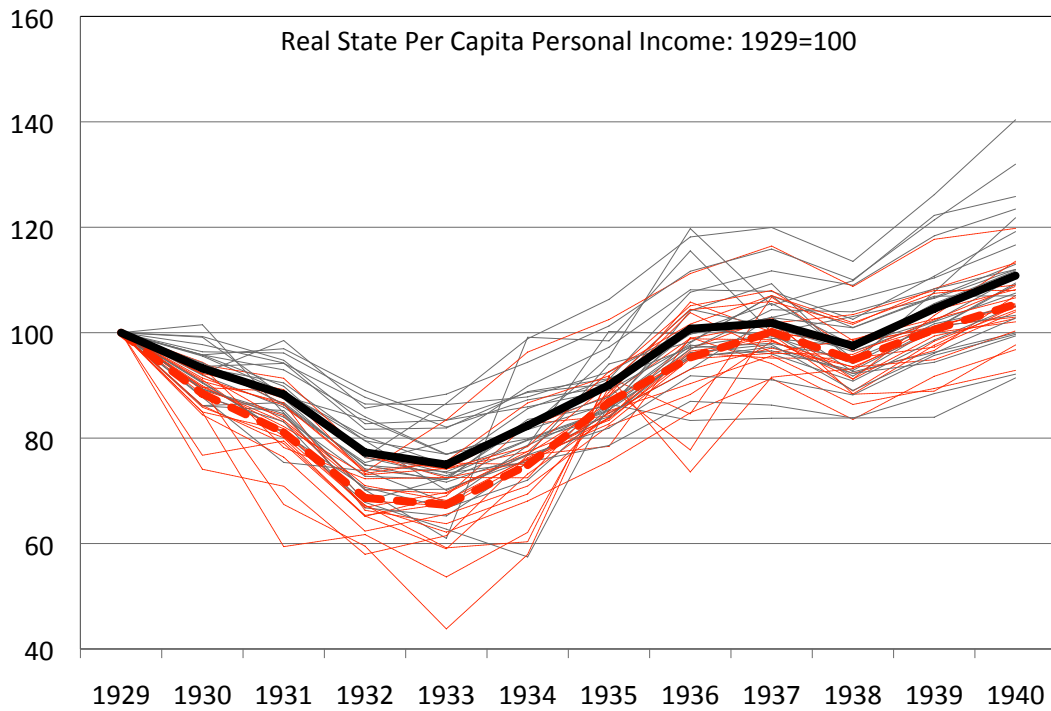
Note: States with a retail sales tax in 1940 are colored red. Nominal state personal income was converted to real values using the U.S. GDP deflator. Sources: BEA, U.S. Department of Commerce (Various Issues).

Figure 1.11: Assessed Property Values and Income Shocks: By State
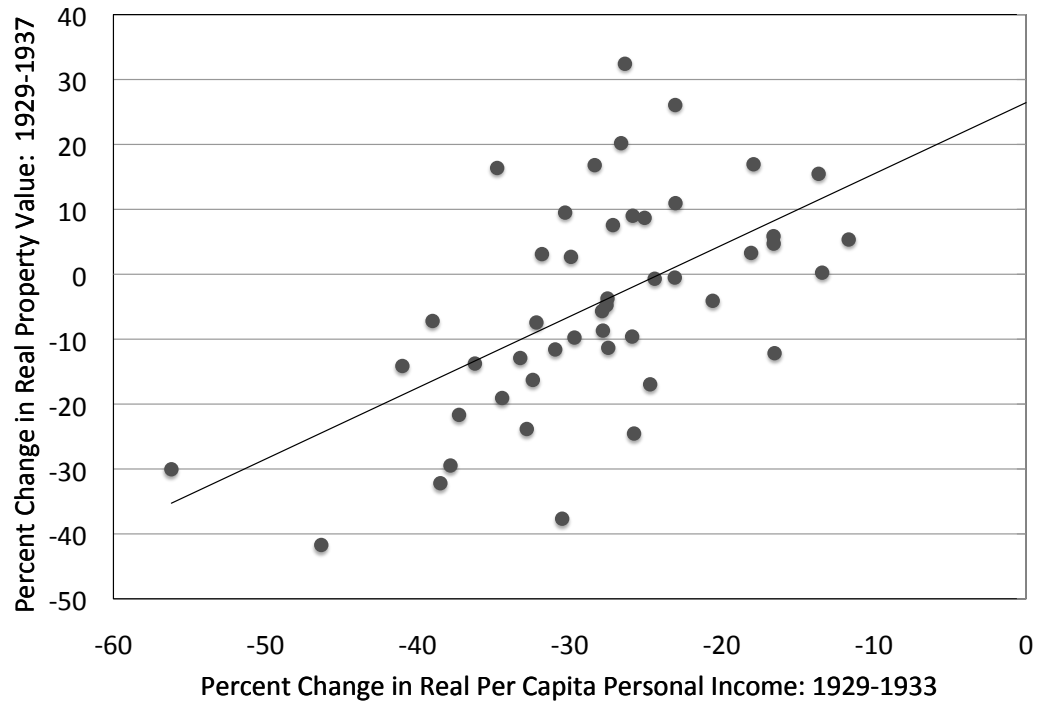
Notes: The straight line gives the OLS estimated linear relationship. Income is state per capita personal income deflated by the U.S. GDP deflator, and property value is the assessed value of general property. Iowa, which experienced a 158 percent increase in the real value of property between 1929 and 1937, is excluded from the figure. Sources: BEA, and U.S. Department of Commerce (Various Issues).

Figure 1.12: Retail Sales Tax Adoption: By Percentage Change in Assessed Property Values
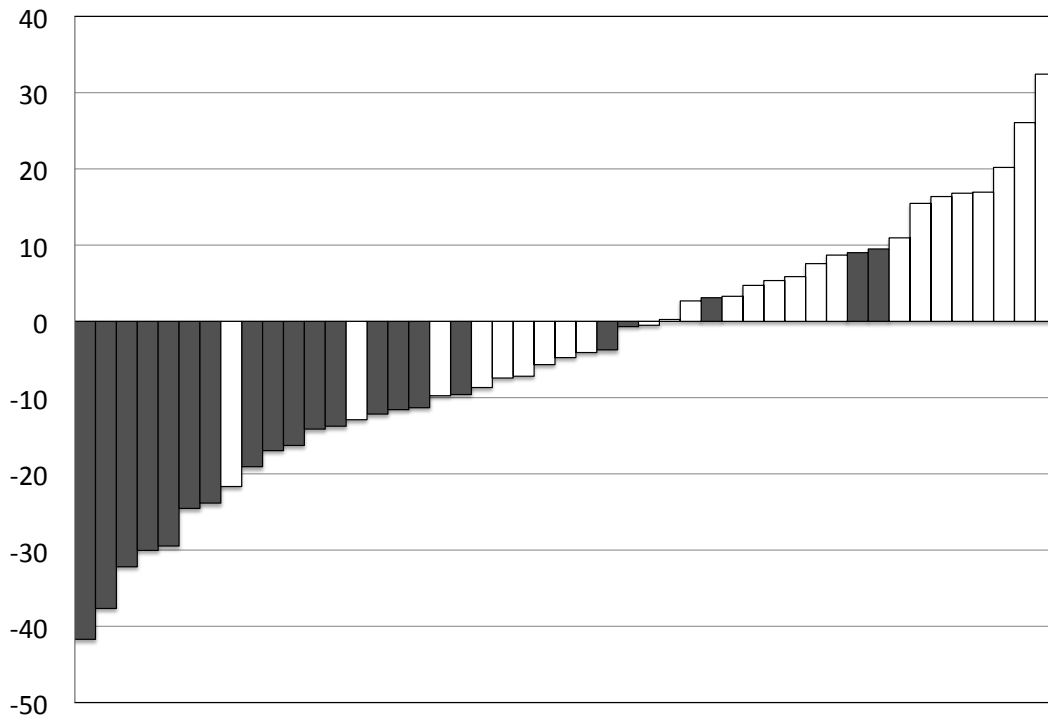
Note: Each column indicates the real percentage change from 1929-1937 in the assessed valuation of property subject to general property tax. States that adopted a retail sales tax by 1940 are shaded black. California, Delaware, Pennsylvania and North Carolina did not levy property tax at the state government level in 1929. Two of those states, California and North Carolina, had adopted a retail sales tax by 1940. Iowa, which experienced a 158 percent increase in the real assessed value of property between 1929 and 1937, is excluded from the figure. Sources: Due and Mikesell (1995), and U.S. Department of Commerce (Various Issues).

1.5 Appendix

1.5.1 Results Omitted From Main Text

1.5.1.1 Default Threshold

For the two-period model of sovereign borrowing outlined in Section 1.2.6, this appendix shows the steps used to find the threshold level of income below which default occurs. Default occurs if second-period utility is higher under default than repayment: $\hat{v}^d_{t+1} \equiv \hat{v}(A_{t+1}(1-\omega), 0, 0, b_{t+1}, b_{t+2}) > \hat{v}(A_{t+1}, D_{t+1}, 0, b_{t+1}, b_{t+2}) \equiv \hat{v}^c_{t+1}$. With some algebra, it can be shown that

$$\hat{v}^c_{t+1} = [\gamma + \phi log\phi] + (1+\phi) log A_{t+1} + (b_{t+1} + \phi) log\left(1 - \tau^c_{t+1}\right), \tag{1.5.1}$$

and

$$\hat{v}^d_{t+1} = [\gamma + \phi log\phi] + (1+\phi) log A_{t+1}(1-\omega) + (b_{t+1} + \phi) log\left(1 - \tau^d_{t+1}\right), \tag{1.5.2}$$

where $\tau^c_{t+1} = \tau(A_{t+1}, D_{t+1}, 0, b_{t+1}, b_{t+2})$ is the optimal tax rate conditional on debt repayment, and $\tau^d_{t+1} = \tau(A_{t+1}(1-\omega), 0, 0, b_{t+1}, b_{t+2})$ is the optimal tax rate conditional on default. Subtracting Equation (1.5.1) from Equation (1.5.2) gives the following expression for the default condition:

$$
\begin{aligned}
(1+\phi) log(1-\omega) &> (b_{t+1} + \phi) log\left(\frac{1 - \tau^c_{t+1}}{1 - \tau^d_{t+1}}\right) \\
&= (b_{t+1} + \phi)\left[log\left(\frac{A_{t+1}b_{t+1} - \left[\xi_{t+1} + \lambda\bar{G} + D_{t+1}\right]}{A_{t+1}(1-\omega)b_{t+1} - \left[\xi_{t+1} + \lambda\bar{G}\right]}\right) + log(1-\omega)\right],
\end{aligned}
\tag{1.5.3}
$$

with the second line following from the condition for the optimal tax rate, given by Equation (1.2.10). With some further re-arrangement it can be shown that default occurs for $A_{t+1} <$

$\bar{A}_{t+1}^d$, where $\bar{A}_{t+1}^d$ is given by Equation (1.2.20). For use in the next section, note that by re-arrangement of Equation (1.5.3) if can be shown that when

$$(1 - \omega)^{\frac{1+\phi}{b_{t+1}+\phi}} \equiv \Gamma_{t+1}(\omega, b_{t+1}) = \frac{1 - \tau_{t+1}^c}{1 - \tau_{t+1}^d} \tag{1.5.4}$$

the government is indifferent between repayment and default.

1.5.1.2 Dependence of the Default Threshold on Tax-Base Breadth

For the two-period model of sovereign borrowing, this appendix section derives the condition under which an increase in second-period fiscal capacity lowers the default threshold $\bar{A}_{t+1}^d$. The utility gain from a marginal increase in second-period tax base breadth for the repayment and default states (evaluated at the default threshold level of income), respectively, is given by:

$$\left. \frac{\partial \hat{v}_{t+1}^c}{\partial b_{t+1}} \right|_{A_{t+1}=A_{t+1}^d} = log\left(1 - \tau_{t+1}^c\right) + \frac{\tau_{t+1}^c}{1 - \tau_{t+1}^c} - \frac{\xi_f'(b_{t+1})}{A_{t+1}\left(1 - \tau_{t+1}^c\right)}, \tag{1.5.5}$$

and

$$\left. \frac{\partial \hat{v}_{t+1}^d}{\partial b_{t+1}} \right|_{A_{t+1}=A_{t+1}^d} = log\left(1 - \tau_{t+1}^d\right) + \frac{\tau_{t+1}^d}{1 - \tau_{t+1}^d} - \frac{\xi_f'(b_{t+1})}{A_{t+1}\left(1 - \omega\right)\left(1 - \tau_{t+1}^d\right)}, \tag{1.5.6}$$

where $\tau_{t+1}^c = \tau\left(A_{t+1}, D_{t+1}, 0, b_{t+1}, b_{t+2}\right)$ and $\tau_{t+1}^d = \tau\left(A_{t+1}\left(1 - \omega\right), 0, 0, b_{t+1}, b_{t+2}\right)$ are the optimal tax rates in the repayment and default states, respectively. The increase in utility in the repayment state relative to the default state is given by

$$\left. \frac{\partial \hat{v}_{t+1}^c}{\partial b_{t+1}} - \frac{\partial \hat{v}_{t+1}^d}{\partial b_{t+1}} \right|_{A_{t+1}=A_{t+1}^d} = log\left(\frac{1 - \tau_{t+1}^c}{1 - \tau_{t+1}^d}\right) + \left[\frac{\tau_{t+1}^c}{1 - \tau_{t+1}^c} - \frac{\tau_{t+1}^d}{1 - \tau_{t+1}^d}\right] \tag{1.5.7}$$

$$- \frac{\xi_f'}{A_{t+1}}\left[\frac{1}{\left(1 - \tau_{t+1}^c\right)} - \frac{1}{\left(1 - \omega\right)\left(1 - \tau_{t+1}^d\right)}\right],$$

which using the condition given by Equation (1.5.4) (that applies at the income level where the government is indifferent between repayment and default) can be re-expressed to yield

$$
\begin{aligned}
\frac{\partial \hat{v}_{t+1}^c}{\partial b_{t+1}} - \frac{\partial \hat{v}_{t+1}^d}{\partial b_{t+1}} \bigg|_{A_{t+1}=A_{t+1}^d} &= log\Gamma_{t+1} + \left(\frac{1}{1-\tau_{t+1}^c}\right)\left[(1-\Gamma_{t+1})\left(1-\frac{\xi_f'}{A_{t+1}^d}\right)\right. \tag{1.5.8} \\
&\quad \left. +\Gamma_{t+1}\left(\frac{\xi_f'}{A_{t+1}^d}\right)\left(\frac{\omega}{1-\omega}\right)\right] \\
&\geq log\Gamma_{t+1} + \left(\frac{1}{1-\tau_{t+1}^c}\right)(1-\Gamma_{t+1})\left(1-\frac{\xi_f'}{A_{t+1}^d}\right) \tag{1.5.9} \\
&\simeq \omega\left(\frac{1+\phi}{b_{t+1}+\phi}\right)\left[\left(\frac{1}{1-\tau_{t+1}^c}\right)\left(1-\frac{\xi_f'}{A_{t+1}^d}\right)-1\right], \tag{1.5.10}
\end{aligned}
$$

where the approximate equality follows from taking a first-order Taylor series approximation about $\omega = 0$ (the output cost experienced upon default). Hence, a marginal increase in tax base breadth raises utility at the default threshold by more in the repayment than the default state if $\tau_{t+1}^c A_{t+1}^d > \xi_f'(b_{t+1})$. This is a weak condition that holds except when the tax base breadth is well-above its optimal level; Equation (1.5.5) implies that, at an optimum for tax base breadth in the repayment state at the default threshold level of income,

$$
\frac{\xi_f'(b_{t+1})}{A_{t+1}^d} = \left(1-\tau_{t+1}^c\right)log\left(1-\tau_{t+1}^c\right) + \tau_{t+1}^c \simeq \frac{1}{2}\left(\tau_{t+1}^c\right)^2 << \tau_{t+1}^c, \tag{1.5.11}
$$

so that $\tau_{t+1}^c A_{t+1}^d >> \xi_f'(b_{t+1})$. Under this condition, following a marginal increase in tax base breadth, repayment gives higher utility than default at the original threshold level A_{t+1}^d. If repayment is preferred to default at A_{t+1}^d then it must also be preferred at all higher levels of income $A_{t+1} > A_{t+1}^d$ (see Section 1.5.1.1). Thus, the default set shrinks following a marginal increase in tax base breadth—and the threshold income level at which the government is indifferent between repayment and default must fall.

1.5.1.3 Effect of Tax Base Breadth on Default Threshold: By Debt Level

To prove the claim made in Section 1.2.6 that $\partial^2 A_{t+1}^d / \partial b_{t+1} \partial D_{t+1} < 0$ first observe that

$$\left. \frac{\partial^2 \hat{v}_{t+1}^c}{\partial b_{t+1} \partial D_{t+1}} \right|_{A_{t+1}=A_{t+1}^d} = \left(\frac{1}{1-\tau_{t+1}^c} \right)^2 \left[\tau_{t+1}^c - \frac{\xi_f'}{A_{t+1}^d} \right] \frac{\partial \tau_{t+1}^c}{\partial D_{t+1}}, \tag{1.5.12}$$

and

$$\left. \frac{\partial^2 \hat{v}_{t+1}^c}{\partial b_{t+1} \partial D_{t+1}} \right|_{A_{t+1}=A_{t+1}^d} = \left(\frac{1}{1-\tau_{t+1}^d} \right)^2 \left[\tau_{t+1}^d - \frac{\xi_f'}{A_{t+1}^d} \right] \frac{\partial \tau_{t+1}^d}{\partial D_{t+1}} = 0 \tag{1.5.13}$$

because τ_{t+1}^d (the tax rate in the default state) does not depends on the level of debt due for repayment. Provided the condition given by Equation (1.5.11) is satisfied,

$$\left. \frac{\partial^2 \hat{v}_{t+1}^c}{\partial b_{t+1} \partial D_{t+1}} \right|_{A_{t+1}=A_{t+1}^d} - \left. \frac{\partial^2 \hat{v}_{t+1}^d}{\partial b_{t+1} \partial D_{t+1}} \right|_{A_{t+1}=A_{t+1}^d} > 0 \tag{1.5.14}$$

because $\partial \tau_{t+1}^c / \partial D_{t+1} > 0$ (see Equation 1.2.10 and recall that $\tau_{t+1}^c = \tau \left(A_{t+1}, D_{t+1}, 0, b_{t+1}, b_{t+2} \right)$). Hence, a marginal increase in tax base breadth makes repayment preferred to default at $A_{t+1} = A_{t+1}^d$—the more so the larger is the level of first period borrowing. Recalling that the difference $\hat{v}_{t+1}^c - \hat{v}_{t+1}^d$ is increasing in A_{t+1}, the greater is the difference $\hat{v}_{t+1}^c \left(A_{t+1}^d \right) - \hat{v}_{t+1}^d \left(A_{t+1}^d \right)$ the larger is the fall in the threshold level of income. Thus, the higher the level of debt, the greater is the reduction in the threshold level A_{t+1}^d following a marginal increase in tax base breadth: $\partial^2 A_{t+1}^d / \partial b_{t+1} \partial D_{t+1} < 0$.

1.5.1.4 Effect of Tax Base Breadth on Total Borrowing Costs: By Debt Level

This section uses the results derived earlier to show that $\partial^2 q_t / \partial b_{t+1} \partial D_{t+1} > 0$. Using the definition in Section 1.2.6,

$$q_t = \frac{1 - F \left(A_{t+1}^d | A_t \right)}{1+r}, \tag{1.5.15}$$

$$\frac{\partial q_t}{\partial b_{t+1}} = -\frac{1}{1+r} f\left(A^d_{t+1}|A_t\right) \frac{\partial A^d_{t+1}}{\partial b_{t+1}}, \tag{1.5.16}$$

and

$$\frac{\partial^2 q_t}{\partial b_{t+1}\partial D_{t+1}} = -\frac{1}{1+r} \frac{\partial f\left(A^d_{t+1}|A_t\right)}{\partial A^d_{t+1}} \frac{\partial A^d_{t+1}}{\partial b_{t+1}} - \frac{1}{1+r} f\left(A^d_{t+1}|A_t\right) \frac{\partial^2 A^d_{t+1}}{\partial b_{t+1}\partial D_{t+1}}. \tag{1.5.17}$$

Next, note that $\partial A^d_{t+1}/\partial b_{t+1} < 0$ (see Section 1.5.1.2), and $\partial^2 A^d_{t+1}/\partial b_{t+1}\partial D_{t+1} < 0$ (see Section 1.5.1.3). Under the assumption that $\partial f\left(A^d_{t+1}|A_t\right)/\partial A^d_{t+1} > 0$, a marginal increase in tax base breadth reduces borrowing costs by more the higher is the level of borrowing: $\partial^2 q_t/\partial b_{t+1}\partial D_{t+1} > 0$.

1.5.2 Quantitative Model

1.5.2.1 Solution Algorithm

The following steps describe how to solve the quantitative model:

1. Discretize the state space, 50 points for the debt state space, and 25 points for the income state space;

2. Set the initial guess for the bond price function to be the risk-free rate: $q_0\left(A_t, D_{t+1}, b_{t-1}\right) = \frac{1}{1+r}$;

3. Taking the bond price function as given, solve the government's maximization problem by value function iteration, to get the government's policy functions, value functions, and repayment and default sets;

4. Update the bond price function given the government's repayment and default sets, to get $q_1\left(A_t, D_{t+1}, b_{t-1}\right)$. If $|q_0\left(A_t, D_{t+1}, b_{t-1}\right) - q_1\left(A_t, D_{t+1}, b_{t-1}\right)| < \delta$, where δ is the convergence tolerance for the bond price function, stop, otherwise iterate over steps 2-4 until convergence.

1.5.2.2 Sensitivity Analysis: Fixed Setup Cost to Increase Tax Base Breadth

Figure 1.13 shows the combinations of debt and income for which an increase in tax base breadth is optimal, for different values of the fixed setup cost parameter F. All other parameter values are the same as in the baseline specification (see Table 1.1).

Figure 1.13: Numerical Solution: Upgrading of Fiscal Capacity: Sensitivity
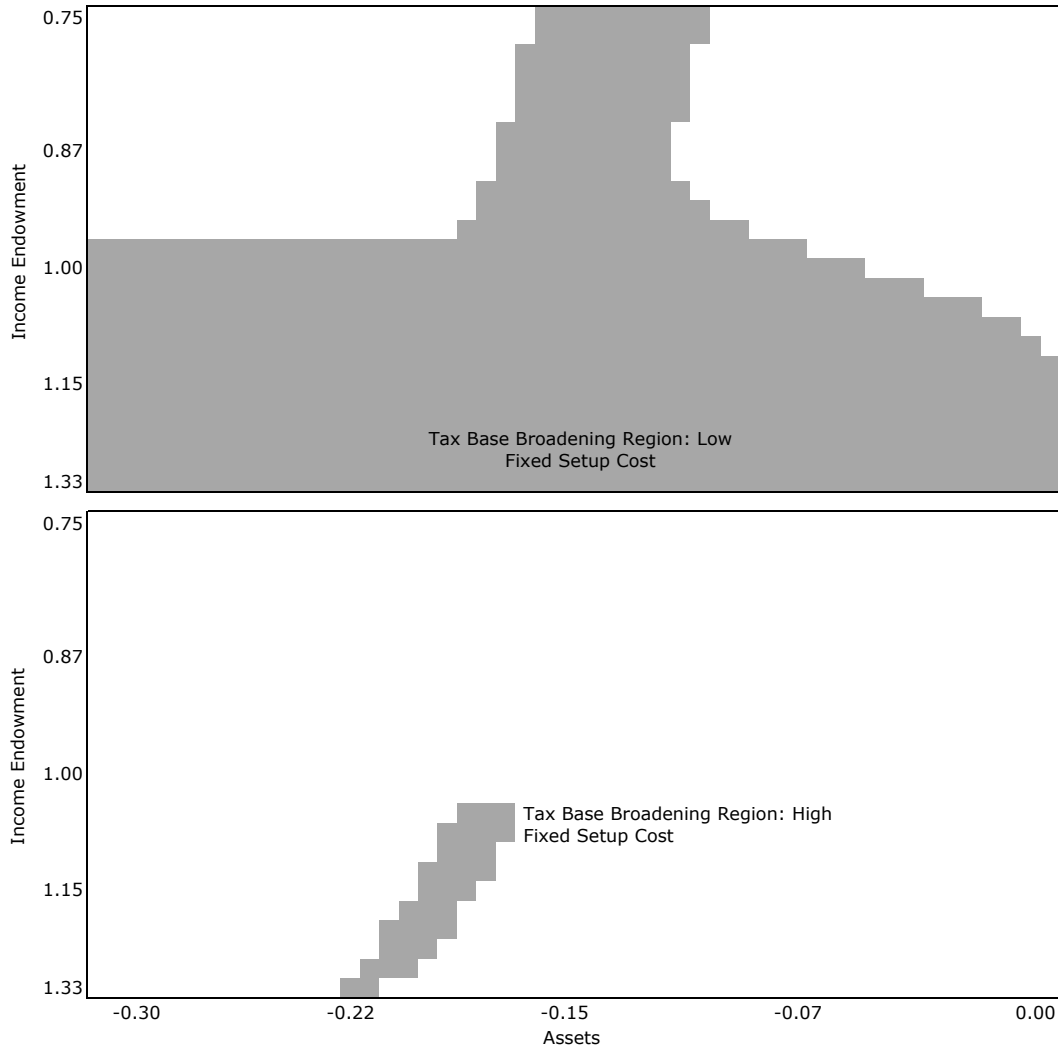
Notes: The shaded regions show combinations of debt (on the x-axis) and income (on the y-axis) for which a government with a low level of fiscal capacity chooses to upgrade to a high level of fiscal capacity: the upper panel shows results when the fixed setup cost is low ($F = 0.006$), and the upper panel when the fixed cost is high ($F = 0.017$).

1.6 Bibliography

Aguiar, Mark and Manuel Amador (Forthcoming). Sovereign Debt. *Handbook of International Economics 4*.

Aguiar, Mark and Gita Gopinath (2006). Defaultable Debt, Interest Rates and the Current Account. *Journal of International Economics* 69(1): 64–83.

Arellano, Cristina (2008). Default Risk and Income Fluctuations in Emerging Economies. *American Economic Review* 98(3): 690–712.

Besley, Timothy, Ethan Ilzetzki, and Torsten Persson (2013). Weak States and Steady States: The Dynamics of Fiscal Capacity. *American Economic Journal: Macroeconomics* 5(4): 205–235.

Besley, Timothy and Torsten Persson (2009). The Origins of State Capacity: Property Rights, Taxation and Politics. *American Economic Review* 99(4): 1218–44.

Besley, Timothy and Torsten Persson (2010). State Capacity, Conflict and Development. *Econometrica 78*: 1–34.

Besley, Timothy and Torsten Persson (2013). Taxation and Development. *Handbook of Public Economics 5*: 51–110.

Bulow, Jeremy and Keneth Rogoff (1989b). Sovereign Debt: Is to Forgive to Forget? *American Economic Review* 79(1): 43–50.

Due, John F. and John L. Mikesell (1995). *Sales Taxation: State and Local Structure and Administration* (2 ed.). Washington DC: Urban Institute Press.

Eaton, Jonathan and Mark Gersovitz (1981). Debt with Potential Repudiation: Theoretical and Empirical Analysis. *Review of Economic Studies* 48(2): 289–309.

English, William B. (1996). Understanding the Costs of Sovereign Default: American State Debts in the 1840s. *American Economic Review* 86(1): 259–275.

Lucas, Robert E. Jr. (1987). *Model of Business Cycles*. New York: Basil Blackwell.

Moody's (1920-1950). *Municipal and Government Bond Manual*. New York: Moody's Investors Service.

Organization for Economic Co-operation and Development (OECD). (2011). Tax Administration in OECD and Selected non-OECD countries: Comparative Information Series (2010). Available at: http://www.oecd.org/dataoecd/2/37/47228941.pdf.

Penniman, Clara (1980). *State Income Taxation*. Baltimore MD: The John Hopkins University Press.

Ratchford, Benjamin U. (1938). American Government and Politics: Constitutional Provisions Governing State Borrowing. *American Political Science Review* 32(4): 694–707.

Ratchford, Benjamin U. (1941). *American State Debts*. Durham NC: Duke University Press.

Rodriguez-Tejedo, Isabel and John J. Wallis (2010). Lessons for California from the History of Fiscal Constitutions. *California Journal of Politics and Policy* 2(3).

Romer, Christina D. and David H. Romer (2013). The Incentive Effects of Marginal Tax Rates: Evidence from the Interwar Era, Working Paper, University of California, Berkeley.

Saez, Emmanuel, Joel Slemrod, and Seth Giertz (2012). The Elasticity of Taxable Income With Respect to Marginal Tax Rates: A Critical Review. *Journal of Economic Literature* 50(1): 3–50.

Slemrod, Joel and Wojciech Kopczuk (2002). The Optimal Elasticity of Taxable Income. *Journal of Public Economics* 84(1): 91–112.

Tilly, Charles (1975). Reflections on the History of European State-Making. In Charles Tilly (Ed.), *The Formation of Nation States in Western Europe*. Princeton NJ: Princeton University Press.

U.S. Department of Commerce (Various Years). *Statistical Abstract of the United States*. Washington DC: United States Government Printing Office, Bureau of Foreign and Domestic Commerce.

Wallis, John J. (2005). Constitutions, Corporations, and Corruption: American States and Constitutional Change, 1842 to 1852. *Journal of Economic History* 65(1): 211–256.

Wilson, John D. (1989). On the Optimal Tax Base for Commodity Taxation. *American Economic Review* 79(5): 1196–1206.

Yitzhaki, Shlomo (1979). A Note on Optimal Taxation and Administrative Costs. *American Economic Review* 69(3): 475–480.

Chapter 2

Consumption Commitments in General Equilibrium

2.1   Introduction

This paper proposes a model of consumption commitments that derives inertial aggregate consumption dynamics in an otherwise standard general equilibrium macroeconomic model. Consumption commitments describe any purchase that triggers an ongoing sequence of consumption and payments that is costly to alter. Housing services, which generates ongoing rental or rental equivalence payments, is the most important source of commitments for most households. Purchase of durable goods, such as autos and furniture, are another important source of consumption commitments. In addition to a service flow that has costs to adjust, these goods and services in turn often trigger further commitments, such as financial insurance and utility payments for housing. Chetty and Szeidl (2007) estimate that about two-thirds of the typical household's consumption basket is subject to commitments.

Because consumption is adjusted for only a subset of goods each period in the model, aggregate consumption responds more slowly to economic news in the presence of commitments than in the standard frictionless model. This allows the model to match the empirical findings of excess smoothness and excess sensitivity in consumption data, generating humped-shaped

dynamics in response to economic news.[1] Aggregate consumption reflects, in part, information available at the time commitments were made, and thus reflects dated information, as in sticky information models.

Households are forward looking when making commitments, taking into account the likelihood that choices made today will constrain utility and spending in future periods. As a consequence, the consumption response to, for example, a transitory real interest rate cut is attenuated relative to a frictionless world, because commitments increase the chance consumption will be remain stuck at a high level long after the real interest rate has returned to its normal level; a corollary is that only long-lasting changes in the expected real interest rate can be used to reliably estimate the structural elasticity of intertemporal substitution in the presence of consumption commitments. Because adjustment costs can delay and attenuate households' consumption response to equity price changes, commitments mask the extent of consumption risk borne by households via their equity holdings, and help to explain some of the apparent equity-premium puzzle.

An important feature of the model is endogenous adjustment of the share of goods over which households maintain commitments. By incurring a monetary cost households can reduce the share of goods for which commitments must be kept each period, with the marginal adjustment cost assumed to be rising in the share of commitments adjusted. In response to small shocks, households optimally choose to maintain a high share of commitments but, faced with large shocks, households choose to pay an increased adjustment cost and reset a large share of commitments. This adjustment mechanism allows aggregate consumption to exhibit a high degree of inertia in the presence of small shocks, but can endogenously generate bigger changes in consumption in response to large shocks.

On several dimensions, the model matches the behavior of the now widely used habit-formation preference specification, but reference dependence arises via consumption adjustment costs, not a behavioral assumption on consumer preferences. A special case of the model is equivalent to consumer habit-formation, providing a microfoundation for habit-formation

---

[1]Consumption is said to be excessively smooth when it responds less than one-for-one to permanent income news shocks, and to display excess sensitivity when it can be predicted by lagged information.

preferences.

This paper shares many qualitative features with the consumption commitments model outlined by Chetty and Szeidl (2010). Households in their two-good economy (one of which is subject to commitments, and the other which is not) are heterogeneous in income risk, leading to variation in the timing of adjustment for the commitment good. This variation in timing creates smooth aggregate consumption dynamics for the commitment good, but the accompanying heterogeneity in asset holdings limits the ability to incorporate their framework into a general equilibrium model. In contrast, the model presented here has a representative agent formulation, and thus can be easily incorporated into a general equilibrium model.

Several related literatures have investigated the implications of consumption commitments. Chetty and Szeidl (2007) show that consumption commitments can raise risk aversion over moderate-stakes gambles: large gambles trigger re-optimization of commitments, limiting their effect on risk preferences; Postlewaite et al. (2008) show that such heightened risk-aversion over moderate-stakes gambles means optimal employment contracts can feature rigid wages, but the prospect of dismissal. In a finance application, Grossman and Laroque (1990) find that in an economy with a single durable good even small transaction costs can lead to infrequent adjustment of consumption, and consequently a low contemporaneous correlation between consumption growth and stock returns; they argue that the apparent equity-premium puzzle is an artifact of the short horizons over which stock returns and consumption growth are typically compared. Flavin and Nakagawa (2008) extend the Grossman and Laroque (1990) model to a two-good housing (which is subject to adjustment costs) and non-housing consumption setting, and find that their model shares many features with the habit-persistence model.

In related work, a range of other frictions have been proposed to explain empirically documented consumption anomalies. Reis (2006) studies the consumption behavior of an agent who finds it costly acquire and process information: households in his model endogenously choose to make consumption plans sporadically, generating delayed and gradual adjustment of aggregate consumption to economic news. Sims (2003) and Moscarini (2004) present

conceptually similar models featuring agents with limited information processing capacity.

In what follows, Section 2.2 lays out the model formally. Section 2.3 discusses the ability of the model the explain the excess smoothness and excess sensitivity properties of aggregate consumption, and discusses the model's implications for the equity-premium puzzle. Section 2.4 solves the model quantitatively, and subjects it to a series of exogenous monetary policy shocks, in order to demonstrate some features of the commitments model in a general equilibrium setting. Section 2.5 provides some concluding remarks.

2.2   Household Problem with Consumption Commitments

2.2.1   Model Overview

The model assumes that households derive utility $U_t = \int_0^1 u\left(c_t\left(j\right)\right) dj$ at time t from a continuum of consumption goods $c_t\left(j\right)$, with $j \, \epsilon \, [0, 1]$. Utility is additively separable within and across time periods for each good, and is concave in the level of consumption for each good. Perfectly competitive final goods firms producing each consumer good face the same costs, implying a relative price of unity for each pair of goods, in all time periods. (The firm side of the model is described in Section 2.4.)

In the absence of frictions, utility is maximized for a given level of spending when consumption in each time period is equated across goods. To see this, consider the expenditure minimization problem for a household that faces no frictions in choosing its consumption basket. In each period, the household solves:

$$\min_{\{c_t(j)\}} \quad \int_0^1 P_t c_t\left(j\right) dj \quad s.t. \quad U_t \geq \overline{U}, \tag{2.2.1}$$

where P_t is the nominal price for each good at time t. Letting λ be the Lagrange multiplier on the constraint, the first-order condition for this expenditure minimization problem is:

$$[\partial c_t\left(j\right)] \quad P_t = \lambda u'\left(c_t\left(j\right)\right), \tag{2.2.2}$$

from which it is clear that at an optimum $c_t(j) = c_t(k) = c_t$ for all $j, k \in [0,1]$. An optimum requires equating spending across goods because each good has the same price P_t and there is diminishing marginal utility in consumption of each good.

The key departure from this frictionless benchmark is an assumption that households face consumption adjustment costs: consumption decisions made today create commitments that are costly to adjust. Consumption commitments may arise for a variety of reasons, as discussed in the introduction. Regardless of their source, the model assumes that commitments create adjustment costs that can be represented in monetary terms. Relative to the benchmark model without commitments, the only new parameters introduced are those related to the adjustment cost function.

When economic news arrives, a household must pay a real cost $D(\theta_t)$ to adjust spending for a $(1 - \theta_t)$-share of its consumption basket. For tractability, when a household chooses to abandon a $(1 - \theta_t)$-share of its commitments, the set of commitments adjusted is assumed to be random. This assumption dramatically simplifies the household's problem by eliminating the need to keep track of the history of consumption choices for each good. The function $D(\theta_t)$ is assumed to satisfy $D(\bar{\theta}) = 0$, $\lim_{\theta_t \to 0} D(\theta_t) = \infty$, $D'(\bar{\theta}) = 0$, and $\lim_{\theta_t \to 0} D'(\theta_t) = -\infty$, where $\bar{\theta} < 1$ is the probability that consumption for a good can be adjusted at no cost: maintaining last period's consumption basket creates no adjustment cost, but the cost to abandon all commitments and choose an entirely new consumption basket is infinite.

Crucially, marginal utility is no longer equated across goods in each time period. Commitments made in good times may remain in place in bad times because they are too costly to abandon, and vice versa. The greater the dispersion in marginal utility across goods, the lower is total utility for a given level of spending. For those goods whose consumption is reoptimized, the same level of consumption is chosen because this maximizes utility for a given level of spending; there are no dynamic considerations that would lead a household to reset consumption to a different level across goods, because the probability a commitment is abandoned is history independent.

The share of goods whose consumption is reset each period is chosen by the household to

balance adjustment costs against the benefit of a more equal distribution of spending, and therefore marginal utility, across goods. Households are forward looking in choosing the share of commitments to adjust, and in making new commitments, taking into account that choices made today are costly to alter in the future.

Total consumption is less responsive to economic news in the presence of commitments than in the standard model: a θ_t-share of the previous period's consumption basket is unchanged, and the concavity of the utility function makes a compensating change in consumption for goods whose consumption is reset undesirable. The following example illustrates the effect of commitments on the response of consumption to economic news. Suppose a household receives a negative permanent income news shock which, in the absence of commitments, the household would respond to by proportionately reducing spending on each of its consumption goods by Δ. In the presence of commitments, it is prohibitively costly to adjust consumption for all goods, so instead the household chooses to keep a θ_t-share of commitments made before the news shock arrived. With spending 'stuck' at a higher than desired level for the θ_t-share of goods subject to ongoing commitments, consumption for the $(1 - \theta_t)$-share of goods whose spending is rechosen would need to decline by $\left(\frac{\theta_t}{1-\theta_t}\right) \Delta$ to achieve the same decline in total spending as in the absence of frictions. But such a large fall in consumption spread across only a $(1 - \theta_t)$-share of goods creates a sharp rise in marginal utility. The household tolerates a distorted intertemporal pattern of consumption—incorporating the news about permanent income into its spending over time, rather than immediately—to limit the unevenness in spending across goods in each time period.

The remaining elements of the household's problem are essentially standard. Households supply labor in a perfectly competitive labor market, and own firms from which they receive flow profits each period via a divided. Households rent capital to firms, and are assumed to pay capital adjustment costs. It is important to include capital in the model because, in the presence of consumption adjustment costs, households use saving to smooth abrupt changes in income.

2.2.2 Household Maximization Problem

Formally, the household's optimization problem at time t can be expressed in terms of the following Bellman equation:

$$V_t\left(b_t, K_t, C_{t-1}, U_{t-1}\right) = \max_{\{c_t, N_t, \theta_t, I_t\}} \left\{U_t - v\left(N_t\right) + \beta E_t V_{t+1}\left(b_{t+1}, K_{t+1}, C_t, U_t\right)\right\}, \qquad (2.2.3)$$

subject to the constraints,

$$b_{t+1} = \frac{1}{\pi_{t+1}}\left[R_{t-1}b_t + R_t^k K_t + w_t N_t + \Pi_t - C_t - I_t - D\left(\theta_t\right)\right], \qquad (2.2.4)$$

$$K_{t+1} = \left(1 - \delta\right) K_t + \Phi\left(\frac{I_t}{K_t}\right) K_t, \qquad (2.2.5)$$

$$C_t = \left(1 - \theta_t\right) c_t + \theta_t C_{t-1}, \qquad (2.2.6)$$

$$U_t = \left(1 - \theta_t\right) u\left(c_t\right) + \theta_t U_{t-1}, \qquad (2.2.7)$$

where C_t is total consumption, c_t is the level of consumption for goods whose consumption is reset, b_t is real bonds held, R_{t-1} is the gross nominal rate-of-return on bonds, $\pi_t = \frac{P_t}{P_{t-1}}$ is the inflation rate on consumption goods, N_t is labor supply, K_t is capital held, R_t^k is the real marginal product of capital, Φ is the capital adjustment cost function, Π_t is profits remitted to households, and $D\left(\theta_t\right)$ is the cost (measured in dollars) of maintaining a θ_t-share of consumption commitments. Each period a household resets consumption for a random $\left(1 - \theta_t\right)$-share of its basket of goods, with θ_t endogenously chosen by the household. Consumption goods can be converted into capital goods according to the function $\Phi\left(\frac{I_t}{K_t}\right) K_t$. Letting δ represent the depreciation rate on capital, the function Φ is assumed to satisfy $\Phi\left(\delta\right) = \delta$, $\Phi'\left(\delta\right) = 1$, and $\Phi''\left(\delta\right) < 0$.

Last period's total spending C_{t-1} and level of utility U_{t-1} are introduced as state variables only for analytical convenience. Making use of the randomization assumption, Equation

(2.2.6) follows from the fact that:

$$C_t = \int C_t(j)\,dj$$

$$= (1 - \theta_t)\,c_t + \theta_t \int C_{t-1}(j)\,dj \tag{2.2.8}$$

$$= (1 - \theta_t)\,c_t + \theta_t C_{t-1}.$$

We only need to keep track of last period's total spending and this period's consumption choice c_t to compute today's total spending; this is because the household buys a θ_t-scale replica of yesterday's consumption basket, and spends c_t on each of the $(1 - \theta_t)$-share of goods whose consumption is reset. By the same reasoning, Equation (2.2.7) follows from the fact that:

$$U_t = \int u\,(C_t(j))\,dj$$

$$= (1 - \theta_t)\,u\,(c_t) + \theta_t \int u\,(C_{t-1}(j))\,dj \tag{2.2.9}$$

$$= (1 - \theta_t)\,u\,(c_t) + \theta_t U_{t-1}.$$

Flow utility is a weighted average of utility from last period's consumption basket, and the utility derived from goods whose consumption is reset today. The dependence of flow utility on lagged consumption choices is due to consumption adjustment costs; there is no reference dependence built into the utility function, as assumed in habit formation models.

The first-order conditions for the household's optimization problem are:

$$[\partial c_t] \quad (1 - \theta_t)\,u'\,(c_t)\left[1 + \beta E_t \frac{\partial V_{t+1}}{\partial U}\right] - (1 - \theta_t)\left[\beta E_t \frac{1}{\pi_{t+1}} \frac{\partial V_{t+1}}{\partial b} - \beta E_t \frac{\partial V_{t+1}}{\partial C}\right] = 0, \tag{2.2.10}$$

$$[\partial N_t] \quad -v'\,(N_t) + \beta E_t \frac{w_t}{\pi_{t+1}} \frac{\partial V_{t+1}}{\partial b} = 0, \tag{2.2.11}$$

$$[\partial \theta_t] \quad (U_{t-1} - u\,(c_t))\left[1 + \beta E_t \frac{\partial V_{t+1}}{\partial U}\right] + (c_t - C_{t-1})\,\beta E_t \left[\frac{1}{\pi_{t+1}} \frac{\partial V_{t+1}}{\partial b} - \frac{\partial V_{t+1}}{\partial C}\right]$$

$$- D'\beta E_t \left[\frac{1}{\pi_{t+1}} \frac{\partial V_{t+1}}{\partial b}\right] = 0, \tag{2.2.12}$$

and

$$[\partial I_t] \quad -\beta E_t \frac{1}{\pi_{t+1}} \frac{\partial V_{t+1}}{\partial b} + \beta E_t \Phi_t' \frac{\partial V_{t+1}}{\partial K} = 0. \tag{2.2.13}$$

The envelope conditions for the state variables b, K, U, and C, respectively, are:

$$\frac{\partial V_t}{\partial b} = \beta E_t \frac{R_{t-1}}{\pi_{t+1}} \frac{\partial V_{t+1}}{\partial b}, \tag{2.2.14}$$

$$\frac{\partial V_t}{\partial K} = \beta E_t \frac{R_t^k}{\pi_{t+1}} \frac{\partial V_{t+1}}{\partial b} + \beta E_t \left[(1 - \delta) + \Phi_t - \frac{I_t}{K_t} \Phi_t' \right] \frac{\partial V_{t+1}}{\partial K}, \tag{2.2.15}$$

$$\frac{\partial V_t}{\partial C} = -\beta E_t \frac{\theta_t}{\pi_{t+1}} \frac{\partial V_{t+1}}{\partial b} + \beta E_t \theta_t \frac{\partial V_{t+1}}{\partial C}, \tag{2.2.16}$$

and

$$\frac{\partial V_t}{\partial U} = \theta_t + \beta E_t \theta_t \frac{\partial V_{t+1}}{\partial U}. \tag{2.2.17}$$

In what follows, these conditions are used to derive expressions for consumption, the share of goods subject to commitments, labor supply, and investment, respectively.

2.2.3 Consumption: First-Order Condition

In the absence of commitments (i.e., when there is no cost to adjust consumption, and thus $\theta_t = 0$ in all time periods) households equate the marginal utility of consumption and the marginal value of wealth:

$$u'(C_t) = \Lambda_t, \tag{2.2.18}$$

where $c_t = C_t$ because consumption is reset for all goods in each period, and Λ_t is used to denote the marginal value of wealth at time t. In the presence of consumption adjustment costs, households take into account the probability that consumption choices made today will remain in place in future periods, resulting in a more general expression than Equation

(2.2.18). Making use of Equations (2.2.16) and (2.2.17), the household's first-order condition for spending on goods whose consumption is reset in the current period (Equation 2.2.10) can be expressed as follows:

$$u'(c_t) = \frac{B_t}{A_t}, \qquad (2.2.19)$$

where

$$A_t = 1 + \beta E_t [\theta_{t+1} A_{t+1}], \qquad (2.2.20)$$

and

$$B_t = \Lambda_t + \beta E_t [\theta_{t+1} B_{t+1}]. \qquad (2.2.21)$$

Total consumption is a weighted sum of consumption for goods whose consumption is reset in the current period, and consumption commitments that are unchanged from the previous period (see Equation 2.2.8). Note that in the steady state, $\frac{B_t}{A_t} = \Lambda_t$ and $c_t = C_t$, in which case Equation (2.2.19) reduces to the standard condition given by Equation (2.2.18). Next, insight is developed by considering a special case of the model in which the share of goods subject to commitments each period is exogenous and constant; the general form of the model with an endogenous and time-varying share of goods subject to commitments is discussed afterward.

2.2.4 Special Case: Exogenous Commitments

2.2.4.1 Euler Equation Representation

When the adjustment cost function $D(\theta_t)$ permits consumption to be reset for a $(1 - \theta)$-share of goods each period at no cost, but requires an infinite cost to reset consumption for a larger set of goods, the share of goods whose consumption can be reset is effectively exogenous

and equal to $\theta < 1$ in each time period. Under this assumption, Equation (2.2.19) can be re-expressed as follows:

$$u'(c_t) = (1 - \beta\theta) E_t \sum_{k=0}^{\infty} (\beta\theta)^k \Lambda_{t+k}. \qquad (2.2.22)$$

Equation (2.2.22) indicates that households set the marginal utility of consumption (for goods whose consumption is reset in the current period) equal to a discounted weighted average of expected future marginal values of wealth. The weights reflect the probability that goods whose consumption is reset today remain subject to commitments in each future period: the higher is the probability that a commitment remains in place in future periods, the larger is the weight placed on distant-in-time expected values of the marginal utility of wealth. Because a pronounced consumption response to a transitory fluctuation in the marginal value of wealth creates consumption commitments that remain in place long after the fluctuation in Λ_t has passed, households optimally respond little to transitory variation in Λ_t when θ is large.

Further insight can be gained by representing consumption behavior in terms of an Euler equation. Continuing to assume that the share of goods subject to commitments each period is constant at $\theta < 1$, and assuming in addition that utility for each good has a constant elasticity of intertemporal substitution σ, the following log-linearized consumption Euler equation holds for aggregate consumption at time t:

$$E_t [\Delta log C_{t+1}] = \theta \Delta log C_t + \sigma (1 - \theta) (1 - \beta\theta) E_t \sum_{k=0}^{\infty} (\beta\theta)^k \left(\widetilde{R_{t+k}} - \widetilde{\pi_{t+k+1}} \right), \qquad (2.2.23)$$

where the term $\widetilde{R_{t+k}} - \widetilde{\pi_{t+k+1}}$ is the percentage deviation in the real interest rate from its steady-state level. (See Appendix 2.6.1 for a step-by-step derivation of this condition.) The standard frictionless consumption Euler equation is a special case: setting the share of goods

subject to commitments equal to zero ($\theta = 0$) yields

$$E_t\left[\Delta log C_{t+1}\right] = \sigma E_t\left[\widetilde{R_t} - \widetilde{\pi_{t+1}}\right]. \tag{2.2.24}$$

That is, when all consumption goods can be adjusted each period at no cost, expected consumption growth is proportional to the expected percentage deviation in the real interest from its steady-state level. But when a θ-share of goods are subject to ongoing commitments each period, expected consumption growth is a weighted sum of a backward looking component (lagged consumption growth) and a forward looking component (expected future deviations in the real interest rate from its steady-state level). The larger is the weight on the backward looking component—equal to the share goods subject to ongoing commitments— the greater is the degree of inertia in aggregate consumption growth. The smaller is the share of goods subject to commitments each period, the larger is the weight on near-term expected deviations in the real interest rate from its steady-state level. As in Equation (2.2.22), the weight placed on expected future realizations of the real interest rate reflect the probability that spending on goods whose consumption is reset today remain unchanged in each future time period. The parameter σ, the intertemporal elasticity of substitution, parameterizes the curvature of the utility function for each good, controlling the willingness of households to vary consumption in response to changes in its intertemporal price: *ceteris paribus*, the larger is σ, the greater is the response of consumption to an expected change in future real interest rates.

An important insight apparent from the Equation (2.2.23) is the dependence of consumption growth on the persistence of expected changes in future real interest rates: persistent changes in the real interest rate meaningfully affect the forward-looking term on the right-hand-side of Equation (2.2.23), but transitory changes do not. Households optimally respond elastically to persistent changes in the real interest rate, and largely ignore transitory changes; mirroring the earlier discussion, a pronounced increase in consumption in response to a short-lived reduction in the real interest rate is undesirable because it is likely to result in spending being 'stuck' at a sub-optimally high level long after the real interest rate has returned to

its normal level, and vice versa. This aspect of consumption behavior suggests that a high degree of inertia in central bank policy interest rate rules may be desirable: in the presence of commitments, aggregate consumption is only meaningfully affected by real interest rate changes that are expected to be long-lasting.

2.2.4.2 Relation to Habit-Formation Preferences

A widely used specification of consumer preferences assumes that utility derived from consumption is reference dependent: the reference level may be either aggregate consumption (external habit) or lagged own consumption (internal habit). Although there is only limited microeconomic evidence that consumer preferences are reference dependent, habit formation preferences are now widely used, especially in general equilibrium macroeconomic models, because they generate consumption behavior that matches several stylized features of aggregate consumption data: excess smoothness and excess sensitivity in response to economic news, and humped-shaped dynamics. Because reference dependence makes consumers unwilling to tolerate abrupt changes in consumption, habit-formation preferences have also become prominent in the asset pricing literature, to reconcile standard estimates of investor risk-aversion and a high equity-premium with an empirically observed low covariance between consumption growth and returns on financial assets.

Critically, the consumption Euler equation implied by the widely-used internal habit preference specification is a special case of the consumption commitments model outlined in this paper. The utility function for internal habit-formation preferences is given by

$$U_t = u\left(C_t - bC_{t-1}\right), \tag{2.2.25}$$

where b is the degree of degree of reference dependence on past own consumption. The larger is b, the greater is the weight placed on past consumption, and the less willing is the consumer to tolerate period-to-period changes in consumption. Utility maximization with this preferences specification (setting $\theta = 0$ and replacing Equation 2.2.7 with Equation

2.2.25), and some rearrangement, results in the following first-order condition for aggregate consumption:

$$u'\left(\hat{C}_t\right) = E_t \sum_{k=0}^{\infty} (\beta b)^k \Lambda_{t+k}, \tag{2.2.26}$$

where $\hat{C}_t \equiv C_t - bC_{t-1}$. Note that this condition closely resembles Equation (2.2.22), the first-order condition with consumption commitments, for consumption on goods whose spending is reset in the current period. As shown in Appendix 2.6.2, the first-order condition for internal habit-formation preferences (Equation 2.2.26) implies a consumption Euler equation identical (up to a first-order approximation) to that for the consumption commitments model, when the share of goods subject to commitments is exogenously equal to $\theta = b$ (see Equation 2.2.23). Accordingly, consumption commitments—with an exogenous and random subset of goods subject to commitments each period—provides a microfoundation for internal habit-formation preferences.

2.2.5 General Case: Endogenous Commitments

Having shown in the previous section that consumption behavior implied by internal habit-formation preferences is matched by a special case of the commitments model, this section returns to the general form of the model in which the share of goods subject to commitments is endogenously chosen by the household each period. This is the key difference between the two models: the degree of reference dependence with habit-formation preferences is a structural (and time-invariant) parameter, whereas the share of goods subject to commitments each period is constrained only by consumption adjustment costs. The ability to vary the share of goods subject to commitments allows consumption to respond asymmetrically to large and small shocks: in the presence of large shocks, it is optimal to incur adjustment costs to reset spending for a high share of goods, and consumption behavior approaches that of the frictionless model. This intuition is formalized in the next two subsections: the implications for consumption dynamics of a household having the ability to vary adjustment of prior

commitments is discussed first, and then the consumer's optimal determination of the share of goods subject to commitments.

2.2.5.1 Consumption Response with Endogenous Adjustment of Commitments

There are two channels via which endogenous adjustment of commitments can magnify the responsiveness of consumption to near-term expected changes in the marginal value of wealth. First, by incurring adjustment cost in the *current* period, the household can increase the set of good whose spending can be reset; recall that the law-of-motion for total consumption is $C_t = \theta_t C_{t-1} + (1 - \theta_t) c_t$, and that the smaller is θ_t the narrower is the set of goods subject to commitments this period. Second, the ability in *future* periods to reset commitments incurred by current spending choices can increase the responsiveness of current consumption to near-term changes in $\tilde{\Lambda}_t$. To illustrate this second channel, it is helpful to take a second-order Taylor series approximation to the right-hand-side of Equation (2.2.19), about the steady-state level for consumption and the steady-state share of goods subject to commitments:

$$u'(c_t) \simeq \Lambda + \Lambda (1 - \beta\theta) \left(1 - \tilde{A}_t\right) \left[\tilde{\Lambda}_t + E_t \sum_{k=1}^{\infty} (\beta\theta)^k \tilde{\Lambda}_{t+k} \left(\prod_{i=1}^{k} \left(1 + \tilde{\theta}_{t+i}\right)\right)\right], \qquad (2.2.27)$$

where

$$\tilde{A}_t \simeq E_t \left[(\beta\theta) \tilde{\theta}_{t+1} + \sum_{k=2}^{\infty} (\beta\theta)^k \tilde{\theta}_{t+k} \left(\prod_{i=1}^{k-1} \left(1 + \tilde{\theta}_{t+i}\right)\right)\right], \qquad (2.2.28)$$

$\widetilde{X_t} = \frac{X_t - X}{X}$ is the percentage deviation of variable X from its steady-state level, and the steady-state condition implies $u'(c) = \Lambda$ (see Appendix 2.6.3 for a derivation of this approximation). Recall that the adjustment cost function permits the household to reset consumption for a $(1 - \bar{\theta})$-share of goods at no cost; whenever adjustment cost is incurred, consumption is reset for a larger share of goods, implying $\tilde{\theta}_{t+k} \leq 0$, for all k: the larger is the set of goods that the consumer expects to reset spending for in period $t + k$, the more negative is $\tilde{\theta}_{t+k}$. If the household expects to reset consumption for a large share of goods in

period $t + k$, then $\tilde{\theta}_{t+k} < 0$ and consumption responds elastically relative to the exogenous commitments benchmark. This can be seen formally by inspection of Equations (2.2.27) and (2.2.28): the more negative is $\tilde{\theta}_{t+k}$, the smaller is the weight on $\left\{\tilde{\Lambda}_{t+i}\right\}_{i=k}^{\infty}$, but the larger is $\left(1 - \tilde{A}_t\right)$, increasing the weight placed on $\left\{\tilde{\Lambda}_{t+i}\right\}_{i=0}^{k-1}$. Intuitively, when the household expects to reset commitments for a broad set of goods in period $t + k$, the constraint imposed on future choices by current spending decisions is small, and it is optimal for consumption to respond elastically to near-term changes in the marginal value of wealth.

Having shown that the consumption response to an economic shock is more elastic the smaller, now and in future periods, is the share of goods subject to commitments, the following discussion considers the household's optimal determination of θ_t.

2.2.5.2 Determination of the Share of Goods Subject to Commitments

The share of commitments maintained each period is chosen by households to balance adjustment costs against the benefit of a more equal distribution of spending, and therefore marginal utility, across goods. Following essentially the same steps used in Section 2.2.3, the household's first-order condition for θ_t (Equation 2.2.12) can be expressed as follows:

$$\left[(C_{t-1} - c_t) - \frac{U_{t-1} - u(c_t)}{u'(c_t)}\right] = -D'(\theta_t)\left[\frac{\Lambda_t}{B_t}\right]. \tag{2.2.29}$$

The left-hand-side of Equation (2.2.29) measures the reduction in the consumption distortion across goods from a marginal decrease in the share of commitments held over from the previous period. This term does not depend on θ_t: the randomization assumption means that the distribution of consumption among goods whose consumption is reset is independent of θ_t. Balancing this benefit, $-D'(\theta_t)$ measures the marginal increase in adjustment costs from a marginal decrease in the share of commitments maintained from the previous period. The term $D'(\theta_t)$ is scaled by $\frac{\Lambda_t}{B_t} \leq 1$, reflecting the fact that incurring cost to adjust consumption for an increased share of goods today may accrue benefit in future periods.

The utility loss arising from an uneven distribution in consumption across goods is second-

order, because commitments affect the composition of a household's consumption basket, but not the lifetime level of consumption. However, the adjustment cost incurred to adjust consumption commitments consumes resources, and has a first-order cost to households. This means that even small adjustment costs can make households unwilling to adjust commitments.[2]

To gain intuition, it is helpful to re-express the left-hand-side of Equation (2.2.29) as follows:

$$\left[(C_{t-1} - c_t) - \frac{U_{t-1} - u(c_t)}{u'(c_t)}\right] = \left[\underbrace{(C_{t-1} - c_t) - \left(\frac{u(C_{t-1}) - u(c_t)}{u'(c_t)}\right)}_{x} + \underbrace{\left(\frac{u(C_{t-1}) - U_{t-1}}{u'(c_t)}\right)}_{y}\right] \qquad (2.2.30)$$

$$\geq \left[\underbrace{(C_{t-1} - c_t) - \left(\frac{u(C_{t-1}) - u(c_t)}{u'(c_t)}\right)}_{x}\right].$$

Term y is the dollar cost of the consumption distortion carried over from the previous period. The inequality follows from the fact that term y is positive: $u(C_{t-1}) \geq U_{t-1}$, because utility for any given level of spending is maximized when consumption is spread evenly across all goods. Term x measures the additional utility loss that arises whenever today's chosen level of consumption, c_t, is unequal to last period's total spending, C_{t-1}. Figure 2.1 graphs x over a range of values for c_t.[3] For c_t close to C_{t-1} the distortion caused by consumption commitments is negligible, but rises at an increasing rate in the gap between c_t and C_{t-1}, shown by the curvature of graph x. The more concave is the utility function for each good,

---

[2]On a technical level, the second-order nature of the utility cost arising from consumption commitments means that a linear approximation to the model's first-order-conditions around the steady-state level of consumption is not sufficient to describe a household's choice of θ_t. When solving the model quantitatively in Section 2.4, a second-order approximation to the model's first-order conditions is used. For the special case where the share of commitments adjusted each period is exogenously fixed at some $\theta_t = \theta < 1$, a linear-approximation is sufficient.

[3]To see that term x is convex, note that $\frac{\partial x}{\partial c_t} = -\frac{u''}{u'^2}[u(c_t) - u(C_{t-1})]$. Hence, $\frac{\partial x}{\partial c_t} > 1$ for $c_t > C_{t-1}$, $\frac{\partial x}{\partial c_t} < 1$ for $c_t < C_{t-1}$, and $\frac{\partial x}{\partial c_t} = 0$ when $c_t = C_{t-1}$.

the greater is the distortion caused for any $c_t \neq C_{t-1}$.

Further intuition can be gained by taking a second-order Taylor series approximation to the left-hand-side of Equation (2.2.29), around the steady-state level of consumption $c_t = \bar{C}$. Doing so yields the following approximate representation of Equation (2.2.29):

$$\underbrace{\frac{1}{2}\left[(\tilde{c}_t)^2 + \sum_{j=1}^{\infty} \omega_j \, (\tilde{c}_{t-j})^2\right]}_{a} - \underbrace{(\tilde{c}_t)\left(\tilde{C}_{t-1}\right)}_{b} \simeq \sigma \underbrace{\left[\frac{\Lambda_t}{B_t}\right]}_{c}\underbrace{\left[\frac{-D'(\theta_t)}{C}\right]}_{d}, \qquad (2.2.31)$$

where $\widetilde{X}_t \simeq \frac{X_t - X}{X}$ is the percentage deviation in variable X_t from its steady-state level X, $\omega_j = \theta_{t-1}\theta_{t-2}\cdots\theta_{t-j+1}\left(1 - \theta_{t-j}\right)$ are the weights on prior commitments, and σ is the intertemporal elasticity of substitution, assuming isoelastic preferences. Term a is proportional to the variance of consumption across goods, relative to a mean level of consumption equal to the steady-state level $c_t = C$. Distant-in-time consumption choices receive less weight than recent choices because they are a smaller share of total consumption. In general, steady-state consumption is not the relevant benchmark from which to measure the dispersion in consumption; term b adjusts the cost measured by term a according to the deviation between desired consumption today and the average level of spending carried over from the previous period. Deviations in past consumption choices from the steady-state are less costly the closer is their mean value to today's desired frictionless level of consumption. For the special case in which consumption today is reset to its steady-state level ($\tilde{c}_t = 0$) term b is equal to zero and the variance in past consumption choices, given by the summation in term a, measures the full cost to households of the dispersion in consumption across goods. The smaller is the intertemporal elasticity of substitution σ, the greater is the cost of any given dispersion in spending across goods, because the degree of curvature in the utility function is higher, and the greater is the adjustment cost that the household is optimally willing to pay to reset spending. Terms c and d, on the right-hand side of Equation (2.2.31), are the same as in Equation (2.2.29), with the exception that term d is now scaled by steady-state consumption.

The following two sub-sections describe the remaining choice variables for the household: labor supply and investment in physical capital. Because these two aspects of the model are standard, these sections are brief.

2.2.6 Labor Supply: First-Order Condition

Frictions affecting household labor supply are undoubtedly important, but their incorporation into a DSGE model is beyond the scope of this paper. Accordingly, the frictionless first-order condition for labor supply (Equation 2.2.11) is standard, implying that at an optimum:

$$\frac{v'(N_t)}{w_t} = \Lambda_t, \tag{2.2.32}$$

where Λ_t is the marginal value of wealth. Equations (2.2.11) and (2.2.14) together imply a standard Euler equation for household labor supply:

$$\frac{v'(N_t)}{w_t} = \beta E_t \left[\frac{R_t}{\pi_{t+1}} \frac{v'(N_{t+1})}{w_{t+1}} \right]. \tag{2.2.33}$$

2.2.7 Investment: First-Order Condition

Letting $q_t \equiv \beta E_t \frac{\partial V_{t+1}}{\partial K}$ denote the discounted shadow value of a unit of installed capital next period, Equations (2.2.13) and (2.2.15) can be used to define q_t recursively:

$$q_t = \beta E_t \left[q_{t+1} \left(R_{t+1}^k \Phi'_{t+1} + (1 - \delta) + \Phi_{t+1} - \frac{I_{t+1}}{K_{t+1}} \Phi'_{t+1} \right) \right]. \tag{2.2.34}$$

The market value of installed capital is increasing in the rate of return on capital, and the degree of convexity in the capital adjustment cost function. It is decreasing in the rate of depreciation, δ. The first-order condition for investment (Equation 2.2.13) is standard:

$$q_t \Phi'_t = \Lambda_t. \tag{2.2.35}$$

2.3 Applications

2.3.1 Excess Smoothness and Excess Sensitivity

The commitments model generates aggregate consumption dynamics that are consistent with two important deviations from the permanent income hypothesis: excess smoothness and excess sensitivity. Consumption is said to be excessively smooth if it responds less than one-for-one to the arrival of news about permanent income (Deaton, 1987), and to be excessively sensitive if it responds to lagged information (Flavin, 1981). Switching costs make it optimal for households to incorporate news about permanent income into their spending over time, rather than immediately, resulting in excess smoothness, while such delayed response makes spending responsive to dated information, and thus exhibit excess sensitivity.

The ability of the consumption commitments model to match these stylized features of aggregate consumption data can be most easily shown analytically for the special case of the model in which the share of goods subject to switching costs is exogenously equal to θ in each period. Suppose the representative household in an economy that has been in steady-state receives favorable news about permanent income at time t, and no further news thereafter. The marginal utility of wealth for the representative consumer in this economy is given by

$$\Lambda_{t+k} = \begin{cases} \Lambda' & for\ k \geq 0; \\ \Lambda & for\ k < 0, \end{cases} \tag{2.3.1}$$

where Λ is the steady-state level of marginal utility before the shock, and $\Lambda' < \Lambda$. In each time period $k > 0$, consumption is reset to an increased level for goods whose spending is not subject to commitments. The first-order condition for consumption on goods whose spending is reset (Equation 2.2.22) implies the following pattern of past and future spending choices:

$$c_{t+k} = \begin{cases} c' & for\ k \geq 0; \\ c & for\ k < 0, \end{cases} \tag{2.3.2}$$

where c is the steady-state level of consumption before the shock and $c' > c$. Consumption responds immediately for goods not subject to commitments but, in each period, consumption for a θ-share of goods is not adjusted, and aggregate consumption incorporates the date t news shock over time, rather than immediately; using the law of motion for aggregate consumption (Equation 2.2.6), total spending in period $k > 0$ after the shock occurs is given by:

$$C_{t+k} = \theta^{k+1} C + c' (1 - \theta) \sum_{j=0}^{k} \theta^j. \tag{2.3.3}$$

Taking a log-linear approximation to this condition around the pre-shock steady-state level of consumption, it can be shown that the change in aggregate consumption between any two periods $k_2 > k_1 > 0$ is approximately equal to

$$logC_{t+k_2} - logC_{t+k_1} \simeq \left[(1 - \theta) \sum_{j=k_1}^{k_2} \theta^j \right] (logc' - logc) \tag{2.3.4}$$

$$\equiv \beta (k_1, k_2, \theta) (logc' - logc). \tag{2.3.5}$$

The term $(logc' - logc)$ is equal to the approximate percentage change in permanent income discovered at time t. In the absence of commitments, $\theta = 0$ and $\beta (0, k_2, \theta) = 1$ in Equation (2.3.4), in which case aggregate consumption responds one-for-one to the news shock at time t. But, in the presence of commitments, $\beta (0, k_2, \theta) < 1$, generating delayed and partial adjustment at time t. Equation (2.3.4) maps directly into a standard empirical test for excess smoothness. Consumption is excessively smooth if the coefficient $\beta_0 < 1$ for any $k_2 > 0$ in the following empirical regression:

$$logC_{t+k_2} - logC_t = \alpha_0 + \beta_0 (logA_{t+k_2} - logA_t) + \varepsilon, \tag{2.3.6}$$

where $logA_{t+k_2} - logA_t$ is the change in permanent income between time t and $t + k$. For an economy subject to the shock studied here (see Equation 2.3.1), $logA_{t+k_2} - logA_t = logc' - logc$. The regression coefficients in Equation (2.3.6) correspond exactly to parameters of the

commitments model: $\alpha_0 = 0$ and $\beta_0 = \beta(0, k_2, \theta) < 1$. The larger is θ the smaller is the value of $\beta(0, k_2, \theta)$, and the greater is the degree of excess sensitivity. Because $\lim_{k_2 \to \infty} \beta(0, k_2, \theta) = 1$ aggregate consumption does eventually fully incorporate the news received at time t.

Similarly, there is evidence of excess sensitivity if $\beta_1 > 0$ in the empirical regression:

$$logC_{t+k_2} - logC_{t+k_1} = \alpha_1 + \beta_1 (logA_s - logA_{s-\Delta}) + \varepsilon, \tag{2.3.7}$$

for any $k_2 > k_1$, $s \leq k_1$ and $\Delta > 0$. For the shock studied here, $logA_s - logA_{s-\Delta} = logc' - logc$, for any $s \geq t + \Delta$; hence, $\beta_1 = \beta(k_1, k_2, \theta) > 0$ in the presence of commitments and there is excess sensitivity. Aggregate consumption exhibits excess sensitivity because adjustment costs limit the ability of households to immediately revise their spending upon receipt of news about permanent income; they reset consumption in response to a news shock over a number of periods, generating a correlation between consumption growth and lagged information.

2.3.2 Asset Pricing Implications

The equity-premium has been an enduring puzzle: reconciling the historical return on stocks over bonds (of about 6 percent per annum) with the empirically observed low covariance between consumption growth and stock returns (of about 0.2 percent per annum) requires investors to have a degree of risk-aversion about an order of magnitude larger than standard estimates in other settings (see Kocherlakota, 1996, for a survey of the literature). This section shows that consumption commitments provide a potential reconciliation: commitments prevent aggregate consumption adjusting fully and immediately in response to asset returns, masking the extent of consumption risk borne by households via their equity holdings. The remainder of this section formalizes this intuition.

First, note that the household's first-order and envelope conditions can be combined to give an expression for the expected return premium for stocks over bonds:

$$E_t \left[(R^s_{t+1} - R_t) \frac{\Lambda_{t+1}}{\Lambda_t} \right] = 0, \tag{2.3.8}$$

where R^s_{t+1} is the return on holding stocks (realized at time $t+1$), R_t is the return on bonds (with the time $t+1$ payoff known at time t), and Λ_t the marginal value of wealth. The term Λ_{t+1}/Λ_t is the stochastic discount factor used to price assets in this representative agent economy. In the standard model, without consumption commitments, households set the marginal utility of consumption equal to the shadow value of wealth in each period: $u'(C_t) = \Lambda_t$. Using this condition, and supposing households have isoelastic utility with constant relative risk-aversion γ, Equation (2.3.8) can be approximated by the following expression:

$$E_t\left[R^s_{t+1}\right] - R_t \quad \simeq \quad -cov_t\left(R^s_{t+1}, \triangle log\Lambda_{t+1}\right) \tag{2.3.9}$$

$$\simeq \quad \gamma cov_t\left(R^s_{t+1}, \triangle log C_{t+1}\right). \tag{2.3.10}$$

Using the historical moments listed above, a coefficient of relative risk aversion of about 30 is implied by Equation (2.3.10)—an order of magnitude larger than standard estimates.

The implications of consumption commitments for asset pricing can be most easily seen for the special case in which the share of goods subject to commitments each period is exogenous: $\theta_t = \theta$. Spending on goods whose consumption is reset today is chosen such that $u'(c_t) = (1 - \beta\theta) E_t \sum_{k=0}^{\infty} (\beta\theta)^k \Lambda_{t+k}$ (see Equation 2.2.22), implying that variation in the current period's marginal value of wealth has a reduced effect on consumption compared to the frictionless model in which $u'(C_t) = \Lambda_t$. In addition, a θ-share of goods have unchanged consumption in the commitments model each period, further dampening the movement of aggregate consumption in response to changes in the marginal value of wealth. In the presence of commitments, the expected equity-premium is:

$$E_t\left[R^s_{t+1}\right] - R_t \simeq \frac{\gamma}{(1 - \beta\theta)(1 - \theta)} cov_t\left(R^s_{t+1}, \triangle log C_{t+1}\right). \tag{2.3.11}$$

See Appendix 2.6.4 for a derivation of this condition. Thus, the presence of consumption commitments reduces the degree of investor risk-aversion required to reconcile the historical

equity-premium with the observed covariance between stock returns and consumption growth by a factor of $(1 - \beta\theta)(1 - \theta) \simeq (1 - \theta)^2$. With a value of $\theta = 0.7$, roughly the share of consumption that Chetty and Szeidl (2007) estimate is unadjusted on an annual basis, there is no apparent equity-premium puzzle. This reconciliation is similar to that proposed by Grossman and Laroque (1990) and Gabaix and Laibson (2001): in each case, there is at least some consumer good (or set of goods) that is not adjusted contemporaneously with news on asset returns.

Because the restriction on consumption imposed by adjustment costs diminishes with time, the commitments model implies that the magnitude of the apparent equity-premium puzzle should decrease when the consumption response to stock market returns is measured over longer time periods. Indeed, there is some evidence that this is the case: Parker (2003) finds that risk borne by households in holding the market portfolio is substantially larger when measured by changes in consumption at horizons of up to about two years, particularly for stockholder households.

The commitments model provides an explanation for the observed high equity-premium without implying a counterfactually high risk-free rate-of-return. Without consumption adjustment costs, a high degree of curvature in consumer utility is required to generate a large equity-premium, but this implies that consumers are also unwilling to experience variation in consumption through time, requiring a high risk-free rate to explain observed mean growth in per capita consumption. This conflict does not arise with the commitments model: utility is assumed to have only a moderate degree of curvature and adjustment costs do not restrict predictable growth in consumption over time.

Interestingly, the commitments model implies that large movements in financial asset returns may be the most informative observations for measuring the extent of consumption risk borne by households via their stock holdings. Large shocks make it optimal for households to adjust spending for a large share of goods, concentrating the necessary adjustment in consumption over a short period of time.

It should be noted that—as is the case with habit-formation preferences—incorporating con-

sumption commitments into an otherwise standard DSGE model is not sufficient to generate the empirically observed level of the equity-premium. The magnitude of consumption risk borne by households from holding capital assets remains low because households have two means through which they can limit variation in the marginal value wealth: i) they can vary labor supply to offset a change in the value of asset holdings; and ii) they can vary saving without bearing large adverse capital gains or losses on holding capital. The marginal product of capital varies little in standard DSGE models, so any adverse capital gains must come from inelasticity in the supply of capital. Microfounding any inelasticity in labor supply and investment supply that acts to raise the equilibrium return premium on stocks is beyond the scope of this paper (see Boldrin et al., 2001, for a two-sector model with habit persistence preferences that matches the empirically observed equity premium).

2.4   Quantitative Model

This section solves the model quantitatively, in order to illustrate the effect of commitments on aggregate consumption dynamics in the presence of a series of monetary policy shocks. This policy exercise demonstrates the ability of the commitments model to generate hump-shaped aggregate consumption dynamics in response to economic news, and shows the potential of the model to reconcile a moderate structural elasticity of intertemporal substitution with an empirically observed low correlation between consumption growth and changes in the real interest rate. Before discussing the results of this simulation, the following sub-sections sketch the remaining elements of the model, and discuss its calibration.

2.4.1   Description of Model

The consumption commitments model outlined in this paper can be easily incorporated into an otherwise standard DSGE model: the assumption that a random set of goods are subject to commitments each period eliminates history dependence that would otherwise complicate solving a general equilibrium model. The model is standard, with one exception: to allow for

nominal rigidity—as is now standard in business-cycle analysis—sticky prices are introduced at the wholesale level, rather than at the retail level. If prices were instead assumed to be sticky at the retail level, then the price paid by households for each good would be different, and consumption of *each* good would be a state variable for the household problem, not just total consumption; the model would be substantially more complicated to solve, for little or no gain in performance.

Each good consumed by households is assembled by a final goods firm (using no labor or capital) from inputs made by intermediate goods firms. Final goods firms are perfectly competitive, but intermediate goods firms are monopolistically competitive. Intermediate goods firms produce inputs using a constant returns to scale production function, employing labor and renting capital from households. These imperfectly competitive intermediate goods firms choose their sale price to maximize profits, subject to a friction that prevents them freely choosing their nominal sale price each period. As is standard in the New-Keynesian DSGE literature, the pricing restriction is assumed to be governed by a Calvo device: each intermediate goods firm can reset its price with state-independent probability $(1 - \gamma)$ each period. The nominal interest rate is determined by a Taylor (1993) monetary policy rule.

The technology used to assemble each final good is assumed to have the same substitutability among inputs. All consumer goods represent the same collection of intermediate inputs that have been costlessly differentiated by, for example, painting the good a different color. This assumption that final goods firms use a common technology implies a common price for each consumer good in every period. Details of the model can be found in Appendix 2.7.

2.4.2 Functional Forms and Calibration

The share of goods whose spending can be reset at no cost each period is chosen to be $(1 - \bar{\theta}) = 0.35$, roughly equal to the expenditure share of consumer goods Chetty and Szeidl (2007) estimate to have no or negligible switching costs. This parameter choice coincides with the internal habit-formation reference weight estimated by Christiano et al. (2005) in their

influential DSGE model of the macroeconomic effects of a monetary policy shock; because the habit-formation model is equivalent to the commitments model when the share of goods subject to switching costs is exogenously fixed, comparing consumption dynamics of the commitments model to the benchmark habit-formation model simply requires examining the special case in which $\theta_t = \bar{\theta} = b$. The adjustment cost function for commitments is given by $D(\theta) = \psi \left[\left(\bar{\theta} - \theta \right)^2 / \left(\theta \bar{\theta} \right) \right]$, which satisfies each of the desired properties listed in Section 2.2.1. The parameter ψ is calibrated so that there is sufficient variation in θ_t to highlight the endogenous response of commitments to shock magnitude; in this dimension, the simulation results are exploratory.

The remaining model parameters and functional forms are essentially standard. Isoelastic functional forms are assumed for labor supply and utility for each consumption good: $u\left(C_t\left(j\right)\right) = \frac{\left(C_t(j)\right)^{1-\frac{1}{\sigma}}}{1-\frac{1}{\sigma}}$ and $v\left(N_t\right) = \phi \frac{N_t^{1+\frac{1}{\eta}}}{1+\frac{1}{\eta}}$, where ϕ parameterizes the disutility of labor supply. The elasticity of intertemporal substitution is chosen to be $\sigma = 0.5$, and the Frisch elasticity of labor supply is set to $\eta = 1$; these are typical parameter values used in business-cycle analysis. The quadratic form $\Phi\left(x\right) = x - \frac{\xi}{2}\left(x - \delta\right)^2$ is chosen for the capital adjustment cost function, in which case $\xi \simeq \frac{d \ln q_t}{d\left(\frac{I}{K} - \delta\right)}$ is the semi-elasticity of the shadow value of capital with respect to deviations in the investment-capital ratio from its steady-state level. Each period in the model corresponds to a quarter; the depreciation rate for capital is set at $\delta = 0.025$, and households' subjective rate of time preference is $\beta = 0.99$. The elasticity of capital to output is $\alpha = \frac{1}{3}$, consistent with empirically observed capital and labor income shares, and the price elasticity of demand for intermediate inputs ϵ is chosen such that intermediate goods firms set a steady-state mark-up over marginal cost of 20 percent. Intermediate goods firms are able to reset their price each period with state-independent probability $(1 - \gamma) = \frac{1}{4}$, resulting in a one year average price duration. The Taylor rule coefficient on inflation is 1.5, a value roughly in line with the coefficient for estimated Taylor rules. The model's calibration is summarized in Table 2.1.

2.4.3 Response to Monetary Policy

The effect of commitments on aggregate consumption dynamics is demonstrated by subjecting the model to a series of exogenous monetary policy shocks. Specifically, the model's Taylor rule is subjected to a shock u_t that decays according to a first-order autoregressive process with autocorrelation coefficient ρ_R:

$$\mu_t = \rho_R \mu_{t-1} + u_t, \tag{2.4.1}$$

with $u_{t+k} = 0$ for $k > 0$. The cumulative size of the shock depends on the autoregressive parameter ρ_R, and is equal to $\sum_{k=0}^{\infty} \rho_R^k u_t = \frac{u_t}{1-\rho}$. Three different shock scenarios are presented to illustrate the role of commitments:

1. A persistent shock with small magnitude: $u_t = 0.01$ and $\rho_R = 0.75$, for a cumulative size 0.04;

2. A persistent shock with large magnitude: $u_t = 0.03$ and $\rho_R = 0.75$, for a cumulative size of 0.12;

3. A transitory shock with large magnitude: $u_t = 0.03$ and $\rho_R = 0.25$, for a cumulative size of 0.04.

Impulse response functions for the model's endogenous variables are presented in response to each of these shock scenarios in Figures 2.2 to 2.4, respectively. In each figure, the model with endogenous consumption commitments corresponds to the set of impulse response functions shown by a blue dotted line; the commitments model with an exogenous and constant share of goods subject to switching costs each period, which is equivalent (to a first-order approximation) to internal habit-formation preferences (with $b = \theta$), is shown by a green line; and the standard model, which is a special case of the commitments model with $\theta_t = 0$, corresponds to the set of impulse response functions shown by a red line.

Figure 2.2 illustrates behavior when the economy is subjected to a persistent, but small in magnitude, expansionary monetary policy shock (simulation number one). Relative to the

standard model (the red set of impulse responses), in the presence of commitments (the blue and green set of impulse responses), households adjust consumption gradually, rather than immediately, generating hump-shaped aggregate consumption dynamics; because the shock magnitude is small, there is negligible difference in consumption behavior between the model with an exogenous (and fixed) share of goods subject to commitments each period (green line) and the model with endogenous (and time-varying) consumption commitments (blue line). In contrast, the simulation shown in Figure 2.3 indicates that, when faced with a large and persistent shock (simulation number two), households abandon a sizable share of commitments: they do this because, in the presence of a large shock, the consumption distortion caused by changing spending for only a small group of goods is large relative to the marginal cost of adjusting commitments. Critically, the consumption response is more elastic when households can reduce the share of goods subject to commitments (compare the blue relative to the green impulse response functions for consumption in Figure 2.3). Thus, the endogenous adjustment mechanism—that is by assumption absent from habit-formation preferences—results in a non-linear response of consumption to shock size.

Because consumption commitments affect the marginal value of wealth, commitments affect labor supply, even though consumption and labor supply are additively separable in households' utility function. Because there are no labor supply frictions in this model, but commitments limit the responsiveness of consumption to economic shocks, saving is a potentially important adjustment mechanism for households: saving is temporarily high whenever earnings respond more elastically than consumption to a favorable economic shock.

For the set of simulated impulse response paths associated with the standard model and the endogenous commitments model, shown in Figures 2.2 to 2.4, estimates of the elasticity of intertemporal substitution are reported in Table 2.2. The estimate for the standard model (i.e., in the absence of consumption adjustment costs) is equal to the structural parameter, $\sigma = 0.5$, up to numerical approximation error. In the presence of commitments, consumption responds gradually to the interest rate shock, continuing to increase after the real interest rate has begun to decline (but is above its normal level), resulting in a low or slightly negative

estimate for the elasticity of intertemporal substitution. The commitments model has the same *structural* elasticity of intertemporal substitution as the standard frictionless model (utility for each good has the same curvature), but commitments break the relationship between *changes* in consumption and the real interest rate implied by the frictionless Euler equation.

This is interesting in light of disagreement in the empirical literature about the magnitude of the elasticity of intertemporal substitution. The seminal work of Hall (1988) estimates an elasticity of intertemporal substitution of close to zero, based on time series data, but cross-sectional evidence (e.g., Gruber 2006), and complementary estimates implied by survey-based attitudes toward risk (e.g., Kimball et al. 2008), typically find much larger estimates. Consumption adjustment costs provide a potential reconciliation: most time-series estimates are based on transitory variation in real interest rates that households may optimally respond little to, because doing so incurs consumption adjustment costs; consumption commitments do not affect a household's willingness to respond to long-lasting variation in real interest rates, potentially explaining why Gruber (2006)—who uses variation in real interest rates across households due to differences in marginal tax rates on interest income—estimates a large value for the elasticity of intertemporal substitution.

2.5   Conclusion

This paper has proposed a model that microfounds inertial aggregate consumption dynamics, based on the assumption that spending on some goods is subject to commitments that are costly to abandon. Household level consumption data, documented elsewhere, indicates that a sizable share of the typical household's consumption basket is subject to such costs, with spending on shelter and autos being the most important sources of commitments for many households.

Commitments magnify the effect on utility of a given change in a household's total spending, because consumption can only be reset for the subset of goods not subject to commitments,

each of which has diminishing marginal utility of consumption. Accordingly, news about permanent income is incorporated into total household spending over time, rather than immediately, making consumption excessively smooth, and responsive to lagged information (excess sensitivity). In response to a temporary change in the real interest rate, consumption displays hump-shaped dynamics; the observed elasticity of intertemporal substitution appears low, despite the model incorporating a moderate structural elasticity, because consumption adjustment costs alter the timing of changes in consumption growth relative to the real interest rate. The model also has implications for the equity-premium puzzle: commitments restrict the ability of households to adjust consumption immediately in response to changes in asset values, masking the extent of consumption risk borne by households via their equity holdings.

The model permits a highly tractable representative agent representation, and can be easily incorporated into general equilibrium macroeconomic models. The widely used (internal) habit-formation preference specification is a special case of the model, and thus the model provides a foundation for reference dependent consumption behavior based on consumption commitments. In contrast to habit-formation preferences, the reference point—due to consumption adjustment costs, not a behavioral assumption—is endogenously determined in the general form of the model. When households face large shocks, they optimally choose to incur adjustment costs necessary to adjust spending for a large share of goods, but in response to small shocks leave most commitments in place: consumption responds non-linearly to the magnitude of economic shocks, approaching the frictionless model for sufficiently large shocks. The model's tractability makes it straightforward to embed in more sophisticated DSGE models.

Table 2.1: Parameter Values: Quarterly

Rate of time preference	$\beta = 0.99$
Elasticity of intertemporal substitution	$\sigma = 0.5$
Labor supply elasticity (hours worked)	$\eta = 1$
Elasticity of Tobin's-Q with respect to investment-capital ratio	$\frac{1}{\delta}$
Capital share of output	$\alpha = \frac{1}{3}$
Commitments adjustment cost parameter	$\psi = 5 \times 10^{-3}$
Steady-state share of goods subject to commitments	$\bar{\theta} = 0.65$
Capital depreciation rate	$\delta = 0.025$
Price reset probability	$(1 - \gamma) = \frac{1}{4}$
Taylor rule coefficient on inflation	$\phi_\pi = 1.5$
Intermediate goods firms' demand elasticity	$\epsilon = 6$

Table 2.2: Elasticity of Intertemporal Substitution

Shock	Model	$\hat{\sigma}$
$u_t = 0.01$ and $\rho = 0.75$	Standard	0.46
	Commitments	-0.07
$u_t = 0.03$ and $\rho = 0.75$	Standard	0.44
	Commitments	-0.01
$u_t = 0.03$ and $\rho = 0.25$	Standard	0.49
	Commitments	0.01

Notes: This table reports estimates of the elasticity of intertemporal substitution, calculated using the simulated impulse response functions for the policy exercises shown in Figures 2.2 to 2.4, respectively.

Figure 2.1: Marginal Utility Cost of Consumption Commitments
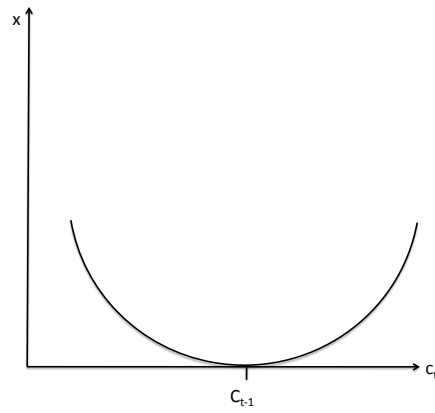
Notes: This figure shows the utility cost, beginning in the steady-state, when spending for goods whose consumption is reset in the current period (c_t) is not equal to spending on goods whose consumption is unchanged from the previous period, due to consumption commitments.

Figure 2.2: Impulse Response Functions to a Monetary Policy Shock: Persistent Shock with Small Magnitude
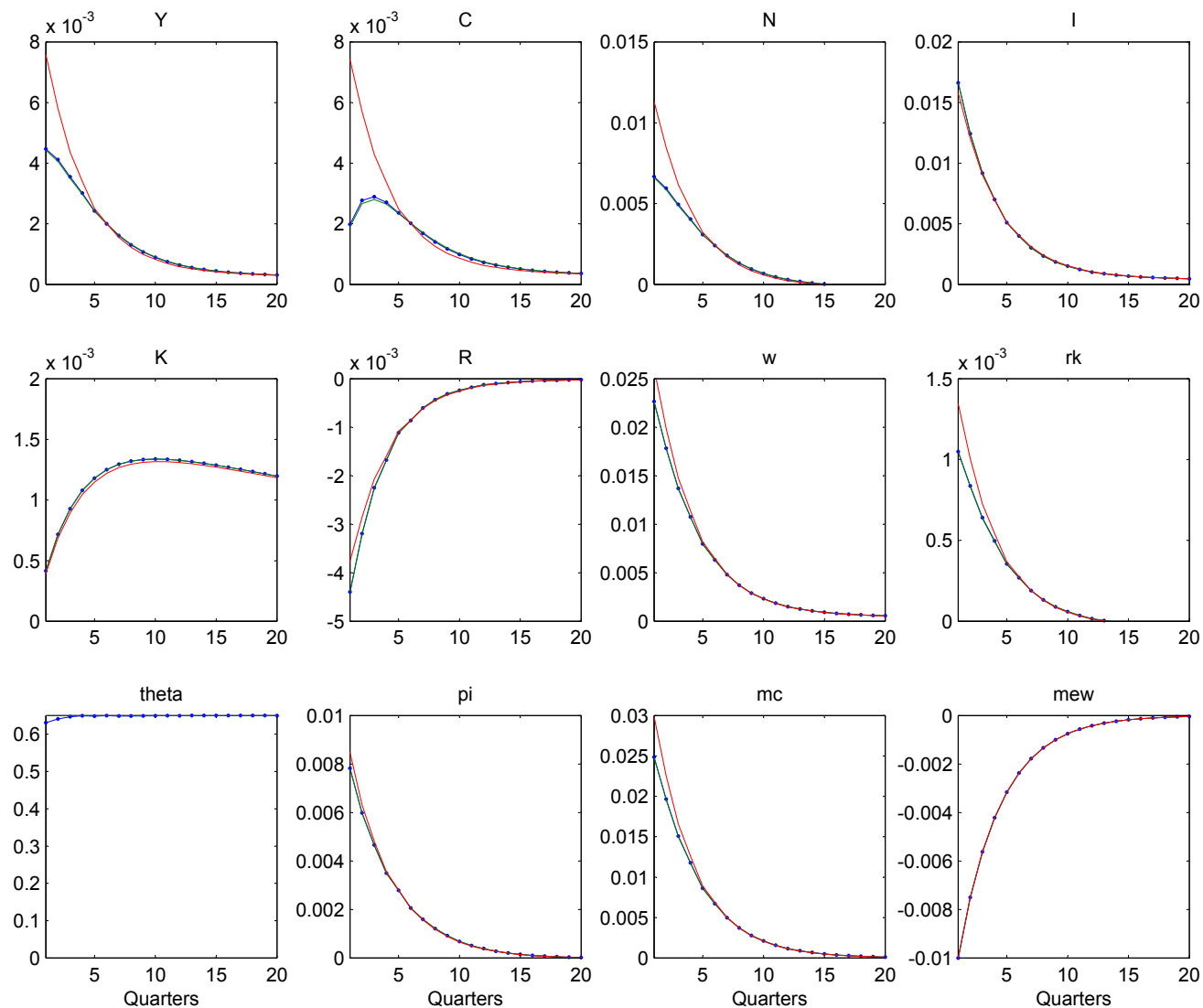
Notes: This figure shows impulse response functions when the model's Taylor rule is subjected to a 100 basis point expansionary interest rate shock, that decays according to an autoregressive process with an AR(1) parameter $\rho_R = 0.75$. The red set of impulse responses is for the frictionless benchmark model, the blue dotted set for the commitments model, and the green set for the commitments model when the share of goods subject to commitments is exogenously held fixed.

Figure 2.3: Impulse Response Functions to a Monetary Policy Shock: Persistent Shock with Large Magnitude

Notes: This figure shows impulse response functions when the model's Taylor rule is subjected to a 400 basis point expansionary interest rate shock, that decays according to an autoregressive process with an AR(1) parameter $\rho_R = 0.75$. The red set of impulse responses is for the frictionless benchmark model, the blue dotted set for the commitments model, and the green set for the commitments model when the share of goods subject to commitments is exogenously held fixed.

Figure 2.4: Impulse Response Functions to a Monetary Policy Shock: Transitory Shock with Large Magnitude

Notes: This figure shows impulse response functions when the model's Taylor rule is subjected to a 400 basis point expansionary interest rate shock, that decays according to an autoregressive process with an AR(1) parameter $\rho_R = 0.25$. The red set of impulse responses is for the frictionless benchmark model, the blue dotted set for the commitments model, and the green set for the commitments model when the share of goods subject to commitments is exogenously held fixed.
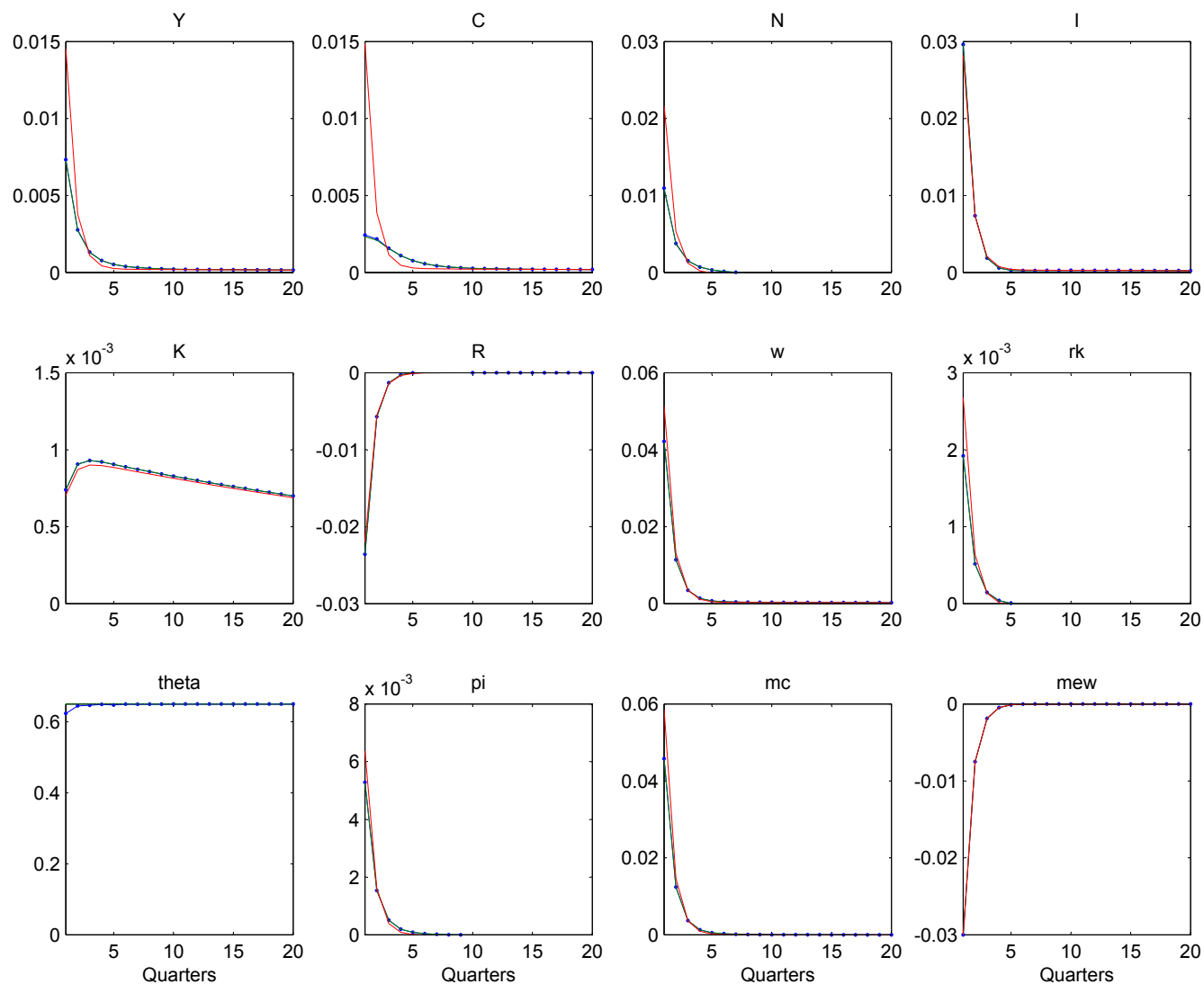
2.6 Appendix A: Proofs Omitted From the Main Text

2.6.1 Euler Equation With Consumption Commitments

This section shows the steps used to derive Equation (2.2.23). Supposing $\theta_t = \theta < 1$, Equation (2.2.19) simplifies to:

$$u'\left(c_t\right) = \left(1 - \beta\theta\right) B_t. \tag{2.6.1}$$

Assuming $u\left(c_t\right)$ exhibits constant relative risk-aversion, with an intertemporal elasticity of substitution σ equal to the reciprocal of the coefficient of relative risk-aversion, Equation (2.6.1) in log-linearized form is:

$$-\frac{1}{\sigma}\widetilde{c_t} = \widetilde{B_t}, \tag{2.6.2}$$

where $\widetilde{X_t} = \frac{X_t - X}{X}$ is the percentage deviation of variable X from its steady-state level. Subtracting this condition from itself advanced one period gives a log-linearized Euler equation for goods whose consumption is reset in the current period:

$$E_t\left[\widetilde{c_{t+1}}\right] = \widetilde{c_t} + \sigma E_t\left[\left(\widetilde{B_t} - \widetilde{B_{t+1}}\right)\right]. \tag{2.6.3}$$

The next step uses the log-linearized condition for B_t, and the household's envelope condition for bond holdings (Equation 2.2.14), to re-express the term in square brackets on the right-hand-side of Equation (2.6.3). Recalling that $\Lambda_t \equiv \frac{1}{R_{t-1}}\frac{\partial V_t}{\partial b}$, the log-linearized versions of these two conditions, respectively, are:

$$\widetilde{B_t} = \left(1 - \beta\theta\right)\widetilde{\Lambda_t} + \left(\beta\theta\right) E_t\left[\widetilde{B_{t+1}}\right], \tag{2.6.4}$$

and

$$\widetilde{\Lambda}_t = \left(\widetilde{R_t} - \widetilde{\pi_{t+1}}\right) + E_t\left[\widetilde{\Lambda_{t+1}}\right].$$ (2.6.5)

Substituting Equations (2.6.4) and (2.6.5) into Equation (2.6.3) gives the following representation of the Euler equation for consumption commitments:

$$E_t\left[\widetilde{c_{t+1}}\right] = \widetilde{c}_t + \sigma\left(1 - \beta\theta\right)E_t\sum_{k=0}^{\infty}\left(\beta\theta\right)^k\left(\widetilde{R_{t+k}} - \widetilde{\pi_{t+k+1}}\right).$$ (2.6.6)

Log-linearizing the law of motion for aggregate consumption (Equation 2.2.6) yields:

$$\widetilde{C}_t = \theta\widetilde{C_{t-1}} + \left(1 - \theta\right)\widetilde{c}_t.$$ (2.6.7)

With some re-arrangement, substitution of this expression into Equation (2.6.6) gives the desired Euler equation for aggregate consumption:

$$E_t\left[\Delta logC_{t+1}\right] = \theta\Delta logC_t + \sigma\left(1 - \theta\right)\left(1 - \beta\theta\right)E_t\sum_{k=0}^{\infty}\left(\beta\theta\right)^k\left(\widetilde{R_{t+k}} - \widetilde{\pi_{t+k+1}}\right),$$ (2.6.8)

where $\Delta logC_t = logC_t - logC_{t-1} \simeq \widetilde{C}_t - \widetilde{C}_{t-1}$.

2.6.2   Habit-Formation Preferences Euler Equation

This section shows that, when the habit parameter (b) is equal to the exogenous and constant share of goods subject to commitments each period (θ), internal habit-formation preferences imply an identical log-linearized consumption Euler equation to the model with consumption commitment. First, note that the log-linearized first-order condition for consumption with internal habit-formation preferences (Equation 2.2.26) is equal to:

$$-\frac{1}{\sigma}\widetilde{\left(\hat{C}_t\right)} = \widetilde{B}_t,$$ (2.6.9)

where $\widetilde{B}_t$ is given by Equation (2.6.4). Subtracting this condition from itself advanced one period, making use of Equations (2.6.4) and (2.6.5), and following essentially the same steps used in the previous section, gives a log-linearized Euler equation representation for consumption with internal habit-formation preferences:

$$E_t \left[\widetilde{\left(\hat{C}_{t+1} \right)} \right] = \widetilde{\left(\hat{C}_t \right)} + \sigma \left(1 - \beta\theta \right) E_t \sum_{k=0}^{\infty} \left(\beta\theta \right)^k \left(\widetilde{R_{t+k}} - \widetilde{\pi_{t+k+1}} \right). \tag{2.6.10}$$

Next, note that $\hat{C}_t \equiv C_t - bC_{t-1} = C_t - \theta C_{t-1}$ in log-linearized form is equal to

$$\widetilde{\left(\hat{C}_t \right)} = \frac{1}{1-\theta} \left[\tilde{C}_t - \theta\tilde{C}_{t-1} \right], \tag{2.6.11}$$

which after substitution into Equation (2.6.10) gives the desired consumption Euler equation representation (see Equation 2.2.23).

2.6.3   Consumption FOC: Approximate Representation

This section details the steps used to find a second-order approximation to Equation (2.2.19). First, take a second-order Taylor series approximation of Equation (2.2.20) to get

$$dA_t \simeq \beta E_t \left[d\theta_{t+1} A + \theta dA_{t+1} + d\theta_{t+1} dA_{t+1} \right], \tag{2.6.12}$$

which after some manipulation, using the steady-state condition $A \left(1 - \beta\theta \right) = 1$, can be re-expressed as

$$\tilde{A}_t \simeq \left(\beta\theta \right) E_t \left[\tilde{\theta}_{t+1} + \tilde{A}_{t+1} + \tilde{\theta}_{t+1}\tilde{A}_{t+1} \right], \tag{2.6.13}$$

where $\widetilde{X_t} = \frac{X_t - X}{X}$ is the percentage deviation of variable X from its steady-state level. Similarly, a second-order Taylor series approximation of Equation (2.2.21) gives

$$dB_t \simeq d\Lambda_t + \beta E_t \left[d\theta_{t+1} B + \theta dB_{t+1} + d\theta_{t+1} dB_{t+1} \right], \tag{2.6.14}$$

which with some manipulation, and using the stead-state condition $B\left(1-\beta\theta\right)=\Lambda$, can be re-expressed as

$$\tilde{B}_t \simeq \left(1-\beta\theta\right)\tilde{\Lambda}_t + \left(\beta\theta\right)E_t\left[\tilde{\theta}_{t+1}+\tilde{B}_{t+1}+\tilde{\theta}_{t+1}\tilde{B}_{t+1}\right]. \tag{2.6.15}$$

Next, let $f\left(A_t,B_t\right)=\frac{B_t}{A_t}$, and using this notation, take a second-order approximation to the right-hand-side of Equation (2.2.19) to get

$$
\begin{aligned}
u'\left(c_t\right) &\simeq f\left(A,B\right)+f_A dA_t+f_B dB_t+\frac{1}{2}f_{AA}dA_t^2+\frac{1}{2}f_{BB}dB_t^2+f_{AB}dA_t dB_t &\tag{2.6.16}\\
&= \frac{B}{A}-\frac{B}{A^2}dA_t+\frac{1}{A}dB_t+\frac{B}{A^3}dA_t^2-\frac{1}{A^2}dA_t dB_t, &\tag{2.6.17}
\end{aligned}
$$

which with some manipulation, and using the steady-state conditions $A\left(1-\beta\theta\right)=1$ and $B\left(1-\beta\theta\right)=\Lambda$ can be re-expressed as follows:

$$
\begin{aligned}
u'\left(c_t\right) &\simeq \Lambda+\Lambda\left[-\tilde{A}_t+\tilde{B}_t+\tilde{A}_t^2-\tilde{A}_t\tilde{B}_t\right] &\tag{2.6.18}\\
&= \Lambda+\Lambda\left(1-\tilde{A}_t\right)\left(\tilde{B}_t-\tilde{A}_t\right). &\tag{2.6.19}
\end{aligned}
$$

Substitution of Equations (2.6.13) and (2.6.15) into Equation (2.6.19) gives the desired representation for the consumer's consumption first-order condition, for goods whose consumption is reset in the current period:

$$u'\left(c_t\right)\simeq\Lambda+\Lambda\left(1-\beta\theta\right)\left(1-\tilde{A}_t\right)\left[\tilde{\Lambda}_t+E_t\sum_{k=1}^{\infty}\left(\beta\theta\right)^k\tilde{\Lambda}_{t+k}\left(\prod_{i=1}^{k}\left(1+\tilde{\theta}_{t+i}\right)\right)\right], \tag{2.6.20}$$

where, given Equation (2.6.13),

$$
\begin{aligned}
\tilde{A}_t &\simeq \left(\beta\theta\right)E_t\left[\tilde{\theta}_{t+1}+\tilde{A}_{t+1}+\tilde{\theta}_{t+1}\tilde{A}_{t+1}\right] &\tag{2.6.21}\\
&= E_t\left[\left(\beta\theta\right)\tilde{\theta}_{t+1}+\sum_{k=2}^{\infty}\left(\beta\theta\right)^k\tilde{\theta}_{t+k}\left(\prod_{i=1}^{k-1}\left(1+\tilde{\theta}_{t+i}\right)\right)\right]. &\tag{2.6.22}
\end{aligned}
$$

2.6.4   Asset Pricing Equations

This section present a more complete derivation of the equations presented in Section 2.3.2.

2.6.4.1   Pricing Kernel

The gross rate-of-return on holding stocks is given by:

$$R_{t+1}^s = \frac{R_{t+1}^k + (1-\delta) P_{t+1}^k}{P_t^k}, \tag{2.6.23}$$

where P_t^k is the price of a unit of capital at time t, and R_{t+1}^k is the rate-of-return from holding a unit of capital from time t to $t+1$. The household's first-order conditions imply that:

$$1 = \beta E_t \left[\left(R_{t+1}^s \right) \frac{\Lambda_{t+1}}{\Lambda_t} \right]. \tag{2.6.24}$$

Recalling that $\Lambda_t \equiv \frac{1}{R_{t-1}} \frac{\partial V_t}{\partial b}$, the envelope condition for bond holdings (Equation 2.2.14) can be used to find a expression for the rate-of-return on a risk-free asset:

$$1 = \beta E_t \left[\left(R_t \right) \frac{\Lambda_{t+1}}{\Lambda_t} \right]. \tag{2.6.25}$$

Subtracting Equation (2.6.25) from Equation (2.6.24) yields the asset pricing kernel given by Equation (2.3.8) in the main text:

$$0 = E_t \left[\left(R_{t+1}^s - R_t \right) \frac{\Lambda_{t+1}}{\Lambda_t} \right]. \tag{2.6.26}$$

Using the fact that $\frac{\Lambda_{t+1}}{\Lambda_t} \simeq 1 + \Delta log\Lambda_{t+1}$, for Λ_t close to its steady-state value Λ, gives:

$$E_t \left[R_{t+1}^s \right] - R_t \simeq -\frac{cov_t \left(R_{t+1}^s, \Delta log\Lambda_{t+1} \right)}{1 + E_t \left[\Delta log\Lambda_{t+1} \right]} \tag{2.6.27}$$

$$\simeq -cov_t \left(R_{t+1}^s, \Delta log\Lambda_{t+1} \right), \tag{2.6.28}$$

with the second approximate equality valid for small values of $\Delta log\Lambda_{t+1}$.

2.6.4.2 Consumption Commitments and Asset Pricing

Assuming utility for each consumption goods exhibits constant relative risk-aversion γ, the log-linearized version of Equation (2.2.22) is:

$$\widetilde{c_t} = -\frac{1}{\gamma}(1 - \beta\theta) E_t \sum_{k=0}^{\infty} (\beta\theta)^k \widetilde{\Lambda_{t+k}}, \qquad (2.6.29)$$

Using the approximation $\widetilde{X_t} = \frac{X_t - X}{X} \simeq log\frac{X_t}{X}$, Equation (2.6.29) can be first-differenced to get an expression for the change in consumption between adjacent time periods for goods whose consumption is reset:

$$E_t[\triangle log c_{t+1}] \simeq -\frac{1}{\gamma}(1 - \beta\theta) E_t \sum_{k=0}^{\infty} (\beta\theta)^k \triangle log\Lambda_{t+k+1}. \qquad (2.6.30)$$

Aggregate consumption evolves according to the law of motion $C_t = (1 - \theta) c_t + \theta C_{t-1}$. Log-linearizing this condition and first-differencing yields the following expression:

$$E_t[\triangle log C_{t+1}] \simeq (1 - \theta) E_t[\triangle log c_{t+1}] + \theta \triangle log C_t. \qquad (2.6.31)$$

Using these conditions together,

$$
\begin{aligned}
cov_t\left(R_{t+1}^s, \triangle log C_{t+1}\right) & \simeq cov_t\left(R_{t+1}^s, (1 - \theta)\triangle log c_{t+1} + \theta \triangle log C_t\right) & (2.6.32) \\
& = (1 - \theta) cov_t\left(R_{t+1}^s, \triangle log c_{t+1}\right) & (2.6.33) \\
& \simeq -\frac{1}{\gamma}(1 - \beta\theta)(1 - \theta) cov_t\left(R_{t+1}^s, \sum_{k=0}^{\infty}(\beta\theta)^k \triangle log\Lambda_{t+k+1}\right) & (2.6.34) \\
& = -\frac{1}{\gamma}(1 - \beta\theta)(1 - \theta) cov_t\left(R_{t+1}^s, \triangle log\Lambda_{t+1}\right), & (2.6.35)
\end{aligned}
$$

where the last line uses the fact that $cov_t\left(R_{t+1}^s, \triangle log\Lambda_{t+k}\right) = 0$ for $k > 1$. Hence,

$$-cov_t\left(R_{t+1}^s, \triangle log\Lambda_{t+1}\right) \simeq \frac{\gamma}{(1 - \beta\theta)(1 - \theta)} cov_t\left(R_{t+1}^s, \triangle log C_{t+1}\right), \qquad (2.6.36)$$

and

$$E_t\left[R_{t+1}^s\right] - R_t \quad \simeq \quad -cov_t\left(R_{t+1}^s, \Delta log\Lambda_{t+1}\right). \tag{2.6.37}$$

$$\simeq \quad \frac{\gamma}{(1-\beta\theta)(1-\theta)}cov_t\left(R_{t+1}^s, \Delta logC_{t+1}\right). \tag{2.6.38}$$

2.7 Appendix B: General Equilibrium Model: Details

This appendix provides details omitted from the main text for the general equilibrium model discussed in Section 2.4.

2.7.1 Firms

2.7.1.1 Final Goods Firms

Each consumer good C_j, for $j \in [0,1]$, is assembled (using no labor or capital) from a continuum of intermediate inputs according to the same constant returns-to-scale (CRS) production technology:

$$C_t\left(j\right) = \left(\int_0^1 x_{t,j}\left(i\right)^{\frac{\epsilon-1}{\epsilon}} di\right)^{\frac{\epsilon}{\epsilon-1}}, \tag{2.7.1}$$

where ϵ is the elasticity of substitution between inputs $x_{t,j}\left(i\right)$ in the production of final good $C_t\left(j\right)$. Firms producing each consumer good $j \in [0,1]$ choose $x_{t,j}\left(i\right)$ to solve:

$$\max_{\{x_{t,j}(i)\}} \quad P_t\left(j\right)C_t\left(j\right) - \int_0^1 \hat{P}_t\left(i\right)x_{t,j}\left(i\right)di, \tag{2.7.2}$$

where $\hat{P}_t\left(i\right)$ is the price faced by final goods firms of type j for input $x_{t,j}\left(i\right)$. Final goods firms' first-order condition for $x_{t,j}\left(i\right)$ implies the demand curve

$$x_{t,j}\left(i\right) = \left(\frac{\hat{P}_t\left(i\right)}{P_{t,j}\left(j\right)}\right)^{-\epsilon}C_t\left(j\right), \tag{2.7.3}$$

for each intermediate input by firms producing consumer good j. Because there is free-entry into the production of consumer goods, final goods firms earn zero profits, and the price of each good $P_t(j)$ is equal to its (constant) marginal cost of assembly:

$$P_t(j) = \left(\int_0^1 \hat{P}_t(i)^{1-\epsilon} \, di \right)^{\frac{1}{1-\epsilon}}. \tag{2.7.4}$$

Symmetry of the production technology across final goods implies $P_t(j) = P_t(k) = P_t$ for all $j, \, k \, \epsilon \, [0, 1]$.

2.7.1.2 Intermediate Goods Firms

Each intermediate goods firm rents capital from households and employs labor to produce its input $x_t(i)$ according to the constant-returns-to-scale technology:

$$x_t(i) = Z_t K_t(i)^\alpha N_t(i)^{1-\alpha}, \tag{2.7.5}$$

where Z_t is the level of total factor productivity, and α is the elasticity of output with respect to capital. The cost minimization problem faced by intermediate goods firm i to produce $\bar{x}_i$ units is:

$$\min_{\{N_t(i), K_t(i)\}} w_t N_t(i) + R_t^k K_t(i) \quad s.t. \quad x_t(i) \geq \bar{x}_i, \tag{2.7.6}$$

with the associated Lagrangian

$$L = - \left[w_t N_t(i) + R_t^k K_t(i) \right] + mc_t(i) \left[Z_t K_t(i)^\alpha N_t(i)^{1-\alpha} - \bar{x}_i \right]. \tag{2.7.7}$$

The first-order conditions for labor and capital demand, respectively, are:

$$[\partial N_t(i)] \quad w_t = (1-\alpha) \frac{x_t(i)}{N_t(i)} mc_t(i), \tag{2.7.8}$$

$$[\partial K_t (i)] \quad R_t^k = \alpha \frac{x_t (i)}{K_t (i)} mc_t (i) . \qquad (2.7.9)$$

Substituting these expressions into the production function (Equation 2.7.5) gives an expression for real marginal cost:

$$mc (x_t (i)) = \frac{1}{Z_t} \left(R_t^k \right)^\alpha (w_t)^{1-\alpha} \left(\frac{1}{\alpha} \right)^\alpha \left(\frac{1}{1-\alpha} \right)^{1-\alpha} , \qquad (2.7.10)$$

from which it is clear that $mc (x_t (i)) = mc (x_t) , \ \forall \ i \ \epsilon \ [0, 1]$.

Intermediate goods producing firms are monopolistically competitive, choosing $\hat{P}_t (i)$ to maximize profits, subject to the demand curve

$$x_t (i) = \int x_{t,j} (i) \, dj = \left(\frac{\hat{P}_t (i)}{P_t} \right)^{-\epsilon} \left(\int C_t (j) \, dj + I_t \right) \qquad (2.7.11)$$

$$= \left(\frac{\hat{P}_t (i)}{P_t} \right)^{-\epsilon} Y_t, \qquad (2.7.12)$$

where the second equality uses the fact that $C_t = \int C_t (j) \, dj$ because each consumer good has the same price P_t. The input price charged by intermediate goods producers is assumed to be sticky, with state-independent reset probability γ. When free to set its input price, the profit maximization problem faced by firm i is:

$$\hat{P}_t^* (i) = \underset{\hat{P}_t^* (i)}{argmax} \, E_t \left[\sum_{k=0}^\infty (\beta\gamma)^k \left[\left(\frac{1}{\Lambda_{t+k} P_{t+k}} \right) \left(\hat{P}_t^* (i) - MC_{t+k} \right) x_t (i) \right] \right]$$

$$= \underset{\hat{P}_t^* (i)}{argmax} \, E_t \left[\sum_{k=0}^\infty (\beta\gamma)^k \left[\left(\frac{1}{\Lambda_{t+k} P_{t+k}} \right) \left(\hat{P}_t^* (i) - MC_{t+k} \right) \left(\frac{\hat{P}_t^* (i)}{P_{t+k}} \right)^{-\epsilon} Y_{t+k} \right] \right],$$

$$(2.7.13)$$

where $MC_t = P_t mc_t$ is nominal marginal cost, and profits are discounted from the household's

point of view. The first-order condition for $\hat{P}^*_t(i)$ is:

$$\hat{P}^*_t(i) = \hat{P}^*_t = \left(\frac{\epsilon}{\epsilon - 1}\right) \frac{E_t\left[\sum_{k=0}^{\infty} (\beta\gamma)^k \left(\frac{1}{\Lambda_{t+k}}\right) P_{t+k}^{\epsilon-1} Y_{t+k} MC_{t+k}\right]}{E_t\left[\sum_{k=0}^{\infty} (\beta\gamma)^k \left(\frac{1}{\Lambda_{t+k}}\right) P_{t+k}^{\epsilon-1} Y_{t+k}\right]}. \tag{2.7.14}$$

Accordingly, the aggregate price level at any point in time is given by:

$$P_t^{1-\epsilon} = (1 - \gamma) \left(\hat{P}^*_t\right)^{1-\epsilon} + \gamma P_{t-1}^{1-\epsilon}. \tag{2.7.15}$$

2.7.2 Monetary Policy

The central bank sets the nominal interest rate each period according to the following rule:

$$R_t = \bar{R} + \phi_\pi (\pi_t - \bar{\pi}), \tag{2.7.16}$$

where R_t is the nominal gross interest rate on bonds, $\bar{R} = \frac{1}{\beta}$ is the steady-state gross nominal interest rate, and $(\pi_t - \bar{\pi})$ is the deviation in inflation from its steady-state level of zero. This interest rate rule is a special case of the Taylor (1993) rule with a zero weight on the output gap. For any value of $\phi_\pi > 1$ the nominal interest rate rises more than one-of-one with an increase inflation, guaranteeing a unique solution to the model.

2.7.3   Market Clearing

Market clearing in goods and factor markets requires that the following conditions hold in each time period:

$$N_t = \int_0^1 N_t(i)\,di, \tag{2.7.17}$$

$$K_t = \int_0^1 K_t(i)\,di, \tag{2.7.18}$$

$$Y_t(i) = x_t(i),\ \forall\, i, \tag{2.7.19}$$

$$Y_t = \int_0^1 Y_t(i)\,di, \tag{2.7.20}$$

$$Y_t = C_t + I_t + G_t. \tag{2.7.21}$$

2.8   Appendix C: Model: First-Order Conditions

This appendix lists the full set of first-order conditions used to quantitatively solve the model. All nominal variables, except the policy interest rate, have been scaled by the nominal price level.

Consumption:

$$(c_t)^{-\frac{1}{\sigma}} A_t = B_t \tag{2.8.1}$$

$$A_t = 1 + \beta E_t\left[\theta_{t+1} A_{t+1}\right] \tag{2.8.2}$$

$$B_t = \Lambda_t + \beta E_t\left[\theta_{t+1} B_{t+1}\right] \tag{2.8.3}$$

$$C_t = (1 - \theta_t)\, c_t + \theta_t C_{t-1}. \tag{2.8.4}$$

Adjustment of commitments:

$$U_t = (1 - \theta_t)\left[\frac{(c_t)^{1-\frac{1}{\sigma}}}{1 - \frac{1}{\sigma}}\right] + \theta_t U_{t-1} \tag{2.8.5}$$

$$\left[(c_t - C_{t-1}) + \left(U_{t-1} - \frac{(c_t)^{1-\frac{1}{\sigma}}}{1 - \frac{1}{\sigma}} \right) (c_t)^{\frac{1}{\sigma}} \right] = \psi \frac{\Lambda_t}{B_t} \left[\frac{(\theta - \bar{\theta})(\theta + \bar{\theta})}{\theta \bar{\theta}} \right]. \tag{2.8.6}$$

Labor supply:

$$\phi \frac{N_t^{\frac{1}{\eta}}}{w_t} = \Lambda_t \tag{2.8.7}$$

$$\frac{N_t^{\frac{1}{\eta}}}{w_t} = \beta E_t \left[\left(\frac{R_t}{\pi_{t+1}} \right) \frac{N_{t+1}^{\frac{1}{\eta}}}{w_{t+1}} \right]. \tag{2.8.8}$$

Capital adjustment costs:

$$\Phi_t = \frac{I_t}{K_t} - \frac{\xi}{2} \left(\frac{I_t}{K_t} - \delta \right)^2 \tag{2.8.9}$$

$$\Phi'_t = 1 - \xi \left(\frac{I_t}{K_t} - \delta \right). \tag{2.8.10}$$

Investment:

$$\Phi'_t q_t = \Lambda_t. \tag{2.8.11}$$

Shadow value of capital:

$$q_t = \beta E_t \left[q_{t+1} \left(R^k_{t+1} \Phi'_{t+1} + (1 - \delta) + \Phi_{t+1} - \Phi'_{t+1} \frac{I_{t+1}}{K_{t+1}} \right) \right]. \tag{2.8.12}$$

Capital accumulation:

$$K_{t+1} = (1 - \delta) K_t + \Phi_t K_t. \tag{2.8.13}$$

Aggregate price index:

$$1 = (1 - \gamma) (\hat{p}^*_t)^{1-\epsilon} + \gamma \pi_t^{\epsilon-1}. \tag{2.8.14}$$

Real marginal cost:

$$mc_t = \frac{1}{Z_t} \left(R_t^k\right)^\alpha (w_t)^{1-\alpha} \left(\frac{1}{\alpha}\right)^\alpha \left(\frac{1}{1-\alpha}\right)^{1-\alpha}. \tag{2.8.15}$$

Optimal reset price:

$$\hat{p}_t^* = \left(\frac{\epsilon}{\epsilon-1}\right) \frac{F_t}{H_t} \tag{2.8.16}$$

$$F_t = Y_t \Lambda_t^{-1} mc_t + \beta\gamma E_t \left[\pi_{t+1}^\epsilon F_{t+1}\right] \tag{2.8.17}$$

$$H_t = Y_t \Lambda_t^{-1} + \beta\gamma E_t \left[\pi_{t+1}^{\epsilon-1} H_{t+1}\right]. \tag{2.8.18}$$

Wage rate:

$$w_t = (1-\alpha) \frac{Y_t}{N_t} \left[(1-\gamma)(\hat{p}_t^*)^{-\epsilon} + \gamma\pi_t^\epsilon\right] mc_t. \tag{2.8.19}$$

Capital rental rate:

$$R_t^k = \alpha \frac{Y_t}{K_t} \left[(1-\gamma)(\hat{p}_t^*)^{-\epsilon} + \gamma\pi_t^\epsilon\right] mc_t. \tag{2.8.20}$$

Goods market clearing:

$$Y_t = C_t + I_t. \tag{2.8.21}$$

Taylor rule:

$$R_t = \bar{R} + \phi_\pi (\pi_t - \bar{\pi}) + \mu_t. \tag{2.8.22}$$

Monetary policy shock process:

$$\mu_t = \rho_R \mu_{t-1} + u_t^\mu. \tag{2.8.23}$$

Technology:

$$Z_t = \rho_Z Z_{t-1} + (1 - \rho_Z) \bar{Z} + u_t^Z. \tag{2.8.24}$$

2.9 Bibliography

Boldrin, Michele, Lawrence J. Christiano, and Jonas D. M. Fisher (2001). Habit Persistence, Asset Returns and the Business Cycle. *American Economic Reivew* 91(1): 149–166.

Chetty, Raj and Adam Szeidl (2007). Consumption Commitments and Risk Preferences. *Quarterly Journal of Economics* 122(2): 831–877.

Chetty, Raj and Adam Szeidl (2010). Consumption Commitments: A Foundation for Reference-Dependent Preferences and Habit Formation, Working Paper, Harvard University.

Christiano, Lawrence J., Martin Eichenbaum, and Charles L. Evans (2005). Nominal Rigidities and the Dynamic Effects of a Shock to Monetary Policy. *Journal of Political Economy* 113(1): 1–45.

Deaton, Angus S. (1987). Life-Cycle Models of Consumption: Is the Evidence Consistent with the Theory? In Truman F. Bewley (Ed.), *Advances in Econometrics, Fifth World Congress, Volume 2*. Cambridge: Cambridge University Press.

Flavin, Marjorie and Shinobu Nakagawa (2008). A Model of Housing in the Presence of Adjustment Costs: A Structural Interpretation of Habit Persistence. *American Economic Review* 98(1): 474–495.

Flavin, Marjorie A. (1981). The Adjustment of Consumption to Changing Expectations About the Future. *Journal of Political Economy* 89(5): 974–1009.

Gabaix, Xavier and David Laibson (2001). The 6D Bias and the Equity-Premium Puzzle. In Ben S. Bernanke and Kenneth Rogoff (Eds.), *NBER Macroeconomics Annual*, Volume 16. Cambridge MA: MIT Press.

Grossman, Sanford J. and Guy Laroque (1990). Asset Pricing and Optimal Portfolio Choice in the Presence of Illiquid Durable Consumption Goods. *Econometrica* 58(1): 25–51.

Gruber, Jonathan (2006). A Tax-based Estimate of the Intertemporal Elasticity of Substitution, NBER Working Paper No. 11945.

Hall, Robert E. (1988). Intertemporal Substitution in Consumption. *Journal of Political Economy* 96(2): 339–357.

Kimball, Miles S., Claudia R. Sahm, and Matthew D. Shapiro (2008). Imputing Risk Tolerance from Survey Responses. *Journal of the American Statistical Association* 103(483): 1028–1038.

Kocherlakota, Narayana R. (1996). The Equity Premium: It's Still a Puzzle. *Journal of Economic Literature* 34(1): 42–71.

Moscarini, Giuseppe (2004). Limited Information Capacity as a Source of Inertia. *Journal of Economic Dynamics and Control* 28(10): 2003–2035.

Parker, Jonathan A. (2003). Consumption Risk and Expected Stock Returns. *American Economic Reivew* 93(2): 376–382.

Postlewaite, Andrew, Larry Samuelson, and Dan Silverman (2008). Consumption Commitments and Employment Contracts. *Review of Economic Studies 75*: 559–578.

Reis, Ricardo (2006). Inattentive Consumers. *Journal of Monetary Economics* 58(8): 1761–1780.

Sims, Christopher A. (2003). Implications of Rational Inattention. *Journal of Monetary Economics* 50(3): 665–690.

Taylor, John B. (1993). Discretion versus Policy Rules in Practice. *Carnegie Rochester Conference Series on Public Policy 39*: 195–214.

Chapter 3

Evidence on Unclaimed Charitable Contributions from the Introduction of Third-Party Information Reporting in Denmark[1]

3.1 Introduction

For the 2008 tax year, Denmark's tax authority (SKAT) introduced third-party information reporting for tax-deductible charitable contributions, where previously these deductions were self-reported and subject to verification only upon an audit. Under the new system, charitable organizations report contributions received from each taxpayer directly to the tax authority. These information reports are used by SKAT to pre-fill charitable deductions on taxpayers' annual declarations (referred to as pre-population). While information reporting is now widely used for sources of income tax return line items in advanced countries, the use of information reporting and pre-population for a tax return *deduction* line item is relatively

---

new.[2]

The effect of the policy change on reported deductions was immediate, large, and in some respects surprising; the number of taxpayers claiming a charitable deduction doubled. The total value of contributions also rose, but by only by 15 percent, due to a fall in the mean charitable tax deduction of 42 percent.

Using data from a recent large-scale audit experiment in Denmark, we document that pre-reform overreporting of charitable contributions was negligible. This is somewhat unexpected, because evasion rates for self-reported sources of income are often large (see Slemrod 2007). The same audit experiment estimated an evasion rate of 42 percent for total self-reported net income, but only 0.3 percent for third-party reported income (see Kleven et al. 2011). There is good reason to trust the accuracy of these audits in identifying overclaiming of charitable deductions: unlike self-reported sources of income, the burden of proof falls on the taxpayer, who under the self-reporting regime was required upon audit to produce receipts to justify all deductions claimed. For reasons discussed in detail in Section 3.3, audits did not appear to identify unclaimed charitable deductions.

Administrative reports on total donations collected by charities enable us to separately identify the effect of the policy change on charitable giving and reporting behavior. We find no evidence of a change in giving behavior coinciding with the introduction of information reporting and pre-population of deductions. Accordingly, we argue that the rise in the value of reported deductible contributions—and the near doubling in the number of reporting contributors—is due to taxpayers with modest tax-deductible contributions who neglected to report their deductions under the self-reporting regime in place before 2008.[3]

We estimate that the average unclaimed charitable tax deduction under the self-reporting regime was worth about DKK786, which, given the one-third subsidy rate, translates to DKK262 in forgone after-tax income.[4] There was little change in the number of tax deduc-

---

[2]See OECD (2006) for a survey of pre-population in OECD countries.

[3]Examining a policy experiment in Finland in the 1990s, Kotakorpi and Laamanen (2013) argue that unclaimed deductions may be particularly prevalent when many sources of income line items are pre-filled for taxpayers.

[4]DKK1 is approximately US$0.18.

tions of more than DKK2,500—indicating that few taxpayers left large sums of money on the table in any given year. But over a period of years, the cumulative amount of foregone benefits appears to have been economically significant for many taxpayers; more than two-thirds of the taxpayers who claimed a deduction in 2008 under the information reporting and pre-population regime, but who did not claim a deduction in either 2006 or 2007 under the self-reporting regime, claimed a deduction in each of the years 2009-2011.

Our finding of negligible charitable overreporting under the self-reporting regime is interesting in light of work by Fack and Landais (2011), who find that reforms in the U.S. and France that tightened enforcement of charitable tax deductions resulted in a fall in reported donations, which they attribute to evasion. In France, a 1983 reform required taxpayers to attach receipts to their tax return for all charitable deductions claimed, whereas previously the tax authority only sought to inspect receipts during an audit. The rule change coincided with a 75 percent fall in the value of charitable tax deductions claimed between 1982 and 1983. In the U.S., a 1969 law change reduced opportunities for top-income earners to use private charitable foundations as a tax sheltering or evasion scheme. Following the law change, creation of private charitable foundations fell by 80 percent. They estimate that 30 percent of charitable tax deductions claimed by the top 0.1 percent of income earners before the policy change was due to tax avoidance or evasion behavior. The reform in France studied by Fack and Landais (2011) is more relevant in our setting because we study behavior for the population of donors, rather than just top-earner taxpayers.

Our findings are consistent with Rehavi (2010), who uses survey reports of U.S. taxpayers to provide suggestive evidence of incomplete claiming of eligible charitable deductions. She goes on to argue that as much as one-third of the response of charitable tax deductions to the subsidy rate is due to changes in reporting rather than giving behavior.[5] In contrast to the survey evidence used by Rehavi (2010), the administrative panel data available to us provides arguably more credible evidence because it is less susceptible to systematic misreporting (providing incorrect information to the tax authority has an expected penalty,

---

[5]Slemrod (1989) finds, based on analysis of audited U.S. income tax returns, that charitable giving overstatement is less sensitive to the subsidy rate than is actual giving behavior.

whereas misreporting on a household survey does not).

A related literature on incomplete enrollment in benefit programs has found evidence of sizable unclaimed benefits. Bhargava and Manoli (2011) estimate that about one-quarter of taxpayers apparently eligible for the U.S. earned-income tax credit (EITC) do not claim the EITC. However, we recognize that the type of taxpayers who are eligible for the EITC and those who make charitable gifts are likely to differ in important ways that affect their claiming behavior. Elsewhere in the literature, stigma is often cited as a reason for incomplete take-up of welfare benefits (see, for example, Besley and Coate, 1992), but there should be no stigma attached to claiming charitable deductions. Pre-population is akin to default enrollment—taxpayers are automatically credited with their eligible charitable tax benefits—and the post-reform surge in charitable tax deductions claimed attests to the power of defaults (see, for example, Carroll et al., 2009 or Chetty et al., 2012).

Some taxpayers may have rationally decided not to claim charitable tax deductions because the private compliance cost exceeds the forgone tax benefits. For 1982 U.S. taxpayers, Pitt and Slemrod (1989) estimated the compliance costs of itemizing deductions by estimating how much taxpayers claiming the standard deduction could have have saved from instead itemizing their deductions. They estimated a compliance cost of \$43, which is, after adjusting their estimate in 1982 dollars for inflation, about double our preferred estimate of the average value of charitable deductions forgone under the self-reporting regime. But the Pitt and Slemrod (1989) estimate of compliance costs should be larger because it measures the compliance costs associated with all deductions for which a taxpayer is eligible, not just charitable contributions; differences in tax-system design between Denmark and the U.S. may also affect the comparability of these estimates.

More generally, this paper contributes to a growing literature that takes optimization frictions seriously: Kleven and Waseem (2013) find that a majority of the income taxpaying population in Pakistan face optimization frictions affecting their taxable income choice of at least 2.5 percent of gross income; Chetty (2012) shows that it is possible to reconcile high-quality intensive-margin labor supply elasticity estimates from the labor and public finance literatures

given an assumption of frictions equal to about one percent of income; and Saez (2010) finds kinks in the U.S. Earned Income Tax Credit insufficiently powerful to create bunching, except at the first kink, and only for self-employed taxpayers.[6] The attenuated response of taxable income to marginal tax rates reflects taxpayer frictions such as inattention, misperception, and inertia, but also adjustment costs faced by taxpayers in finding employers offering desired combinations of hours of work and compensation. Unsurprisingly, the magnitude of frictions affecting claiming of charitable tax deductions appear to be much smaller than is required in other recent work to reconcile observed behavior of taxable income with a frictionless benchmark.

Theoretical work by Kleven and Kopczuk (2011) argues that differential response by type of taxpayer to compliance costs can be exploited to discriminate between deserving and undeserving welfare program recipients. If hassle costs are more burdensome for undeserving than for deserving applicants, the introduction of a hassle cost, such as a paperwork requirement, may facilitate a higher benefit level, that in the absence of the hassle cost would induce substantial additional take-up by undeserving applicants. The reform we study is interesting in light of this (mostly) theoretical literature because it provides empirical evidence on taxpayer response to a change in a paperwork requirement (taxpayers had to maintain receipts and process their own charitable deductions under the self-reporting regime). Although we find the increase in the share of taxpayers claiming a deduction in the post-reform period to be particularly large for some groups of taxpayers, this variation appears related to underlying giving propensity, rather than a differential effect of hassle cost across taxpayer types.

Our findings suggest that the use of information reports to pre-populate tax-deduction line items may result in a loss in revenue. The use of information reports alone need not though: a tax authority could use third-party reports to flag for further investigation taxpayers who overclaim on their charitable contributions, but not amend tax returns for underclaiming. Unlike pre-population of sources of income line items, automatic crediting of deductions increases tax expenditures on taxpayers who would otherwise neglect to claim deductions

---

[6]Saez (2010) attributes the bunching of self-employed taxpayers at the first kink in the EITC schedule to tax evasion.

for which they are eligible. The introduction of information reporting and pre-population of charitable deductions in Denmark coincided with an increase in the value of charitable tax expenditures of DKK35.4 million. Absent a change in giving behavior, the reform is socially desirable only if it lowers taxpayer compliance costs by more than an appropriately weighted sum of the increase in administrative expense and tax benefits. We outline a simple modeling framework that can be used to weight these costs and benefits.

In what follows, section 3.2 provides background information on relevant aspects of Denmark's tax system, section 3.3 uses data from a pre-reform tax audit experiment to investigate reporting behavior before the policy change, and section 3.4 discusses the change in reporting behavior when information reporting and pre-population of charitable deductions was introduced in 2008. Section 3.5 presents evidence indicating that there was no change in charitable giving—as opposed to reporting of charitable gifts—around the time of the policy change, and section 3.6 uses a notch created by the pre-2012 charitable gift eligibility rules to investigate taxpayer awareness of incentives for charitable giving. In section 3.7 we outline a simple normative model of the optimal reporting regime for charitable tax deductions, using our empirical findings to inform judgment on the social desirability of the policy change. We offer some concluding remarks in section 3.8.

3.2   Background

Denmark's individual-income tax system features broad use of information reporting across various sources of income. For most taxpayers, information reports made by third parties for the tax year ending in December arrive at the tax authority for processing by late January. Most information reports correspond to payments from which tax has been withheld, but some do not. Making use of the information in these reports, and other known information such as place of residence, SKAT prepares pre-populated (pre-filled) returns that are mailed to taxpayers each year in mid-March.[7]   Taxpayers have until May 1 to amend their pre-

---

[7]Taxpayers can also access their pre-filled tax returns electronically.

populated return to reflect sources of income not subject to information reports, any income for which information reports were not received in a timely manner by the tax authority, and any self-reported deductions for which the taxpayer is eligible.[8] All income-tax-liable people in Denmark are required to file a tax return, which is approximately 88 percent of the population (Kleven et al., 2011).[9]

All taxpayers file as individuals, unlike in the U.S. where married couples generally elect to pool their income and file a joint tax declaration. The subsidy rate for tax deductible charitable contributions varies only (slightly) by region of residence—and so does not depend on a taxpayer's marginal tax rate. Assuming married couples live in the same tax region, this means that there is no tax advantage gained from shifting the claiming of charitable deductions between husband and wife depending on who faces the higher marginal tax rate. Because there is no difference in tax treatment of charitable deductions between singles and couples, our unit of analysis is individual taxpayers. Even if taxpayers have no tax liability, they are able to receive tax benefits for their charitable contributions.

According to government documents, the principal stated motivation for the introduction of information reporting and pre-population for charitable deduction was a desire to limit perceived abuse of charitable deductions and to lower taxpayer compliance costs. The tax authority also appears to have been aware that pre-population would lead to some taxpayers receiving tax benefits they previously neglected to claim. No net change in charitable tax expenditures was expected prior to the reform. To ease their transition to the new policy regime, charitable organizations received a subsidy for expenses associated with implementing the new compliance procedures.

Charitable deductions fall into three tax-relevant categories, each with different requirements for tax favored treatment. The bulk of charitable contributions are regular gifts, for which there was a somewhat complicated eligibility requirement before 2012. Only total annual gifts to each eligible charity of DKK500 or more qualified for tax deductibility, and in calculating

---

[8]Taxpayers can amend their pre-populated return electronically, by phone, or by mail. Self-employed filers have until July 1 to submit their final tax return.

[9]The bulk of those not required to file are children under the age of 16.

the total tax deduction for each taxpayer the first DKK500 in gifts was excluded.[10] We discuss the incentives created by this policy design in detail in Section 3.6. Information reporting *and* pre-population of deductions for regular gifts was introduced in 2008. With the introduction of information reporting and pre-population for regular gifts, the tax authority also *locked* this line item for most taxpayers. This means taxpayers are prevented from changing the charitable deduction recorded on their pre-populated return. If the taxpayer finds an error on their pre-populated return they must contact the relevant charity and request a revised message be sent to the tax authority. Deductions are also capped, and thus so is the maximum value of charitable tax benefits. The maximum value of regular deductions eligible for tax deducibility has increased over time: from 1997-2004 the cap was DKK5,000, but the cap was lifted to DKK6,400 in 2005, and to DKK6,600 in 2006; in 2007 the upper threshold more than doubled to DKK13,600, and has increased modestly since, to DKK14,000 in 2008, and to its current DKK14,500 level in 2011.

The second category of charitable donations corresponds to giving contracts with a minimum 10-year length, for which information reporting and pre-population of deductions was also introduced in 2008. This category permits donors to deduct the larger of DKK15,000 or 15 percent of taxable income each year. A third category was introduced in 2008 for gifts to cultural and research organizations, and for which information reporting was introduced in 2010. Because this type of gift was not tax deductible before 2008, we exclude this category from our analysis entirely. For the two categories of gift we study (regular and long-term charitable gifts), only cash contributions are eligible for a tax deduction.[11]

In 2011 the number of taxpayers claiming a deduction for regular, long-term, and cultural and research gifts was 360,527, 44,399, and 23,477, respectively. Total gifts for each category was DKK747m, DKK261m, and DKK20m, respectively. Before 2008 regular and long-term

---

[10]In 2012 the lower threshold was abolished, making gifts of less than DKK500 eligible for tax deductibility. In addition, the 2012 reform no longer requires subtracting the first DKK500 in gifts from total eligible deductions.

[11]Non-monetary gifts to cultural organizations have been eligible for a tax deduction since 2005. There was no upper threshold for these gifts, but as for regular gifts only contributions with a value greater than DKK500 were eligible to receive tax-deductibility. In 2008, the first year in which we observe data specifically for cultural and research organization gifts, there were only 11 such gifts made.

gifts were self-reported together on one tax return line item, but from 2008 forward each category corresponds to a separate line item. Because we do not observe each category of donation separately before 2008, we group regular and long-term gifts together to form one consistent series for charitable giving.

There was little change in the number of charities reporting charitable gifts in the years before and after the 2008 policy change. In 2008 there were 796 organizations approved by SKAT to receive tax-deductible contributions, only slightly higher than 790 in 2007 (see Table 3.1). This represents the equal second smallest year-to-year increase for the years 1998-2011. In both 2007 and 2008 the fraction of eligible organizations making an annual report to the tax authority was 93 percent. This fraction has been stable, but had an upward trend over our sample period.

Most donations were collected from the following groups of charitable organizations: international aid organizations (e.g., UNICEF, Red Cross); religious organizations (e.g., Catholic Church); national social and humanitarian organizations (e.g., Blue Cross Denmark); nature, environment, and animal welfare organizations (e.g., Danish Society for Nature Conservation); and disease fighting and disability organizations (e.g., Cancer Society).

In the next section we investigate reporting behavior prior to the introduction of information reporting and pre-population of deductions.

3.3   Pre-Reform Misreporting of Charitable Gifts

Before investigating the effect of the policy change in the next section, we first draw on data from the Kleven et al. (2011) audit experiment to ascertain the level of misreporting of charitable gifts in Denmark prior to the reform. A random sample of about 20,000 taxpayers was subjected in 2007 to an unannounced extensive and thorough audit of their 2006 tax returns. The overall misreporting rate for charitable contributions was small: of the 872 taxpayers in the audit sample who reported any charitable contribution, only 7 percent were found upon audit to have overclaimed charitable deductions, while 3 percent were

found to have underclaimed, combining, with rounding, to give a gross misreporting rate of 11 percent. For the 7 percent of taxpayers who overclaimed, the median value of excess charitable deductions reported was DKK1,100, and for the 3 percent of taxpayers in the audit sample found to have underclaimed, the median value of missing deductions was DKK1,975. The value of underclaiming offset about half the value of overclaiming, giving a net evasion rate (net overclaiming as a share of deductions that should have been claimed) of 2.3 percent conditional on having initially reported a non-zero charitable gift, and about 0.1 percent as a share of all taxpayers in the audit sample. This evasion rate is trivial compared to the 37 percent evasion rate found by Kleven et al. (2011) for self-reported sources of income. Evidently, those seeking to evade income taxes do not view overstatement of charitable contributions as a high expected benefit-to-cost evasion opportunity.

In light of these audit results, our finding of a surge in reported tax-deductible charitable contributions following the introduction of third-party information reporting and pre-population may seem surprising. If so many taxpayers neglected to claim their tax deductible contributions under the self-reporting regime, why did the auditors in the Kleven et al. (2011) study detect such little underclaiming? We have ascertained from discussions with SKAT officials that auditors did not investigate line items for which no deductions were claimed.[12] This is most probably a sensible audit policy rule for the tax authority: the social value of finding unclaimed deductions for taxpayers is arguably less than the social cost of auditors' time. But it means that the Kleven et al. (2011) audit sample results cannot be used to accurately measure the fraction of taxpayers with unclaimed tax-deductible charitable gifts. The only way in which the audit process could have resulted for a taxpayer in a higher post-audit than pre-audit charitable deduction was if the audit process prompted the taxpayer to review their records and discover charitable deductions they had not reported. However, we have been told by SKAT that some audits in the Kleven et al. (2011) audit study involved only computerized cross-checking of information reports, in which case the taxpayer was unaware that their tax return had been audited; for example, a taxpayer with no self-reported income

---

[12]We would like to thank Søren Pedersen for sharing this detail of SKAT's audit procedure with us.

or deductions would have had their third-party reported information cross-checked electronically, but would have only been contacted as part of the audit process if a discrepancy was discovered.

Having established that there was negligible charitable evasion under the pre-reform self-reporting regime, in the next section we describe the change in reporting behavior due to the introduction of information reporting and pre-population of charitable deductions in 2008.

3.4 Effect of the Reform on Reporting Behavior

3.4.1 Aggregate Data

Figure 3.1 reports the number and average size of charitable tax deductions reported over the period 1997-2011.[13] As foreshadowed in section 3.1, the introduction of information reporting and pre-population for charitable deductions coincided with a near doubling in the number of taxpayers claiming a charitable tax deduction: 150,311 taxpayers reported a charitable tax deduction in 2007 under the self-reporting regime, and 300,122 taxpayers had a charitable deduction in 2008 following the policy change (see Table 3.2). There was an accompanying rise in the value of tax deductions claimed between 2007 and 2008, but the rise was a relatively modest 15.3 percent. As discussed in detail below, we find that the bulk of the new claims were small in value. Accordingly, the mean value of tax deductions claimed fell sharply between 2007 and 2008, from DKK4,671 to DKK2,697 (see Table 3.2).

Interestingly, the mean value of contributions was higher in the year before the reform than in earlier years. Between 1997 and 2006 the mean value of charitable tax deductions claimed was between DKK3,859 and DKK4,029, lower than the DKK4,671 mean value recorded in 2007. This change can be mostly explained by a relaxation in the upper threshold for eligible regular gifts: in 2007 taxpayers were permitted to deduct up to DKK13,600 in regular

---

[13]Before 1997 charitable gifts were reported on the same tax return line item as a standard deduction available to fishermen, and a special childcare deduction. Since 1997 these deductions have been reported separately from charitable gifts.

charitable tax deductions, compared to only DKK6,600 in 2006. As described in section 3.2, this threshold has increased over time, but the 2007 increase was by far the largest over our sample period. The number of taxpayers with total tax deductions greater than DKK10,000 rose by 6,344 between 2006 and 2007, and there was a corresponding 6,350 fall in the number of taxpayers with total tax deductions in the range DKK5001-10,000. There was a further modest rise in the upper eligibility threshold for regular tax deductions in 2008, but this does not meaningfully affect our analysis. The bulk of the increase in tax deductions due to the policy reform were small in value, so our focus is on the lower tail of the distribution of claims, that is largely unaffected by changes to the upper eligibility threshold.

We have access to taxpayer level microdata beginning in 2006, and can compute the median tax deduction reported in each year (see Table 3.2). Because most claims are small in value, the median value of claims is only a little more than half the mean contribution for the years 2007-2011. The relaxation in the upper threshold for regular gifts in 2007 was relevant for a relatively small number of taxpayers making large donations, explaining why the median deduction rose by only DKK70 between 2006 and 2007, compared to the DKK638 rise in the mean value of contributions.

To gain further insight on the effect of the reform, we investigate changes in tax deductions reported by claim size. Table 3.3 reports these data for each year 2006-2011, and Figure 3.2 presents these data graphically. Note that claim size is the tax deductible amount on individual tax returns, not the total value of contributions, which is larger because of the exemption limits that existed before 2012. For example, a taxpayer who gave a total of DKK600 to one charity would qualify for a tax deduction of DKK100 and be counted in the category DKK0-500 in Table 3.3 and Figure 3.2. As previewed earlier, the surge in the number of charitable deductions claimed between 2007 and 2008 were primarily small in value; there was an almost ten-fold increase in the number of claims less than DKK500, and a more than doubling in the number of claims in the range DKK500-DKK1,500. In contrast, there was little change in the number of claims larger than DKK3,000.

For the two years before and after the policy change, Figure 3.3 presents a finer picture

for the distribution of claims less than DKK5,000. The surge in small claims in 2008 when information reporting and pre-population of deductions was introduced is particularly evident here. Abstracting from the policy change, the distribution of claims is very stable: Figure 3.3 shows that the pre-reform 2006 and 2007 distribution of tax deductions claimed are almost identical, as are the post-reform 2008 and 2009 distributions. This makes us confident that the pronounced change in the left tail of the claim distribution between 2007 and 2008 is not in part accounted for by regular variation in the distribution of claims over time.

If we attribute all the change in charitable tax deductions between 2007 and 2008 to a decline in unreported claims, the value of forgone charitable deductions in 2007 was DKK717. However, this is an imprecise estimate of the value of deductions forgone under the self-reporting regime. Any change in the number of large tax deductions between 2007 and 2008 is probably unrelated to the policy change: those with large deductions forgo a substantial amount of money from not reporting their eligible deductions and so are unlikely to have not done so under the self-reporting regime. Informed by the distribution of claims data presented in Figure 3.3, we estimate the value of forgone deductions under the self-reporting regime by restricting our attention only to the increase in claims less than DKK2,500. Between 2007 and 2008 the total number of tax deductions claimed amounting to less than DKK2,500 increased from 77,046 to 226,855, and the total value of these deductions increased from DKK116 million to DKK234 million. This implies an average value of DKK786 for forgone deductions, which corresponds to DKK262 in after-tax income. This calculation is not particularly sensitive to the upper threshold of DKK2,500 used in this calculation (see Figure 3.11 in the appendix, and the notes therein for details on this calculation). Had the reform not occurred, our estimated value of previously unreported deductions implies that we would have observed a mean value of tax deductions equal to DKK4,601 in 2008, rather than the actual value of DKK2,697.

These estimates implicitly assume that there would have been no change in average giving behavior had the reform not occurred, which absent a control group (the reform affected all taxpayers at the same time) we cannot formally test. Although this assumption is almost

certainly violated, the magnitude of the change in reporting behavior pre- and post-reform is several orders of magnitude larger than the usual year-to-year variation in reporting behavior (see Figures 3.1 and 3.3); hence, any error in our estimate due to trend changes in average giving behavior is likely to have only a minor effect on our estimate of the change in reporting behavior due to the reform.

Interestingly, the bulk of the increase in charitable deductions claimed after 2008 appear to be associated with regular, rather than occasional, donors who did not claim their eligible tax benefits under the prior self-reporting regime. Of the 152,857 taxpayers who claimed a charitable tax deduction in 2008 (under the information reporting and pre-population regime) but not in 2006 or 2007 (under the self-reporting regime), 68 percent claimed a deduction in each subsequent year 2009-2011. The share claiming zero, one, and two further tax deductions between 2009 and 2011 was 13, 9, and 10 percent, respectively (see Table 3.8). This suggests that foregone tax benefits under the self-reporting regime were concentrated among regular donors who systematically did not claim eligible charitable deductions, rather than a larger group of donors who occasionally did not claim their eligible deductions. Although the typical amount of forgone tax benefits appears to have been modest in any given year, our finding that many taxpayers repeatedly failed to claim eligible tax benefits indicates that the cumulative amount of forgone deductions and tax savings may have been substantial for a sizable fraction of charitable donors.

3.4.2 Effect of the Reform by Type of Taxpayer

In this section we look for evidence of differential response to the policy change by type of taxpayer. We present estimates for the following OLS panel data regression using the universe of tax returns for the period 2006-2011:

$$D_{it} = \sum_{j=1}^{k} \beta_j X_{ijt} + \gamma_j post_t X_{ijt} + \varepsilon_{ijt}, \tag{3.4.1}$$

where $D_{it} = \{0, 1\}$ is an indicator for person i claiming a charitable deduction in year t, X_{ijt} is characteristic j for taxpayer i in year t, and *post* is an indicator variable taking the value unity in the post-reform period 2008-2011.[14] The vector of characteristics X_{ijt} includes the following variables: age, personal income (the sum of labor income, transfers, pensions, and other adjustments), gender, marital status, self-employment status, a Copenhagen location dummy variable, and a linear time trend. We do not include a taxpayer fixed effect because many of the covariates of interest are constant or vary little at the taxpayer level over our data sample. We report robust standard errors and, because we have access to the universe of tax returns, all but a few point estimates are highly statistically significant. The full set of regression results is reported in Table 3.7 in the appendix.

The coefficient on the *post* variable, shown in Figure 3.4, indicates the estimated pre- to post-reform change in probability of claiming a charitable tax deduction, for a taxpayer with the baseline set of characteristics (the baseline set of characteristics represents a male taxpayer aged 46-65, in 50-75th income percentile, single, residing outside Copenhagen, and not self-employed); the coefficients on the *post × income* interaction terms, also shown in Figure 3.4, indicate estimated variation in post-reform claiming behavior by income percentile. There is a clear positive income gradient evident for the *post × income* interaction terms shown in Figure 3.4, indicating that the increase in the share of high-income taxpayers claiming a charitable deduction in the post-reform period was large relative to low income groups. But, because high income earners were also more likely to claim a charitable deduction in the pre-reform period (shown by the main effect coefficients in Figure 3.4), the proportional increase in likelihood of claiming a deduction following the reform is similar for high income groups; the regression estimates are consistent with a roughly constant fraction of taxpayers in high-income groups neglecting to claim eligible deductions in the pre-reform period. For below-median income earners, the regression results indicate a small fall in the probability of claiming a deduction post-reform; this most likely reflects their underlying very low

---

[14]We restrict our analysis to the sample of taxpayers who filed a tax return in each year 2006-2011 (only two percent of taxpayers who claimed a charitable tax deduction in 2008 did not file a tax return in each year 2006-2011).

propensity to claim a charitable deduction, and variation unrelated to the reform.

Figure 3.5 displays the analogous coefficient estimates by age category. The likelihood of claiming a charitable deduction post-reform increased for each age category: the increase is estimated to have been largest for young taxpayers (under 25), and smallest for taxpayers aged 26-45. The change in likelihood of claiming a deduction post-reform for selected other taxpayer characteristics is reported in Figure 3.6. The increase in post-reform claiming probability was particularly large for female taxpayers, and those residing in Copenhagen— both groups with a high propensity relative to other taxpayers of claiming a deduction in the pre-reform period (indicated by the main effect coefficients shown in Figure 3.6). Post-reform claiming behavior appears unrelated to employment status, and married taxpayers were a little less likely to claim a charitable deduction in the post- than pre-reform period.

In summary, the increase in the share of taxpayers claiming a charitable deduction in the post-reform period was particularly large for high income groups, female taxpayers, and those residing in Copenhagen. But because these groups of taxpayers had an above-average likelihood of claiming a deduction in the pre-reform period, the share of unclaimed deductions under the self-reporting regime is unlikely to have been particularly large for these groups of taxpayers.

3.5   Charitable Giving Propensity

To this point, we have not addressed the possibility that the policy change coincided with— or caused—a change in actual giving behavior, rather than the reporting propensity. One possibility is that the introduction of information reporting and pre-population of charitable deductions reduced the compliance cost for taxpayers, and so the effective cost of charitable giving, by enough to induce an increase in actual donations. To investigate whether there was a change in giving propensity coinciding with the policy change in 2008, we make use of annual administrative reports received by SKAT from charities eligible to collect tax-favored contributions. These filings are required in order for charities to maintain their tax-favored

status, and contain, among other information, reports on the total value of contributions received and the number of contributing members (donations) for each charity. These data correspond to donations that, provided they were of at least DKK500, qualify for a charitable tax deduction.

Given our main finding—that there was a surge in the *number*, but less so in the *value*, of charitable tax deductions following the policy reform—we first investigate whether there was any change in the number of contributing members reported by charities following the policy change. We restrict our attention here to the ten largest charities, measured by the *number* of information reports received by SKAT over the period 2008-2011. These ten charities together represent about 60 percent of the information reports received from all charities. We exclude small charities to avoid our findings being influenced by potentially misleading reporting behavior of some small charities: a few small organizations did not file reports in each year, and, in some circumstances, reported implausible year-to-year changes in their number of donors. The line labeled "Tax Return Data" in Figure 3.7 shows the number of information reports received (aggregated by charity for each taxpayer) from the top-ten charities for the period 2008-2011 (the information reporting period), and the line labeled "Charity Data" in Figure 3.7 reports the number of charitable donors reported by these top-ten charities for the period 2001-2011.[15] The number of donors reported by these charities exceeds the number of information reports received by SKAT from these organizations, most probably because some charitable donors do not provide their tax identification together with their gift; for gifts less than DKK500 this is not surprising: they do not result in a tax deduction. A few other factors are likely to contribute to the divergence between these series: transfer of funds via cell phone SMS (short message service) has become widespread in Denmark for popular giving campaigns, for which donations appear in charity records, but not tax records; "tin rattling" and church day donations are collected without tax identification; and some taxpayers may prefer to give anonymously. Between 2007 and 2008, when information reporting and pre-population was introduced, the number of charitable tax deductions claimed doubled, but, as

---

[15]All results that follow are qualitatively the same if we consider instead the 25 largest charities, measured by the number of information reports received from each charity (per taxpayer) over the period 2008-2011.

Figure 3.7 shows, the number of donations received by large charities was almost unchanged. This is consistent with the notion that the surge in the *number* of tax deductions claimed in 2008 was due to a change in reporting behavior, not actual giving behavior.

We are further persuaded that the policy change affected reporting but not giving behavior by the fact that there was no apparent change in the trend value of donations collected before and after the policy change. Mirroring Figure 3.7, the line labeled "Tax Return Data" in Figure 3.8 shows the total value of charitable contributions reported on information reports sent to SKAT by the top-ten charities (with charity size measured by the number of donors, as above), and the line labeled "Charity Data" shows the total value of donations collected by the top-ten charities for each year 2001-2011. Apart from the spike in donations in 2005 (see Figure 3.8), most likely due to giving campaigns following the Indian Ocean tsunami in December 2004, growth in the total value of donations has been stable. The fraction of total donations reported to SKAT via information reports has also been stable over the information reporting period 2008-2011. Given that was almost no change in the *number* of donations made pre- and post- reform, the data in Figure 3.8 indicate that there was no intensive margin giving response coinciding with the policy change either.

Supporting our claim that the reform did not affect giving behavior, there was little difference in the growth rate of mean charitable deductions in the post-reform period between taxpayers who claimed a deduction in the pre-reform period and those who claimed for the first time in 2008. For the group of taxpayers who claimed a charitable tax deduction in 2008 (the first year of the reform), but not in either of 2006 or 2007 (the pre-reform period), growth in mean contributions over the period 2008-2011 averaged 2.2 percent, only slightly more than the 0.8 percent average growth rate for the group of taxpayers who claimed a charitable deduction in 2008 and in at least one of the two pre-reform years 2006 or 2007.[16]

Having established that there was no meaningful change in giving propensity around the time of the policy change, we attribute the surge in charitable tax deductions claimed between 2007 and 2008 to a change in reporting behavior. Before 2008, many taxpayers appear to have

---

[16]The calculation includes those who did not claim a charitable deduction in some years 2009-2011, for both groups. We also restrict the sample to those taxpayers who filed a return in each year 2006-2011.

neglected to report their tax deductible charitable contributions, but since 2008 information reports have been used to automatically credit charitable deductions on taxpayers' behalf. Recall that the randomized audit experiment found only negligible amounts of charitable overclaiming.

3.6   Awareness of Giving Incentives

Our finding of substantial underclaiming of eligible charitable tax benefits points to the existence of pervasive frictions affecting reporting behavior. One potential friction is a lack of awareness of the tax incentives created by charitable giving. We investigate this further by examining an aspect of Denmark's charitable giving rules, in existence before 2012, that created a region of strictly dominated giving choices.

We begin by formally describing the incentives created by the pre-2012 regime, under which only total annual gifts per charity of DKK500 or more were eligible to tax deductibility, and in calculating the total amount of eligible tax deductions, the first DKK500 in contributions was excluded. Supposing taxpayer i can donate to N charities eligible for regular charitable deductions, the amount of their total charitable deductions, up to a maximum of 14,500, is given by

$$S_i = max \left\{ \left[\sum_{n=1}^{N} g_{i,n} 1 \left(g_{i,n} \geq 500 \right) - 500 \right], 0 \right\}, \tag{3.6.1}$$

where $g_{i,n}$ is taxpayer i's total annual gifts to charity n, and $1\left(\cdot\right)$ is an indicator function taking the value one for gifts of DKK500 or more. The amount of tax benefits received is the tax deductible amount multiplied by the one-third subsidy rate.[17] The examples provided in Table 3.5 are provided in order to help clarify this formula. For simplicity, we assume there are $N = 3$ charities in this example. Taxpayer A's gift is less than DKK500, so she receives no tax deductions for her charitable contributions. Taxpayer B makes one gift of

---

[17]In the text we refer to a one-third subsidy rate for simplicity, but there is slight variation based on the taxpayer's place of residence.

DKK700, exceeding the DKK500 threshold, and so is eligible to receive tax preferences for this gift, but because the first DKK500 in gifts receives no tax benefit she has only DKK200 in charitable tax deductions. Taxpayer C is eligible to receive tax preferences on both her gifts of DKK500, and receives a total tax deduction of DKK500, after taking the exemption limit into account. Even though taxpayer D gave an additional DKK400 to charity number three compared to taxpayer C, and has given more than DKK500 in total, she receives no more tax deductions than taxpayer C because her gift to charity number three is less than DKK500.

For a taxpayer contemplating a gift to a single charity, the S_i function reduces to a kinked subsidy scheme with a DKK500 threshold. But once a taxpayer has made at least one charitable gift of DKK500 or more they face a *notched* subsidy for gifts to all other charities. The first gift meets the DKK500 exemption threshold, so all subsequent gifts to other charities are eligible for full tax deductibility if each gift is DKK500 or more. Suppose that a taxpayer's largest gift is $g_1 \geq 500$, Figure 3.9 shows the budget set facing the taxpayer for all *subsequent* gifts in the current tax year. Any second or subsequent gift to the value of $g \, \epsilon \, (\underline{g}, \overline{g})$ is strictly dominated because a gift of $\overline{g} = 500$ affords a higher level of charitable contributions at no, or less, cost to the taxpayer. With the tax subsidy rate $\tau = \frac{1}{3}$ and $\overline{g} = 500$ then the lower limit on the strictly dominated region is $\underline{g} = \overline{g} \, (1 - \tau) = DKK333$.

To illustrate the incentives created by this notched subsidy scheme with an example, consider taxpayer D in Table 3.5, whose gift of DKK400 to charity number three is a dominated choice: either of her first two gifts meets the DKK500 exemption threshold, so each subsequent gift is eligible for tax deductibility provided it is to the value DKK500 or more. If she raised her donation to charity number three by DKK100 to DKK500, this gift would be eligible for tax deductibility, giving her a tax saving of DKK166 (given the one-third subsidy rate), leaving her with DKK66 more in after-tax income (plus any utility gain from higher charitable contributions).

Fortunately, under the information reporting regime charities report to the tax authority all gifts above *and below* the DKK500 eligibility threshold for each taxpayer, allowing us to

investigate taxpayer awareness of the incentives created by the kinked-and-notched subsidy scheme. Figure 3.10 plots the number of charitable gifts made in 2011 by claim size for taxpayers with a maximum gift of DKK500 or more. The distribution for the years 2008-2010 is similar to the distribution shown in Figure 3.10 for 2011. All of these taxpayers face the budget set shown by Figure 3.9: each second or subsequent gift qualifies for full tax deductibility if it is DKK500 or more. The black bars in Figure 3.10 indicate the number of gifts made in the strictly dominated region. Only a few taxpayers made more than one dominated giving choice, so almost all these observations represent unique taxpayers; in total, 11,624 taxpayers made a gift in the strictly dominated region in 2011. There is a clear mass point at DKK500, at the upper limit of the notch, suggesting that many taxpayers understood the budget set created by the subsidy scheme, and were induced to raise their donations to DKK500. As a share of all taxpayers claiming a charitable deduction, only about 2 to 3 percent of taxpayers made strictly dominated giving choices in each year 2008-2011. However, the number of gifts in the dominated region DKK333-500 in 2011 was about one-quarter the number in the range DKK500-666, and a little less in earlier years.

A clustering of donations in DKK100 multiples is evident, with the mass point at DKK600 even larger than that at DKK500. Because many taxpayers make gifts via automatic deduction on a monthly basis, we conjecture that the DKK600 mass point corresponds to taxpayers choosing an integer DKK50 per month charitable deduction: DKK50 is the smallest multiple of 10 that results in annual contributions qualifying for a subsidy, suggesting that the location of this mass point is influenced by the notch.

The economic significance of these dominated giving choices depends on the frequency with which individual taxpayers make such errors. Making a dominated choice in any one year results in a relatively small loss, and a taxpayer may make a mistake in any given year for idiosyncratic reasons. But for taxpayers making repeated mistakes, the cost may cumulate to a substantial amount, providing perhaps more persuasive evidence of ignorance of tax incentives for giving. To examine the frequency of dominated giving choices, Table 3.6 reports, for the data sample available 2008-2011, the number of taxpayers who made dominated choices

in each given and subsequent year. For example, in 2008 5,927 taxpayers made a dominated choice, and of those 2,050 also made a dominated choice in 2009; 1,878 made a dominated choice in each year 2008-2010, and so on. For each year on the diagonal, about one-third of the taxpayers making a dominated choice do so again the following year. And of those taxpayers making a dominated choice in 2008, about 25 percent made a dominated choice in each of the next three years.

Taken together, these results provide evidence that a sizable minority of taxpayers did not understand the complex giving incentives created by the notched subsidy scheme in place before 2012. A non-trivial fraction of those making dominated choices did so repeatedly. However, a majority many taxpayers made giving choices just above the dominated region, indicating a high degree of awareness of the complex giving incentives in place before 2012.

3.7   Optimal Reporting Regime

Was the rise in the value of charitable tax deductions claimed as a result of the reform—given no increase in actual charitable giving— socially desirable or undesirable? On the one hand, between 2007 and 2008 (before and after the policy change), there was an increase in charitable tax expenditures of about DKK35m. However, the reduction in net revenue collected alone, even conditional on no change in giving behavior, tells us nothing about the welfare implications of the reform. The reduction in net revenue represents a transfer among citizens, not lost resources. Given an unchanged government revenue requirement, there is a social cost if the revenue shortfall must be raised by reliance on a distortionary tax instrument. On the other hand, the introduction of information reporting and pre-population is likely to have reduced compliance costs borne by taxpayers.

To examine these arguments more closely, we next outline a simple normative model, drawing on Mayshar (1991), that can be used to evaluate the social welfare implications of moving from a self-reporting regime for charitable gifts to an information reporting and pre-population regime. The model is written with reference to charitable deductions but, with some minor

relabeling, it could be used to inform the choice of reporting regime for a variety of other tax deduction line items.

We consider a social planner choosing tax-system parameters affecting charitable giving and reporting behavior to maximize social welfare. The chosen tax system parameters are summarized by the policy vector $\theta \, \epsilon \, \{R, \tau\}$, where R is the reporting regime, and τ is the subsidy rate on eligible charitable contributions. We let $R_{i,p}$ denote an information reporting and pre-population of deductions regime for charitable gifts, and R_s a self-reporting regime.

Taxpayer i is assumed to derive warm-glow utility from their own charitable gifts $g_i(\theta)$ that depend on the tax-system policy θ, the total amount of public goods funded by charitable gifts $G(\theta) = \int g_i(\theta) \, di$, and consumption of all other goods. For simplicity we assume quasi-linear utility, with consumption of all other goods the numeraire. This assumption is not limiting because most taxpayers make only small charitable contributions, in which case income effects due to changes in charitable giving are not important. There is an exogenous government revenue requirement unaffected by variation in charitable giving tax system parameters, and all charitable tax-expenditures and administrative costs must be funded using a distortionary tax instrument.[18] These assumptions allow us to specify the following simple money-metric utilitarian social welfare function given tax policy θ:

$$W(\theta) = \alpha + bG(\theta) - m(\theta) - (MECF - 1) TE(\theta) - (MECF) A(\theta), \qquad (3.7.1)$$

where α is the component of social welfare unrelated to charitable giving and unaffected by the charitable giving policy vector θ, b is the constant marginal social value of all charitable contributions made, $m(\theta)$ is the dollar cost borne by taxpayers in complying with tax-system θ, $TE(\theta)$ are tax expenditures on charitable tax deductions claimed, $A(\theta)$ are administrative costs, and $MECF$ is the marginal efficiency cost of funds for the tax instrument used to raise revenue to fund charitable tax expenditures and administrative costs. The marginal social

---

[18]We exclude the possibility of levying lump-sum taxes by assumption as in, for example, the Ramsey optimal commodity tax literature. More generally, administrative and compliance cost considerations can endogenize the absence of lump-sum tax instruments, even where there are no distributional concerns (see, for example, Yitzhaki 1979).

value of charitable contributions, b, captures both private (warm-glow) and public utility gained from charitable giving. Tax expenditures represent a transfer between households, so the social cost of raising revenue to fund charitable deductions comprises only the efficiency cost of raising that revenue: $(MECF - 1) TE(\theta)$. In contrast, administrative expenses have a resource cost, and reduce our money-metric social welfare function by $(MECF) A(\theta)$. The term $m(\theta)$ comprises compliance costs borne by individual taxpayers directly and by charities on behalf of taxpayers; this is important for the policy change we study because part of the reduction in compliance costs from the introduction of information reporting may have been shifted onto charities.

Some taxpayer may engage in evasion by overclaiming charitable tax deductions. Total charitable tax expenditures are equal to:

$$TE(\theta) \equiv \tau G^r(\theta), \qquad (3.7.2)$$

where $G^r(\theta)$ is the value of all *reported* deductions, and τ the subsidy rate.

Letting $\Delta X(\theta_1, \theta_0) \equiv X(\theta_1) - X(\theta_0)$ represent the change in variable X due to a discrete change in the tax-system policy vector θ, the welfare effect of moving from a self-reporting regime to an information reporting and pre-population regime for charitable deductions is equal to:

$$\begin{aligned} \Delta W(\theta_{i,p}, \theta_s) = \; & b\Delta G(\theta_{i,p}, \theta_s) - \Delta m(\theta_{i,p}, \theta_s) - (MECF - 1)\Delta TE(\theta_{i,p}, \theta_s) \\ & - (MECF)\Delta A(\theta_{i,p}, \theta_s), \end{aligned} \qquad (3.7.3)$$

where $\theta_{i,p} = \{R_{i,p}, \tau\}$ and $\theta_s = \{R_s, \tau\}$. Introducing an information reporting and pre-population regime increases social welfare if $\Delta W(\theta_{i,p}, \theta_s) > 0$. By rearrangement of Equation (3.7.3), it can be seen that welfare rises if:

$$b\Delta G(\theta_{i,p}, \theta_s) - \Delta m(\theta_{i,p}, \theta_s) \geq (MECF - 1)\Delta TE(\theta_{i,p}, \theta_s) + (MECF)\Delta A(\theta_{i,p}, \theta_s). \quad (3.7.4)$$

The first term on the left-hand-side of Equation (3.7.4) is the social benefit from any increase in charitable giving, and the second term is the social benefit from any reduction in compliance costs. On the right-hand-side, the first term measures the increase in the social cost of transferring income among taxpayers via charitable tax expenditures, and the second term measures the social cost of an increase in administrative expenses. Administrative costs receive a higher social weight than compliance costs because they are funded from after-tax revenue raised using distortionary taxes.

We find no evidence of an increase in charitable giving accompanying the introduction of information reporting and pre-population of deductions in 2008, in which case $\Delta G\left(\theta_{i,p}, \theta_s\right) \simeq$ 0. Hence, the introduction of information reporting and pre-population of deductions is socially desirable only if the reduction in compliance costs for taxpayers is greater than the appropriately weighted sum of any additional administration cost and the net social cost of funding increased transfers among taxpayers via charitable tax deductions.

Denmark's tax system has long featured widespread information reporting, so the marginal increase in administrative costs from extending its use to charitable giving is likely to be very small. Similarly, any reduction in audit costs from the use of information reporting for charitable tax deductions is likely to be negligible, relative to the other terms in Equation (3.7.4). Less than one percent of tax returns are routinely examined by an auditor, and the fraction of all taxpayers reporting any charitable tax deductions under the self-reporting regime was only about 3.5 percent. Compared to some sources of income line items, auditing charitable tax deductions is not particularly costly for the tax authority, requiring only cross-checking of taxpayer provided receipts against reported deductions. Together, these considerations imply that any change in administrative costs due to the policy change is an order of magnitude smaller than the social benefit of reduced compliance costs and the social cost of increased charitable tax expenditures. Hence, our framework indicates that the policy change increased social welfare iff $-\frac{\Delta m}{\Delta TE} \gtrsim (MECF - 1)$.

The $MECF$ has been estimated in numerous empirical studies (see, for example, Ballard et al., 1985, Gruber and Saez, 2002) and we have estimated the change in charitable tax

expenditures due to the reform, but we must estimate the change in compliance costs. One measure of the reduction in compliance costs borne by taxpayers due to the policy change is the average value of forgone tax deductions under the self-reporting regime. This approach, used by Pitt and Slemrod (1989) to estimate the magnitude of compliance costs for the U.S. income tax system, provides a valid estimate under the assumptions that i) taxpayers correctly estimate their cost of reporting charitable tax deductions; ii) the magnitude of compliance costs is unrelated to the size of a taxpayer's eligible charitable deduction, and iii) taxpayers only fail to report their deductions when it is privately optimal to do so (when compliance costs exceed tax benefits forgone). Under these assumptions, the *average* value of forgone tax deductions provides an estimate of the *average* compliance cost borne by those who did claim their charitable tax benefits under the self-reporting regime: in section 3.4, we estimated the average value of forgone tax deductions to be DKK262 per taxpayer. Were there no increase in compliance costs borne by charities, then the aggregate reduction in compliance costs is estimated to be $-\Delta m = DKK262 \times 150,311 \simeq DKK39.4m$ (the average value of compliance costs multiplied by the number of taxpayers claiming a deduction under the self-reporting regime in 2007). The estimated increase in the value of tax expenditures due to the reform is the increase in the after-tax value of tax deductions claimed between 2007 and 2008: $\Delta TE \simeq DKK35m$. Hence, $-\frac{\Delta m}{\Delta TE} \simeq 1.1$ and the optimality condition $-\frac{\Delta m}{\Delta TE} \gtrsim (MECF - 1)$ is easily satisfied for standard estimates of the $MECF$, indicating that the policy change was desirable under these assumptions.

Some of the reduction in compliance costs directly borne by taxpayers was undoubtedly shifted to charities. Nevertheless, the shift in compliance cost from taxpayers to charities is likely to have resulted in a reduction in overall compliance costs because of charities' expertise compared to taxpayers, and any economies of scale, in record keeping. Any marginal increase in compliance costs imposed on charities by the information reporting requirement is likely to be small because charities routinely record the names of donors and the size of their gifts to aid their own fundraising efforts and, in any case, charities eligible to receive tax-favored contributions had to maintain similar records prior to the policy change in order to comply

with SKAT's annual reporting requirements. Perhaps more problematically, the value of forgone deductions may overstate the magnitude of compliance costs borne by taxpayers under the self-reporting regime: taxpayers may fail to claim eligible deductions for reasons other than a rational cost-benefit calculation, such as inattention, which do not provide evidence on the magnitude of compliance costs for those who did claim charitable deductions under the self-reporting regime.

3.8   Conclusion

This paper provides evidence of substantial underclaiming of charitable tax deductions under the self-reporting regime that existed in Denmark before 2008; the introduction of information reporting and pre-population of charitable deductions coincided with a doubling in the number of deductions claimed. We estimate the after-tax value of unclaimed charitable tax deductions to have been about DKK262 per taxpayer per-year, but that the total value of forgone benefits to be larger because many taxpayers systematically did not claim their eligible deductions under the self-reporting regime. We document that there was negligible evasion under the self-reporting regime, and that there was no change in giving behavior at the time of the reform. Most taxpayers making multiple charitable gifts appear to have understood the giving incentives created by the notched subsidy scheme in place before 2012, but a still sizable minority made dominated giving choices, in some cases repeatedly. Our results caution researchers using tax return data to measure real behavioral response to be aware of simultaneous (and possibly endogenous to the behavioral response) changes in reporting behavior; we have demonstrated that this is an important concern for low-value tax deductions.

For tax administrators, perhaps the most surprising finding is that the introduction of information reporting for a tax deduction line item can result in a loss in revenue—unlike sources of income line items, for which information reporting has proven to be very successful at limiting evasion opportunities and thus raising revenue collections (see Kleven et al., 2011,

and Slemrod, 2007). Nevertheless, even if it generates a revenue loss, such a reform may be socially desirable, even if it has no effect on charitable giving; we have laid out a simple modeling framework that shows how to appropriately weight the social cost of changes in compliance and administrative cost due to such tax reporting reform.

Table 3.1: Number of Charitable Organizations

	Approved Organizations	Reporting Organizations
2000	695	579
2001	716	642
2002	752	682
2003	772	704
2004	778	702
2005	792	715
2006	756	698
2007	790	736
2008	796	743
2009	813	780
2010	817	782
2011	833	809

Notes: *Approved Organizations* refers to the number of organizations SKAT recognizes as eligible to receive tax deductible charitable gifts. *Reporting Organizations* refers to the subset that made an annual declaration to SKAT in each year.

Table 3.2: Taxpayers Reporting a Charitable Deduction: Summary Statistics

	2006	2007	2008	2009	2010	2011
Number of Deductions	162,983	150,311	300,122	325,525	365,167	388,976
Regular gifts			270,826	294,912	336,571	360,527
Long-Term Contracts			44,381	46,069	44,676	44,399
Mean Value (DKK)	4,033	4,671	2,697	2,689	2,650	2,593
Regular gifts			2,026	2,071	2,098	2,074
Long-Term Contracts			6,009	5,740	5,850	5,879
Median Value (DKK)	2,400	2,470	1,500	1,500	1,500	1,375
Regular gifts			1,350	1,375	1,300	1,280
Long-Term Contracts			2,526	2,400	2,500	2,500
Total Value (DKK, '000)	657,310	702,103	809,429	875,337	967,693	1,008,615
Regular gifts			548,693	610,763	706,126	747,733
Long-Term Contracts			260,676	264,436	261,355	261,010

Notes: *Number of Deductions* is the number of taxpayers reporting a charitable deduction in each year shown. For 2008 and after, charitable gifts were reported in two categories. Information reporting and pre-population of deductions was introduced in 2008 for both regular and long-term gifts. The total number of taxpayers claiming a charitable tax deduction in each year is less than the sum of the two groups because some taxpayers claimed deductions in both categories.

Table 3.3: Number of Tax Deductible Claims: By Claim Size

Claim Size (DKK)	2006	2007	2008	2009	2010	2011
0-500	8,931	7,356	80,170	88,181	103,558	110,701
501-1,500	34,468	30,276	71,103	75,896	83,344	93,948
1,501-3,000	60,536	56,085	89,407	96,082	103,297	105,273
3,001-5,000	24,379	21,931	25,260	27,547	32,122	34,467
5,001-10,000	25,434	19,084	18,027	20,482	23,838	25,123
> 10,000	9,235	15,579	16,155	17,337	19,008	19,464
Mean	4,033	4,671	2,697	2,689	2,650	2,593
Median	2,400	2,470	1,500	1,500	1,500	1,375

Notes: Claim size is the amount of tax deductions received. Information reporting and pre-population was introduced in 2008.

Table 3.4: Charitable Tax Deductions Claimed: 2008

	Number
Total deductions claimed in 2008	300,122
Filed a return 2006-2011	293,134
No deduction 2006-2007	152,857
3 deductions 2009-2011	104,197
2 deductions 2009-2011	15,571
1 deductions 2009-2011	13,961
0 deductions 2009-2011	19,128

Notes: *Filed a return* is the number of taxpayers who claimed a charitable tax deduction in 2008 and filed a tax return in each year 2006-2011. *No deduction 2006-2007* is the subset who did not claim a charitable tax deduction in 2006 or 2007. The *No deduction 2006-2007* group is split into four mutually exclusive groups according to the number of charitable tax deductions claimed in the years 2009-2011.

Table 3.5: Tax Value of Regular Gifts

Taxpayer	Charity			Tax Deductible
	1	2	3	Amount
A	400	0	0	0
B	0	700	0	200
C	500	500	0	500
D	500	500	400	500

Notes: This table shows the amount of regular tax deductions received by four hypothetical taxpayers. Only annual gifts of DKK500 or more per charity qualified for a tax deduction before 2012, and the first DKK500 in total gifts is excluded in calculating the total value of regular tax deductions. The value of charitable deductions is equal to the deductible amount multiplied by the one-third subsidy rate.

Table 3.6: Dominated Giving Choices

|	2008	2009	2010	2011
2008	5,927	2,050	1,878	1,480
2009		7,350	2,421	1,925
2010			9,743	3,168
2011				11,624
Total	270,826	294,912	336,571	360,527

Notes: The diagonal elements report the number of taxpayers making a dominated giving choice in that year; the off-diagonal elements report the number of those taxpayers who made a dominated giving choice in each subsequent year. For example, 5,927 taxpayers made a dominated choice in 2008, and of those 1,878 also made a dominated choice in 2009 and 2010. *Total* is the number of taxpayers claiming a regular charitable tax deduction in each year.

Figure 3.1: Number and Average Value of Charitable Deductions Claimed
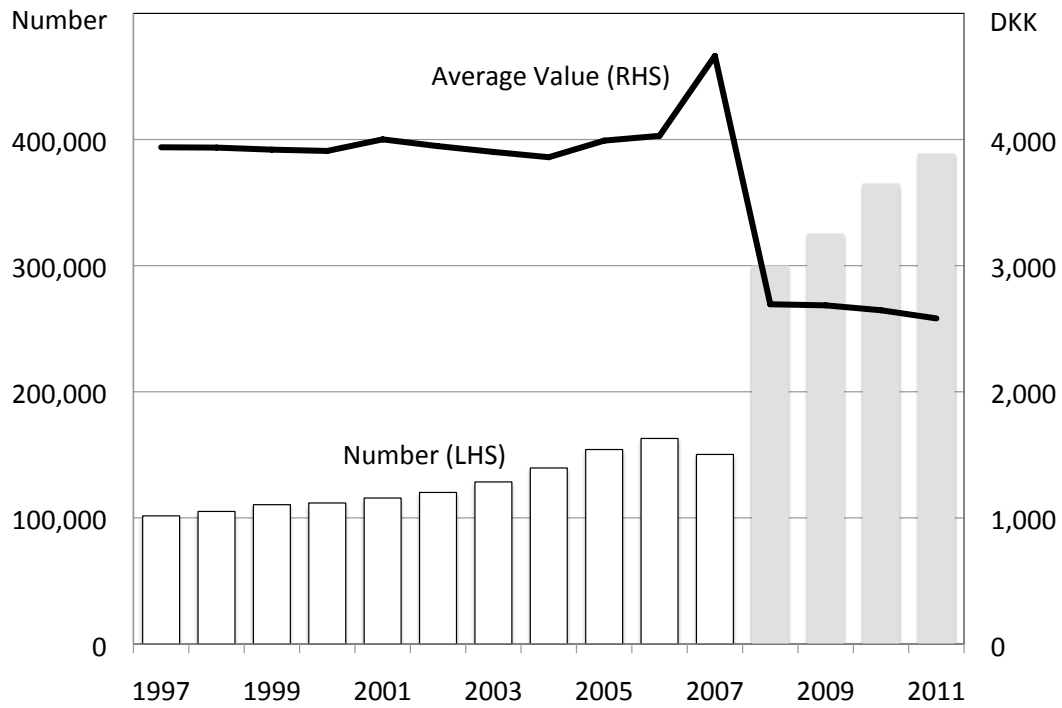
Notes: The columns in this figure show the number of taxpayers reporting a charitable deduction for the years 1997-2011, on the left-hand scale. The line shows the mean value of tax deductions claimed, on the right-hand scale. The shaded columns are for years in which there was information reporting and pre-population of deductions for regular and long-term gifts.

Figure 3.2: Number of Tax-Deductible Claims: By Claim Size and Year
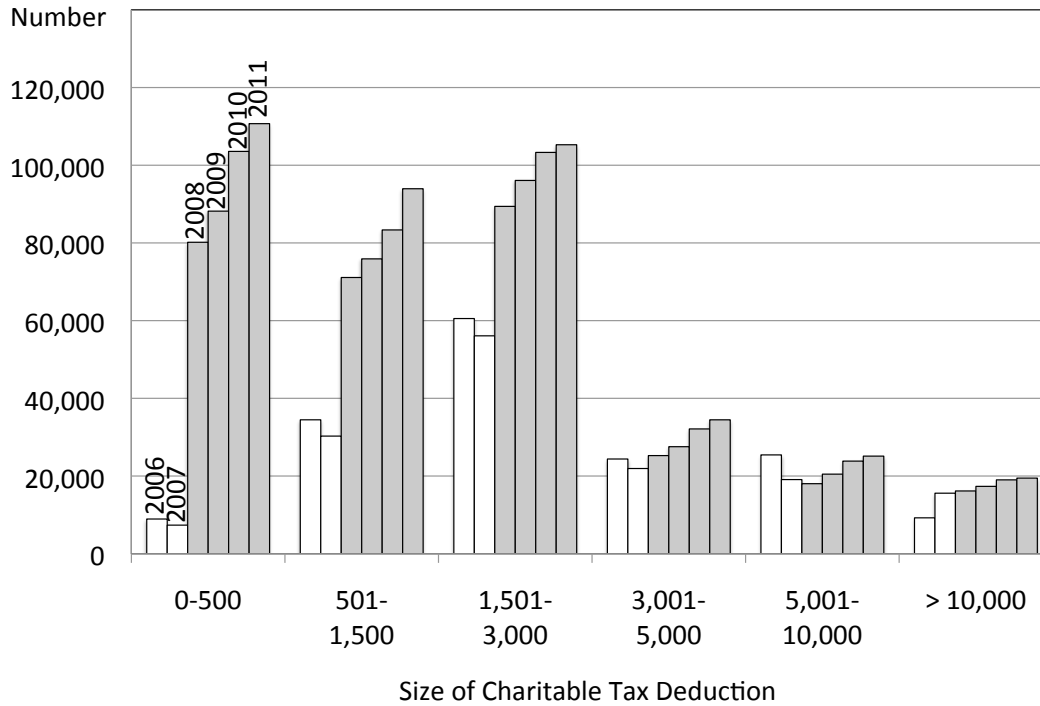
Notes: This figure shows the number of taxpayers reporting a charitable deduction for the years 2006-2011, by size of reported tax deduction. The claim size on the x-axis is the amount of tax deduction claimed, not the total value of charitable gifts made. Years for which there was information reporting and pre-population of deductions for regular and long-term gifts correspond to the shaded bars.

Figure 3.3: Distribution of Tax Deductions Claimed
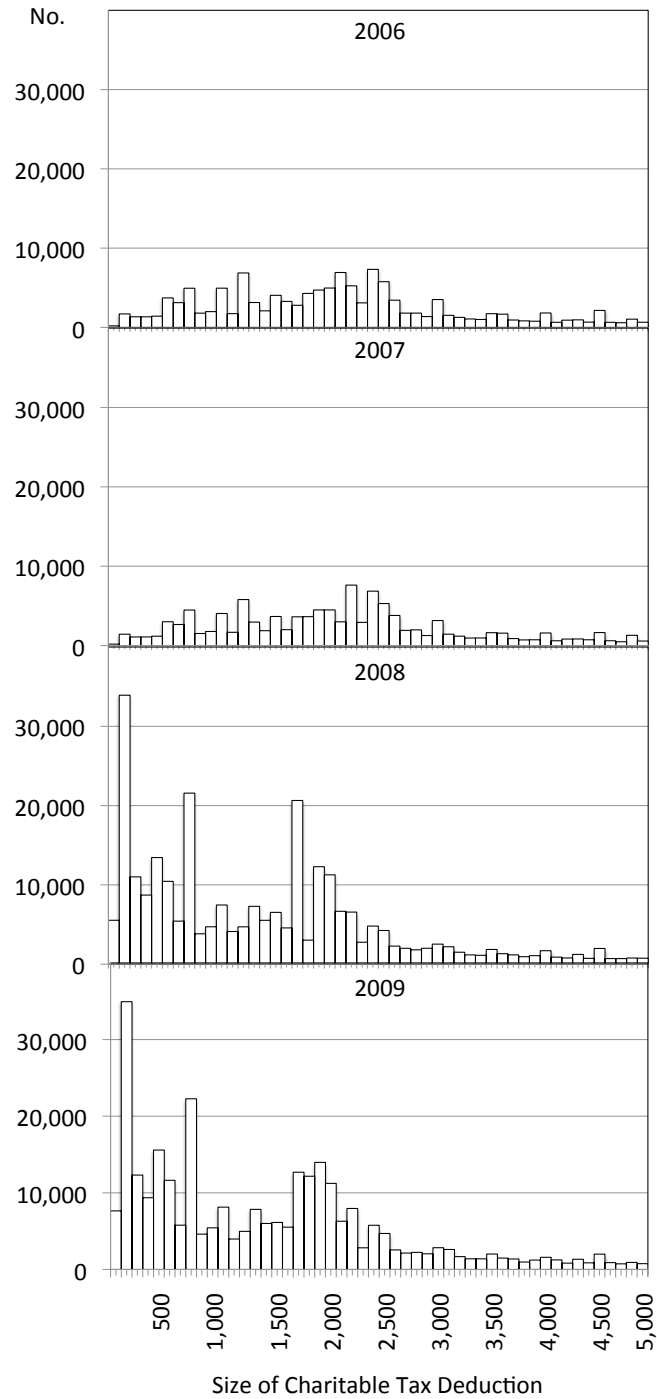
Notes: This figure shows the distribution of tax deductions claimed for the years 2006-2009. Information reporting and pre-population for regular and long-term charitable gifts was introduced in 2008.

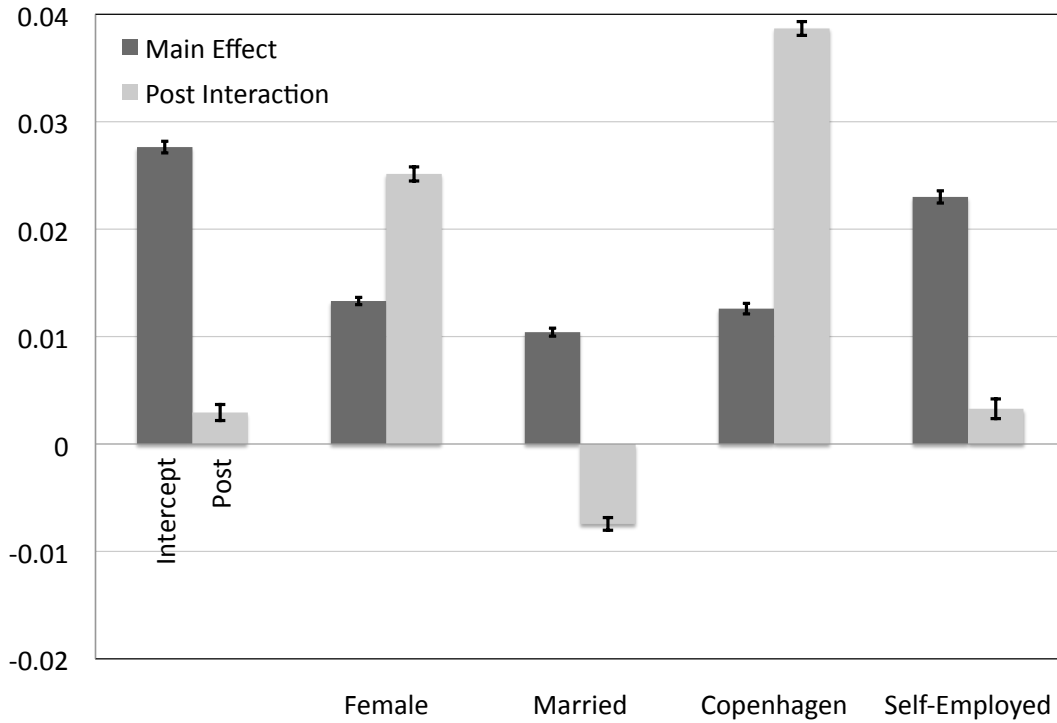Figure 3.4: Regression Parameter Estimates: Income Percentile
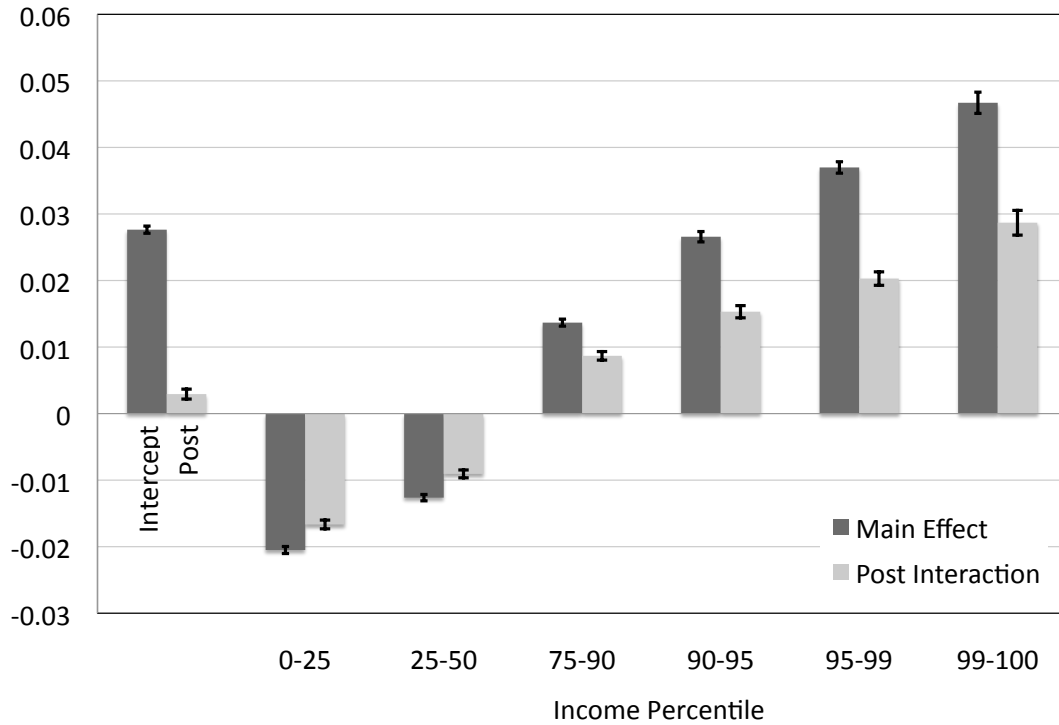
Notes: This figure reports OLS parameter estimates for the regression specification shown by Equation (3.4.1). The error bars show a 95 percent confidence interval for each parameter estimate. The intercept term indicates the probability that a taxpayer with the baseline set of characteristics claimed a charitable tax deduction in the pre-reform period: the baseline set of characteristics is a male taxpayer aged 46-65, in 50-75th income percentile, single, residing outside Copenhagen, and not self-employed.

Figure 3.5: Regression Parameter Estimates: Age
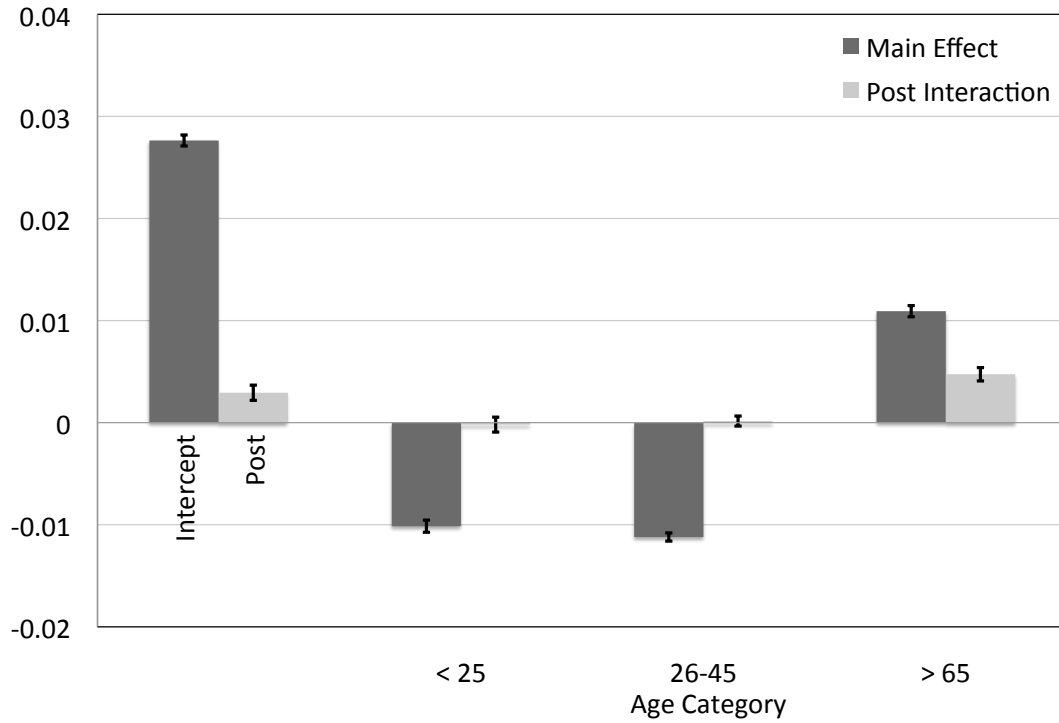
Notes: This figure reports OLS parameter estimates for the regression specification shown by Equation (3.4.1). The error bars show a 95 percent confidence interval for each parameter estimate. The intercept term indicates the probability that a taxpayer with the baseline set of characteristics claimed a charitable tax deduction in the pre-reform period: the baseline set of characteristics is a male taxpayer aged 46-65, in 50-75th income percentile, single, residing outside Copenhagen, and not self-employed.

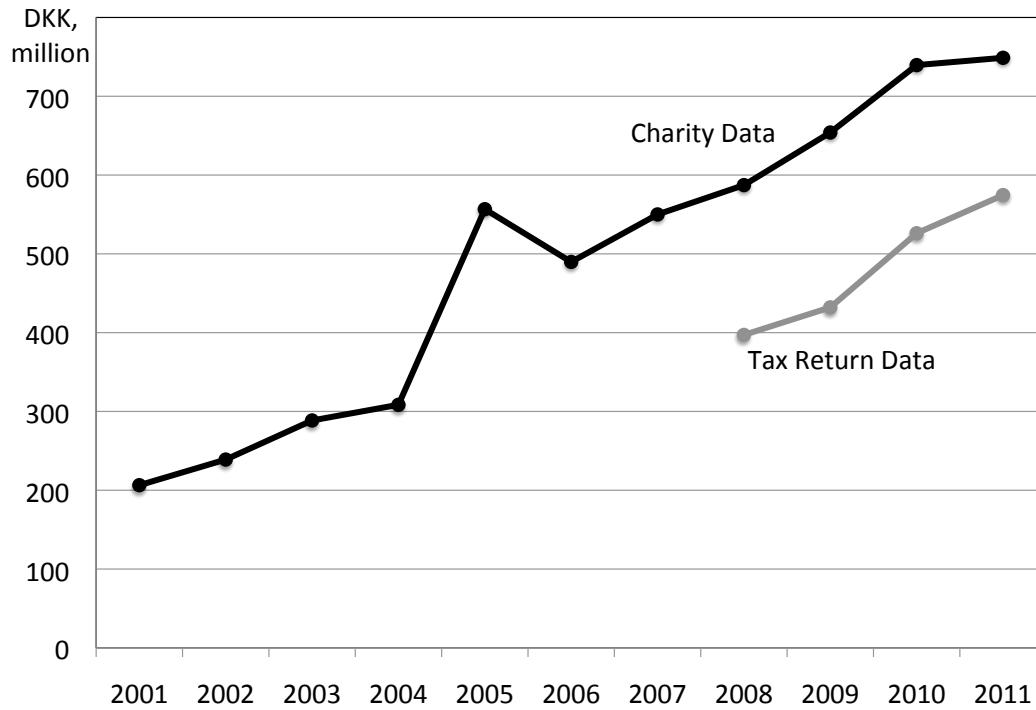Figure 3.6: Regression Parameter Estimates: Selected Characteristics

Notes: This figure reports OLS parameter estimates for the regression specification shown by Equation (3.4.1). The error bars show a 95 percent confidence interval for each parameter estimate. The intercept term indicates the probability that a taxpayer with the baseline set of characteristics claimed a charitable tax deduction in the pre-reform period: the baseline set of characteristics is a male taxpayer aged 46-65, in 50-75th income percentile, single, residing outside Copenhagen, and not self-employed.

Figure 3.7: Number of Charitable Donations: Ten Largest Charities
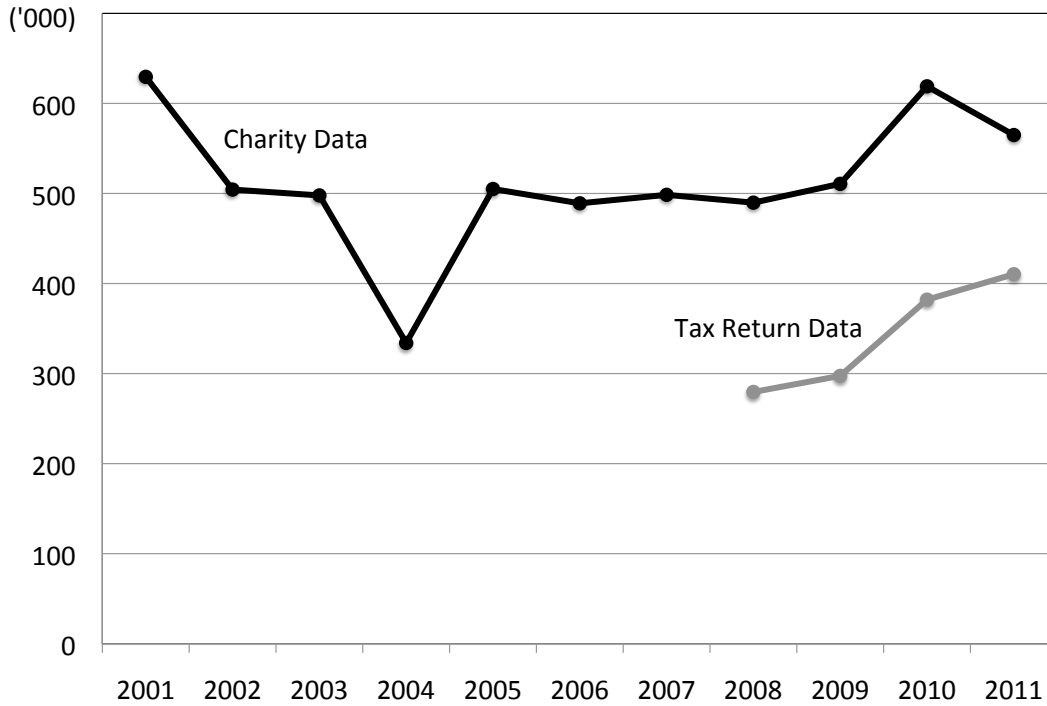

Notes: The *Tax Return Data* line indicates the total number of information reports received by SKAT from the 10 largest charities (aggregated by charity for each taxpayer), where charity size is measured by the total number of information reports received by SKAT over the period 2008-2011 (information reporting and pre-population for regular and long-term charitable gifts was introduced in 2008). The *Charity Data* line indicates the number of contributing members reported by those 10 charities. The dip in 2004 is due to a sharp drop in the number of donors reported by one large charity. Because there was no accompanying drop in the value of donations reported, we suspect this to be a reporting error.

Figure 3.8: Value of Charitable Donations: Ten Largest Charities

DKK,
million

700

600

500

400

300

200

100

0

2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011

Charity Data

Tax Return Data

Notes: The *Tax Return Data* line indicates the total value of charitable donations contained in information reports received by SKAT from the 10 largest charities, where charity size is measured by the total number of information reports received by SKAT over the period 2008-2011 (information reporting and pre-population for regular and long-term charitable gifts was introduced in 2008). The *Charity Data* line indicates the total value of donations collected by those 10 charities.

Figure 3.9: Notched Budget Set
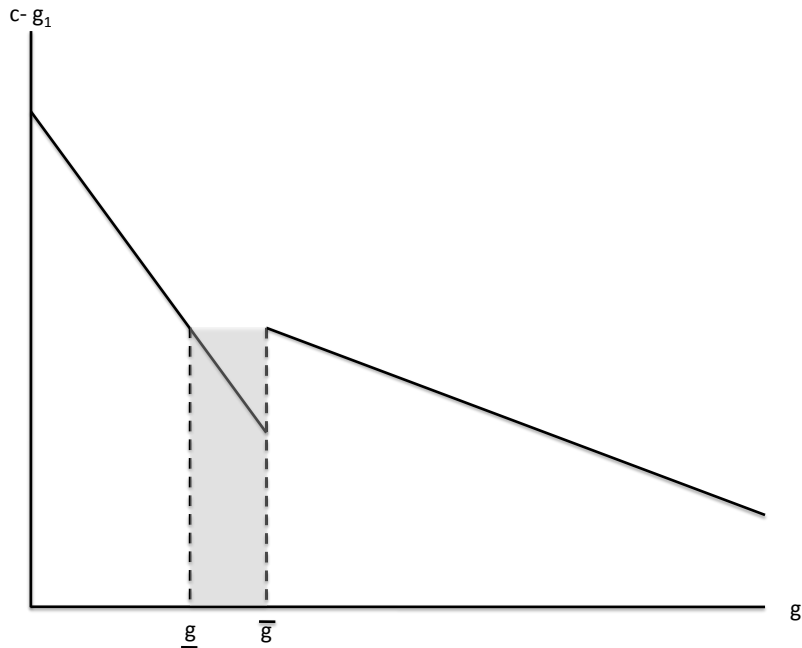
Notes: This figure shows the budget set for regular gifts, for a taxpayer with total annual gifts of DKK500 or more to a particular charity. All subsequent gifts to other charities qualify for tax deductibility provided they are of DKK500 or more per year. Any gift in the shaded region $g \in [\underline{g}, \overline{g}]$ is a strictly dominated choice for a taxpayer because a gift of $\overline{g}$ results in a higher level of charitable contributions and either the same or a higher level of consumption of all other goods. At the one-third subsidy rate, $\overline{g} = 500$ and $\underline{g} = 333$. The y-axis measures consumption on all non-charitable items, less the largest charitable donation in excess of the DKK500 threshold (g_1).

Figure 3.10: Distribution of Charitable Gifts in 2011: Dominated Choices
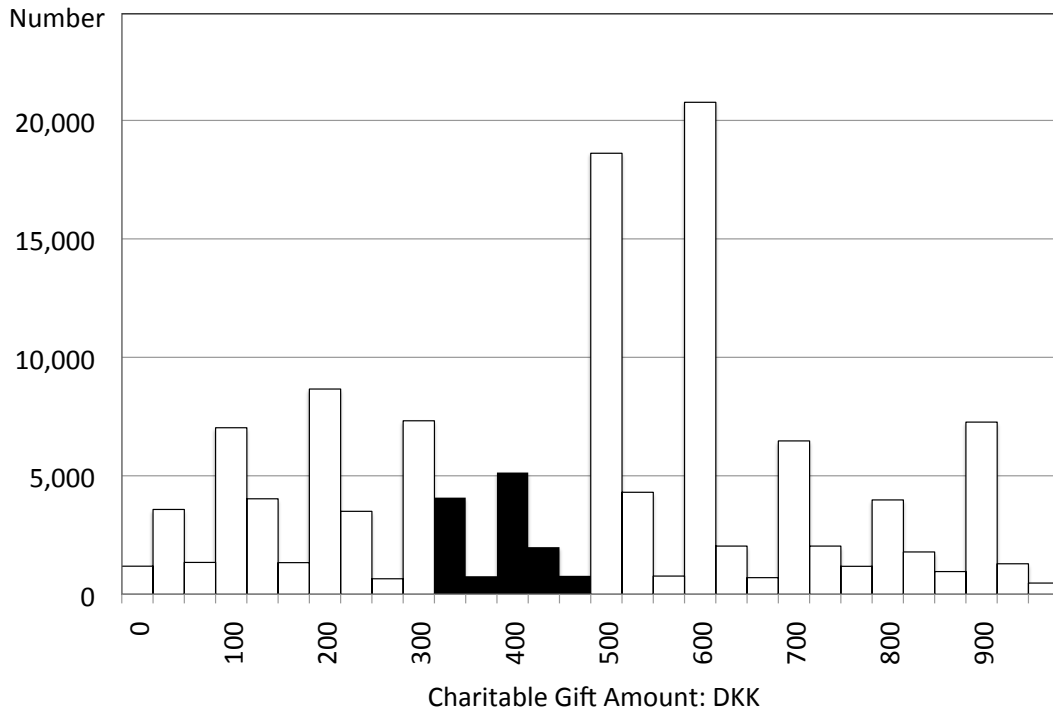
Notes: For the group of taxpayers with a maximum regular gift greater than or equal to DKK500, this figure shows the number of other regular gifts made in 2011 (on the y-axis) by gift amount (on the x-axis). Gift amounts are in bins of DKK33.3, with tick mark labels corresponding to the lower limit of each bin. The solid bars show the number of strictly dominated charitable gift choices made in 2011. A taxpayer makes a strictly dominated choice if they make total annual gifts to at least one charity of DKK500 or more, and any further total annual gifts to other charities of more than DKK333 but less than DKK500. Raising any gift strictly inside the range DKK333-500 to DKK500 affords a higher level of charitable contributions at either no or less cost to the taxpayer. A few taxpayers made more than one strictly dominated choice, each of which is shown in the figure. The distribution is similar for the years 2008-2010 in which data are available.

3.9 Appendix

Figure 3.11: Average Value of Unclaimed Deductions

Largest Unclaimed Charitable Tax Deduction: DKK
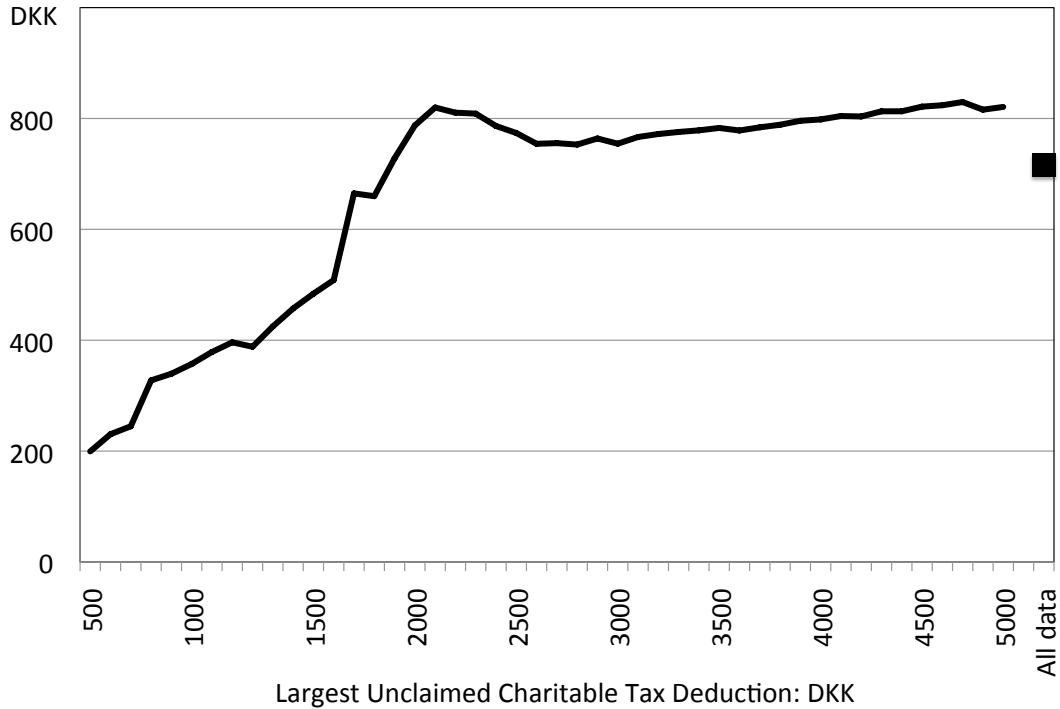
Notes: The black line shows the average value of the change in charitable deductions claimed between 2007 and 2008 for claims having a value no more than the upper limit shown on the x-axis. That is, the mean value (m) of net new contributions between 2007 and 2008 conditional on claimed gifts (g) being no more than x is $(m|g < x) = [(V_{2008}|g < x) - (V_{2007}|g < x)] / [(N_{2008}|g < x) - (N_{2007}|g < x)]$, where $(V_t|g < x)$ is the total value of tax deductions less than x in value claimed in year t, and $(N_t|g < x)$ is the number of tax deductions with a value no more than x claimed in year t. The solid dot sets x to its maximum observed value: $x = x_{max}$.

Table 3.7: Regression Results

Dependent Variable: Claimed a Tax Deduction			
	Coefficient	Std. Error	t-Statistic
Intercept	0.02764	0.00027068	102.11
Female	0.01331	0.00016898	78.77
Married	0.01041	0.00018638	55.86
Copenhagen	0.01260	0.00024514	51.39
Self-Employed	0.02300	0.00028487	80.74
Time	-0.00398	0.00016266	-24.45
Age: <25	-0.01014	0.00029713	-34.14
Age: 26-45	-0.01120	0.00020158	-55.55
Age: >65	0.01092	0.00027426	39.81
Income: 0-25th Percentile	-0.02050	0.00026347	-77.82
Income: 25-50th Percentile	-0.01263	0.00023914	-52.83
Income: 75-90th Percentile	0.01367	0.00026402	51.78
Income: 90-95th Percentile	0.02658	0.00039014	68.13
Income: 95-99th Percentile	0.03699	0.00042990	86.05
Income: Top Percentile	0.04671	0.00079829	58.51
Post	0.00293	0.00037333	7.86
Post × Female	0.02514	0.00020635	121.84
Post × Married	-0.00744	0.00022603	-32.93
Post × Copenhagen	0.03868	0.00029907	129.35
Post × Self-Employed	0.00328	0.00034702	9.46
Post × Time	0.00961	0.00017064	56.30
Post × Age: <25	-0.00019246	0.00036645	-0.53
Post × Age: 26-45	0.00015791	0.00024668	0.64
Post × Age: >65	0.00474	0.00032730	14.48
Post × Income: 0-25th Percentile	-0.01666	0.00032806	-50.78
Post × Income: 25-50th Percentile	-0.00906	0.00029509	-30.71
Post × Income: 75-90th Percentile	0.00868	0.00032131	27.03
Post × Income: 90-95th Percentile	0.01531	0.00045837	33.39
Post × Income: 95-99th Percentile	0.02029	0.00050629	40.07
Post × Income: Top Percentile	0.02868	0.00092872	30.88

Notes: This table reports OLS regression output for Equation (3.4.1). The data consists of the universe of taxpayers (4.37 million) observed over the years 2006-2011. *Time* is a linear time trend, and the R-squared statistic for the regression is 0.0265. The omitted category represents a male taxpayer aged 46-65, in 50-75th income percentile, single, residing outside Copenhagen, and not self-employed. Robust standard errors have been used.

3.10 Bibliography

Ballard, Charles L., John B. Shoven, and John Whalley (1985). General Equilibrium Computations of the Marginal Welfare Costs of Taxes in the United States. *American Economic Review* 75(1): 128–138.

Besley, Timothy and Stephen Coate (1992). Understanding Welfare Stigma: Taxpayer Resentment and Statistical Discrimination. *Journal of Public Economics* 48(2): 165–183.

Bhargava, Saurabh and Dayanand S. Manoli (2011). Why Are Benefits Left on the Table? Assessing the Role of Information, Complexity, and Stigma on Take-up with an IRS Field Experiment. Working Paper, University of California, Los Angeles.

Carroll, Gabriel D., James J. Choi, David Laibson, Brigitte C. Madrian, and Andrew Metrick (2009). Optimal Defaults and Active Decisions. *Quarterly Journal of Economics* 124(4): 1639–1674.

Chetty, Raj (2012). Bounds on Elasticities with Optimization Frictions: A Synthesis of Micro and Macro Evidence on Labor Supply. *Econometrica* 80(3): 969–1018.

Chetty, Raj, John N. Friedman, Soren Leth-Petersen, Torben Nielsen, and Tore Olsen (2012). Active vs. Passive Decisions and Crowdout in Retirement Savings Accounts: Evidence from Denmark, NBER Working Paper No. 18565.

Fack, Gabrielle and Camille Landais (2011). Charitable Giving and Tax Policy in the Presence of Cheating: Theory and Evidence from the U.S. and France, Working Paper, London School of Economics.

Gruber, Jonathan and Emmanuel Saez (2002). The Elasticity of Taxable Income: Evidence and Implications. *Journal of Public Economics* 84(1): 1–32.

Kleven, Henrik J., Martin B. Knudsen, Claus T. Kreiner, Soren Pedersen, and Emmanuel Saez (2011). Unwilling or Unable to Cheat? Evidence from a Tax Audit Experiment in Denmark. *Econometrica* 79(3): 651–692.

Kleven, Henrik J. and Wojciech Kopczuk (2011). Transfer Program Complexity and the Take Up of Social Benefits. *American Economic Journal: Economic Policy* 3(1): 54–90.

Kleven, Henrik J. and Mazhar Waseem (2013). Using Notches to Uncover Optimization Frictions and Structural Elasticities: Theory and Evidence from Pakistan. *Quarterly Journal of Economics* 128(2): 669–723.

Kotakorpi, Kaisa and Jani-Petri Laamanen (2013). Complexity, Salience and Income Tax Filing Behavior: Evidence From a Natural Experiment, Working Paper, University of Turku.

Mayshar, Joram (1991). Taxation with Costly Administration. *Scandinavian Journal of Economics* 93(1): 75–88.

Organization for Economic Co-operation and Development (OECD). (2006). Using Third Party Information Reports to Assist Taxpayers Meet their Return Filing Obligations: Country Experiences With the Use of Pre-Populated Personal Tax Returns. Paris, France.

Pitt, Mark M. and Joel Slemrod (1989). The Compliance Cost of Itemizing Deductions: Evidence from Individual Tax Returns. *American Economic Review* 79(5): 1224–1232.

Rehavi, Marit (2010). Partial Reporting: An Example from Charitable Giving, Working Paper, University of British Columbia.

Saez, Emmanuel (2010). Do Taxpayers Bunch at Kink Points? *American Economic Journal: Economic Policy* 2(3): 180–212.

Slemrod, Joel (1989). Are Estimated Tax Elasticities Really Just Tax Evasion Elasticities? The Case of Charitable Contributions. *Review of Economics and Statistics* 71(3): 517–522.

Slemrod, Joel (2007). Cheating Ourselves: The Economics of Tax Evasion. *Journal of Economic Perspectives* 21(1): 25–48.

Yitzhaki, Shlomo (1979). A Note on Optimal Taxation and Administrative Costs. *American Economic Review* 69(3): 475–480.