Essays on Savings, Investment, and Monetary Policy

by

Matthew Rognlie

Submitted to the Department of Economics
on May 15, 2016, in Partial Fulfillment of the
Requirements for the Degree of
Doctor of Philosophy in Economics

Abstract

This thesis consists of three chapters on savings, investment, and monetary policy. The first chapter studies optimal monetary policy in an environment where negative nominal interest rates are possible, as demonstrated by recent central bank actions worldwide. I identify the central tradeoff: negative rates help stabilize aggregate demand, but at the cost of an inefficient subsidy to paper currency. Near 0%, the first side of this tradeoff dominates, and negative rates are generically optimal whenever output averages below its efficient level. In a benchmark scenario, negative rates are sufficient to undo most welfare losses relative to the first best. Credible commitment by the central bank is essential to implementing optimal policy, which backloads the most negative rates. The option to set negative nominal rates lowers the optimal long-run inflation target, and abolishing paper currency is only optimal when currency demand is highly elastic.

The second chapter documents and interprets new facts about the functional distribution of income. It shows that the net capital share of income initially fell in the postwar era, before rising from the 1970s onward, with the entire net increase coming from the housing sector. Accounts of the capital share that emphasize capital accumulation are not consistent with the data: they require empirically improbable elasticities of substitution, and they presume a correlation between the capital-income ratio and capital share that is not visible in the data. A more limited narrative that stresses scarcity and the increased cost of housing better fits the data, as I clarify using a new multisector model of factor shares.

The third chapter, joint with Adrien Auclert, studies the possibility that feedbacks between sovereign bond spreads and governments' desire to default may lead to multiple equilibria in sovereign debt markets. We show that such multiplicity does not exist in the infinite-horizon model of Eaton and Gersovitz (1981), a widely adopted benchmark for quantitative analyses of these markets. Our proof may be important to understand Euro government bond markets, and calls for renewed attention on the theoretical conditions that are needed for sovereign debt models to generate multiple equilibria.

Thesis Supervisor: Iván Werning
Title: Robert M. Solow Professor of Economics

Thesis Supervisor: Daron Acemoglu
Title: Elizabeth and James Killian Professor of Economics
Acknowledgments

I am indebted to many others who helped me throughout graduate school and in the writing of this thesis.

Iván Werning has been a tremendous advisor: brilliant and prolific, yet also ready to spend an entire afternoon talking about research on the spur of the moment. I have learned an immense amount from watching him work over the past several years, and I'm grateful to him for his readiness to treat us as equal partners in the enterprise. My mind boggles when I think of the diverse and deep contributions to macroeconomics for which I've had a front-row seat—and how despite all that, Iván somehow found the time to engage so deeply with my work as well, bringing always his formidable intuition and insatiable curiosity.

Daron Acemoglu has unmatched energy and dedication to his students, and is responsible for more "how can one human know so much?" moments than anyone else I've ever known. He's put up cheerfully with all the wild changes of direction in my research, and somehow always finds an idea to push my work to the next level—no matter whether it's in his area or not even close.

Alp Simsek has enriched all of us at MIT by returning here. His level of commitment to students and advisees is far beyond any junior faculty I've ever seen, and it will be an inspiration to me as I set out along the same path. He's always ready to meet and talk, and always thoughtful and engaged about finance and macro. It has been a pleasure to be his TA, advisee, and coauthor.

My knowledge of economics has been greatly influenced by the other macro faculty at MIT as well, each of whom brings a distinct worldview and research style: Marios Angeletos, Ricardo Caballero, Rob Townsend, and also Guido Lorenzoni, who I'll be privileged to rejoin at Northwestern. In addition, I would like to thank Jim Poterba—whose public finance class and brief stretch employing me as an RA have proven remarkably fruitful in my understanding of public policy—Andrei Shleifer, and David Autor. And I would be remiss not to mention Gary King, who effectively runs the department and has rescued me from administrative disaster time and time again.

I've had a great group of classmates here at MIT. In the macro community, I have enjoyed discussions with Nicolas Carapao, Sebastian Fanelli, Dana Foarta, Greg Howard, Ernest Liu, Juan Passadore, Pascual Restrepo, Daan Struyven, Ludwig Straub, and Olivier Wang, and I am excited that Rodrigo Adao and Dejanir Silva will be heading along with me to Princeton this coming year. I've also had a lot of fun talking with Vivek Bhattacharya (who took the same path from Duke to MIT but has been far better at every step), Steve Murphy (who shares some of my political idiosyncrasies, and who has the kind of intuition about economics that's nearly extinct these days), and Giovanni Reggiani (my partner in studying for macro and theory generals and in
always-entertaining conversation).

I cannot possibly say enough about Adrien Auclert’s role in my growth as an economist and in the work underlying this thesis. Without his above-and-beyond support, I literally would not have been able to go on the academic job market, or to graduate. I am consistently blown away by his vision, his discipline, and his engagement with economics. We share many ideas and interests—including, oddly, a passion for hoarding obscure books from every corner of the field—but Adrien somehow manages to also be outstanding in all the ways that I am deficient. He is the formal coauthor of the third chapter here, in addition to two other joint papers, but informally he has been a coauthor for this entire thesis.

I owe everything to my wife Joo-Young, who has been the single most important person in my life over our now nearly five years of marriage. Every day I’m a better person knowing that I’ll come home to someone as kind and delightful and extraordinary as her. Her love and encouragement have carried me through the past several years, and I know we are both very excited to begin the next chapter of our lives together.

Finally, nothing would be possible without the tireless love and support of my family—especially my mother, Jane; and my father, Kevin, who I will miss dearly. I dedicate this thesis to them.
## Contents

1 What Lower Bound? Monetary Policy with Negative Interest Rates 11
   1.1 Introduction ................................................. 11
   1.2 Model and assumptions on cash ................................. 19
      1.2.1 Zero lower bound and cash demand ....................... 19
      1.2.2 Benchmark model ........................................... 28
   1.3 Optimal policy and negative rates .............................. 32
   1.4 Revisiting ZLB traps .......................................... 43
      1.4.1 Calibration ................................................... 44
      1.4.2 Partial commitment case .................................... 46
      1.4.3 Full commitment case ....................................... 49
      1.4.4 Full commitment case, with a positive natural rate after the trap 50
   1.5 Interaction with other policies ................................ 52
      1.5.1 Trend inflation .............................................. 52
      1.5.2 Abolishing cash ............................................. 55
      1.5.3 Multiple denominations .................................... 58
   1.6 Conclusion ................................................... 61
   1.7 Appendix: Proofs ............................................... 62

2 Deciphering the Fall and Rise in the Net Capital Share: Accumulation, or Scarcity? 71
   2.1 Introduction ................................................... 71
   2.2 Evidence on factor income shares in developed countries .... 76
      2.2.1 Conceptual issues ........................................... 76
3.4 Extensions of the benchmark model ........................................ 157
  3.4.1 Bound on savings ................................................. 158
  3.4.2 Stochastic market reentry ....................................... 162
  3.4.3 Other variations on the model and multiplicity results .... 164
3.5 Conclusion ........................................................................... 167
3.6 Appendix: Proofs for section 3.2 ....................................... 167
  3.6.1 Existence of Markov perfect equilibrium ....................... 167
  3.6.2 Uniqueness of subgame perfect equilibrium ................... 175
3.7 Appendix: Proofs for section 3.4 ....................................... 181
  3.7.1 Proof of lemma 3.4.1 ................................................ 181
  3.7.2 Proof of proposition 3.4.3 ......................................... 182
  3.7.3 Proof of proposition 3.4.4 ......................................... 184
Chapter 1

What Lower Bound? Monetary Policy with Negative Interest Rates

1.1 Introduction

Can nominal interest rates go below zero? In the past two decades, the zero lower bound (ZLB) on nominal rates has emerged as one of the great challenges of macroeconomic policy. First encountered by Japan in the mid-1990s it has, since 2008, become a constraint for central banks around the world, including the Federal Reserve and the European Central Bank. These central banks’ perceived inability to push short-term nominal rates below zero has led them to experiment with unconventional policies—including large-scale asset purchases and forward guidance—in order to try to achieve their targets for inflation and economic activity, with incomplete success.

Events in the past year, however, have called into question whether zero really is
a meaningful barrier. Central banks in Switzerland, Denmark, and Sweden have targeted negative nominal rates with apparent success, and without any major changes to their monetary frameworks. Policymakers at other major central banks, including the Federal Reserve and the ECB, have recently alluded to the possibility of following suit.12

In this paper, I consider policy in this new environment, where negative nominal rates are a viable option. I argue that these negative rates, though feasible, are not costless: they effectively subsidize paper currency, which now receives a nominal return (zero) that exceeds the return on other short-term assets. Policymakers face a tradeoff between the burden from this subsidy and the benefits from greater downward flexibility in setting rates. This paper studies the tradeoff in depth, exploring the optimal timing and magnitude of negative rates, as well as their interaction with other policy tools.

The traditional rationale behind the zero lower bound is that the existence of money, paying a zero nominal return, rules out negative interest rates in equilibrium: it would be preferable to hoard money rather than lend at a lower rate. This view was famously articulated by Hicks (1937):

If the costs of holding money can be neglected, it will always be profitable to hold money rather than lend it out, if the rate of interest is not greater than zero. Consequently the rate of interest must always be positive.

Of course, this discussion presumes that money pays a zero nominal return, which is not true of all assets that are sometimes labeled “money”. Bank deposits can pay positive interest or charge the equivalent of negative interest through fees; similarly,

---

1In response to a question while testifying before Congress on November 4, 2015, Federal Reserve Chair Janet Yellen stated that if more stimulative policy were needed, “then potentially anything, including negative interest rates, would be on the table.” (Yellen 2015.) In a press conference on October 22, 2015, ECB President Mario Draghi stated: “We’ve decided a year ago that [the negative rate on the deposit facility] would be the lower bound, then we’ve seen the experience of countries and now we are thinking about [lowering the deposit rate further].” (Draghi 2015.)

2By some measures, the ECB has already implemented negative rates, since the Eurosystem deposit facility (to which Draghi 2015 alluded) pays -0.20%. Excess reserves earn this rate, which has been transmitted to bond markets: as of November 20, 2015, short-term government bond yields are negative in a majority of Euro Area countries. Since the ECB’s benchmark rate officially remains 0.05%, however, I am not classifying it with Switzerland, Denmark, and Sweden.
central banks are free to set the interest rate on the reserves that banks hold with them. The one form of money that is constrained to pay a zero nominal return is paper currency—which in this paper I will abbreviate as “cash”. The traditional argument for a zero lower bound, therefore, boils down to the claim that cash yielding zero is preferable to a bond or deposit yielding less—and that any attempt to push interest rates below zero will lead to an explosion in the demand for cash.

In light of recent experience, I argue that this claim is false: contrary to Hicks’s assumption, the costs of holding cash cannot be neglected. I write a simple model of cash use in which these costs make it possible for interest rates to become negative. These very same costs, however, make negative rates an imperfect policy tool: since cash pays a higher return, households hold it even when the marginal costs exceed the benefits. The distortionary subsidy to cash creates a deadweight loss. This is the other side of a mainstay of monetary economics, the Friedman rule, which states that nominal rates should be optimally set at zero, and that any deviation from zero creates a welfare loss. The Friedman rule has traditionally been used to argue that positive nominal rates are suboptimal, but I argue the same logic captures the loss from setting negative rates—and this loss may be of far greater magnitude, since cash demand and the resulting distortion can grow unboundedly as rates become more negative.

I integrate this specification for cash demand into a continuous-time New Keynesian model. With perfectly sticky prices, nominal interest rates determine real interest rates, which in turn shape the path of consumption and aggregate output. The challenge for policy is to trade off two competing objectives—first, the need to set the nominal interest rate to avoid departing too far from the equilibrium or “natural” real interest rate, determined by the fundamentals of the economy; and second, the desire to limit losses in departing from the Friedman rule. Optimal policy navigates these two objectives by smoothing interest rates relative to the natural rate, to an extent determined by the level and elasticity of cash demand. These results echo earlier results featuring money in a New Keynesian model, particularly Woodford (2003b), though my continuous-time framework provides a fresh look at several of
these previous insights, in addition to a number of novel findings.

I then provide a reinterpretation of the ZLB in this new framework. Under my standard specification of cash demand, motivated by the evidence from countries setting negative rates, the ZLB is not a true constraint on policy, though it is possible to consider optimal policy when it is imposed as an exogenous additional constraint. I argue that this optimal ZLB-constrained policy is equivalent to optimal policy in a counterfactual environment, where the net marginal utility from cash is equal to zero for any amount of cash above a satiation point. Central banks that act as if constrained by a ZLB, therefore, could be motivated by this counterfactual view of cash demand.

In the baseline case where cash demand does not explode at zero, I show that it is generally optimal to use negative rates. The key observation is that the zero bound is also the optimal level of interest rates prescribed by the Friedman rule. In the neighborhood of this optimum, any deviation leads to only second-order welfare losses, which are overwhelmed by any first-order gains from shaping aggregate demand. These first-order gains exist if, over any interval that begins at the start of the planning horizon, the economy will on average (in a sense that I will make precise) be in recession. Far from being a hard constraint on rates, therefore, zero is a threshold that a central bank should go beyond whenever needed to boost economic activity.

With this in mind, I revisit the standard “liquidity trap” scenario that has been used in the literature to study the ZLB. As in Eggertsson and Woodford (2003) and Werning (2011), I suppose that the natural interest rate is temporarily below zero, making it impossible for a ZLB-constrained central bank to match with its usual inflation target of zero. With negative rates as a tool, it is possible to come much closer to the optimal level of output, but this response is mitigated by the desire to avoid a large deadweight loss from subsidizing cash.

In the simplest case, I assume that the natural rate reverts to zero after the “trap” is over, and that it is impossible to commit to time-inconsistent policies following the trap. Solving the model for optimal policy with negative rates, the key insight that emerges is that the most negative rates should be backloaded. Relative to the cost
of violating the Friedman rule, which does not vary over time, negative rates have the greatest power to lift consumption near the end of the trap. The optimal path of rates during the trap, in fact, starts at zero and monotonically declines, always staying above the natural rate. If full commitment to time-inconsistent policies is allowed, it becomes optimal to keep rates negative even after the trap has ended and the natural rate is no longer below zero—taking backloading one step further, and effectively employing forward guidance with negative rates.

Quantitatively, I compare the outcomes of ZLB-constrained and unconstrained policy using my benchmark calibration. Freeing the policymaker to set negative rates closes over 94% of the gap between equilibrium utility and the first best. A second-order approximation to utility, which is extremely accurate for the benchmark calibration, offers insight into the forces governing the welfare improvement: negative rates offer greater gains when the trap is long and the welfare costs of recession are high, but they are less potent when the level and elasticity of cash demand are large.

I also consider the case where, following the trap, the natural rate reverts to a positive level. This allows a ZLB-constrained central bank to engage in forward guidance, continuing to set rates at zero after the trap. In this environment, I show that the optimal ZLB-constrained and unconstrained policies produce qualitatively similar results: they both use forward guidance to create a boom after the trap, which limits the size of the recession during the trap. ZLB-constrained policy, however, produces far larger swings in output relative to the first-best level, in both the positive and negative directions. With negative rates, it is possible to smooth these fluctuations by more closely matching the swings in the natural rate.

I next relax the assumption of absolute price stickiness, assuming instead that prices are rigid around some trend inflation rate, which can be chosen by the central bank. This allows me to evaluate the common argument that higher trend inflation is optimal because it allows monetary policy to achieve negative real rates despite the zero lower bound (see, for instance, Blanchard, Dell'Ariccia and Mauro 2010). I show that once negative nominal rates are available as a policy tool, the optimal trend inflation rate falls, as inflation becomes less important for this purpose. The
ability to act as a substitute for inflation may add to negative nominal rates' popular appeal.

Finally, I consider supplemental policies that limit the availability of cash. The most extreme such policy is the abolition of cash, frequently discussed in conjunction with the zero lower bound (see, for instance, Rogoff 2014). This policy is equivalent of imposing an infinite tax on cash, and in that light can be evaluated using my framework: the crucial question is whether the distortion from subsidizing cash when rates are negative is large enough to exceed the cost from eliminating cash altogether. I argue that this depends on the extent of asymmetry in the demand for cash with respect to interest rates, and I describe a simple sufficient condition that makes it optimal for policymakers to retain cash. As an empirical matter, I conclude that it is probably not optimal to abolish cash—but this does depend on facts that are not yet settled, including the extent to which cash demand rises when rates fall below levels that have thus far been encountered. One possible intermediate step is the abolition of larger cash denominations, which have lesser holding costs and are demanded more elastically than small denominations. In an extension of my cash demand framework to multiple denominations, I show that it is always optimal to eliminate these large denominations first.

Related literature. This paper relates closely to several literatures.

The literature on negative nominal interest rates has seen considerable growth in the past decade. In contrast to my paper, this literature generally makes the same presumption as Hicks (1937): it assumes that cash demand becomes infinite once cash offers a higher pecuniary return than other assets. When this is true, major institutional changes are required before negative rates are possible. Buiter (2009) summarizes the options available: cash can either be abolished or made to pay a negative nominal return. The former option, the abolition of cash, has been explored in detail by Rogoff (2014). The latter option, a negative nominal return, can be implemented either by finding some way to directly tax cash holdings, or by decoupling cash from the economy's numeraire.
The idea of taxing cash originated with Gesell (1916), who proposed physically stamping cash as proof that tax has been paid. At the time, this proposal was influential enough to be cited by Keynes (1936). More recently, similar ideas have been explored by Goodfriend (2000), who proposes including a magnetic strip in each bill to keep track of taxes due; by Buiter and Panigirtzoglou (2001, 2003), who integrate a tax on cash into a dynamic New Keynesian model; and more whimsically by Mankiw (2009), who suggests that central banks hold a lottery to invalidate cash with serial numbers containing certain digits.

The idea of decoupling cash from the numeraire originated with Eisler (1932), who envisioned a floating exchange rate between cash and money in the banking system, with the latter as the numeraire. This floating rate makes it possible to implement negative nominal interest rates in terms of the numeraire, even as cash continues to pay a zero nominal rate in cash terms, by engineering a gradual relative depreciation of cash. More recently, Buiter (2007) has resurrected this approach, and Agarwal and Kimball (2015) provide a detailed guide to its implementation and possible advantages.

Each of these approaches makes negative rates unambiguously feasible, but at the cost of major changes to the monetary system: either abolishing cash, taxing it via a tracking technology, or removing its status as numeraire. My paper, by contrast, primarily focuses on the consequences of negative rates within the existing system, as they are currently being implemented in Switzerland, Denmark, and Sweden. For policymakers who are not yet ready or politically able to make major reforms to the monetary system, the paper provides a framework for understanding negative rates; by clarifying the costs of negative rates within the existing system, it also provides a basis for comparison to the costs of additional reforms.

Some very recent work explores the practical side of the negative rate policies now in effect. Jackson (2015) provides an overview of recent international experience with negative policy rates, and Jensen and Spange (2015) discuss the pass-through to financial markets and impact on cash demand from negative rates in Denmark. Humphrey (2015) evaluates ways to limit cash demand in response to negative rates.
This paper is also closely related to the modern zero lower bound literature, which began with Fuhrer and Madigan (1997) and Krugman (1998) and subsequently produced a flurry of papers. I revisit the “trap” scenario contemplated in much of this work—notably Eggertsson and Woodford (2003) and Werning (2011)—in which the natural rate of interest is temporarily negative and cannot be matched by a central bank subject to the zero lower bound. One particularly important theme—both in the ZLB literature and in this paper—is forward guidance, which is the focus of a large emerging body of work that includes Levin, López-Salido, Nelson and Yun (2010), Campbell, Evans, Fisher and Justiniano (2012), Del Negro, Giannoni and Patterson (2012), and McKay, Nakamura and Steinsson (2015). I also consider the interaction of the ZLB, negative rates, and the optimal rate of trend inflation, which has been covered by Coibion, Gorodnichenko and Wieland (2012), Williams (2009), Blanchard et al. (2010), and Ball (2013), among others.

At its core, this paper uses the canonical New Keynesian framework laid out by Woodford (2003a) and Galí (2008), but since price dynamics are not a focus, for simplicity I replace pricesetting à la Calvo (1983) with the assumption of fully rigid prices. I follow Werning (2011) by using a continuous-time version of the model, which permits a sharper characterization of both cash demand and the liquidity trap. In adding cash to the model, the paper is reminiscent of much of the New Keynesian literature with money, including Khan, King and Wolman (2003), Schmitt-Grohé and Uribe (2004b), and Siu (2004). It perhaps comes closest to Woodford (1999) and Woodford (2003b), which also find that smoothing interest rates is optimal in the model with money—though this smoothing takes a particularly stark form in the continuous-time framework I provide.

This paper is deeply connected with the literature on the Friedman rule, since it emphasizes deviation from the Friedman rule—in a novel direction—as the reason why negative rates are costly. This literature began eponymously with Friedman (1969), and was exhaustively surveyed by Woodford (1990). The seminal piece opposing the Friedman rule was Phelps (1973), which argued that a government minimizing the overall distortionary burden of taxation should rely in part on the inflation tax as
a source of revenue; much subsequent work has investigated this claim. The key intuition for why the Friedman rule may be optimal, even when alternative sources of government revenue are distortionary, is that money is effectively an intermediate good, facilitating transactions: versions of this idea are in Kimbrough (1986), Chari, Christiano and Kehoe (1996), and Correia and Teles (1996).

As Schmitt-Grohé and Uribe (2004a) and others point out, however, positive nominal interest rates may be optimal as an indirect tax on monopoly profits. Inversely, da Costa and Werning (2008) find that negative rates may be preferable due to the complementarity of money and work effort, although they interpret this finding as showing that the Friedman rule is optimal as a corner solution, under the presumption that negative rates are not feasible. In this paper I sidestep much of the complexity in the literature by taking a simple model where the government has a lump-sum tax available, and the Friedman rule is therefore unambiguously optimal absent nominal rigidities. If, in a richer model, the optimum nominal rate is positive or negative instead, much of the analysis in the paper still holds, except that zero no longer has the same special status as a benchmark.

1.2 Model and assumptions on cash

1.2.1 Zero lower bound and cash demand

Why should zero be a lower bound on nominal interest rates? Traditionally, the literature has held that negative rates imply infinite money demand, which is inconsistent with equilibrium.

For instance, the influential early contribution by Krugman (1998) models money demand using a cash-in-advance constraint. Once this constraint no longer binds, the nominal interest rate falls to zero—but it cannot fall any further, because individuals prefer holding money that pays zero to lending at a lower rate. Similarly, Eggertsson and Woodford (2003) posit that real money balances enter into the utility function, and that marginal utility from money is exactly zero once balances exceed some
satiation level. Again, rates can fall to zero, but no further: once the marginal utility from money is zero, holding wealth in the form of money is indistinguishable from holding it in the form of bonds, and if bonds pay a lower rate there will be an unbounded shift to money.

Many traditional models of money demand similarly embed this zero lower bound. In the Baumol-Tobin model (Baumol 1952 and Tobin 1956), for instance, the interest elasticity of real money demand is $-1/2$. As the nominal interest rate $i$ approaches 0, money demand $M/P \propto i^{-1/2}$ approaches infinity. The same happens in any model where the interest elasticity of money demand is bounded away from zero in the neighborhood of $i = 0$, including many of the specifications in the traditional empirical money demand literature, which assume a constant interest elasticity—see for instance, Meltzer (1963).3

In contrast, other empirical studies of money demand, dating back to Cagan (1956), assume a constant interest semielasticity—see, for instance, Ball (2001) and Ireland (2009). With this specification, money demand does not explode as $i \to 0$; indeed, if extended to cover negative $i$, the specification continues to imply finite money demand.

Figure 1-1 displays three possible shapes for the demand curve for money with respect to interest rates. The first is a curve featuring a constant elasticity of demand, such that money demand explodes as $i \to 0$. The second is a curve featuring a constant semielasticity, such that money demand remains finite even as $i$ becomes negative. The third is a modification of the second curve along the lines of Eggertsson and Woodford (2003) and much of the other zero lower bound literature, where money demand is unbounded at $i = 0$ even though it remains finite in the limit $i \to 0$. The first and third cases feature a zero lower bound, while the second does not.

As argued by Ireland (2009), modern experience with low nominal interest rates contradicts the first case in figure 1-1: money demand does not explode in inverse

---

3 This feature has played a prominent role in welfare calculations: under specifications assuming a constant interest elasticity, Lucas (2000) finds that the costs of moderate departures from the Friedman rule are significant, while under specifications assuming a constant interest semielasticity, the costs are much smaller. Roughly speaking, when assuming a constant elasticity, the explosion in money demand as $i \to 0$ means that the deadweight loss from setting $i > 0$ is much larger.
proportion to rates near zero. It has been an open question, however, whether money demand more closely resembles the second or third case: does it smoothly expand as rates dip below zero, or does it abruptly become infinite at zero? The modern zero lower bound literature has generally assumed the latter, either implicitly (when the bound is imposed as an ad-hoc constraint) or explicitly (when the bound is micro-founded using money demand).

Cash vs. other forms of money. At this point, it is useful to distinguish between different forms of "money". Inside money, consisting of bank deposits and other liquid liabilities of private intermediaries, is not subject in principle to any zero lower bound: it can pay negative interest as well as positive interest, and sometimes does so implicitly through account fees. There may be frictions in adjusting to negative rates, but these are highly specific to the institution and regulatory regime, and are not central to the zero lower bound as a general notion.\footnote{For instance, McAndrews (2015) mentions the dilemma of retail and Treasury-only money market mutual funds in the US, which as currently structured would "break the buck" and be forced to disband in an environment with negative rates. Money market mutual funds elsewhere, however, have successfully adapted to negative rates.}

Most central bank liabilities can also pay negative interest: for instance, a central bank can charge banks who hold reserve balances with it. In fact, this is exactly what central banks that implement negative rates do. In a world where all liabilities of the central bank could pay negative interest, there would be no hint of a lower bound.\footnote{In fact, in the canonical treatment of the New Keynesian model in Woodford (2003a, p. 68), the lower bound on interest rates \( i_t \) is derived to be the interest \( i_t^* \) paid on money by the central bank.}
The difficulty is that one central bank liability, paper currency, has a nominal return that is technologically constrained to be zero. If nominal interest rates on other assets are negative, the concern is that demand for paper currency—which I abbreviate as cash—will become infinite. If this is true, zero does serve as an effective lower bound on interest rates. Interpreting figure 1-1 as depicting alternative possible shapes for the cash demand function, the crucial question is therefore whether the second or third possibility is more accurate.

New evidence: successful implementation of negative rates. In the last year, three central banks have set their primary rate targets at unprecedently negative levels: both Switzerland and Denmark at -0.75%, and Sweden at -0.30%. This is depicted in figure 1-2.

Implementation has been successful: in line with the targets, market short-term nominal interest rates have fallen well into negative territory. Indeed, expectations that the negative rate policy will be continued in Switzerland are sufficiently strong that even the 10-year Swiss government bond yield has been negative for much of 2015.

This novel policy experiment provides a useful test of whether negative market interest rates are consistent with bounded cash demand. Thus far, the verdict has been clear: not only has cash demand remained finite, but its response to negative rates has been quite mild. For Switzerland, where monthly data on banknotes outstanding is publicly available, figure 1-3 shows the total value of cash in circulation against the path of the Swiss target rate. Following the decline to -0.75% at the beginning of 2015, there has been little perceptible break in the trend. For Denmark, Jensen and Spange (2015) have similarly noted little increase in cash demand.

Among the possibilities depicted in figure 1-1, therefore, the empirical cash de-

---

6 As discussed in section 1.1, there have been proposals to remove this constraint through changes in technology: for instance, the idea of Goodfriend (2000) to embed a magnetic strip in paper currency that tracks taxes paid on it.

7 As before, the argument in Ireland (2009) rules out the first possibility: cash demand appears not to explode as nominal interest rates asymptote to zero.

8 For example, as of November 20, 2015, one-month government bond yields are -0.89% in Switzerland, -0.70% in Denmark, -0.39% in Sweden.
(Target rates are 3-month Libor CHF for Switzerland, Danmarks Nationalbank certificates of deposit rate for Denmark, and Riksbank repo rate for Sweden.)
mand schedule appears to most closely resemble the middle case, with no discontinuity at \( i = 0 \). My analysis will build upon this observation.

**Cost of negative rates: a subsidy to cash.** If negative rates do not lead to infinite cash demand, and are therefore feasible, is there potentially any reason to avoid them? Yes.

To build intuition, it is useful to consider an extreme case: suppose that setting \( i = -1\% \) leads cash demand to increase by a factor of 100. Since this is not quite an infinite increase, it is still feasible in equilibrium, but there is a considerable cost. If the central bank holds short-term bonds on the asset side of its balance sheet, for instance, then its cash liabilities will pay 0\% while its assets earn -1\%. The effective 1\% subsidy to cash, relative to the market interest rate, will cost the central bank greatly—and on a massively expanded base of cash, leading to annual losses equal in magnitude to the entire prior level of cash in circulation.

The public will both benefit from this subsidy and ultimately pay the cost of providing it, via a larger tax burden. Under certain assumptions, which I will use in this paper, this cost and benefit will cancel to first order in \( i \).\(^9\) But there will be a second-order net cost—which, if cash demand increases by a factor of 100, will be quite large—since the subsidy leads the public to demand more cash than is socially optimal.

The intuition for this second-order cost is similar to that for any subsidy. If the public decides to hold more cash when the subsidy is 1\% than when it is 0\%, there must be a net nonpecuniary cost to the marginal unit of cash: the inconvenience of holding wealth in the form of cash exceeds, at the margin, the liquidity benefits. Once the public pays for the subsidy through taxes, all that remains is this inefficiently high level of cash demand—with, perhaps, a large drain of resources going to the manufacturers of safes.

This is the inverse of the traditional story, in which positive nominal interest rates

---

\(^9\)This first-order cancellation arises because 0\% is the Friedman rule optimum. More generally, whether or not the Friedman rule holds depends on assumptions about distributive effects, fiscal instruments available to the government, and so on. See the discussion of the literature in section 1.1.
act as a tax, leading the public to demand inefficiently little cash. The idea that the optimal level of nominal interest rates is zero—with neither a tax nor a subsidy on cash—is called the *Friedman rule*, in recognition of Friedman (1969). Generally, only one side of the Friedman rule has been discussed: prior to recent events, negative rates were not viewed as a feasible option, and it made little sense to talk about the inefficiency from too much cash demand.

But this inefficiency, in fact, is at the center of the policy tradeoff with negative rates. The prior consensus that negative rates were infeasible, due to an explosion in cash demand at zero, can be interpreted as just an extreme form of the same point: as cash demand becomes more and more elastic with respect to negative interest rates, the inefficiency increases until negative rates become infinitely costly in the limit. More generally, it is plausible that the cost of deviating from the Friedman rule is much more severe on the negative side than on the traditional, positive one, because the rise in cash demand is potentially unbounded.

**Interpretation in a simple model of cash demand.** In section 1.2.2, I will integrate cash demand into a simple infinite-horizon New Keynesian model by including concave flow utility $v(m(t))$ from real cash balances $m(t)$ into household preferences (1.2). The opportunity cost of holding wealth in the form of cash rather than bonds is the nominal interest rate $i$ on bonds, and real cash demand $M^d(i, c)$ as a function of nominal interest rates $i$ and consumption $c$ is given by the optimality condition

$$v'(M^d(i, c)) = iu'(c)$$

where $u'(c)$ is marginal utility from consumption.

If there is finite cash demand when $i = 0$, its level $m^* = M^d(0, c)$ is given by $v'(m^*) = 0$; and if cash demand continues to be finite for negative $i$ as well, then $v'$ must be strictly declining at $m^*$. It follows that $v(m^*)$ is a global maximum of $v$. This is depicted in figure 1-4.

Positive $i$ corresponds to $v' > 0$ and to an inefficiently low level of cash demand.
\( m < m^* \), while negative \( i \) corresponds to \( \nu' < 0 \) and an inefficiently high level of cash demand \( m > m^* \). The utility shortfall relative to \( v(m^*) \) can be obtained in consumption terms by integrating marginal utility \( \nu' \), which according to (1.1) is proportional to the cash demand curve times \( u'(c^*) \). This is visualized in figure 1-5, which shows as shaded areas the loss from setting positive \( i \) (the standard case) and the loss from setting negative \( i \) (the new case). Figure 1-6 shows the second-order Harberger triangle approximation to the loss from negative \( i \), which depends on the level of cash demand at \( i = 0, M^d(0, c) = m^* \), and crucially the local semielasticity of cash demand \( \partial \log M^d(0^-, c) / \partial i \). When the semielasticity is higher and cash demand grows more rapidly as \( i \) falls below 0, the loss is more severe—and in the limit as the semielasticity becomes infinite, the cost becomes infinite as well, leading in effect to a zero lower bound.

As figures 1-5 and 1-6 illustrate, therefore, the cost of setting negative rates fits squarely into the standard microeconomic analysis of distortions. With this view in mind, I now turn to a dynamic framework, studying how this cost trades off against the other objectives of monetary policy.
Figure 1-5: Loss from violating Friedman rule: integrating under the demand curve

Loss for positive $i$

Loss for negative $i$

Figure 1-6: Approximate cost of deviating from Friedman rule: Harberger triangle

$$M^d(0, c) \cdot \frac{\partial \log M^d(0^-, c)}{\partial i} \cdot i$$
1.2.2 Benchmark model

In this section, I describe the basic infinite-horizon continuous-time model that will be used for the analysis, with a particular focus on the specification for cash demand.

**Households.** Households have the objective

\[
U({c(t), n(t), M(t), P(t)}) = \int_0^\infty e^{-\int_0^t \rho(u)du} \left( u(c(t)) - \chi(n(t)) + \nu \left( \frac{M(t)}{P(t)} \right) \right) dt \quad (1.2)
\]

where \(c(t)\) is consumption, \(n(t)\) is labor supplied, \(M(t)\) is the level of cash held by the household, \(P(t)\) is the price of the consumption good, and \(\rho(t)\) is the time-varying rate of time preference. In general, I will assume that both utility from consumption \(u\) and disutility from labor \(x\) are isoelastic, denoting the elasticity of intertemporal substitution in consumption by \(\sigma\) and the Frisch elasticity of intertemporal substitution in labor by \(\psi\):

\[
u(c) = \frac{c^{1-\sigma} - 1}{1 - \sigma^{-1}} \quad \chi(n) = \gamma \frac{n^{1+\psi} - 1}{1 + \psi^{-1}} \quad (1.3)
\]

The assumption that \(\nu\) is separable from the rest of the utility function is in line with much of the New Keynesian literature featuring money in the utility function. Here, the assumption is made primarily for analytical convenience, but calibrated studies have generally found that (for instance) ignoring the possible complementarity between consumption and money does not have significant quantitative ramifications.

Households have access to two stores of value, cash \(M\) and bonds \(B\), and face the nominal flow budget constraint

\[
\dot{M}(t) + B(t) + P(t)c(t) = i(t)B(t) + W(t)n(t) + \Pi(t) + T(t) \quad (1.4)
\]

where \(i(t)\) is the nominal interest rate paid on bonds and \(W(t)\) is the nominal wage paid for labor by firms. \(\Pi(t)\) is firms’ profit, and \(T(t)\) is net lump-sum transfers by the government, both of which will be specified later. Cash is assumed to pay zero interest in (1.4). As discussed in section 1.2.1, other liabilities of the central bank—
such as electronic reserves—can pay nonzero interest. Here I am abstracting away from the difference between these liabilities and bonds $B(t)$, since both are short-term interest-paying liabilities of the government.

Dividing by $P(t)$, the real flow budget constraint becomes

$$
\dot{m}(t) + \dot{b}(t) + c(t) = r(t)b(t) + w(t)n(t) + \frac{\Pi(t)}{P(t)} + \frac{T(t)}{P(t)}
$$

where $m(t)$ and $b(t)$ are real cash and bonds, respectively, $w(t)$ is the real wage rate, and $r(t) \equiv i(t) - \dot{P}(t)/P(t)$ is the real interest rate, and $\Pi(t)/P(t)$ and $T(t)/P(t)$ are real transfers. Integrating (1.5) and imposing a no-Ponzi condition gives the infinite-horizon version of the budget constraint:

$$
\int_0^\infty e^{-\int_0^s r(s)ds} \left( c(t) + i(t)m(t) \right) dt = \int_0^\infty e^{-\int_0^s r(s)ds} \left( w(t)n(t) + \frac{\Pi(t)}{P(t)} + \frac{T(t)}{P(t)} \right) dt
$$

(1.6)

Given paths \{i(t), r(t), w(t)\} for prices and \{\Pi(t)/P(t), T(t)/P(t)\} for transfers, the household’s problem is to choose \{c(t), n(t), m(t)\} to maximize (1.2) subject to (1.6).

**Firms.** A continuum of monopolistically competitive firms $j \in [0, 1]$ produce intermediate goods using labor as the only input, subject to a potentially time-varying productivity parameter $A(t)$:

$$
y_j(t) = A(t)f(n_j(t))
$$

(1.7)

I will also generally assume that $f$ is isoelastic, with $1 - \alpha$ as the constant elasticity of output with respect to labor $n$:

$$
f(n) = \frac{n^{1-\alpha}}{1-\alpha}
$$

(1.8)

These firms’ output is aggregated into production $y(t)$ of the final consumption good by a perfectly competitive final good sector, which operates a final constant elasticity
of substitution production technology \( y(t) = \left( \int_0^1 y_j(t)^{1-\epsilon} \, dj \right)^{1/\epsilon} \). Demand by this sector for firm \( j \)'s output is \( y_j(t) = (P_j(t)/P(t))^{-\epsilon} y(t) \), where \( P(t) = \left( \int_0^1 P_j(t)^{1-\epsilon} \, dj \right)^{1/(1-\epsilon)} \) is the aggregate price index. Market clearing for labor requires that firms' total demand for labor equals household labor supply: \( n(t) = \int_0^1 n_j(y) \, dj \).

I consider two possible specifications of firms' pricesetting. In the benchmark flexible price case, they choose prices at each \( t \) to maximize profits

\[
\Pi_j(t) = \max_{P_j(t)} P_j(t) \left( \frac{P_j(t)}{P(t)} \right)^{-\epsilon} y(t) - C(y_j(t); t)
\]

where \( C(y; t) \equiv f^{-1}(y/A(t)) \, W(t) \) is the nominal cost of producing \( y \) at time \( t \). Profits (1.9) are maximized when \( P_j(t) \) is set at a markup of \( \epsilon / (\epsilon - 1) \) over marginal cost:

\[
P_j(t) = \frac{\epsilon}{\epsilon - 1} C_y(y; t) = \frac{\epsilon}{\epsilon - 1} \frac{W(t)}{f'(f^{-1}(y/A(t)))}
\]

It follows that all firms \( j \) set the same price at time \( t \) and produce the same output, and that real wages are given by

\[
w(t) = \frac{\epsilon - 1}{\epsilon} A(t) f'(n(t))
\]

In the sticky price case, by contrast, prices are rigid at \( P(t) \equiv \bar{P} \) for all \( t \). This simple assumption will create an aggregate demand management role for the monetary authority, generating the tradeoff at the heart of this paper: the distortionary costs from setting interest rates below zero, versus the benefits of bringing output closer to its optimal level.

In both cases, I assume that aggregate profits \( \Pi(t) = \int_0^1 \Pi_j(t) \, dj \) are immediately rebated to the household, as seen earlier in (1.4).

**Government.** The government, representing both the fiscal and monetary authorities, has two liabilities, bonds \( B(t) \) and cash \( M(t) \). Nominal interest \( i(t) B(t) \) is earned on bonds, while the nominal interest rate on cash is fixed at zero. A lump-sum transfer \( T(t) \) to households, which can be positive or negative, is also available.
The government's nominal flow budget constraint is then

\[ \dot{M}(t) + \dot{B}(t) = i(t)B(t) - T(t) \]  

(1.11)

which, when normalized by \( P(t) \) and integrated subject to a no-Ponzi condition, becomes

\[ \int_0^\infty e^{-\int_0^t r(s)ds}i(t)m(t)dt = \int_0^\infty e^{-\int_0^t r(s)ds}T(t)dt \]  

(1.12)

which states that the net present value of real seignorage \( i(t)m(t) \) must equal that of real transfers \( T(t)/P(t) \) to the public.

**Equilibrium.** With these ingredients in place, I am now ready to define equilibrium.\(^{10}\)

**Definition 1.2.1.** A flexible-price equilibrium consists of quantities

\[ \{c(t), n(t), y(t), M(t), \Pi(t), T(t)\}_{t=0}^\infty \]

and prices

\[ \{i(t), W(t), P(t)\}_{t=0}^\infty \]

such that households optimize intertemporal utility (1.2) subject to (1.4), firms optimize profits (1.9), the government satisfies its budget constraint (1.11), and goods, factor, and asset markets all clear. In a sticky-price equilibrium, profit optimization is replaced by a sticky-price constraint \( P_f(t) = \bar{P} \).

**Natural rate.** The real interest rate achieved in flexible-price equilibrium—which is uniquely pinned down by fundamentals \( \{A(t), \rho(t)\} \)—will prove useful as a benchmark for sticky-price equilibrium as well. Following common usage, I call it the natural rate.

\(^{10}\)Note that for economy of notation, this definition of flexible-price equilibrium assumes that all firms set the same price, so that there is no need to carry around the distribution of individual prices as an equilibrium object. This is true given my assumptions on firms.
Lemma 1.2.2. In flexible-price equilibrium, $c(t)$, $y(t)$, $n(t)$, and $w(t)$ are uniquely determined by the two equations

$$
\frac{u'(n(t))}{u'(c(t))} = w(t) = \frac{\epsilon - 1}{\epsilon} A(t)f'(n(t))
$$

$$
c(t) = y(t) = A(t)f(n(t))
$$

Assuming isoelastic preferences (1.3) and technology (1.8), the equilibrium real interest rate, which I denote by $r^*(t)$, is then given by

$$
r^*(t) = \rho(t) + \frac{1 + \psi}{\sigma + \psi + (\sigma - 1)\psi \alpha} \frac{\dot{A}(t)}{A(t)}
$$

(1.13)

Definition 1.2.3. The natural rate $r^*(t)$ is the flexible-price equilibrium real interest rate in (1.13).

Note that the natural rate reflects both the rate of pure time preference $\rho(t)$ and the rate of productivity growth $\dot{A}(t)/A(t)$.

1.3 Optimal policy and negative rates

In this section, I set up the optimal policy problem and discuss the implications for negative rates.

Characterizing equilibria. The equilibrium concept in definition 1.2.1 is such that the paths for real quantities $\{c(t), n(t), y(t), m(t)\}$ and prices $\{w(t), i(t)\}$ are uniquely characterized by a much smaller set of paths.

For flexible-price equilibrium, lemma 1.2.2 already shows that $c(t)$, $n(t)$, $y(t)$, and $m(t)$ are determined by exogenous fundamentals. Given the nominal interest rate $i(t)$, the quantity of cash is then given by $m(t) = M^d(i(t), c(t))$.

In contrast, the real quantities and prices in sticky-price equilibrium are not pinned down by nominal interest rates alone. Instead, conditional on nominal interest rates $\{i(t)\}$ there is a single degree of indeterminacy in the consumption path. This inde-
terminacy can be indexed by the level of consumption at some selected time, which I choose to be $t = 0$ for simplicity.

With this in mind, given any path $\{i(t)\}$ for the nominal interest rate and the time-0 level of consumption $c(0)$, consumption at any time $t$ can be obtained by integrating the household’s consumption Euler equation

$$\frac{\dot{c}(t)}{c(t)} = \sigma(i(t) - \rho(t)) \log c(t) = \log c(0) + \int_0^t \sigma(i(s) - \rho(s)) ds$$

(1.14)

With $c(t)$ known, output $y(t) = c(t)$ and labor input $n(t) = f^{-1}(y(t)/A(t))$ are given by market clearing and the production function. The quantity of cash is given by $m(t) = M^d(i(t), c(t))$.

The following proposition summarizes these observations.

**Proposition 1.3.1.** Given any path $\{i(t)\}_{t=0}^\infty$ for nominal interest rates, real quantities

$$\{c(t), n(t), y(t), m(t)\}_{t=0}^\infty$$

and prices

$$\{w(t), i(t)\}_{t=0}^\infty$$

are uniquely determined in flexible-price equilibrium. Additionally, given the level $c(0)$ of consumption at time 0, these real quantities and prices are uniquely determined in sticky-price equilibrium as well.

By offering a straightforward characterization of equilibria, proposition 1.3.1 simplifies the search for equilibria that are optimal from a household welfare (1.2) standpoint.

**Optimal policy: definition and solution under flexible prices.** I assume that the policymaker can freely choose between equilibria, as characterized by proposition 1.3.1. For flexible-price equilibria, this is natural, since the nominal interest rate path $\{i(t)\}$ chosen by the government is sufficient to characterize the equilibrium.

For sticky-price equilibria, this is slightly less natural, since the time-0 level $c(0)$
of consumption must also be specified. To pin down a particular level for \( c(0) \)—and, by extension, the entire path \( \{c(t)\} \)—the government requires some additional policy tool, which I show in the Online Appendix can be a Taylor-style rule for \( i(t) \) off the equilibrium path. Here, I simply assume that the policymaker is capable of choosing \( c(0) \).

**Definition 1.3.2.** Optimal policy for flexible-price equilibrium is the choice of path \( \{i(t)\}_{t=0}^{\infty} \) for nominal interest rates such that the flexible-price equilibrium characterized by proposition 1.3.1 maximizes household utility (1.2).

Optimal policy for sticky-price equilibrium is the choice of \( \{i(t)\}_{t=0}^{\infty} \), along with time-0 consumption \( c(0) \), such that the sticky-price equilibrium characterized by proposition 1.3.1 maximizes household utility (1.2).

Note that optimal policy by this definition is not necessarily time consistent, and that I am therefore assuming full commitment by the policymaker. I will relax this assumption in section 1.4.2.

The flexible-price case turns out to be extremely simple. Since consumption \( c(t) \) and labor supply \( n(t) \) are already pinned down by fundamentals as per lemma 1.2.2, the only quantity entering into household utility (1.2) that can be affected by policy is real cash \( m(t) \). The \( v(m(t)) \) term is maximized under the Friedman rule \( i(t) = 0 \).

**Proposition 1.3.3.** Optimal policy for flexible price equilibrium is given by \( i(t) = 0 \) for all \( t \).

With optimal policy for flexible price equilibrium characterized, I will focus on sticky price equilibrium for the remainder of the paper.

**Optimal policy under sticky prices.** The sticky-price case, by contrast, involves a nontrivial tradeoff: as before, the nominal interest rate affects the level of cash, but it also directly affects the path of consumption in (1.14). Optimal policy now requires balancing the first force against the second.

This can be formulated as an optimal control problem with state \( c(t) \) and control \( i(t) \). Letting \( \mu(t) \) be the costate on \( \log c(t) \), the current-value Hamiltonian is (dropping
dependence on \( t \) for economy of notation):

\[
H = g(c; A) + v(M^d(i, c)) + \mu \sigma (i - \rho) \tag{1.15}
\]

where \( g(c; A) = u(c) - x(f^{-1}(c/A)) \) is defined to be the net utility from consumption \( c \) minus the disutility from the labor required to produce that consumption.

It follows from the maximum principle that \( i \) must maximize (1.15), and therefore that

\[
v'(M^d(i, c)) \cdot \frac{\partial M^d(i, c)}{\partial i} + \mu \sigma = 0 \tag{1.16}
\]

The law of motion for the costate \( \mu \) is

\[
\frac{\dot{\mu}}{\mu} = -c \mu^{-1} \left( g'(c; A) + v'(M^d(i, c)) \cdot \frac{\partial M^d(i, c)}{\partial c} \right) + \rho \tag{1.17}
\]

Since I assume that the policymaker can optimally choose \( c \) at time 0, \( c(0) \) is free and the corresponding costate is zero:

\[
\mu(0) = 0 \tag{1.18}
\]

Together with the Euler equation \( \dot{c}/c = \sigma (i - \rho) \), conditions (1.16), (1.17), and (1.18) characterize optimal policy.

**Simplifying optimal policy.** Define \( \hat{\mu} \equiv \mu/u'(c) \), which is the costate in consumption equivalent terms. Dividing (1.16) by \( u'(c) \), and using \( v'(M^d(i, c))/u'(c) \), I obtain

\[
\hat{\mu} \sigma = i \cdot m \cdot - \frac{\partial \log M^d}{\partial i} \tag{1.19}
\]

Also note that \( \dot{\hat{\mu}}/\hat{\mu} = \hat{\mu} = \hat{\mu} + \sigma^{-1} \hat{c}/c = \hat{\mu} + i - \rho \), which allows (1.17) to be rewritten as

\[
\frac{\dot{\hat{\mu}}}{\hat{\mu}} = -\hat{\mu} \left( c g'(c; A) + \frac{v'(M^d(i, c))}{u'(c)} \cdot \frac{\partial M^d}{\partial \log c} \right) + i \tag{1.20}
\]

Now, let \( \tau(c; A) \equiv 1 - \frac{x(f^{-1}(c/A))}{x(f^{-1}(c/A))} \) denote the labor wedge, defined as one minus the ratio of the marginal rate of substitution between leisure and consumption \( \chi'/u' \).
to the marginal product of labor $f'$. Since $g'(c; A) = u'(c) - \frac{\chi(f^{-1}(c/A))}{f'(f^{-1}(c/A))}$, it follows that $\tau(c; A) = g'(c; A)/u'(c)$. Using this result and again $v'(M^d(i, c))/u'(c) = i$, and rearranging:

$$i\hat{\mu} - \hat{\mu} = c\tau + i \cdot m \cdot \frac{\partial \log M^d}{\partial \log c} \quad (1.21)$$

It is useful to pause and interpret the terms in the above expression. The costate $\hat{\mu}$ gives the present discounted value, in terms of current consumption, from proportionately increasing consumption at all future dates. This value includes two terms, visible on the right side of (1.21).

The first term, $c\tau$, captures the effect on net utility $g$ from increasing consumption. If, for instance, the labor wedge $\tau$ is positive—meaning that consumption is low relative to the first best—this value is positive, because increasing consumption is beneficial. The second term, $i \cdot m \cdot \frac{\partial \log M^d}{\partial \log c}$, captures the effect on utility from cash. For instance, if $i$ is positive—meaning that cash is low relative to the first best—then this term is positive, because the increase in cash demand induced by a rise in consumption brings the household closer to the first best.

Under the assumption in (1.2) of separable utility from cash, an additional simplification of (1.21) is possible. Differentiating $v'(M^d(i, c)) = iu'(c)$ with respect to $i$ and $\log c$ gives

$$v''(M^d(i, c)) \cdot \frac{\partial M^d}{\partial i} = u'(c) \quad \text{and} \quad v''(M^d(i, c)) \cdot \frac{\partial M^d}{\partial \log c} = -i\sigma^{-1}u'(c)$$

respectively. It follows that

$$\frac{\partial \log M^d}{\partial \log c} = -i\sigma^{-1} \frac{\partial \log M^d}{\partial i}$$

Substituting this identity into (1.21) and applying (1.19) gives

$$i\hat{\mu} - \hat{\mu} = c\tau - i\sigma^{-1} \left(i \cdot m \cdot \frac{\partial \log M^d}{\partial i}\right) = c\tau + i\hat{\mu}$$

36
and cancelling the $i\dot{\mu}$ on both sides, the law of motion (1.21) simplifies to just

$$
\dot{\mu} = -c\tau
$$

This cancellation reflects the equality of two forces in the optimal policy problem: discounting in the law of motion for $\mu$, and the interaction of $\log c$ with the inefficiency in cash demand. Without separable utility from cash, this equality no longer holds, but the results are most likely robust to the presence of complementarities of plausible magnitude. Full cancellation also depends on the assumption of perfectly sticky prices: since discounting depends on the real interest rate while cash demand depends on the nominal interest rate, nonzero inflation would lead to another term, which I will derive once inflation is introduced in section 1.5.1.

To sum up, sticky price equilibrium under optimal policy is characterized by the following system:

$$
\begin{align*}
\dot{c} &= \sigma(i - \rho) \\
\dot{\mu} &= i \cdot m - \frac{\partial \log M^d}{\partial i} \\
\dot{\mu} &= -c\tau \\
\mu(0) &= 0
\end{align*}
$$

The basic tradeoff: demand management versus the Friedman rule. The great advantage of (1.24) is that it permits an especially simple characterization of the optimal policy tradeoff. Integrating (1.24) forward using the initial condition $\mu(0) = 0$ from (1.18) gives

$$
\mu(t) = -\int_0^t c(s) \tau(s) ds
$$

Substituting this into (1.23) gives

$$
\sigma \int_0^t c(s) \tau(s) ds = i \cdot m \cdot \frac{\partial \log M^d}{\partial i}
$$

which characterizes the basic optimal policy tradeoff.
(1.26) can be interpreted as equating the benefits and costs of an decrease in interest rates at time t. Holding consumption from time t onward constant, decreasing \( i(t) \) raises the path of consumption prior to t, providing benefits of \( \sigma \int_0^t c(s)\tau(s) \). If this integral is positive, which (loosely speaking) means that consumption is on average too low over the interval \([0, t]\), then the right side of (1.26) must be positive as well; since the interest semielasticity \( \partial \log M^d/\partial i \) of cash demand is negative, this means that the nominal rate must be negative.

**Smoothing and the natural rate.** Optimal interest rate policy is characterized here by *smoothing*. One striking manifestation of this feature is the continuity of optimal \( \{i(t)\} \).

**Proposition 1.3.4.** Under optimal policy, \( i(t) \) is continuous.

This continuity holds regardless of any discontinuities in the fundamentals \( \rho \) or \( A \). It emerges as a feature of the optimum because (1.26) trades off the benefit from reshaping the overall path of consumption—which changes continuously—against the cost of departing from the Friedman rule.

The costs of departing from the Friedman rule, however, depend on cash’s importance in preferences (1.2). As cash becomes less important, the right side of (1.26) diminishes in magnitude, allowing interest rate policy to more closely match the natural rate.

This can be formalized by introducing the parameter \( \alpha \), and writing

\[
v(m; \alpha) \equiv \alpha \Upsilon(\alpha^{-1}m)
\]

Here, cash demand is proportional to \( \alpha \): \( M^d(i, c; \alpha) = \alpha M^d(i, c; 1) \).

**Proposition 1.3.5.** Under optimal policy, \( i(t) \to r^a(t) \) for all \( t \) as \( \alpha \to 0 \).

Together, these two propositions reflect the two sides of optimal policy: proposition 1.3.4 capturing the tendency toward smoothing, and proposition 1.3.5 showing how this tendency weakens as cash demand shrinks.
Figure 1-7: Optimal $i$ relative to $r^n$, for varying cash demands

Figure 1-7 illustrates the contest between these two forces, by taking a simple example where the natural rate is -1% prior to $t = 0$ and 1% afterward, and considering optimal policy over several different levels of cash demand. In all cases, proposition 1.3.4 holds: despite the discontinuous natural rate, the optimal policy rate varies continuously. Yet the smoothing is much stronger in the $M^d(0) = 1.0$ case than the $M^d(0) = 0.01$ case—and in the latter, policy comes much closer to matching the natural rate.

ZLB-constrained optimal policy. As already discussed, zero has a special role as a benchmark for nominal interest rates: it is the optimal level of rates prescribed by the Friedman rule. Proposition 1.3.3 shows that zero rates are, in fact, optimal in the flexible-price case, where the path of interest rates only affects welfare by changing the level of cash demand. This does not carry over to the sticky-price case, and indeed figure 1-7 provides an example where optimal policy involves both a path for nominal rates with both strictly negative and strictly positive values.

Until recently, however, zero was significant for a different reason: it was the perceived lower bound on nominal interest rates, and central banks did not attempt to target rates beneath it. To consider the effects of this perceived bound, I will define the concept of ZLB-constrained optimal policy. This is identical to the original notion of optimal policy from definition 1.3.2, except that the constraint $i(t) \geq 0$ is exogenously imposed.
Definition 1.3.6. **ZLB-constrained optimal policy under sticky prices** is the choice of \( \{i(t)\}_{t=0}^{\infty} \), along with time-0 consumption \( c(0) \), such that the sticky-price equilibrium characterized by proposition 1.3.1 maximizes household utility (1.2), subject to the constraint that \( i(t) \geq 0 \) for all \( t \).

**Proposition 1.3.7.** ZLB-constrained optimal policy, given \( v \), is identical to (unconstrained) optimal policy under the alternative utility function from cash

\[
\tilde{v}(m) = \begin{cases} 
  v(m) & m \leq m^* \\
  v(m^*) & m \geq m^*
\end{cases}
\tag{1.27}
\]

where \( m^* \) is given by \( v'(m^*) = 0 \).

Proposition 1.3.7 provides one way that ZLB-constrained optimal policy can be interpreted: as optimal policy under an alternative hypothesis about the utility from cash. Figure 1-8 depicts the difference between the original \( v \) and the \( \tilde{v} \) defined in (1.27). The modified utility \( \tilde{v} \) flattens out at \( m^* \), which corresponds to a zero nominal interest rate; since \( \tilde{v}' \) never becomes strictly negative, a strictly negative nominal interest rate is not possible in equilibrium.

It is natural to ask when this implicit misapprehension matters: when is ZLB-constrained optimal policy different from unconstrained optimal policy—or, equivalently, when does unconstrained optimal policy feature negative nominal interest
rates? The next proposition provides a simple characterization in terms of the ZLB-constrained optimal policy.

**Proposition 1.3.8 (Optimality of negative rates).** Unconstrained optimal policy features negative nominal rates if and only if under ZLB-constrained optimal policy, there is some $t$ for which

$$
\int_0^t c^{ZLB}(s) r^{ZLB}(s) ds > 0
$$

According to proposition 1.3.8, negative rates are optimal when the ZLB-constrained solution features, at any time $t$, a positive (consumption-weighted) average labor wedge between 0 and $t$. Loosely speaking, this means that negative rates are optimal if there is any $t$ at which the economy has on average, to date, been in a slump rather than a boom.

To build intuition for this result, take some $t$ where (1.28) holds, and consider a small downward perturbation $-\Delta i$ to the interest rate over the small interval $[t, t+\Delta t]$. The welfare impact of this perturbation working through the path of consumption is approximately

$$
\left( \int_0^t c^{ZLB}(s) r^{ZLB}(s) ds \right) \Delta i \Delta t > 0,
$$

which is positive and first order in $\Delta i$.

If $i(t) > 0$, this perturbation also brings us closer to the Friedman rule and is therefore unambiguously optimal—contradicting the assumption that we start at the ZLB-constrained optimal policy.

Consider alternatively the case where $i(\cdot) = 0$ on the interval $[t, t + \Delta t]$. Here we start at the Friedman rule, and the downward perturbation to interest rates moves us away from it—but the cost of this deviation is second order in $\Delta i$, at approximately

$$
- \left( \int_t^{t+\Delta t} \frac{\partial \log M^d(0, c^{ZLB}(s))}{\partial i} ds \right) \frac{1}{2} (\Delta i)^2 \Delta t < 0
$$

For sufficiently small $\Delta i$, the first-order benefit in (1.29) from increasing consumption over the interval $[0, t]$ dominates the second-order cost in (1.30) from deviating
from the Friedman rule over the interval \([t, t + \Delta t]\). Hence a perturbation toward negative rates offers a welfare gain, and ZLB-constrained optimal policy does not coincide with unconstrained optimal policy.

The foundation of this argument is the fact that the zero lower bound coincides with the Friedman rule. Pushing interest rates below zero, assuming that cash demand does not become infinite, creates a distortion—but since zero is the Friedman rule optimum, the resulting welfare loss is second-order. As long as negative rates bring the economy closer to an optimal level of output, creating first-order benefits, they are warranted.

Although proposition 1.3.8 is conceptually important, the condition (1.28) may be unwieldy to verify. The following corollary offers a much simpler test for when negative rates are optimal.

**Corollary 1.3.9.** Unconstrained optimal policy features negative nominal rates if and only if under ZLB-constrained optimal policy, there is some \(t\) for which \(i^{ZLB}(t) = 0\) and \(\tau^{ZLB}(t) \neq 0\).

This corollary demonstrates that negative rates are generically optimal whenever ZLB-constrained optimal policy features zero interest rates. The only exception is when consumption is at precisely its first-best level, \(\tau^{ZLB}(t) = 0\), for all \(t\) where \(i^{ZLB}(t) = 0\).

The following proposition expands upon propositions 1.3.8 and 1.3.9 by offering a striking, novel characterization of how ZLB-constrained policy and unconstrained policy differ.

**Proposition 1.3.10.** If unconstrained optimal policy features negative nominal rates, then \(i^{ZLB}(t) \leq \max(i(t), 0)\), with strict inequality whenever \(i(t) > 0\).

In short, if the zero lower bound constraint is ever binding, then a ZLB-constrained policymaker optimally sets interest rates lower at every \(t\) where it is feasible to do so. These lower rates are used to compensate for the higher-than-optimal rates during periods when the zero lower bound binds.
1.4 Revisiting ZLB traps

Following the general results in section 1.3, in this section I consider a more specific scenario: a zero lower bound "trap", featuring a negative natural real rate over some period of time.

Specifically, I suppose that in the interval \([0, T]\)—the "trap"—the natural rate takes some strictly negative value \(-\bar{r}\), followed by a return to a nonnegative steady state value \(r^{ss} \geq 0\).

\[
    r^n(t) = \begin{cases} 
    -\bar{r} & 0 \leq t < T \\
    r^{ss} & T \leq t 
    \end{cases}
\]

To start, I assume that \(r^{ss} = 0\), which greatly simplifies characterization of the solution and facilitates some useful analytical results. This path for the natural rate is depicted in figure 1-9. I will later consider the case where \(r^{ss} > 0\) in section 1.4.4.11

Exercises of this form are ubiquitous in the literature on the zero lower bound—for instance, a stochastic trap is in Eggertsson and Woodford (2003), and a deterministic trap similar to my own is in Werning (2011). The negative natural rate trap is popular because it epitomizes the problems created by the zero lower bound: when interest rates cannot be set low enough to match the natural rate, the level of output during the trap—relative to the first-best level—must fall below the level expected after the trap. If the central bank is expected to target first-best output after the trap, then this means that output during the trap is inefficiently low: there is a zero lower bound recession. If, on the other hand, the central bank can commit at the beginning of the trap to policy after the trap, then it optimally engages in "forward guidance"—using low interest rates to generate a boom once the trap is over, lifting up the level of economic activity during the trap as well.

Once negative rates are available as a policy tool, however, this standard analysis

\footnote{For simplicity, I will assume that productivity is constant, so that \(r^n(t) = \rho(t)\) according to (1.13). Variation in the time preference \(\rho(t)\) of the representative household can be interpreted as reflecting variation in the effects of idiosyncratic uncertainty and incomplete markets in an underlying heterogeneous-agents model; see, for instance, Werning (2015). For the dynamics of interest rates and the output gap, to a first-order approximation it does not matter whether variation in \(r^n(t)\) is driven by time preference \(\rho(t)\) or productivity growth \(A(t)/A(t)\), but the assumption that \(r^n(t) = \rho(t)\) is needed for the fully nonlinear solution here.}
of the trap no longer applies. The central bank can now, in principle, set rates to match the natural rate at every point—but given the costs of setting negative rates, of course, this policy is not optimal. My goal in this section is to study the structure of optimal policy under negative rates in detail and contrast its outcomes with the traditional ZLB-constrained policy, with a particular focus on the extent to which negative rates can close the welfare gap relative to the first best.

1.4.1 Calibration

Now that I am interested in a more quantitative analysis, I need to specify a calibration of the model. Aside from cash, there are three parameters: the elasticity of intertemporal substitution \( \sigma \) and Frisch elasticity \( \psi \) in (1.3), and the elasticity of output \( 1 - \alpha \) with respect to labor input in (1.8).

Since \( 1 - \alpha \) is also the labor share in the model, I calibrate it on this basis at \( 1 - \alpha = 0.56 \), to match the labor share of factor income in the United States in 2014.\(^{12}\) I calibrate the Frisch elasticity of labor supply to be \( \psi = 0.86 \), reflecting the Frisch elasticity for aggregate hours obtained from studies with micro identification in the Chetty, Guren, Manoli and Weber (2013) meta-analysis. Finally, I calibrate the elasticity of intertemporal substitution to be \( \sigma = 0.50 \), which the meta-analysis in Havránek (2013) identifies as the mean value in the literature, and Hall (2009)

\(^{12}\)This is taken from NIPA Table 1.10, as the ratio of compensation of employees to gross domestic income minus production taxes net of subsidies.
Table 1.1: Calibration

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Elasticity of intertemporal substitution $\sigma = 0.5$</td>
<td>Havránek (2013), Hall (2009)</td>
</tr>
<tr>
<td>Frisch elasticity of labor supply $\psi = 0.86$</td>
<td>Chetty et al. (2013)</td>
</tr>
<tr>
<td>Elasticity of output to labor input $1 - \alpha = 0.56$</td>
<td>Labor share in US, 2014</td>
</tr>
<tr>
<td>Cash demand at 0% interest rates $m^* = 0.075$</td>
<td>Cash/GDP in US, 2014</td>
</tr>
<tr>
<td>Interest semielasticity of cash demand $\frac{\partial \log M^d/\partial i}{5}$</td>
<td>Ball (2001)</td>
</tr>
</tbody>
</table>

describes as the "most reasonable" choice for the parameter.\textsuperscript{13}

My calibrated functional form for the utility from cash is given by

$$v'(m) = -\frac{u'(c^*)}{b} \log \left(\frac{m}{m^*}\right)$$  \hspace{1cm} (1.31)

This function implies a roughly constant interest semielasticity $\frac{\partial \log M^d/\partial i}{5}$ of cash demand, which is exactly constant and equal to $-b$ when consumption $c$ is at its first-best level $c^*$. Cash demand at $i = 0$ is given by $m^*$, which I calibrate to match the current ratio of cash in circulation to GDP in the United States (where $i \approx 0$), which is 0.075.\textsuperscript{14}

The literature on the interest semielasticity has produced varied results, and it generally looks at demand for M1—including both cash and demand deposits—rather than isolating cash. I tentatively adopt the estimate from Ball (2001), drawn from the postwar United States, of an interest semielasticity equal to $-5$. Although this estimate does not cover a period with negative interest rates, it nevertheless appears consistent with the behavior of Swiss cash demand in response to negative rates as displayed in figure 1-3: with a semielasticity of $-5$, the recent drop in Swiss target rates from 0% to -0.75% would be expected to produce just shy of a 4% increase

\textsuperscript{13}Havránek (2013) argues that this mean value is inflated somewhat due to publication bias. On the other hand, since I am interpreting consumption $c$ in this model as a measure of the overall level of economic activity—which also includes the much more interest-sensitive category of fixed investment—the relevant EIS here should be higher than the estimates obtained in the literature for private consumption alone. There are also aggregate redistributive effects that boost the consumption response to interest rates, as identified in Auclert (2015). Altogether, for my benchmark calibration, I assume that these biases roughly offset each other.

\textsuperscript{14}I am normalizing first-best aggregate output, $c^*$, to 1.
in cash demand, similar to the slight increase in cash demand actually observed in Switzerland relative to trend.

Table 1.1 summarizes this calibration. My benchmark scenario features a natural rate of \(-\bar{r} = -2\%\) during the trap, and a trap length of \(T = 4\). This is intended to generate a moderately severe recession under ZLB-constrained policy, with output starting the trap at 4\% below its first-best level, as seen in the next section.

1.4.2 Partial commitment case

To facilitate comparison with the zero lower bound literature—which often emphasizes the case where the monetary authority lacks commitment—I will start by modifying the assumption of full commitment from section 1.3. Dropping commitment entirely, however, is not a viable option in this environment: in the limit where policy is continually reoptimized, nominal interest rates are simply set to 0\% at all times to satisfy the Friedman rule.

Instead, I consider a simple case with partial commitment, where optimal policy is reoptimized at \(t = T\), the end of the trap. Figure 1-10 shows the results. Under both ZLB-constrained and unconstrained policy, the nominal rate is set to zero—equal to the natural rate—following \(T = 4\), and consumption is stabilized at its natural level, thereby achieving the first best from \(T = 4\) onward.

Prior to \(T = 4\), a recession ensues in both cases, and \(i(t)\) exceeds the natural rate for all \(t\). Unsurprisingly, however, this recession is far more severe in the ZLB-constrained case; with negative rates, the gap between \(i(t)\) and the natural rate \(-\bar{r}\) shrinks substantially.

The path \(i(t)\) of interest rates in the unconstrained case in figure 1-10 exhibits three distinctive features that are, in fact, general—as summarized by the following proposition.

**Proposition 1.4.1.** In the trap under partial commitment, \(i(0) = 0\), \(i(T^-) > -\bar{r}\), and \(i(t)\) is strictly decreasing on the interval \([0, T)\).

All three features of \(i(t)\) result from the central tradeoff in the model—the tradeoff
Figure 1-10: Optimal policy under partial commitment: with and without ZLB

Output gap: $\log(c/c^*)$

between the costs of departing from the Friedman rule by setting negative rates and the benefits of increasing consumption during the trap.

A lower rate at time $t$ raises consumption over the interval $[0, t]$, and the overall welfare gains naturally depend on the length of this interval. At $t = 0$, the length is zero, implying that the tradeoff is resolved entirely in favor of the Friedman rule: $i(0) = 0$. As $t$ increases, the benefits grow relative to the costs, implying that optimal $i(t)$ decreases. But $i(t)$ does not decrease so steeply that it falls below $\bar{r}$: if it did, there would be a boom in the latter part of the trade episode $[0, T]$, and the policymaker could increase welfare by smoothing the path of $i(t)$ and eliminating this boom.

This backloading of the most negative rates is an important feature of optimal policy. This differs sharply from the path of $i(t)$ implied by, for instance, a Taylor rule where $i$ depends on the output gap—which would be strictly increasing. Hence, although the case for negative rates is often illustrated by an appeal to the negative rate implied by a Taylor rule\textsuperscript{15}, this view misses a distinctive feature of the optimal policy.

\textsuperscript{15}See, for instance, Rudebusch (2010), contrasting the path of nominal interest rates implied by a simple linear policy rule—which falls to nearly 6% in 2009–10—with the actual ZLB-constrained path.
To what extent do negative rates close the utility gap? Let $V_{ZLB}^*$ denote household utility (1.2) under ZLB-constrained optimal policy, $V^*$ denote household utility under unconstrained optimal policy, and $V_{FB}^*$ denote household utility in the first best.

One natural measure of how unconstrained optimal policy improves upon ZLB constrained optimal policy is the extent to which it shrinks the gap in utility relative to the first best: the ratio $(V^* - V_{FB}^*)/(V_{ZLB}^* - V_{FB}^*)$. As displayed in figure 1-10, in the benchmark scenario under partial commitment this ratio is extremely small, at 5.6%: the ability to set negative rates eliminates the vast majority of the utility shortfall.

This ratio can be analytically characterized as a function of primitives, up to a second-order approximation, as revealed by the following proposition.

**Proposition 1.4.2.** The following is a second-order approximation for the decline in the welfare gap:

$$\frac{V^* - V_{FB}^*}{V_{ZLB}^* - V_{FB}^*} = 3 \left( \frac{1}{T\sqrt{A}} \right)^2 + O \left( \left( \frac{1}{T\sqrt{A}} \right)^3 \right)$$

where

$$A \equiv \sigma^2 \times \left( \frac{\partial \log \tau(c)}{\partial \log c} \right) \times \frac{\partial \log \lambda(0,c^*)}{\partial i} \times m^*$$

(1.32) implies that the decline in the welfare gap is more dramatic when $T$ is large. This is no surprise: when the trap is longer, negative rates can lift output over a longer period, making them a more useful tool.

The decline is also increasing in the composite parameter $A$. As (1.33) reveals, $A$ is increasing in the elasticity of intertemporal substitution $\sigma$ (which determines the influence of negative rates on consumption) and the elasticity $\partial \log \tau(c)/\partial \log c$ of the labor wedge with respect to consumption (which determines the magnitude of welfare loss from the output gap). It is decreasing in the interest semielasticity of cash demand and the level of cash demand $m^*$ at $i = 0$, which as depicted in figure 1-6 determine the costs from deviating from the Friedman rule.
In addition to these qualitative insights, (1.32) is remarkably accurate from a quantitative standpoint, as long as $T \sqrt{A}$ is relatively large. In the benchmark parameterization above, for instance, the approximation is

$$3 \left( \frac{1}{T \sqrt{A}} \right)^2 = 3 \times \frac{\partial \log \mathcal{M}^*(0, c^*)}{\partial \bar{h}} \times m^* = 3 \times \frac{5 \times 0.075}{4^2 \times 0.5^2 \times 4.86} = 0.058 \quad (1.34)$$

which is very close to the actual ratio 0.056 obtained in the simulation.

It is clear from (1.34) how various features of the calibration in section 1.4.1 contribute to closing the utility gap. A higher interest semielasticity of cash demand, for instance, would result in a larger utility gap under unconstrained policy—but even if the semielasticity of 5 was replaced by 25, negative rates would still cut the utility gap relative to ZLB-constrained policy by over two-thirds.

### 1.4.3 Full commitment case

Now I revert to the original assumption of full commitment. Figure 1-10 then becomes figure 1-11.

The path of ZLB-constrained optimal policy does not change. Since the natural rate after the trap is zero, it is not possible to create a boom by committing to hold
rates below the natural rate after the trap, as in Eggertsson and Woodford (2003) or Werning (2011).

The path of unconstrained optimal policy, meanwhile, involves negative rates even after the trap ends at $T$. This leads to a boom in consumption in the neighborhood of time $T$, which pulls up the entire path of consumption over the interval $[0, T]$ and brings output closer to its first-best level. The relevant features of the solution are general, and summarized in the following proposition.

**Proposition 1.4.3.** The optimal solution under full commitment features a path:

- for $i(t)$ that is decreasing from $0$ to $T'$ (where $T' < T$), reaches a minimum at $T'$, and is increasing from $T'$ onward, such that $i(0) = 0$, $i(t) \to 0$ as $t \to \infty$, and $i(t) > -\bar{r}$ for all $t$.

- for $c(t)$ that is increasing from $0$ to $T$ and decreasing from $T$ onward, with $c(0) < c^*$, $c(T) > c^*$, and $c(t) \to c^*$ as $t \to \infty$.

The key insight from the full commitment case is that negative rates are not just an alternative to forward guidance. Instead, the two are complements, in the sense that it is optimal to do forward guidance with negative rates. Quantitatively, however, this is not of great importance: as figure 1-11 reveals, with full commitment the utility gap is reduced to 5.1% of its ZLB-constrained level, not much better than the 5.6% achieved with partial commitment in figure 1-10.

1.4.4 Full commitment case, with a positive natural rate after the trap

The assumption that $r^{ss} = 0$, and therefore that $r^{ss}(t) = 0$ for all $t \geq T$, simplified characterization of optimal policy, but it made forward guidance in the ZLB-constrained case impossible. I now relax this assumption, considering $r^{ss} > 0$ instead. Figure 1-12 shows the paths that result when $r^{ss} = 1.5\%$.\(^{16}\)

\(^{16}\)The utility gap measure reported in previous cases is no longer meaningful, since the assumption that the steady-state natural rate is above zero implies that there is inevitably a departure from first-best utility in the steady state, which creates a large wedge between actual intertemporal utility (1.2), over the interval $[0, \infty)$, and the first best.
Letting $i(t)$ and $c(t)$ denote interest rates and consumption with unconstrained optimal policy, and $i_{ZLB}(t)$ and $c_{ZLB}(t)$ denote these with ZLB-constrained optimal policy, the following proposition summarizes the general features of the solution.

**Proposition 1.4.4.** In the optimal solution under full commitment, $i(t)$ starts below $i_{ZLB}(t)$ and crosses it once. $c(t)$ starts above $c_{ZLB}(t)$ and crosses it once.

As figure 1-12 depicts, outcomes under ZLB-constrained policy and unconstrained policy are now qualitatively similar. Interest rates are set below the natural rate after the trap, generating a boom (in which output exceeds its first-best level) in the neighborhood of $T = 4$. Interest rates exceed the natural rate during the trap, leading to a bust (in which output falls short of its first-best level) for the majority of the trap.

The key difference between the two cases is quantitative: the output gap is much closer to zero, and less variable, when negative rates are available. At the same time, interest rates are much more volatile. Effectively, once the zero lower bound constraint is lifted, optimal policy accepts more interest rate volatility in order to stabilize output. In contrast, under the zero lower bound, rates are artificially stable, as policymakers are forced to maintain $i = 0$ for a prolonged period after the trap in order to lift economic activity during the trap.
1.5 Interaction with other policies

1.5.1 Trend inflation

One limitation of the framework in the model thus far is the simplifying assumption of perfectly sticky prices. With inflation fixed at zero by assumption, it is impossible to evaluate the common idea—proposed by Blanchard et al. (2010), and evaluated formally by Coibion et al. (2012) and others—that higher trend inflation alleviates the limitations on policy imposed by the zero lower bound.

In this section, I relax the assumption of perfect price stickiness, allowing for nonzero inflation. To preserve the parsimony of the model, however, I continue to make a strong assumption on prices: I replace sticky prices with sticky inflation, where the path of prices is constrained to take the form \( P(t) = e^{\pi t} P(0) \) for some trend inflation rate \( \pi \). This embeds my earlier case, which corresponds to \( \pi = 0 \). It is intended to capture the role of trend inflation in the simplest way possible, and it can also be understood as a stylized representation of well-anchored inflation expectations under an inflation targeting regime.

The Euler equation characterizing the path of consumption becomes

\[
\frac{\dot{c}}{c} = \sigma (i - \pi - \rho)
\]  

(1.35)

Optimal policy is now characterized by the system

\[
\hat{\mu} \sigma = i \cdot m \cdot \frac{\partial \log M^d}{\partial i}
\]  

(1.23)

\[
\dot{\hat{\mu}} = -cT - \pi \hat{\mu}
\]  

(1.36)

\[
\hat{\mu}(0) = 0
\]  

(1.25)

where (1.23) and (1.25) are unchanged from the formulation in section 1.2.2, while (1.36) is a modified version of (1.24) with the additional term \( \pi \hat{\mu} \). When comparing optimal policy under distinct trend inflation rates \( \pi \), I will make the dependence on \( \pi \) explicit by introducing it as an argument: for instance, \( i(t; \pi) \).
As trend inflation \( \bar{\pi} \) increases, the average level of nominal interest rates over time has to increase along with it; otherwise, the Euler equation implies that consumption will diverge unboundedly to either 0 or \( \infty \). But this does not imply a parallel shift in the path of nominal interest rates \( i(t) \): optimal policy accommodates inflation in a way that minimizes the cost of deviating from the Friedman rule.

For instance, figure 1-13 shows how the optimal path of nominal interest rates \( i(t) \) from the scenario in section 1.4.3 changes as trend inflation is set to \( \bar{\pi} = 1\% \).\(^{17}\) In the ZLB-constrained case, rates do not change at all during and immediately after the trap, staying at zero; instead, the planner takes advantage of trend inflation in order to implement a lower real interest rate and lessen the severity of the recession. In the unconstrained case, \( i(0) \) remains at 0\%, although rates shift up by approximately \( \bar{\pi} \) at later \( t \).

Given the subtlety of the adjustment in figure 1-13, is it possible to make any general statements about how optimal policy responds to trend inflation? The following proposition shows that it is.

**Proposition 1.5.1.** The path of optimal nominal interest rates under optimal policy

\(^{17}\)Note that the dashed line in figure 1-13 shows the nominal natural rate \( r^n(t) + \bar{\pi} \), which is shifted up when \( \bar{\pi} > 0 \).
is, for all $t > 0$, strictly increasing in trend inflation:

$$\bar{\pi}' > \bar{\pi} \implies i(t; \bar{\pi}') > i(t; \bar{\pi})$$

In short, although the response of optimal nominal interest rates to rising trend inflation is uneven—with the Fisher equation holding only on average—it is unambiguously true that interest rates will rise at each $t$.

**Optimal trend inflation.** How do different levels of $\bar{\pi}$ affect household utility under optimal policy—and, in particular, how does this differ depending on whether the zero lower bound is imposed? Let $W(\bar{\pi})$ denote household utility (1.2) under unconstrained optimal policy, and $W_{ZLB}(\bar{\pi})$ denote the same under ZLB-constrained optimal policy.

**Proposition 1.5.2.** There exists some $\bar{\pi}_u$ such that $W_{ZLB}(\bar{\pi}) - W(\bar{\pi})$ is strictly increasing in $\bar{\pi}$ for all $\bar{\pi} < \bar{\pi}_u$, and $W_{ZLB}(\bar{\pi}) = W(\bar{\pi})$ for all $\bar{\pi} \geq \bar{\pi}_u$.

This proposition states that the zero lower bound constraint is complementary to higher inflation: as long as the constraint is binding, the utility cost of imposing the constraint becomes smaller as trend inflation rises. Eventually, a high enough level of inflation $\pi_u$ is reached that the zero lower bound is no longer binding at all, and $W_{ZLB}(\pi_u) = W(\pi_u)$. This follows from the usual intuition: higher inflation allows more negative real rates to be realized without setting negative nominal rates.

Let $\pi^* = \arg\max W(\bar{\pi})$ and $\pi_{ZLB}^* = \arg\max W_{ZLB}(\bar{\pi})$ be the levels of inflation that maximize household utility. If we interpret the sticky inflation rate $\bar{\pi}$ as the result of a long-term policy of anchoring inflation expectations, then these are the policymaker’s optimal choices of trend inflation, as characterized by the following corollary to proposition 1.5.2.

**Corollary 1.5.3.** Optimal inflation with unconstrained policy is below optimal inflation with the zero lower bound: $\pi^* \leq \pi_{ZLB}^*$. This inequality is strict, and $\pi^* < \pi_{ZLB}^* < \pi_u$, except in the special case $r^*(t) = \bar{r}$ where the natural rate is constant over all $t$, in which case $\pi^* = \pi_{ZLB}^* = \pi_u = -\bar{r}$.
In general, optimal inflation under unconstrained policy is lower than optimal inflation under ZLB-constrained policy ($\pi^* < \pi^*_{ZLB}$), because inflation is necessary to achieve negative real rates when there is a zero lower bound.

The corollary also states that $\pi^*_{ZLB} < \pi_u$; in other words, that inflation should not be set so high as to negate the zero lower bound constraint altogether. For levels of $\pi$ immediately below $\pi_u$, the zero lower bound barely binds, and the cost imposed by the constraint is second-order—which is overridden by the first-order benefits from lowering the trend rate of inflation and thereby bringing the average level of nominal interest rates slightly closer to the Friedman rule optimum.

The special case $r^n(t) = \tilde{\pi}$ where the natural rate never varies is an exception, and it features particularly simple optimal policy. By setting trend inflation equal to minus the natural rate, the policymaker can match the natural rate while setting the nominal interest rate equal to zero at all times. This achieves the first best: it implements a first-best level of output while also satisfying the Friedman rule.

The crucial message of corollary 1.5.3 is that once negative nominal interest rates become available as a tool, trend inflation should be brought down. This weakens a longstanding argument—dating back at least to Summers (1991) and Fischer (1994)—that low inflation is dangerous due to the limitations of interest rate policy. It also hints at a source of political appeal: although negative rates may not seem attractive on their own, they can facilitate a low inflation target, which is a broadly popular idea.

1.5.2 Abolishing cash

As discussed in section 1.1, one policy that has been commonly suggested as a response to the zero lower bound is the abolition of cash—see, for instance, Rogoff (2014) for details. Once negative rates are permissible, this original rationale no longer holds in the same form.

Since negative rates are not costless, however, it remains possible that abolishing cash is optimal, if the cost from the subsidy to cash under negative rates exceeds the benefits from having cash. Figure 1-14 depicts this possibility. There is some level
of cash $\bar{m}$ at which $v(\bar{m}) = v(0) = 0$, beyond which it would be optimal to replace $m > \bar{m}$ with $m = 0$.

Figure 1-14 additionally shows the interest rate $\bar{i} < 0$ such that $\bar{iu}'(c^*) = v'(\bar{m})$, and therefore $\bar{m} = M^d(\bar{i}, c^*)$, where $c^*$ is the first-best level of consumption. If consumption is at its first-best level, and the nominal interest rate is kept at or above $\bar{i}$, flow utility is higher when cash is kept rather than abolished.

This observation gives rise to the following proposition, which offers a simple sufficient condition for keeping cash (rather than abolishing it) to be optimal given some sequence of fundamentals.

**Proposition 1.5.4.** Let $c^* = \min_t c^*(t)$, where $c^*(t)$ is the first-best level of consumption at time $t$, and write $\bar{i} \equiv v'(\bar{m})/u'(c^*)$. Then it is optimal to keep cash if the natural nominal rate is bounded from below by $\bar{i}$: $\bar{i} + r^n(t) \geq \bar{i}$ for all $t$.

**Proof.** One feasible policy is to set $i(t) = \bar{i} + r^n(t)$ and $c(t) = c^*(t)$ for all $t$. Then since $i(t) = \bar{i} + r^n(t) \geq \bar{i}$,

$$m(t) = M^d(i(t), c^*(t)) \leq M^d(\bar{i}, c^*) = \bar{m}$$

and therefore $v(m(t)) \geq v(\bar{m}) = v(0)$. It follows that this policy achieves weakly better utility than that from the no-cash $m = 0$ case. $\square$

The idea behind proposition 1.5.4 is straightforward. Suppose that the natural rate is always above the interest rate threshold $\bar{i} < 0$ at which the cumulative utility from cash might become negative. Then one option for the policymaker is to set the nominal interest rate to equal the natural nominal rate at all $t$, achieving the first-best level of economic activity at all $t$ while never exceeding the level $\bar{m}$ at which $v(\bar{m}) = v(0)$. With this option available, the policymaker would never opt to abolish cash.

What is a plausible level of $\bar{i}$? This depends on the shape of $v$, which can be mapped onto measurable objects like the interest semielasticity.
Lemma 1.5.5. Suppose that $M^d(i, c^*)$ has a constant semielasticity of $-b$ with respect to $i$. Then $\hat{i} = -1/b$.

Lemma 1.5.6. Let $v$ and $\tilde{v}$ be two alternative utility functions for cash, with corresponding cash demand schedules $M^d$ and $\tilde{M}^d$. Suppose that the interest semielasticity of cash demand for $M^d$ is always smaller than that of $\tilde{M}^d$:

$$\frac{\partial \log M^d(i, c)}{\partial i} < \frac{\partial \log \tilde{M}^d(i, c)}{\partial i} \text{ for all } i, c$$

Then $\hat{i} < \tilde{i}$.

Lemma 1.5.5 shows how a constant semielasticity maps onto the lower bound $\hat{i}$ in proposition 1.5.4. Note that $\hat{i}$ is more negative when the semielasticity is smaller. Intuitively, this is because when cash demand is less interest-elastic, the inframarginal benefits of cash are larger compared to the marginal losses, and the level of cash demand has to rise higher—corresponding to a much more negative interest rate—until these inframarginal benefits are wiped out. Lemma 1.5.6 formalizes this point in the general case, showing that when cash demand is less interest elastic, $\hat{i}$ is lower.

Calibrating lemma 1.5.5 using the values for the interest semielasticity from section 1.4.1, the implied $\hat{i}$ is extremely low. Given $b = 5$, for instance, $\hat{i} = -20.0\%$. 

57
In practice, the semielasticity probably becomes much higher for such low nominal interest rates: indeed, at \(i = -20.0\%\), cash demand may be virtually infinite.

But even with a much higher semielasticity—intended to reflect the more elastic response of cash demand in the relevant range of negative interest rates—\(i\) is still quite low. At \(b = 50\), for instance, lemma 1.5.5 implies \(i = -2\%\); and since the empirical semielasticity is much lower than this when nominal interest rates are positive or mildly negative, a more precise calculation would obtain a lower \(i\). Regardless, based on estimates in the literature, it seems plausible that the natural nominal rate \(\bar{r} + r^n(t)\) has always, or almost always, been above \(-2\%\) in the United States.\(^\text{18}\) Proposition 1.5.4 would then indicate that it is not optimal to abolish cash.

More generally, proposition 1.5.4—in conjunction with lemmas 1.5.5 and 1.5.6—tells us that the path of the natural rate and the semielasticity of cash demand are key considerations in calculating whether or not to keep cash. Since these are both matters of some empirical controversy, further work will be needed to obtain a more definitive answer.

### 1.5.3 Multiple denominations

One alternative to abolishing cash altogether is to selectively eliminate certain denominations of cash. A number of observers have proposed eliminating high-value denominations (for instance, the $100 bill) to curtail cash demand under negative rates, with the reasoning that these denominations have lower holding costs per unit value and are thus particularly likely to be hoarded. I evaluate this idea by extending the model such that cash comes in more than one denomination.

For simplicity, I consider a case with only two denominations, letting \(z(m_h, m_l)\) de-

\(^{18}\)Côrdia (2015) estimates that the \(r^n(t)\) reached a minimum of slightly above \(-4\%\) in the aftermath of the Great Recession, which corresponds to a natural nominal rate \(\bar{r} + r^n(t)\) of slightly above \(-2\%\) under trend inflation of \(\bar{r} = 2\%\). Del Negro, Giannoni, Cocci, Shahanaghi and Smith (2015) directly provide figures for the natural nominal rate and show that it fell to slightly above \(-2\%\) for much of the Great Recession, though it dipped below \(-2\%\) for a brief period around the beginning of 2013. Laubach and Williams (2015)—using a statistical model rather than the structural models in the other two papers—find much higher natural rates, with a real natural rate of roughly 0% with the benchmark methodology and a real natural rate of slightly below \(-2\%\) using an alternative methodology for estimating the output gap.
note the utility from holding $m_h$ in high-denomination cash and $m_l$ in low-denomination cash. I then define

$$v_{bd}(m) \equiv \max_{m_h + m_l = m} z(m_h, m_l)$$

$$v_{hd}(m) \equiv z(m, 0)$$

$$v_{ld}(m) \equiv z(0, m)$$

Here $v_{bd}$ denotes the utility from having cash $m$ that can be spread between both denominations, while $v_{hd}$ denotes the utility from cash only in the high denomination and $v_{ld}$ denotes the utility from cash only in the low denomination. I take these as representing alternative policy choices: when setting up the monetary system, the government can choose to provide both denominations or only one. Whatever choice is adopted, the household utility function (1.2) then uses the corresponding $v$.

Let $M_{bd}^d$, $M_{hd}^d$, and $M_{ld}^d$ be the corresponding demand functions for cash. I then make the following assumption.

**Assumption 1.5.7.** When $i = 0$, cash demand is highest when both denominations are available and lowest when only the low denomination is available:

$$M_{ld}^d(0, \cdot) < M_{hd}^d(0, \cdot) < M_{bd}^d(0, \cdot)$$

Furthermore, for any $c$,

$$\frac{M_{bd}^d(i, c)}{M_{hd}^d(i, c)} \quad \text{and} \quad \frac{M_{ld}^d(i, c)}{M_{hd}^d(i, c)}$$

are strictly increasing functions in $i$.

Both parts of assumption 1.5.7 are natural for a model with multiple denominations. The second part, in particular, captures a key distinction between high and low denominations: since the holding costs for high denominations are lower, the cost of foregone nominal interest $i$ looms larger, making demand for these denominations more elastic with respect to the nominal interest rate. (For instance, the nominal

---

19I show in the Online Appendix how a utility function from cash of this form can be microfounded.
interest rate is much more important when deciding how many $100 bills to carry than when deciding how many $1 bills to carry.)

The following example shows how a simple functional form for $z$ can produce cash demand functions that satisfy assumption 1.5.7. I show in the Online Appendix how such a functional form can be microfounded in a model of cash transactions.

**Example 1.5.8.** Suppose that

$$z(m_h, m_l) = \frac{A_h m_h^{1-\zeta}}{1-\zeta} + \frac{A_l m_l^{1-\zeta}}{1-\zeta} - \alpha_h m_h - \alpha_l m_l$$

where $\alpha_h < \alpha_l$ and $A_h/\alpha_h > A_l/\alpha_l$. Then the conditions of assumption 1.5.7 are satisfied.

Now define $W_{bd}$, $W_{hd}$, $W_{ld}$, and $W_{nd}$ to be household utility (1.2) under optimal policy with both denominations, the high denomination only, the low denomination only, and no cash, respectively.

**Proposition 1.5.9.** If eliminating the low denomination increases utility under optimal policy, then utility is increased further by either eliminating the high denomination instead, or abolishing cash altogether:

$$\text{if } W_{hd} > W_{bd}, \text{ then } W_{ld} > W_{hd} > W_{bd} \text{ and } W_{nd} > W_{hd} > W_{bd}$$

This proposition states that it is never optimal to eliminate only the low denomination: if this leads to an improvement, then it is even better to either eliminate the high denomination or to abolish cash.

The key force driving proposition 1.5.9 is the assumption of increasing ratios (1.37). Demand for the high denomination is relatively greater when interest rates are lower. The only reason to eliminate denominations is to cut down on excessive cash demand when interest rates are negative; hence, if anything, the high denomination should be axed, since its demand increases most disproportionately at these negative rates.
1.6 Conclusion

This paper studies, for the first time, the use of negative nominal interest rates as part of optimal monetary policy—without any major changes to the monetary system. I show that negative rates are costly in this environment because they imply an inefficient subsidy to cash, violating the Friedman rule in the opposite of the traditional direction. I replace the zero lower bound, which is imposed as an ad-hoc constraint in many New Keynesian models, with a more flexible, microfounded tradeoff between the distortionary costs of negative rates and their benefits in raising aggregate demand.

The first insight that emerges from this framework is that when the economy is, on average, in a slump, negative rates are always optimal to some degree: the first-order benefits of boosting aggregate output outweigh the second-order costs of deviating from the Friedman rule optimum. An effective zero lower bound only emerges in a limit case, where cash demand becomes infinitely elastic at zero—a case that is not consistent with recent evidence from currencies with negative rates.

Revisiting the liquidity trap scenarios that are studied in the zero lower bound literature, I show that negative rates bring significant improvements. In a benchmark scenario, introducing negative rates brings utility 94% closer to the first best. Optimal policy dictates that the most negative rates should be backloaded, and indeed that negative rates should continue even after the trap has ended. Although ZLB-constrained policy can mitigate the worst consequences of the trap through forward guidance, negative rates facilitate a path for output that is much closer to the first best.

Negative rates are most useful as a tool when cash demand is relatively inelastic, because this is when the distortion from violating the Friedman rule is least severe. Policies that can constrain cash demand in this region are therefore important complements to negative rates. In the most extreme case, cash can be abolished; but more limited measures, such as the elimination of larger denominations, may also be worthwhile. An important task of future research will be to study the properties of cash demand at very low interest rates, and to search for new measures that can
contain this demand—ensuring that negative rates realize their great potential as a policy tool.

1.7 Appendix: Proofs

Proof of proposition 1.5.1

Proof. Since \(i\) is strictly increasing in \(\dot{\mu}\) by (1.23), it suffices to show that \(\dot{\mu}(t; \pi') > \dot{\mu}(t; \bar{\pi})\) for all \(t > 0\).

Suppose to the contrary that the set \(A = \{\dot{\mu}(t; \pi') \leq \dot{\mu}(t; \bar{\pi}); t > 0\}\) is nonempty, and let \(t' = \inf A\). We have \(\dot{\mu}(t'; \pi') = \dot{\mu}(t'; \bar{\pi})\) and \(\dot{\mu}(t'; \pi') \leq \dot{\mu}(t'; \bar{\pi})\):

- If \(t' = 0\), then \(\dot{\mu}(t'; \pi') = \dot{\mu}(t'; \bar{\pi}) = 0\), and since \(A\) contains \(t\) arbitrarily close to \(0\) such that \(\dot{\mu}(t; \pi') \leq \dot{\mu}(t; \bar{\pi})\), we must have \(\dot{\mu}(0; \pi') \leq \dot{\mu}(0; \bar{\pi})\).

- If \(t' > 0\), then since \(\dot{\mu}(t; \pi') > \dot{\mu}(t; \bar{\pi})\) for all \(t < t'\) by construction, it must be that \(\dot{\mu}(t'; \pi') = \dot{\mu}(t'; \bar{\pi})\) and \(\dot{\mu}(t'; \pi') \leq \dot{\mu}(t'; \bar{\pi})\).

From (1.36) it follows that \(c(t'; \pi') \leq c(t'; \bar{\pi})\).

I argue now that we must have \(A = [t', \infty)\). Suppose to the contrary that \(\inf([t', \infty) \setminus A) = t''\). Then \(i(t'; \pi') \leq i(t; \bar{\pi})\) for all \(t \in [t', t'']\), and it follows from the Euler equation and \(c(t'; \pi') \leq c(t'; \bar{\pi})\) that \(c(t; \pi') < c(t; \bar{\pi})\) for all \(t \in [t', t'']\). Integrating (1.36) from \(t'\) to \(t''\) then implies that \(\dot{\mu}(t''; \pi') < \dot{\mu}(t''; \bar{\pi})\), which is a contradiction.

But if \(A = [t', \infty)\), then \(i(t; \pi') \leq i(t; \bar{\pi})\) for all \(t > t'\), implying that \(c(t; \pi')/c(t; \bar{\pi}) \leq c(t; \pi')/c(t; \bar{\pi}) + (\pi' - \bar{\pi})\), and hence that \(\log c(t; \pi') - \log c(t; \bar{\pi}) \leq - (\pi' - \bar{\pi})t\), leading consumption under the two rates of trend inflation to diverge without bound. This violates the transversality condition.

I conclude that \(A = \emptyset\) and hence that \(\dot{\mu}(t; \pi') > \dot{\mu}(t; \bar{\pi})\) for all \(t > 0\). \(\Box\)
Proof of lemma 1.5.5

Proof. If \( M^d(i, c^*) \) has a constant semielasticity of \(-b\) with respect to \( i \), i.e. \( \frac{\partial \log M^d(i, c^*)}{\partial i} = -b \), then integrating

\[
\log M^d(i, c^*) = \log M^d(0, c^*) + \int_0^i \frac{\partial \log M^d(i', c^*)}{\partial i'} di' = \log m^* - bi
\]

Rearranging:

\[
i = \frac{\log M^d(i, c^*) - \log m^*}{b} \quad v'(m) = iu'(c^*) = -\frac{\log m - \log m^*}{b} u'(c^*)
\]

(1.38)

Now, integrating \( v'(m) \) starting from the initial condition \( v(0) = 0 \) gives

\[
v(m) = \int_0^m v'(m') dm' = -\frac{u'(c^*)}{b} \int_0^m \log m' - \log m^* dm' = -\frac{u'(c^*)}{b} (\log m - \log m^* - 1)
\]

and \( \bar{m} \) is given by \( \log \bar{m} - \log m^* = 1 \). Plugging this into (1.38) implies \( i = \frac{\log \bar{m} - \log m^*}{b} = -\frac{1}{b} \) as desired. \( \square \)

Verification of example 1.5.8

Solving for cash demand:

\[
M^d_{hd}(i, c) = \left( \frac{A_h}{\alpha_h + iu'(c)} \right)^{1/\kappa}
\]

\[
M^d_{id}(i, c) = \left( \frac{A_i}{\alpha_i + iu'(c)} \right)^{1/\kappa}
\]

\[
M^d_{bd}(i, c) = M^d_{hd}(i, c) + M^d_{id}(i, c)
\]

63
It follows that $\frac{M^{d}_{hd}(0,\cdot)}{M^{d}_{ld}(0,\cdot)} = \left( \frac{\frac{A_{h}}{\alpha_{h}}}{\frac{A_{l}}{\alpha_{l}}} \right)^{1/\zeta} > 1$, and also 

$$\frac{M_{ld}(i, c)}{M_{hd}(i, c)} = \left( \frac{A_{l}}{A_{h}} \right)^{1/\zeta} \left( \frac{\alpha_{h} + iu'(c)}{\alpha_{l} + iu'(c)} \right)^{1/\zeta}$$

which is strictly increasing in $i$ since $\alpha_{h} < \alpha_{l}$. Hence $\frac{M_{ld}(i, c)}{M_{hd}(i, c)} = (M_{ld}(i, c) + M_{ld}(i, c))/M_{hd}(i, c) = 1 + M_{ld}(i, c)/M_{hd}(i, c)$ is strictly increasing as well, and the conditions of assumption 1.5.7 are satisfied.
Bibliography


Draghi, Mario, "Introductory Statement to the Press Conference (with Q&A)," October 2015.


Chapter 2

Deciphering the Fall and Rise in the Net Capital Share: Accumulation, or Scarcity?

2.1 Introduction

How is aggregate income split between labor and capital? Ever since Ricardo (1821) pronounced it the "principal problem of Political Economy," this question of distribution has puzzled and inspired economists.

Views differ. In one popular interpretation, the division between labor and capital remains remarkably stable over time: Keynes (1939) called this "one of the most surprising, yet best-established, facts in the whole range of economic statistics," and Kaldor (1957) immortalized it as one of the stylized facts of economic growth. In contrast, another tradition has emphasized variation in income shares: Solow (1958) was famously skeptical, disputing the labor share's status as "one of the great constants

Prepared for the Spring 2015 Brookings Papers on Economic Activity. For helpful comments and conversations I thank my discussants Robert Solow and Brad DeLong, the editors David Romer and Justin Wolfers, the participants at the Spring 2015 Brookings Panel on Economic Activity, as well as Adrien Auclert, Loukas Karabarbounis, Stephen Murphy, Thomas Piketty, Alp Simsek, Ludwig Straub, Ivan Werning, and Gabriel Zucman.
of nature.” Recently, Solow’s view has experienced a resurgence, with the labor share apparently trending downward. Elsby, Hobijn and Sahin (2013) carefully document this decline for the US, and Karabarbounis and Neiman (2014b) describe a broad, worldwide retreat of labor income in favor of capital.

Some influential recent narratives of this shift adopt what I call the accumulation view: capital’s share has risen, and will continue to rise, because of capital accumulation. According to Piketty (2014) and Piketty and Zucman (2014), several forces are driving up aggregate savings relative to income, and the resulting growth in the ratio of the capital stock to income has led to a rise in capital’s share. Alternatively, Karabarbounis and Neiman (2014b) stress the role of falling prices for investment goods; in their account, lower prices lead to more aggregate capital investment and ultimately more capital income. Although these two narratives specify different initial shocks, the subsequent channel is common to both: accumulation of capital through investment leads to growth in capital income, because the rising quantity of capital is not fully offset by a fall in the returns per unit of capital.

This paper argues against the accumulation view, on both empirical and theoretical grounds. Empirically, it reveals that the long-term increase in the capital’s net share of income in large developed countries has consisted entirely of housing. Outside of housing, capital’s rise in recent decades has merely reversed a substantial earlier fall, and in neither direction have there been a parallel movement in the value of capital—all facts that are difficult to reconcile with the accumulation view.

From a more theoretical perspective, the accumulation view is only successful when the elasticity of substitution between labor and capital is sufficiently high. Clarifying the distinction between elasticities gross and net of depreciation, this paper argues that the elasticity required is much higher than the existing literature suggests (particularly in the Piketty (2014) case).

Moving beyond the canonical one-sector model to a multisector model that explicitly acknowledges some important dimensions of capital heterogeneity—for instance, the distinction between housing and non-housing, as well as the distinction between equipment and structures—I continue to find little support for either version of the
accumulation view. Instead, a more viable (albeit incomplete) explanation of recent trends is that residential investment has become more expensive, and land scarcer. Although this has lowered the quantity of housing, there has been a more than offsetting rise in net rents per unit of housing, pushing up the contribution of housing to capital's net share of income. In short, the data and theory support a scarcity view: the net capital share is rising in part because some forms of capital are becoming relatively more scarce, not more abundant.

I begin the paper with a look at the evidence on factor income shares for large developed economies over the postwar period 1948–2010. Several conceptual issues are crucial, especially the distinction between gross and net shares. Although both gross and net concepts are worthwhile when interpreted properly, I argue that the net viewpoint—much less common among recent entries in the literature—is more directly applicable to the discussion of distribution and inequality, because it reflects the resources that individuals are ultimately able to consume. I also restrict attention to the private sector, and in light of the severe measurement difficulties for proprietor's income identified by Elsby et al. (2013) and others, I apply the net shares from the corporate sector to the non-housing sector as a whole.

This measurement reveals a striking discrepancy in the long-term behavior of gross and net shares, echoing the claims of Bridgman (2014). It shows that the net capital share generally fell from the beginning of the sample through the mid-1970s, at which point the trends reversed. In the long run, there is a moderate increase in the aggregate net capital share, but this owes entirely to the housing sector. Indeed, housing's average portion of the aggregate net capital share rose from roughly 3% to 9% over the sample period, even as the private sector fell from 23% to 20%. This essential role of housing is notably absent from previous discussions of the factor distribution of income, and represents an important new contribution of this paper. It parallels a large (though less dominant) role for rising housing wealth in the aggregate wealth-income ratio, which has been documented by Piketty and Zucman (2014), Bonnet, Bono, Chapelle and Wasmer (2014), and others. Although these two trends are sometimes conflated, their alignment is not preordained: in fact, section
2.5.2 finds that a shock to savings should push them in opposite directions.

Outside of housing, there is a pronounced U-shape in the net capital share, with a steep fall in the 1970s and a more recent recovery. At shorter horizons, there is also a strong cyclical element, as long acknowledged by observers ranging from Mitchell (1913) to Rotemberg and Woodford (1999). To gauge whether the long-term fall and rise is consistent with the accumulation view, I contrast it with the time series for the capital-income ratio, finding that there is little similarity between the two. Using US data on the value of the three major components of non-housing capital—equipment, structures, and land—I perform a simple decomposition of the net capital share into returns on these components, plus a residual that can be interpreted as representing firm markups over cost. Markups are responsible for most of the change in shares, in both directions; in particular, accumulation of equipment or structures cannot explain the recent rise.

With these facts in mind, I next ask whether the accumulation view is viable theoretically. First, I look at the canonical single-sector model with a production function \( F(K, N) \) that combines capital and labor. Here, for both the Piketty (2014) and Karabarbounis and Neiman (2014b) narratives, the key parameter is the elasticity of substitution for \( F \). One important oversight in past discussions, however, has been the distinction between gross and net \( F \): the elasticity for gross production is always higher than the elasticity for net. The Piketty (2014) hypothesis—accumulation through aggregate savings driving up the net capital share—is only viable if the net elasticity of substitution is greater than 1, which I argue is out of line with most existing evidence. The related conjecture of rising \( r - g \) requires even more unlikely levels of substitutability. By contrast, the Karabarbounis and Neiman (2014b) hypothesis only calls for a gross elasticity above 1, which I argue is more plausible but still unlikely.

Given the limitations of the single-sector model, to better confront the data and formulate an alternative to the accumulation view I build a multisector model that incorporates key distinctions between sectors (housing and non-housing) and types of capital (equipment, structures, and land). When calibrated to match the structure of
the US economy, the model continues to contradict Piketty (2014). For any choice of lower-level elasticities near the range suggested by the literature, an increase in savings results in a lower net capital share. By contrast, the mechanism in Karabarbounis and Neiman (2014b) remains theoretically viable when labor and equipment are close substitutes, but it works by increasing the value of equipment relative to total income, which is not consistent with the time series evidence.

The multisector model offers better support for the scarcity view. If, as most evidence suggests, consumers’ demand for housing is sufficiently inelastic, the rising price of residential investment and growing scarcity of land can account for most of the growth in housing’s portion of capital income. Although this does not resolve all aspects of the time series—especially the fall and rise in the corporate sector—it does explain a sizable portion of the long-term contribution of housing.

Before the recent preeminence of the accumulation view, there were varied attempts to explain a falling labor share. Elsby et al. (2013) highlight the role of offshoring, while other papers emphasize additional structural and institutional forces.1 This literature does not apply directly here, since it uses gross concepts rather than net. Nonetheless, given the diverse accounts that have been proposed, it is no surprise that this paper fails to find a single mechanism that can explain the recent behavior of factor shares in its entirety.

The paper proceeds as follows. Section 2 discusses the conceptual basis of factor shares and provides evidence on the postwar path of the net capital share among G7 economies, including a decomposition that isolates the role of housing. Section 3 uses a simple decomposition to analyze the trends in net capital share further, restricting itself to the US to make use of more detailed data on capital stocks. Section 4 examines the canonical single-sector model, clarifying the different between net and gross elasticities. Section 5 integrates data and theory by building a multisector model, which refutes the accumulation view more definitively and supports the scarcity view as a partial alternative.

1See, for instance, Azmat, Manning and Reenen (2012), who address the role of privatization, and Arpaia, Perez and Pichelmann (2009), who draw attention to capital-skill complementarity.
2.2 Evidence on factor income shares in developed countries

2.2.1 Conceptual issues

The notion of a "labor" or "capital" share is not monolithic; there are several ways to define and measure these concepts, and different choices lead to strikingly different interpretations of the data.

Decomposing gross value added. In the national accounts, the gross value added of a sector at market prices—the value of its gross output, minus the intermediate inputs used in production—can be divided into three components: labor income (which includes both wages and supplementary compensation), taxes on production, and gross capital income (usually called "gross operating surplus" in the national accounts). Since the second component, taxes on production, does not accrue to either labor or capital, when analyzing the distribution of income between factors it is often convenient to subtract this component, leaving us with gross value added at factor cost. This can then be divided entirely into labor and gross capital shares, which sum to 1. Since I focus in this paper on the division of income between capital and labor, I will generally use this approach.

It is important to recognize that the split of value added between labor and capital is only the initial distribution. Labor income goes both to wages and to supplementary benefits, and a sizable share of wage income is subsequently paid to the government in taxes. Capital income is ultimately apportioned between many recipients, including the government (in the form of corporate and proprietor income taxes) and both debt and equity investors.

For instance, consider a sawmill. The gross value added at factor cost is the difference between its sales of lumber and the cost of logs, excluding taxes on production.

---

2This decomposition potentially applies at many levels of aggregation: for instance, the "sector" may be the entire domestic economy, in which case gross value added at market prices is called gross domestic product (GDP).
Once all compensation of employees at the sawmill is subtracted, the remainder is its gross capital income. Some of this capital income will be paid to lenders in the form of interest, some will be paid to the government in taxes on profits, and the rest may be retained on the balance sheet of the sawmill or distributed as dividends to shareholders. Gross capital income is thus a very broad concept, encompassing funds that are ultimately paid out to many different recipients—it is unaffected, for instance, by the split in financing between debt and equity.  

Gross versus net: concepts. An alternative to gross value added is net value added, which subtracts depreciation. This can be divided into labor and net capital income, the latter being gross capital income minus depreciation. Whether a gross or net measure is more appropriate depends on the question being asked: the allocation of gross value added between labor and gross capital more directly reflects the structure of production, while the allocation of net value added between labor and net capital reflects the ultimate command over resources that accrues to labor versus capital.

For instance, in an industry where most of the output is produced by short-lived software, the gross capital share will be high, evincing the centrality of capital’s direct role in production. At the same time, the net capital share may be low, indicating that the returns from production ultimately go more to software engineers than capitalists—whose return from production is offset by a loss from capital that rapidly becomes obsolete.

Both measures are important: indeed, a rise in the gross capital share in a particular industry is particularly salient to an employee whose job has been replaced by software, and it may proxy for an underlying shift in distribution within aggregate labor income—for instance, from travel agents to software engineers. The massive reallocation of gross income in manufacturing from labor to capital, documented by

---

3 This invariance can be very useful in analyzing trends—for instance, when high inflation pushes up nominal interest rates, a large share of capital income is often paid to bondholders in the form of nominal interest. As Modigliani and Cohn (1979) memorably observed in the context of late-1970s inflation, this causes recorded profits to dramatically understate true profits, since they do not reflect the gain from real depreciation in nominal liabilities.
Elsby et al. (2013), has certainly come as unwelcome news to manufacturing workers. But when considering the ultimate breakdown of income between labor and capital, particularly in the context of concern about distribution in the aggregate economy, the net measure is likely more relevant. This point is affirmed by Piketty (2014), who uses net measures; the welfare relevance of net concepts is elucidated by Weitzman (1976).

**Gross versus net shares: measurement and history.** Historically, the study of income shares has spanned both gross and net concepts: indeed, the famous quote by Keynes (1939) about the stability of labor's share referred to data on net shares, as did Kaldor (1957)'s influential stylized fact.

More recently, however, the vast majority of work on the topic—including Karabarbounis and Neiman (2014b)'s well-known documentation of the declining global labor share—has examined gross shares. To a large extent, this is because gross shares are easier to measure and interpret: as economists since Kalecki (1938) have observed, net income inherently involves a somewhat arbitrary computation of depreciation. High-quality data on gross shares is available for more countries, more years, and more levels of aggregation within each country.

Recently, debate has intensified about the empirical importance of this distinction. Bridgman (2014) argues that the inclusion of depreciation—and, to a lesser extent, taxes on production—in the denominator of the labor share has caused economists to greatly overstate the magnitude and novelty of the labor share's decline. Augmenting their global dataset with information on depreciation, Karabarbounis and Neiman (2014a) argue to the contrary that gross and net labor shares have mainly moved together, and that moving from gross to net shares at most moderately attenuates the downward trend. In my data analysis, I will focus on net shares, finding that the concerns in Bridgman (2014) are valid, especially in the years preceding the start of the Karabarbounis and Neiman (2014a) sample.
Mixed income and other concerns. The distinction between gross and net is not the only concern when computing income shares. Another crucial problem is how to allocate "mixed" income—income earned by the self-employed that is recorded in the national accounts as going to capital. The central difficulty is that this income includes both returns to labor and returns to the capital investments made by the self-employed, with no data available to disentangle the two. This was an essential question for early students of the labor share in the US: as Johnson (1954) and others pointed out, the dramatic rise in workers’ share of income in the first half of the twentieth century was in large part due to the shift from entrepreneurial income (often on farms) to formal labor income.

One solution is to disregard the entrepreneurial sector of the economy, limiting attention (for instance) to the labor share within the corporate sector. In any attempt to measure the labor share for the economy as a whole, however, some approach to dividing mixed income must be chosen—and this choice can matter a great deal. Indeed, Elsby et al. (2013) demonstrate that the "headline" measure provided by the BLS most likely exaggerates the decline in the US gross labor share, due to weaknesses in its approach to imputing labor income for the self-employed. This approach assumes that the self-employed receive the same average compensation per hour as all other workers—an imputation that, although popular and tractable, has some unlikely implications for the US data.

Alternative approaches to dividing mixed income, discussed by Gollin (2002), take several forms: they may do a more sophisticated estimation of labor income for the self-employed based on personal characteristics, or assume that the entrepreneurial sector has the same division between labor and capital as either some other sector or the economy as a whole. I follow Piketty and Zucman (2014) in adopting a form of the latter imputation, assuming that the non-corporate sector (excluding housing) has the same net capital share as the corporate sector.

Finally, another difficult point is the treatment of general government, as well as any other sectors whose output is valued in the national accounts "at cost"—meaning that gross value added is set equal to labor and depreciation costs—rather than by
the market. Here, net capital income equals zero by construction; regardless, it is unclear what net capital income would mean in the context of government.

2.2.2 Income shares in the G7

To better understand the recent evolution of factor shares, I turn to a panel with national accounts data from the G7—which consists of the US, Japan, Germany, France, the UK, Italy, and Canada, currently the seven largest advanced economies by nominal GDP. Most of the data for the panel is derived from the Piketty and Zucman (2014) database, which in turn is taken directly from each country’s national accounts publications.

Although this is a much narrower selection of countries than in the global panels of Karabarbounis and Neiman (2014a,b), it has several offsetting advantages. Most importantly, it covers a longer timespan: five countries have data starting in 1960 or earlier, and three countries have data starting in 1950 or earlier.\footnote{The full set of start dates is 1948 (France, UK, US), 1955 (Japan), 1960 (Canada), 1990 (Italy), and 1991 (Germany). Data for France, the UK, and the US is available starting even earlier, but I focus on 1948 onward because that is when the necessary data starts becoming available for my subsequent, more detailed exercise for the US in section 2.3. This also keeps the focus on postwar dynamics, detached from the sizable dislocations associated with depression and wartime, and mostly postdates the transition from agricultural self-employment to formal employment that bedeviled older analysts like Johnson (1954).} By contrast, Karabarbounis and Neiman (2014a,b)’s dataset starts in 1975, and for many small and developing countries data only starts becoming available much later. Since the net labor share in most countries was close to its postwar peak in the mid-1970s, this offers an incomplete view of the overall trend. The dataset here also permits greater disaggregation, particularly along a dimension that will turn out to be crucial (housing versus the rest of the economy). By focusing on developed economies, it loses some generality but stays closer to the contemporary debate about inequality and income distribution, which has mostly dealt with the developed world.

Estimated average shares. To summarize the evolution over time of various income share measures \( s_{i,t} \), I follow Karabarbounis and Neiman (2014a,b) by running
panel regressions of the form

$$s_{i,t} = \phi_i + \alpha_t + \epsilon_{i,t}$$

for countries $i$ and years $t$. I then display the yearly fixed effects $\alpha_t$, normalizing them such that the fixed effect for the first year of the sample, $\alpha_{1948}$, equals the average share in the dataset in 1948.\textsuperscript{5} I run both unweighted and weighted regressions; the weight for a country is its share of the sample’s aggregate GDP in that year, as measured at PPP by version 8.0 of the Penn World Table.\textsuperscript{6} (For convenience, I will refer to these normalized time fixed effects as yearly “averages”.)

Unlike in the usual presentation, I deal with the capital share rather than its complement, the labor share. Of course, since I deal with value added at factor cost, the capital share is always one minus the labor share; I focus on the former because I will emphasize the composition of capital income.

Overall capital shares: net and gross. First I consider average capital shares for the private economy (excluding government, whose net capital share is zero by construction). As discussed in section 2.2.1, I deal with the problem of self-employment income by following Piketty (2014) and Piketty and Zucman (2014) in the assumption that the net capital share in non-corporate, non-housing sector equals the net capital share in the corporate sector.\textsuperscript{7}

Figures 2-1 and 2-2 report the average net and gross capital shares, respectively. As figure 2-1 demonstrates, the postwar behavior of the net capital share is characterized not so much by a secular rise as by a precipitous fall in the 1970s, which preceded a steady rebound. In this light, it is clear why Karabarbounis and Neiman

\textsuperscript{5}When countries have different trends in $s_{i,t}$, there will be an artifactual discontinuity in $\alpha_t$ when a country enters the sample, which in principle could deliver a misleading impression of the actual year-to-year changes in $s_{i,t}$. In practice, this does not seem to be much of an issue here, and alternative approaches—for instance, averaging the first differences $\Delta s_{i,t}$ across countries in the sample for each year $t$, then plotting the cumulative average first difference over time—deliver similar results.

\textsuperscript{6}See Feenstra, Inklaar and Timmer (2013).

\textsuperscript{7}There are two exceptions: the Canadian national accounts already provide a decomposition of mixed income into labor and capital, which I use; and the Japanese national accounts do not fully break out the corporate sector, necessitating some additional imputations.
(2014a,b)—with a sample starting in 1975, the year in which the unweighted estimate for the net capital share hits its minimum—observe such a dramatic and pervasive rise in capital income relative to labor.

Although Piketty (2014) and others have documented an overall U-shaped trend in the capital share, the claims about timing are quite different: for instance, Piketty (2014) observes that capital’s aggregate valuation and share of income fell greatly in the first half of the twentieth century, during the depression and two world wars. The postwar period is characterized as a period of recovery from this decline. Yet figure 2-1 shows that if anything, the first half of the postwar era experienced a fall in the net capital share, and we are only today returning to levels achieved in the immediate aftermath of the war.

Set against figure 2-1, figure 2-2 reveals that there is a remarkable difference between the long-run behavior of net and gross shares, echoing the results of Bridgman (2014): since average depreciation as a share of gross value added has risen, the gross capital share displays much more of a long-term upward trend. Crucially, much of this disparity emerges before the mid-1970s, perhaps explaining why Karabarbounis and Neiman (2014a) do not detect such an important role for depreciation in their sample. Given the unreliability of depreciation figures at high frequencies, the sudden rise in depreciation prior to the mid-1970s (which causes the divergence between gross and net) should not be given too much credence. The long-term rise in depreciation, however, appears much more robust. As Koh, Santaulalia-Llopis and Zheng (2015) discuss, it is due in part to rapidly depreciating intellectual property—especially software—included in the capital stock.

As argued in section 2.2.1, net shares are likely most relevant for discussions of distribution and inequality. Still, figure 2-1 paints a perhaps ambiguous picture of the net capital share: the recent rise might be in part just a recovery from the anomalously low levels of the 1970s, but the capital share is now reaching and even surpassing the heights previously achieved in the 1950s and 60s. To what extent, then, is the current high share of capital income a truly novel phenomenon? This question is best addressed by disaggregating further along an important dimension,
distinguishing between capital income from housing and capital income from the rest of the economy.

**Composition of the net capital share: the role of housing.** Figure 2-3 subdivides the aggregate net capital share from figure 2-1 into two components: net capital income originating in the housing sector, and net capital income from all other sectors of the economy. It reveals that the aggregate net capital share originating in sectors other than housing has seen only a partial recovery since the 1970s; it remains below the levels of the 1950s, and slightly below or at par with the levels of the 1960s. In contrast, housing’s contribution to net capital income has expanded enormously, from roughly 3pp in 1950 to nearly 10pp today.

Housing’s central role in the long-term behavior of the aggregate net capital share has, to my knowledge, not been emphasized elsewhere. It demands careful scrutiny. Income from housing is unlike most other forms of capital income recorded in the national accounts: in countries where homeownership is dominant, most output in the housing sector is recorded as *imputed* rent paid by homeowners to themselves. It may not be a coincidence that Germany, which table 2.2 reveals to have by far the lowest housing component of net capital income, also has the lowest homeownership rate in the G7. Indeed, imputed rents from owner-occupied housing should arguably be treated as a form of mixed income akin to self-employment income: in part, they reflect labor by the homeowners themselves. Figure 2-3 may therefore exaggerate the level of true “capital” income originating in the housing sector.

Nevertheless, even if figure 2-3 exaggerates the *level* of capital income from the housing sector, this does not necessarily explain the vast *increase* in housing capital income—unless the bias is greater today than in the past. One possible contributor to the trend could be a rise in the rate of homeownership; but this has not been nearly dramatic enough to account for a more than 3x increase in housing capital income.

---

8 For Canada and Japan, the “housing” sector is actually the owner-occupied housing sector due to data limitations. Importantly, Canada and Japan do not drive the trend here: to the contrary, from 1960 (when Canada enters the sample) to 2010, the average contribution of housing to net capital income in Canada and Japan increases by 3pp, while in France, the UK, and the US it increases by 4.5pp.
Another distinct source of bias could be rent control: if the rents imputed for homeowners in the national accounts improperly reflect controlled rents in the tenant-occupied sector, then the ebb and flow of rent regulations will have an inflated impact on income in the housing sector as a whole.

These possible biases notwithstanding, the main thrust of figure 2-3 is that housing has a pivotal role in the modern story of income distribution. Since housing has relatively broad ownership, it does not conform to the traditional story of labor versus capital, nor can its growth be easily explained with many of the stories commonly proposed for the income split elsewhere in the economy—the bargaining power of labor, the growing role of technology, and so on.

The divergence between housing and other forms of capital is also hard to reconcile with the accumulation view: in the Piketty (2014) narrative, for instance, it is not clear why a rise in the aggregate wealth-to-income ratio should be channeled entirely into a rise in the housing component of the net capital share, while the non-housing component stagnates.\(^\text{10}\)

**Net capital share within the corporate sector.** For additional clarity, figure 2-4 plots the average net capital share within the corporate sector. Restricting attention to the corporate sector is a common way to deal with perceived conceptual and measurement difficulties elsewhere in the economy—including ambiguity in the labor/capital split of mixed income, as well as the crucial role of housing. Figure 2-4 echoes the behavior of the non-housing component in figure 2-3, with a sustained fall until the 1970s and a partial recovery in the decades since. Indeed, this resemblance is no coincidence: as discussed above, figure 2-3 imputes the net capital share in the non-housing, non-corporate sector to be the same as in the corporate sector, so that movement in all non-housing capital income is fundamentally driven by the corporate

---

\(^{8}\)See, e.g., Andrews and Sanchez (2011) for some discussion of trends in homeownership.

\(^{10}\)In fact, computations using the multisector model in section 2.5.2 will show that this is backward: the fall in \(r\) induced by savings should lead to a concentrated _decline_ in the housing component of the net capital share.
capital share visible in figure 2-4.\(^{11}\)

Although it does not show any decisive, long-term trend, figure 2-4 does clash with the Kaldor (1957) view of stable income shares. It also contrasts with the relatively steady upward creep of housing capital income in figure 2-3. Fluctuations in the average corporate capital share have been rapid and macroeconomically significant—dropping from a high around 26% in 1950 to a trough around 18% in the 1970s and 80s, then rebounding to a peak of 24% in the 2000s. Indeed, the fall in the unweighted average share from 26.4% in 1950 to 17.7% in 1975, all else equal, contributed nearly half a percentage point annually to growth in corporate labor compensation during that interval. In contrast, the rapid rise from 17.7% in 1975 to 23.6% in 1988 subtracted slightly more than half a percentage point of annual compensation growth.\(^{12}\)

Yet over the long term, the role of fluctuating corporate income shares is comparatively quite mild. For both the weighted and unweighted averages, the impact on annual compensation growth from the 1948-2010 change in net shares is roughly three-hundredths of a percentage point.\(^{13}\) The overall message is clear, and arguably consistent with the Kaldor (1957) perspective on long-run growth: changes in the distribution of corporate income—even systematic ones spread across several countries—can have a marked effect on the short-to-medium-run growth of paychecks. The impact on long-run labor compensation, however, appears to be little more than a rounding error when set against trend growth.\(^{14}\)

There is also a pronounced cyclical pattern in figure 2-4. This is has long been recognized: the labor share tends to rise late in expansions and fall late in recessions.

\(^{11}\) As described in footnote 7, different imputations are used for Canada and Japan. Furthermore, since separate data for the corporate sector is not available in Japan, figure 2-4 displays the overall capital share for Japan instead.

\(^{12}\) Explicitly, \(((1 - .177)/(1 - .264))^{1/25} - 1 \approx .45\%\) and \(((1 - .236)/(1 - .177))^{1/13} - 1 \approx -.57\%\).

\(^{13}\) Explicitly, for unweighted: \(((1 - .214)/(1 - .229))^{1/62} - 1 \approx .03\%\). For weighted: \(((1 - .231)/(1 - .245))^{1/62} - 1 \approx .03\%\).

\(^{14}\) To be clear, the long-run impact in individual countries can be larger. Perhaps the most extreme example is Japan, which table 2.2 shows to have experienced a decline in the average non-housing share of aggregate capital income from 31% in the 1960s to 20% in the 2000s, implying an annualized contribution to wage growth of roughly three-tenths of a percentage point. But Table 2.3 does not suggest any long-run tendency for corporate capital shares in different countries to diverge from each other; the distinct paths across countries are therefore probably best interpreted as mean-reverting variations around an apparently trendless average.
The economic explanation for this pattern, however, is somewhat harder to discern. Conventional wisdom is that low unemployment puts upward pressure on real wages and hence the labor share, while high unemployment keeps real wage growth subdued. This story, however, implicitly involves variation in markups: as Mitchell (1941) observes, “a problem still remains: Why cannot businessmen defend their profit margins against the threatened encroachment of costs by marking up their selling prices?” Answering this challenge, the business cycle literature offers an abundance of proposed explanations for the cyclical pattern of markups, of which Rotemberg and Woodford (1999) provides an excellent summary.

2.3 Decomposing the capital share

2.3.1 Bringing in the value of capital

Section 2.2 provided some preliminary insights into the structure of the net capital share, by distinguishing it into housing and non-housing components. It found that the housing component has seen a steady increase, while the non-housing component has experienced a dramatic fall and then rise. To better understand these movements, it is important to look at another piece of evidence: the value of the capital stock itself.

Both the Piketty (2014) and Karabarbounis and Neiman (2014b) versions of the accumulation view, for instance, explain the recent rise in the capital share through a rise in the value of reproducible capital relative to aggregate income. In fact, this is a central feature of virtually any narrative that stresses capital accumulation: if capital is earning a larger share because we are building more of it, then data on the value of capital should reveal that it has indeed grown relative to income.

Furthermore, this should be true within sectors: for instance, if accumulation explains the rise in the non-housing capital share over the last few decades, then we should see a rising value of capital within the non-housing sector, relative to sectoral value added. This is a simple but crucial check. Elaborating upon it, we can try to
disentangle the roles of different influences on the capital share—the observed value of capital itself; the net user cost of that capital; and firms' markups over cost that lead to additional capital income, not attributable to the user cost of the measured capital stock.

**Theory.** Formally, let $K_1, \ldots, K_n$ be different types of capital, and let

$$Y = F(N, K_1, \ldots, K_n)$$

be a constant-returns-to-scale production function that takes labor $N$ and capital $K_1, \ldots, K_n$ as inputs. Suppose that output $Y$ is sold at a price $P$ that represents a markup of $\mu \geq 1$ over the cost of production\(^{15}\), such that the share of what I will call “pure profits”—capital income above and beyond the user cost of capital $K_1, \ldots, K_n$—in gross income is $\pi \equiv 1 - \mu^{-1}$. I allow for a potentially time-varying markup $\mu$ in part because of the discussion of the corporate capital share in section 2.2.2, which notes a pronounced cyclical pattern that has been explained in the literature through markup variation.

Letting $W_N$ denote the wage paid to labor and $W_{K_1}, \ldots, W_{K_n}$ denote the user costs of capital, we have

$$(1 - \pi)PY = W_N N + \sum_{i=1}^{n} W_{K_i} K_i$$

(2.1)

Suppose further that the model is cast in continuous time (suppressing time subscripts for convenience), and that the flow real cost of funds is $r$. Capital $K_i$ has real price $P_i$, with expected real growth rate $g_{P_i}$, as well as a flow depreciation rate of $\delta_i$. The user cost $W_{K_i}$ is then

$$W_{K_i} = P_i (r + \delta_i - g_{P_i})$$

(2.2)

reflecting the real cost $P_i r$ of financing each unit of capital and the expected combined effect $P_i (\delta_i - g_{P_i})$ of depreciation and price growth on the value of capital held.

\(^{15}\)Since $F$ is constant-returns-to-scale, marginal and average costs are equal, so I will refer to them both as "cost".

87
Combining (2.1) with (2.2), we see that we can divide net output into labor income $W_NN$ and net capital income; the latter can further be divided into a share $\pi PY$ of profits and a component $(r - g_P)P_iK_i$ corresponding to each type of capital $i$.

$$PY - \sum_{i=1}^{n} \delta_i P_i K_i = W_N N + \pi PY + \sum_{i=1}^{n} (r - g_P)P_i K_i \quad (2.3)$$

Letting $Y_{net}$ denote net output on the left of (2.3), we can divide through by $Y_{net}$ to write (2.3) in terms of shares:

$$1 = \frac{W_N N}{Y_{net}} + \frac{\pi PY}{Y_{net}} + \sum_{i=1}^{n} \frac{(r - g_P)(P_iK_i/Y_{net})}{Y_{net}} \quad (2.4)$$

(2.4) illustrates formally how we can divide the net capital share into components that reflect the ratio $P_iK_i/Y_{net}$ of the value of capital of each type $i$ to net income. As discussed earlier, this allows us to evaluate a central element of the accumulation view—namely, that changes in $P_iK_i/Y_{net}$ have played a key role in the evolution of the net capital share.

**Discussion of implementation.** Suppose that in practice we have disaggregated capital into $n$ types, for which we have data on the value $P_iK_i$, and we want to divide the observed capital share of net income $Y_{net}$ into the components identified in (2.4). First, expected price growth $g_P$ is needed; this is very difficult to obtain in principle, since we rarely observe agents' individual expectations of price growth, but it can be roughly approximated by assuming that $g_P$ matches the trend rate of growth over some interval.

The most difficult parts of (2.4) are $\pi$ and $r$: with knowledge of one, we can infer the other, but neither is readily available in the data. In principle, $r$ could be obtained from financial markets, perhaps as some function of bond and equity prices. But this is a notoriously hard problem: it is challenging to know how exactly the costs of borrowing or equity finance map onto the effective cost of funds faced by
an enterprise. Furthermore, since this $r$ is pretax while returns on bonds or equity are after corporate taxes, a time-varying tax adjustment would be needed to infer $r$ directly from market returns.

2.3.2 Implementation: decomposing the net corporate capital share in the US, 1948–2013

I first attempt the disaggregation in (2.4) for the net capital share in the US corporate sector, at an annual frequency for the years 1948 through 2013.\(^{16}\) I disaggregate fixed capital into its three most important components: structures, equipment, and land (denoted by $i = s, e, l$), and I obtain the values $P_tK_i$ for the corporate sector from the flow of funds.\(^{17}\) I assume that the expected price growth $g_{P_i}$ of each form of capital is its actual average real price change from the end of 1947 to the end of 2013. I then try several approaches to resolving the difficulties identified at the end of section 2.3.1.

**Evaluating the accumulation view:** assuming constant $r$. One way to implement the decomposition in (2.4) is to simply impose constant $r$. Taken literally, this is probably not a viable assumption, but it is a straightforward approach to testing the accumulation view: if we rule out variation in $r$ as a source of change in (2.4), how much of the time series can $P_tK_i/Y^{net}$ itself explain? How well do movements in $P_tK_i/Y^{net}$ correlate with changes in the net capital share, and what role can they play quantitatively when $r$ is chosen to be of reasonable size?

Note that in this exercise, the “pure profit” term $\pi PY/Y^{net}$ is effectively just a residual. The goal, for now, is not to provide a complete and convincing decomposition of the net capital share into changes in $\pi$, $r$, and $P_tK_i/Y^{net}$; but instead, to see what role $P_tK_i/Y^{net}$ alone can play. This exercise, though similar, is more informative than mere inspection of the paths of $P_tK_i/Y^{net}$ relative to the path of the net capital share,

\(^{16}\)Ideally, this exercise would extend to all seven of the G7 countries covered in section 2.2, but the additional data required makes this difficult.

\(^{17}\)Since the flow of funds provides end-of-year values for capital, I average the adjacent end-of-year values to obtain the effective capital stock used in production during each year.
because it provides some indication of magnitude. For instance, if $P_i K_i / Y^{net}$ moves together with the net capital share for most $i$, this pattern would appear consistent with the accumulation view; but to see whether this support is quantitatively viable, it is necessary to map the changes in $P_i K_i / Y^{net}$ onto their contributions to the net capital share. This is the role of (2.4), together with some choice of constant $r$.

First, I assume that $r$ takes a constant value over the sample period 1948–2013 such that the average profit share $\pi$ of corporate revenue over the sample is zero. This implies $r \approx 11\%$.\(^{18}\) Effectively, the assumption here is that in the long run, there are no pure profits in the corporate sector—on average, net capital income reflects a return on equipment, structure, or land. This is consistent with Chamberlinian monopolistic competition, where entry drives monopoly profits to zero, on average, in the long run.

Figure 2-6 shows how the net capital income for the US corporate sector in figure 2-5 breaks down into the four components in (2.3) under this assumption. Though there are some fluctuations in each component’s contribution, both the U-shaped pattern and the cyclical fluctuations in the corporate capital share in figure 2-5 appear dominated by the residual component of “pure profits” $\pi$. In other words, contrary to the accumulation view, time series shifts in the capital share in the corporate sector cannot be explained by parallel shifts in the measured value of capital.

Consequences for the falling investment prices hypothesis. As figure 2-5 further reveals, the contribution from equipment in particular is if anything the inverse of the U-shaped pattern in the corporate net capital share in figure 2-5: it rises in the 1970s and 1980s, and then later trends downward. Since equipment is the component of fixed capital that has experienced a decline\(^{19}\) in real price, this is hard to

\(^{18}\)Although this seems high for a real return, note that it is a pretax return: the return before taxes are applied either to corporate profits or distributions of interest or dividends. Interestingly, it is slightly lower than the constant return in figure 2-7 estimated using my alternative approach, which is roughly 12.8%. As explained below, this return is higher because according to the flow of funds, the total market value of the corporate sector in the US has actually been lower than the book value, on average, in the postwar era—suggesting that pure profits are, if anything, negative, and that the assumption that pure profits are zero on average is not misattributing these profits to an exaggerated return $r$ on capital.

\(^{19}\)The equipment investment deflator has relative to the GDP deflator at an annualized rate of 1.5% during the sample period, as opposed to a 1.1% average rise in the deflator for nonresidential structures.
reconcile with a central role for falling investment prices in the dynamics of capital's share, the hypothesis emphasized by Karabarbounis and Neiman (2014b). Without a structural model, of course, this exercise is not decisive: falling investment prices might contribute to a rising capital share via some more indirect causal channel, and indeed Karabarbounis and Neiman (2014a) suggest one such possibility. I address these concerns with a multisector model in section 2.5.2, where I generally do not find a major role for such indirect mechanisms.

Surprisingly, my finding here is consistent with the result of a closely related exercise in section IV.B of Karabarbounis and Neiman (2014b), who also decompose non-labor income into a component reflecting the return on accumulated capital and a component reflecting markups, under the assumption of a constant real interest rate. Although it is not their focus, Karabarbounis and Neiman (2014b) remark that they generally do not find increases in the share of the former component. This implies that the fall in labor share comes in the aggregate from the rise in markups, rather than returns on measured capital.

At face value, this contradicts the emphasis on capital accumulation as a source of the falling labor share; but Karabarbounis and Neiman (2014b) point out that if their elasticity estimate is valid, it remains correct to say that counterfactually, the labor share would be higher if not for the role of falling investment prices in encouraging investment. Of course, if this is true, it follows that there must be two other unidentified forces influencing the factor distribution of income: (1) some force of similar magnitude that offsets their mechanism in the aggregate by pushing investment downward, and (2) another force leading to the rise in markups, which accounts for the entire aggregate fall in the labor share. With these forces in play, the accumulation view only plays a secondary role regardless.

**Smaller r.** Figure 2-5 can also be constructed assuming a smaller $r$, under the assumption that pure profits in the corporate sector are not all dissipated in the long run. This does not materially change the conclusion that the measured value of capital is unable to account for the major shifts in the net capital share. (Indeed, a
smaller $r$ in (2.4) directly leads to a lower weight on $P_iK_t/Y^{net}$.

**Structural approach: identify time path for $r$ from market minus book value.** In an attempt to more convincingly disentangle the roles of $r$ and $\pi$, I turn to a more elaborate approach for estimating $r$. The basic idea is that the difference between the market value of corporations and the value of their fixed assets should reflect the expected stream of future pure profits $\pi PY$ (up, possibly, to some stochastic pricing error). We can use this observation as a strategy to estimate the implied $r$: for instance, if market value is much higher than the value of the firm's assets, the expected stream of pure profits $\pi PY$ is high, and $r$ in the future must be low enough that there are pure profits left over in (2.3) after the direct return from capital $\sum_i(r-g_P)P_iK_i$ is subtracted.\(^\text{20}\)

**Description of the method.** Appendix 2.7.3 provides the technical details, along with the specific theoretical assumptions in a continuous-time model that are needed to make the procedure valid. The core equation implied by the theory is (see (2.33) and (2.34)):

$$E \left[ \frac{\phi(t)}{\text{output between } t-1 \text{ and } t} \times (OMV(t) - \text{discount} \times OMV(t+1)) \right] = E \left[ \frac{\phi(t)}{\text{output between } t-1 \text{ and } t} \times \text{pure profits between } t \text{ and } t+1 \right] \tag{2.5}$$

where $OMV(t)$ denotes the difference between the market value and book value of corporations recorded at time $t$, and $\phi(t)$ is an arbitrary time-dependent function. Implicit in (2.5) is a (nonstochastic) time path $r(t)$ for the real interest rate, which is needed to calculate profits $\pi(t)P(t)Y(t)$ as a residual in (2.3) and to calculate the proper discount factors.

The interpretation of (2.5) is straightforward: it states that the expected difference

---

\(^{20}\)For simplicity, I will call the total value of the firm's fixed assets its "book value", even though this is not necessarily book value in the usual sense: I will define it to exclude financial assets—these are instead subtracted from the market value, which includes net financial liabilities—and it uses values from the flow of funds for real estate and equipment, which are updated to reflect changes in price.
between the present value of next year's excess market value $OMV(t + 1)$ and this year's excess market value $OMV(t)$ reflects expected pure profits between $t$ and $t + 1$. This relation continues to hold, in expectation, when both sides are normalized by the previous year's recorded output, which I do to render values comparable across time; and it also holds when both sides are multiplied by any choice of the time-dependent function $\phi(t)$.

Technically speaking, equation (2.5) can be used as a moment condition to estimate $r(t)$. If we have $n$ functions $\{\phi_1(t), \ldots, \phi_n(t)\}$, we obtain $n$ distinct moment conditions (2.5), and can enforce these conditions in the sample to solve for an $n$-parameter functional form for $r(t)$. I choose $\phi_1(t) = 1$, $\phi_2(t) = t$, and $\phi_3(t) = t^2$, and estimate three specifications for $r(t)$: a constant value $r(t) = \bar{r}$, a linear trend $r(t) = a_0 + a_1 t$, and a quadratic trend $r(t) = a_0 + a_1 t + a_2 t^2$, using the moment conditions implied by $\{\phi_1(t)\}$, $\{\phi_1(t), \phi_2(t)\}$, and $\{\phi_1(t), \phi_2(t), \phi_3(t)\}$, respectively.

Effectively, I am solving for the constant $\bar{r}$ such that the expression

\[
OMV(t) - \text{discount} \times OMV(t + 1) - \text{pure profits between } t \text{ and } t + 1
\]

output between $t - 1$ and $t$

(2.6)

equals zero on average throughout the sample; and I am also solving for the linear $r(t) = a_0 + a_1 t$ and the quadratic $r(t) = a_0 + a_1 t + a_2 t^2$ such that (2.6) does not have any linear or quadratic trends over time, respectively.

When calculating $OMV$, the difference between the market value of the corporate sector and the book value of its fixed capital, I interpret the "market value" to be the total value of all financial claims on a corporation—both its equity market capitalization and its net financial liabilities—in order to be consistent with the computation of capital income in the national accounts, which includes income that ultimately goes to both shareholders and bondholders.\(^2\) Both market and book value are taken from the flow of funds.

\[^2\]This causes some anomalies in the early postwar years, when the corporate sector was left with large cash balances and relatively little debt, making net liabilities negative while equity valuations were already quite low, and leading to an extremely low market relative to book value. To avoid undue influence from this period, I exclude data from prior to 1955 in the benchmark results displayed here; otherwise, there is an even more dramatic estimated downward trend in $r(t)$.  

93
**Estimated paths for r.** Figure 2-7 shows the estimated constant, linear, and quadratic time trends for the corporate rate of return \( r(t) \) following the procedure above. The most striking feature of these plots is the general downward trend in \( r(t) \): according to this procedure, the required return on capital for the US corporate sector has fallen over the postwar era. This reflects the fact that the market value of corporations has grown relative to book value over this period, albeit unevenly, as can be seen in figure 2-9. The estimation infers from this that pure profits are trending upward, so that the required return on capital \( r(t) \) itself must be declining.

Another interesting feature of figure 2-7 is that estimated constant \( \bar{r} \), at roughly 12.8%, is actually higher than the \( r \) chosen in my benchmark decomposition to set the average share of pure profits to zero. This reflects the fact that according to the flow of funds, on average, the aggregate market value of corporations has actually been slightly below the book value during the sample period, as depicted in figure 2-9. This suggests that the assumption of zero average pure profits for the benchmark decomposition was not too far out of line: corporations, on average, have not been worth more than the underlying value of their assets.

Since I am only estimating parametric trends for \( r(t) \) here, I am not allowing \( r(t) \) to vary at high frequencies with the business cycle; market prices at high frequencies are too noisy and volatile to permit credible estimation of \( r(t) \) using the method above. This means that I still cannot address, for instance, the role played by cyclical fluctuations in \( r(t) \) in driving cyclical fluctuations in the capital share. But by allowing for a long-term trend in \( r(t) \), I can disentangle the long-term effects of \( r \) from the effects of changing capital-income ratios \( P_i K_i / Y_{net} \), and obtain a better assessment of the role of pure profits \( \pi PY / Y_{net} \).

Since long-term trend in the corporate net capital share is U-shaped, with a large fall and recovery, I will emphasize the results from the quadratic estimated trend \( r(t) \). To the extent that varying \( r \) is partly responsible for the U-shaped trend, quadratic \( r(t) \) can capture much of its impact.
**Implications of quadratic trend in \( r \).** Redoing the decomposition in figure 2-6, using the quadratic trend for \( r(t) \) rather than a constant, produces figure 2-8. The impact of the change in \( r(t) \) is unsurprising. Relative to figure 2-6, figure 2-8 initially attributes a larger share of returns to fixed capital, offset by substantial negative pure profits; over time, the return on fixed capital falls, and the role of pure profits grows substantially. As in figure 2-6, pure profits play a central role in the U-shaped path for the overall corporate net capital share—but these movements come in addition to broad offsetting trends, in which pure profits have replaced income from fixed assets in (2.3).

It is difficult to say how literally these trends should be interpreted. Given the methodology for identifying \( r(t) \), they are ultimately the consequence of the long-term rise in the ratio of market value to book value in the US corporate sector, as seen in figure 2-9; and this, in turn, may be the result of other, unmodeled changes in financial markets, not a rise in \( \pi \). Nevertheless, figure 2-8 is certainly suggestive, and it casts additional doubt on the accumulation view, since it indicates that (contrary to the assumption of relatively stable returns per dollar of capital) \( r(t) \) has, if anything, experienced a sizable decline.

### 2.3.3 Extending the decomposition: the net capital share for the private economy

I now extend the decomposition in section 2.3.2 to the net capital share for the private domestic economy as a whole—excluding the non-housing government and NPISH (nonprofit institutions serving households) sectors, which have zero net capital share by construction in the national accounts.

Due to the inherent difficulties in apportioning mixed income between labor and capital, as discussed in section 2.2.1, this requires some imputations. I will assume that both the rate of return \( r \) and the pure profit share \( \pi \) are the same in the non-housing, non-corporate sector and the corporate sector, and use the estimated
quadratic path for $r$ from the previous section.\textsuperscript{22} For the housing sector, I will assume that there is no pure profit, and allow $r$ to vary over time in (2.3) such that net housing capital income always equals $(r - g_{P_2})P_{s2}K_{s2} + (r - g_{P_2})P_{t2}L_2$, where $P_{s2}K_{s2}$ is the value of residential structures and $P_{t2}L_2$ is the value of residential land.

The results are displayed in figure 2-11, which decomposes the net capital share displayed in figure 2-10. Figure 2-11 is noisy, and for the most part it combines the lessons from sections 2.2.2 and 2.3.2: there is a strong, long-term upward trend in net capital income from housing, and the volatile capital share elsewhere in the economy is driven principally by pure profits.

There are, however, some additional insights in the figure 2-11 decomposition. For instance, the rise in net income for the housing sector has come both from residential structures and land, but figure 2-11 attributes a larger portion of the increase (and of the level) to structures.

This may come as a surprise, since one plausible hypothesis for the growth of net housing income is the rising scarcity of land. In part, the secondary role of residential land here comes from its more rapid price appreciation. Since I assume that the net rate of return \textit{including} expected capital gains is equalized between residential structures and land, the net rate of return \textit{excluding} expected capital gains—which is used in the decomposition, because income in the national accounts also excludes capital gains—is significantly lower for land. In a sense, then, the lesser role of land is due to the idiosyncrasies of national accounting; and an alternative definition of net capital income that included some form of expected capital gains would show a larger impact from land. (With this in mind, it is remarkable that housing plays such a large aggregate role in section 2.2.2 already: if the G7 national accounts data were modified to include capital gains, housing’s centrality would only increase.)

Another interesting feature of figure 2-11 is that there has been a sizable decline in the role of capital income from non-residential land over time, from roughly 10%...
of net private value added in the first half of the sample to an (erratic) average of roughly 2.5% today. In other words, there has been a shift in net capital income from non-residential to residential land—but the decline in the former has been far larger than the growth in the latter, suggesting that the direct contribution of land to net capital income in the US has actually fallen.

2.4 Capital share theory: one-sector model

2.4.1 One-sector, one-good model.

I now take a step back from the decomposition in section 2.3—with its multiple capital goods—to recount the simplest, traditional model of income shares, with a single production sector and a single good. This offers a first-pass test of the theoretical viability of the accumulation view: all else equal, should we expect a larger capital-income ratio to cause an increase or decrease in capital’s share?

Let $F(K, N)$ be a constant returns to scale production function, with capital $K$ and labor $N$ as factor inputs, and positive but diminishing returns in each factor. Assume that this is a one-good model, where the relative price of capital and output is fixed at one. The elasticity of substitution $\sigma$ between $K$ and $N$ is defined as

$$\sigma \equiv -\left( \frac{d(\log(F_K/F_N))}{d(\log(K/N))} \right)^{-1} \quad (2.7)$$

This gives us the (inverse) elasticity of the ratio $F_K/F_N$ of marginal products to the ratio $K/N$ of capital. Equivalently, $\sigma$ tells us the extent to which a cost-minimizing producer’s relative demand for $K/N$ will change if there is a change in the relative cost $R/W$ of using capital and labor as inputs.

From the definition (2.7), one can show that $\sigma$ also gives the inverse elasticity of $F_K$ with respect to a change in the capital-output ratio $K/F$:

$$\sigma = -\left( \frac{d(\log F_K)}{d(\log(K/F))} \right)^{-1} \quad (2.8)$$
which implies that the elasticity of the capital income share \( F_K K/F \) with respect to the capital-output ratio \( K/F \) is

\[
\frac{d(\log(F_K K/F))}{d(\log(K/F))} = 1 - \frac{1}{\sigma}
\] (2.9)

This indicates the critical importance of the threshold \( \sigma = 1 \). If \( \sigma > 1 \), the elasticity is positive, so that the capital income share will increase as \( K/F \) rises. Inversely, if \( \sigma < 1 \), the capital income share will fall as \( K/F \) rises. In the important special case \( \sigma = 1 \), diminishing returns exactly offset the increased quantity of capital, and the share remains constant.

Indeed, one of the original motivations behind Cobb and Douglas (1928)'s eponymous production function was the apparent constancy of capital and labor shares in the data; this is guaranteed by the Cobb-Douglas production function \( F(K, N) = K^\alpha N^{1-\alpha} \), which has a constant elasticity of substitution \( \sigma = 1 \).

**Net versus gross.** Thus far, I have been ambiguous about whether the function \( F \) gives gross production, or production net of depreciation. In principle, either interpretation is legitimate—especially since this is a one-good model, where the relative price of capital and output is fixed at one, and losses from capital depreciation can reasonably be included as part of the production function.

If \( F \) is gross production, then \( 1 - 1/\sigma \) is the elasticity of gross capital income with respect to the ratio of capital to gross output. If \( F \) is net production, then \( 1 - 1/\sigma \) is the elasticity of net capital income with respect to the ratio of capital to net output. As discussed in section 2.2.1, both measures are useful, but net concepts are probably more meaningful when studying income distribution.

It is important to recognize that \( \sigma \) depends greatly on which measure is used—a subtlety that is often overlooked. Suppose \( F(K, N) \) is the gross production function, with an elasticity of substitution of \( \sigma \). Then the net production function is \( \tilde{F}(K, N) = \)
\( F(K, N) - \delta K \), and from (2.8) the elasticity of substitution for \( \tilde{F} \) is

\[
\tilde{\sigma} = \frac{d(\log(\tilde{F}/K))}{d(\log \tilde{F}_K)} = \frac{d(F/K - \delta)/(F/K - \delta)}{d(F_K - \delta)/(F_K - \delta)} = \frac{d(F/K)/(F/K)}{\sigma \cdot (F_K - \delta)/(F - \delta K)} \frac{F_K/K/F}{F/K} \tag{2.10}
\]

Hence the elasticity of substitution \( \tilde{\sigma} \) for the net production function ("net elasticity") equals the elasticity of substitution \( \sigma \) for the gross production function ("gross elasticity") times the ratio of the net capital share \( (F_K K - \delta K)/(F - \delta K) \) and the gross capital share \( F_K K/F \). Since the net capital share is always less than the gross capital share, it follows that the net elasticity is always below the gross elasticity.

Why, intuitively, is the net elasticity always lower? The net return on capital \( \tilde{F}_K \) is less than the gross return \( F_K \) by a constant—the depreciation rate \( \delta \)—meaning that a given change in \( F_K \) translates into an equal absolute, and a larger relative, change in \( \tilde{F}_K \). For instance, if \( \delta = 5\% \), and \( F_K \) declines from 10% to 8%, \( \tilde{F}_K \) will decline from 5% to 3%. A 20% decline in the gross return becomes a 40% decline in the net return, and the ratio of the two is (A). As we increase capital relative to labor, the net marginal product of capital declines more rapidly than the gross—in short, capital is less substitutable for labor from a net perspective.

**Calibrating (2.10).** To obtain an illustrative calibration, I take the data from section 2.3.3, where pure profits are estimated using the quadratic path for \( r(t) \). I exclude pure profits and land from the capital share, since they are not reproducible forms of capital—and relevant question for the Piketty (2014) hypothesis is whether adding more reproducible capital through investment increases or decreases capital's share of income.

In the most recent year in the sample, 2013, the resulting US private net capital share (excluding pure profits and land) was 25.6%, while the US private gross capital share (excluding pure profits and land) was 34.5%. This results in a ratio of
approximately 0.74, and (2.10) implies

\[ \bar{\sigma} \approx 0.74 \times \sigma \]  

(2.11)

so that the net elasticity is slightly less than three-quarters the gross elasticity.

If the decomposition in section 2.3.3 is performed assuming a lower rate of return \( r \), such that a more significant share of net capital income is attributed to pure profits rather than returns on measured capital, then the ratio can be appreciably lower than in (2.11). For instance, in an alternative estimate where \( r \) is chosen to be roughly 5.5% for the corporate sector—implying that half of long-run net capital income is attributable to pure profits—the private net and gross capital shares (excluding pure profits and land) become 15.5% and 25.6%, respectively, resulting in a ratio of approximately 0.60.

**Empirical implications.** Ever since Arrow, Chenery, Minhas and Solow (1961) first proposed the constant elasticity of substitution (CES) production function, researchers have attempted to estimate the key elasticity parameter. These studies have virtually always looked at the elasticity of substitution in the gross production function.

The literature is vast and its conclusions muddled, but one consistent theme has been the rarity of high elasticity estimates. Chirinko (2008) provides an excellent summary of the empirical literature, listing estimates from many different sources and empirical strategies; table 2.1 displays the estimates compiled there, both in their original gross terms and converted to net terms, where the conversion factor of 0.74 from (2.11) is used.

Of the 31 sources\(^\text{23}\) listed, table 2.1 reveals that only 5 show a gross elasticity above 1, and only 2 imply a net elasticity above 1. From (2.9), it follows that a rise in the capital-income ratio, holding the production function constant, most likely will cause a decline in the net share of capital income. This is inconsistent with the Piketty

\(^{23}\text{For a few sources that list a range of elasticities, I take the midpoint. This has minimal effect on the distribution.}\)
Table 2.1: Distribution of elasticity estimates compiled by Chirinko (2008): in gross terms (as originally stated) and in net terms (converted using (2.11)).

<table>
<thead>
<tr>
<th></th>
<th>[0, 0.5)</th>
<th>(0.5, 1)</th>
<th>(1, 1.5)</th>
<th>(1.5, 2)</th>
<th>[2, 4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frequency of gross $\sigma$</td>
<td>14</td>
<td>12</td>
<td>3</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Frequency of net $\tilde{\sigma}$</td>
<td>21</td>
<td>8</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

(2014) and Piketty and Zucman (2014) version of the accumulation view, which holds that a rise in the capital-income ratio has led—and will lead going forward—to a rise in capital's net share.

Implications for $r-g$. A closely related theme in Piketty (2014) is the gap $r - g$ between the real return $r$ on capital and the real growth rate $g$ of the economy. This gap, for instance, gives the rate at which a wealthy dynasty can withdraw capital income for consumption purposes without decreasing its wealth relative to the size of the economy. More generally, when $r - g$ is higher, “old” accumulations of wealth become more important relative to “new” ones. Higher $r - g$ generally implies that the power law tail of the wealth distribution has a smaller exponent—so that there is more inequality of wealth at the top, and extreme levels are more likely. Many readers take the dynamics of $r - g$ to be the central theme of the book.

Both Piketty (2014) and Piketty and Zucman (2013) make heavy use of the identity

$$\frac{K}{Y^{net}} = \frac{s}{g}$$

(2.12)

where $s$ is the net savings rate and $g$ is the growth rate, which is dubbed the “Second Fundamental Law of Capitalism”. This identity only holds asymptotically—if $s$ or $g$ changes, convergence to the new value of $K/Y^{net}$ does not happen instantaneously—and it is unlikely that $s$ is exogenous and invariant to changes in $g$. Nevertheless, Piketty (2014) argues that it is useful to explore the implications of this identity given exogenous $s$, particularly the fact that $K/Y^{net}$ rises as $g$ falls, which is central to the projection that the capital-income ratio will rise in the future.\(^{24}\)

\(^{24}\)There is some conflict between the assumption of exogenous $s$ for all income and the emphasis on $r - g$. If only this fraction $s$ of capital income $r$ is saved, then existing fortunes will grow at the rate $s \cdot r - g$, not $r - g$; and for plausible values of $s$ as a share of all income, $s \cdot r - g$ is likely to be
If \( r = F_K \), then (2.8) shows that the elasticity of \( r \) with respect to \( K/Y^{net} \) is simply \(-\tilde{\sigma}^{-1}\), where \( \tilde{\sigma} \) is the net elasticity of substitution. For exogenous \( s \), (2.12) indicates that the elasticity of \( K/Y^{net} \) with respect to \( g \) is \(-1\), implying that the elasticity of \( r \) with respect to \( g \) is \( \tilde{\sigma}^{-1} \). It follows that

\[
\frac{\partial(r - g)}{\partial g} = \frac{r}{g} \tilde{\sigma}^{-1} - 1 \tag{2.13}
\]

This expression is positive if \( r/g > \tilde{\sigma} \). Again taking data from section 2.3.3, in 2013 the average return on measured capital was 7.5%. Taking this to be the \( r \) in (2.13), and taking \( g \) to be 2.5% (approximate trend real GDP growth in the US in the last 25 years), we have \( r/g = 3 \), in which case the derivative in (2.13) is positive as long as the net elasticity \( \tilde{\sigma} \) is less than 3.

The evidence in table 2.1 indicates that this is overwhelmingly likely. Indeed, converting via (2.11), a net elasticity of 3 corresponds to a gross elasticity of \( \sigma = \tilde{\sigma}/0.74 = 4.05 \), which is above every estimate listed in Chirinko (2008) and virtually every estimate in the wider literature. Hence a decline in \( g \) will result in a decline in \( r - g \): the decline in \( g \) itself is less than the decline in \( r \) that it induces through capital accumulation and diminishing returns. Given the assumption (2.12) on capital accumulation, the prediction in Piketty (2014) that \( r - g \) will rise as \( g \) falls is especially hard to reconcile with empirically plausible degrees of substitutability.

### 2.4.2 One-sector, two-good model.

The canonical model in section 2.4.1 can be enriched slightly by allowing the price \( P_K \) of capital relative to the output good to vary. This modification is central to the account in Karabarbounis and Neiman (2014b), who attribute the rise in the gross capital share to high capital demand induced by a fall in \( P_K \).

To be more explicit, take the net required return \( r \) on capital as given. Ignoring quite negative, implying the rapid erosion of existing wealth.
expected capital gains, demand for capital is pinned down by the condition

\[ F_K(K, N) = P_K(r + \delta) \]  (2.14)

The elasticity of the gross capital/output ratio \( K/F \) with respect to \( P_K \) is then

\[ \frac{\partial(\log(K/F))}{\partial(\log P_K)} = \frac{d(\log(K/F))}{d(\log F_K)} \cdot \frac{\partial(\log F_K)}{\partial(\log P_K)} = -\sigma \]  (2.15)

where \( \frac{\partial(\log F_K)}{\partial(\log P_K)} = 1 \) follows directly from (2.14), and

\[ \frac{d(\log(K/F))}{d(\log F_K)} = -\sigma \]

follows from (2.8). The elasticity of the gross capital share with respect to \( P_K \) becomes

\[ \frac{\partial(\log(F_K/F))}{\partial(\log P_K)} = \left(1 + \frac{d(\log(K/F))}{d(\log F_K)}\right) \cdot \frac{\partial(\log F_K)}{\partial(\log P_K)} = 1 - \sigma \]  (2.16)

implying that a decline in the relative price \( P_K \) of capital will increase the gross capital share if \( \sigma > 1 \).

Meanwhile, the elasticity of the net capital share with respect to \( P_K \) is can be obtained through a somewhat more involved computation.\(^\text{25}\) The result, first derived by Karabarbounis and Neiman (2014a), is

\[ \frac{\partial(\log((F_K - \delta P_K)K/(F - \delta P_K K)))}{\partial(\log P_K)} = (1 - \sigma) \cdot \frac{F}{F - \delta P_K K} \]  (2.17)

Note that \( \sigma = 1 \) is still the critical threshold: a decline in the relative price \( P_K \) of capital increases both the net and gross capital shares if \( \sigma > 1 \). This consistency is a noteworthy contrast with the distinction (2.10) between gross and net elasticities

\(^{25}\) For instance, one can write

\[ \frac{\partial(\log((F_K - \delta P_K)K/(F - \delta P_K K)))}{\partial(\log P_K)} = \frac{\partial(\log(F_K - \delta P_K))}{\partial(\log P_K)} + \frac{\partial(\log(K/F))}{\partial(\log P_K)} - \frac{\partial(\log(1 - \delta P_K K/F))}{\partial(\log P_K)} \]

\[ = 1 - \sigma + \frac{\delta P_K K/F}{\delta(\log P_K)} (1 - \sigma) + (1 - \sigma) \frac{\delta P_K K}{F - \delta P_K K} = (1 - \sigma) \frac{F}{F - \delta P_K K} \]
of substitution, where a rise in the capital-output ratio could produce an increase in
the gross capital share and a decrease in the net capital share. From an intuitive
standpoint, this is unsurprising: since we are holding \( r \) constant, the ratio \( r/(r + \delta) \)
of net to gross capital income is fixed, and the two move in the same direction in
response to a change in \( P_K \).

Karabarbounis and Neiman (2014a) stress the role of (2.17), which shows that
their focus on the role of changes in \( P_K \) can potentially account for simultaneous
changes in both the gross and net capital shares, assuming that the gross elasticity of
substitution \( \sigma \) is greater than 1. In light of the estimates compiled in table 2.1 (26
out of 31 of which find \( \sigma < 1 \)), \( \sigma > 1 \) still appears unlikely, but it is somewhat more
plausible than \( \sigma > 1 \).

2.5 Capital share theory: a multisector model

2.5.1 Design of the multisector model

The theory in section 2.4 enables a first-pass analysis of how the distribution of
income is affected by various forces. It shows that accumulation of capital—all else
equal—will likely result in a decline in the net capital share, since the net elasticity
of substitution is almost certainly below one. This counters the central hypothesis of
Piketty (2014). It also shows that a decline in the relative price \( P_K \) of capital, holding
the required return \( r \) constant, will result in an increase in the net capital share if the
gross elasticity of substitution is above 1—a claim that is still hard to reconcile with
the bulk of empirical evidence, but for which Karabarbounis and Neiman (2014b)
mount a spirited case.

Nevertheless, the one-sector model in section 2.4 is in many ways unsatisfactory as
a model of the distribution between capital and labor. For instance, sections 2.2 and
2.3 demonstrated the decisive role of the housing sector in the long-term trajectory
of the net capital share—but a one-sector model is by construction unable to account
for a shift toward housing. Indeed, Piketty (2015) has recently voiced discomfort
with the one-sector interpretation of the rising capital share, arguing that “the right model to think about rising capital-income ratios and capital shares in recent decades is a multi-sector model of capital accumulation.” In this section I will construct a tentative version of such a model.

**Nested framework.** Given the central role of housing in sections 2.2 and 2.3, it is first important to distinguish between non-housing and housing output. If household preferences are homothetic in these two types of output, the household objective can be written as a monotonic transformation of a constant returns to scale aggregator \( Z(Y_{nh}, Y_h) \) that takes non-housing and housing services as inputs. We can view \( Z \) as the “top-level” production function for the economy.

For the non-housing sector, it will be useful to model the production process in a way that reflects the different types of capital studied in section 2.3 (equipment, structures, and land), so that the results from that disaggregation exercise can be used to inform the model. One natural approach is to assume that structures and land together provide “real estate” services that serve as an input to production, while labor and equipment together provide all other services. This approach enables me to draw upon several empirical literatures, which estimate the relevant elasticities of substitution—for instance, the elasticity of substitution between structures and land in the production of real estate services, or the elasticity of substitution between housing and non-housing in consumer preferences.

Concretely, let \( H(N, K_e) \) be a constant returns to scale aggregator combining labor \( N \) and equipment \( K_e \), and let \( G_1(K_{s1}, L_1) \) be another constant returns to scale aggregator combining nonresidential structures \( K_{s1} \) and land \( L_1 \). Finally, let \( F \) be another constant returns to scale aggregator that combines \( H \) and \( G_1 \), so that the consolidated production function for the non-housing sector takes the form

\[
Y_{nh} = F(H(N, K_e), G_1(K_{s1}, L_1)) \tag{2.18}
\]

Following section 2.3, I assume that gross output in the non-housing sector is sold at
some markup $\mu$ over marginal cost.

Similarly, suppose that residential structures $K_{s2}$ and land $L_2$ are combined by an aggregate $G_2(K_{s2}, L_2)$ to provide housing services, so that the production function for the housing sector takes the form

$$Y_h = G_2(K_{s2}, L_2)$$  \hspace{1cm} (2.19)

Finally, as already mentioned, $Z$ combines $Y_{nh}$ and $Y_h$ into an aggregate that reflects household preferences:

$$Y = Z(Y_{nh}, Y_h)$$ \hspace{1cm} (2.20)

This multisector economy captures the distinction between the non-housing and housing sectors, as well as all five forms of capital analyzed in section 2.3: equipment ($K_e$), nonresidential structures ($K_{s1}$), nonresidential land ($L_1$), residential structures ($K_{s2}$), and residential land ($L_2$).

The aggregate, nested structure of production in the economy is depicted in the tree below.

\[ Z(F, G_2) \]
\[ F(H, G_1) \quad G_2(K_{s2}, L_2) \]
\[ H(N, K_e) \quad G_1(K_{s1}, L_1) \]

\[ \text{Services from labor and equipment} \quad \text{Non-housing real estate} \]

\textbf{Elasticities of substitution.} The response of the multisector model to various shocks is influenced by the local (gross) elasticities of substitution $(\sigma_Z, \sigma_F, \sigma_{G_1}, \sigma_{G_2}, \sigma_H)$ for each of the five constant-returns-to-scale production functions $(Z, F, G_1, G_2, H)$ in the model above.
Although there are extensive empirical literatures that study many of these elasticities, a convincing research design is often elusive, and there is rarely strong consensus around a single point estimate. In the absence of such consensus, I will draw upon each literature to obtain plausible ranges for each elasticity, and study the implications of choosing different values within each range. The objective is to see which, if any, conclusions emerge robustly from the multisector model despite allowing for some uncertainty about the $\sigma$s. Another goal is to investigate which $\sigma$s matter most to aggregate outcomes, both to clarify thinking and to direct future research toward the most crucial targets.

Surveying the relevant literatures, I find:

- $\sigma_Z$ equals the elasticity of demand for housing services (as a share of total output) with respect to its price (relative to the aggregate price index for $Z$). Closely related elasticities of demand for housing have been studied in the literature, which has generally obtained relatively low values. For instance, in a review of the literature Ermisch, Findlay and Gibb (1996) state that “price elasticity estimates are less dispersed than the income elasticity measures, yielding results between 0.5 and 0.8”; and themselves provide an estimate of 0.4.\textsuperscript{26} I set a range of $\sigma_Z \in [0.4, 0.8]$.

- $\sigma_F$, the elasticity of substitution between real estate and other services in the non-housing sector, does not map closely onto any empirically studied elasticity. In the absence of direct evidence, I set a wide range of $\sigma_F \in [0.5, 1.5]$.

- $\sigma_{G1}$ and $\sigma_{G2}$ are the elasticities of substitution between structures and land in the non-housing and housing sectors, respectively. These elasticities play an important role in the urban economics literature, where substitutability between structures and land in the provision of real estate services is of great practical and theoretical interest.

\textsuperscript{26}$\sigma_Z < 1$ is strongly supported by casual observation as well. For instance, as the real price of housing services has risen in the US over the last several decades, its share of consumption has increased slightly; there is also a well-known tendency for consumers to spend a larger share of their budgets on housing in areas where housing is expensive.
The more voluminous literature is for housing, $\sigma_{G_2}$, with a widely cited early entry by Muth (1971), who estimates $\sigma_{G_2} = 0.5$ using several approaches. More recently, Thorsnes (1997) surveys the literature and finds that recent estimates have generally been below 1, in the range $[0.5, 1]$; but he also argues that some of these estimates may be biased downward due to measurement error, and that the true elasticity may not be much below 1. This claim is seconded by Ahlfeldt and McMillen (2014). In light of these findings, I set a range of $\sigma_{G_2} \in [0.5, 1]$.

The literature for non-housing real estate, $\sigma_{G_1}$, is more scattered, with a range of elasticity estimates similar to that for housing—generally below one, but with concerns about bias from measurement error. For instance, Clapp (1979) obtains elasticities from high-rise office data mostly in the range $[0.5, 0.75]$, but in a tentative attempt to correct for measurement error finds that elasticities closer to 1 may be appropriate. Interpretation is complicated by the fact that non-housing real estate is much more heterogenous than housing real estate, spanning everything from high-rise office towers to farmland. Amid this uncertainty, I also set the range $\sigma_{G_1} \in [0.5, 1]$.

$\sigma_H$ is the elasticity of substitution between equipment and labor. This is of great speculative interest—there are frequent discussions about the extent to which automation, for instance, can replace existing workers, and $\sigma_H$ governs the extent to which the decline in equipment prices documented by Karabarbounis and Neiman (2014b) will lead to substitution away from labor. In his survey, Chirinko (2008) reports a wide range of relevant estimates; the majority are still below one, but several are above one as well, and he suggests that the elasticity for equipment may be higher than the aggregate elasticity. Cummins and Hassett (1992), for instance, obtain implied elasticities of 0.93 for equipment but only 0.28 for structures, and estimates listed by Chirinko (2008) that use computer investment obtain values as high as 1.58. I therefore set a range $\sigma_H \in [0.5, 1.5]$. 

108
2.5.2 Response of the net capital share to exogenous shocks

General methodology. I now study the elasticity of the net capital share with respect to various shocks, in the multisector model whose structure is described in the previous section.

I assume that the quantities \((N, L_1, L_2)\) of labor and both types of land are exogenous. I take final output from \(Z\) to be the numeraire, and assume that the prices \(P_e, P_{s1},\) and \(P_{s2}\) of reproducible capital in terms of this numeraire are exogenously fixed by technology.\(^{27}\) As in (2.2), the user cost of reproducible capital for \(i \in \{e, s_1, s_2\}\) are

\[
W_{Ki} = P_i(r + \delta - g_P)
\]

where the required return \(r\), the depreciation rate \(\delta\), and the expected real change in prices \(g_P\) are also all assumed to be exogenous. As in section 2.3, \(r\) may differ between the non-housing and housing sectors. The quantities \((K_e, K_{s1}, K_{s2})\) of reproducible capital are then given endogenously by demand at this user cost.

I will consider exogenous shocks to either the quantities \((N, L_1, L_2)\) or to either the prices \((P_e, P_{s1}, P_{s2})\) or \(r\), which jointly determine the user costs \((W_{Ke}, W_{Ks1}, W_{Ks2})\).

The elasticity of factor shares in the model with respect to either these shocks depends only on the initial gross and net shares and the local elasticities \((\sigma_Z, \sigma_F, \sigma_{G1}, \sigma_{G2}, \sigma_H)\) of substitution at each level of production; with these in hand, it can be obtained numerically. (Unfortunately, unlike in Oberfield and Raval (2014), elasticities here cannot be expressed in closed form as a weighted average of the individual elasticities \((\sigma_Z, \sigma_F, \sigma_{G1}, \sigma_{G2}, \sigma_H)\). Analytically, this is due to the fact that I assume more than one exogenous quantity.)

I calibrate the initial shares to match the decomposition of the US economy in section 2.3.3 for the final year in the sample, 2013. Table 2.4 displays the resulting gross and net shares of each factor as a fraction of total income, while table 2.5 shows

\(^{27}\)Since housing is probably not an input to the production of equipment or structures, it would be slightly more natural to assume that these prices are fixed relative to the price of non-housing output \(F\); I assume they are fixed relative to \(Z\) for convenience, and in general the relative prices of \(F\) and \(Z\) do not change enough that this has a sizable impact on the results.
the gross shares of each factor as a fraction of the parent aggregate.

Implementation and results. I focus on the elasticity of the net capital share with respect to four specific exogenous shocks:

- A shock to the required rate of return $r$.
- A shock to the price of equipment investment $P_e$.
- A shock to the price of residential structures investment $P_{e2}$.
- A shock to the quantity of residential land $L_2$.

As discussed in greater detail below, the first and second correspond to the Piketty (2014) and Karabarbounis and Neiman (2014b) versions of the accumulation view, respectively. The third and fourth shocks, which relate to residential housing, correspond to my proposed alternative of a “scarcity view”.

The core results are summarized in tables 2.6, 2.7, and 2.8. For table 2.6, I calculate the elasticity of the net capital share with respect to each shock over the full range of $\sigma_i$ deemed plausible in the previous section

$$(0.4, 0.5, 0.5, 0.5, 0.5) \leq (\sigma_Z, \sigma_F, \sigma_{G1}, \sigma_{G2}, \sigma_H) \leq (0.8, 1.5, 1.0, 1.0, 1.5)$$

and report the minimum and maximum elasticities of the net capital share obtained for any combination of $\sigma_i$ in this range. I also calculate the elasticity of the net capital share at a set of “benchmark” $\sigma_i$, which I choose to be the midpoint of the range:

$$(\sigma_Z, \sigma_F, \sigma_{G1}, \sigma_{G2}, \sigma_H) = (0.6, 1.0, 0.75, 0.75, 1.0).$$

Table 2.7 provides additional insight into how different assumptions on $\sigma_i$ combine to produce an aggregate response to shocks. For each shock, the table shows the sensitivity (partial derivative) of the net capital share elasticity to changes in each of the underlying $\sigma_i$, starting from the benchmark values. Essentially, table 2.7 shows the gradient of the values in the “benchmark” column of table 2.6 with respect to perturbations in the $\sigma_i$.  

110
For instance, in the case of a shock to $P_e$, the second row of table 2.7 shows small sensitivities to all $a_i$ except $a_H$, for which the sensitivity is -0.29. This means that if $a_H$ is increased slightly from its benchmark value—say, from $a_H = 1.0$ to $a_H = 1.1$—the elasticity of the net capital share with respect to $P_e$ will decline by 0.029. The intuition in this case is straightforward: when $a_H$ is higher, it is easier to replace equipment with labor in response to higher equipment prices, meaning that a rise in $P_e$ will result in a smaller increase in (or greater decline in) net capital income.

Finally, table 2.8 decomposes the elasticity of the net capital share, at the benchmark $a_i$, into contributing changes in each source of capital income. Each row of table 2.8 sums to the elasticity for the corresponding shock in the “benchmark” column of table 2.6, with one exception: an extra row is included for a shock to $P_e$, showing the decomposition in the “high elasticity” case where each elasticity $a_i$ is chosen to be at the maximum of the range. (This is because there is virtually no effect from the shock to $P_e$ at the benchmark $a_i$.) For instance, for a shock to the price $P_{s2}$ of residential structures investment, the contribution of residential structures $K_{s2}$ is 0.09, out of a total elasticity (from table 2.6) of 0.07; this means that when the cost of residential investment rises, more than 100% of the resulting increase in the net capital share is due to a rise in income from residential structures themselves.

I now discuss and interpret the results for each shock.

**Shock to the required rate of return $r$.** This case tests the Piketty (2014) hypothesis that a rise in savings will push up the net capital share. In general equilibrium, increased savings influences capital income by pushing down the real interest rate; hence, to learn the sign of the effect of savings on the net capital share, it suffices to study the partial equilibrium effect of a change in the real interest rate.

Since the decomposition of the US economy in section 2.3.3 allows $r$ in the non-housing and housing sectors to be different, I define a “shock to $r$” to be a parallel shift $dr$ in these two rates of return. I then define the elasticity of the net capital
share with respect to this shock to be

\[
\frac{\partial (\text{net capital share})/(\text{net capital share})}{\partial r/r^{\text{ave}}}
\]

where \(r^{\text{ave}}\) is the average return on capital across the economy as a whole, including both the non-housing and housing sectors.

Table 2.6 shows that for all \(\sigma_i\) within range (2.21), the response of the net capital share to \(r\) is positive: barely so at minimum (0.04) and strongly so at maximum (0.54). This is inconsistent with the Piketty (2014) hypothesis that a decline in \(r\) can produce an increase in the net capital share, and it corroborates the findings from the single-sector model in section 2.4.

Table 2.7 reveals that the response of the net capital share to \(r\) depends primarily on three elasticities, all negatively: \(\sigma_Z\), \(\sigma_F\), and \(\sigma_H\), each with a sensitivity of about \(-0.20\). Each of these elasticities governs the extent to which an aggregate that includes labor (which is unaffected by \(r\)) can be substituted for an aggregate that does not include labor. But even when these elasticities are chosen at the maximum level in the range \((\sigma_Z = 0.8, \sigma_F = 1.5, \sigma_H = 1.5)\), the response of the net capital share to \(r\) remains slightly positive.

Table 2.8 shows that the vast majority of the response to \(r\) comes from residential structures: at benchmark \(\sigma_i\), a contribution of 0.23 out of an overall elasticity of 0.26. This is for two reasons. First, since both housing \(G_2\) and aggregate consumer demand \(Z\) have \(\sigma_i\)s below 1, the direct positive impact of rising \(r\) on income from residential structures outweighs the negative effect of substitution—much more so than for nonresidential structures or equipment. Second, since section 2.3.3 finds a lower \(r\) for the housing sector than the non-housing sector, a parallel shift in these rates has a disproportionate effect on housing. This reinforces the centrality of housing to any assessment of the Piketty (2014) narrative.

Finally, table 2.8 indicates the importance of a crucial distinction—namely, the distinction between (A) the ratio of housing capital to aggregate income and (B) the share of housing capital income in aggregate income. In response to rising \(r\),
(A) falls: higher $r$ pushes down the demand for residential structures relative to aggregate income\textsuperscript{28}, and since residential land's share of income remains roughly constant in table 2.8, higher $r$ will push down the valuation of this land relative to aggregate income. At the same time, as already discussed, (B) rises dramatically. Hence a shock to $r$ pushes (A) and (B) in different directions, making it important to document (A) and (B) separately.

**Shock to the price of equipment investment $P_e$.** This case tests the Karabarbounis and Neiman (2014b) hypothesis that declining investment prices—which have been concentrated in equipment—will push up the net capital share. As table 2.6 shows, this remains ambiguous for the range of $\sigma_i$ specified in (2.21), which are consistent with either a positive or negative relationship between $P_e$ and the net capital share.

Table 2.7 makes clear the source of this ambiguity: the response of the net capital share to $P_e$ depends almost entirely on the elasticity of substitution $\sigma_H$ between labor and equipment. When $\sigma_H$ is near the top of the $[0.5, 1.5]$ range, falling $P_e$ leads to a rise in the net capital share, consistent with Karabarbounis and Neiman (2014b); when $\sigma_H$ is near the bottom, the opposite is true.

The “high elasticity” row in table 2.8, however, provides cause for skepticism of the Karabarbounis and Neiman (2014b) channel. Here, $P_e$ has a substantial negative effect on the net capital share. But this effect comes almost exclusively from the net capital income of equipment itself—which, in this partial equilibrium exercise, moves in parallel with the value of the equipment stock—rather than through some less direct channel. Section 2.3 found that the value of equipment (which has recently fallen) has followed a path quite distinct from the path of the net capital share (which has recently risen). This is not consistent with a major role for $P_e$.

For the $P_e$ hypothesis to be consistent with the data, it would be necessary for declining $P_e$ to push up the net capital share through some channel other than a rise in the value of the equipment stock. Karabarbounis and Neiman (2014a) sketch

\textsuperscript{28}This occurs in the model, but not directly visible in tables 2.6 through 2.8.
one such possibility, where falling $P_e$ can lead to an increase in the net capital share despite an actual decline in the aggregate value of equipment, but the multisector model here does not corroborate their mechanism.\footnote{Karabarbounis and Neiman (2014a) devise a model where two types of capital, high-depreciation (which can be interpreted as equipment) and low-depreciation (which can be interpreted as structures) combine to form a capital aggregate; the elasticity of substitution between these types of capital is less than 1, while the elasticity of substitution between the capital aggregate and labor is greater than 1. A decline in the price of equipment lowers the price of the capital aggregate, which induces substitution from labor to the capital aggregate; but since the elasticity of substitution between equipment and structures is less than 1, this also causes a decline in equipment relative to structures. With the right parameters, it is possible for a decline in the price of equipment to increase the net capital share while net capital income from equipment itself actually declines.}

**Shock to the price of residential structures investment $P_{e2}$.** In table 2.6, a rise in $P_{e2}$ leads to a rise in the net capital share for benchmark $\sigma_i$; for other choices of $\sigma_i$ within range (2.21), there is at worst roughly no effect. According to table 2.7, the effect is most sensitive to the elasticity $\sigma_Z$ of substitution between housing and non-housing output; and according to table 2.8 it works almost entirely through the net capital income of residential structures themselves. The mechanism here is relatively simple: when the ability to substitute away from housing is limited, costlier residential investment leads to a higher-value housing stock and a larger share of income accruing to housing.

**Shock to the quantity of residential land $L_2$.** This is similar to the previous case. In table 2.6, a decline in the quantity of residential land $L_2$ leads to a rise in the net capital share for benchmark $\sigma_i$; for other choices of $\sigma_i$ within range (2.21), there is at worst roughly no effect. Again, according to table 2.7, the effect is most sensitive to $\sigma_Z$; now, however, the effect is smaller and works mainly through the net capital income earned by residential land.

**Summary of results and conclusion.** I have examined the response of the multi-sector model to four exogenous shocks. The first two shocks correspond to versions of the accumulation view: a shock to $r$ captures the general equilibrium channel through which the rise in savings postulated by Piketty (2014) affects factor shares, while a
shock to $P_e$ is central to the Karabarbounis and Neiman (2014b) narrative.

In both cases, the results do not support the proposed mechanism. For all choices of $\{\sigma_i\}$ within the range considered, a fall in $r$ leads to a fall in the net capital share, in contrast with Piketty (2014). Meanwhile, although a fall in $P_e$ can produce a rise in the net capital share, it only does so by pushing up the net capital income from equipment itself, which is at odds with the evidence from section 2.3.

The latter two shocks both embody some form of the scarcity view, which is more successful in the multisector model. Either a rise in the price $P_{s2}$ of residential investment or a fall in the quantity $L_{s2}$ of residential land leads to a rise in the net capital share, for the vast majority of $\{\sigma_i\}$ in the range (2.21). In both cases, the mechanism works through increasing the net capital income earned by housing, consistent with the dramatic rise in the contribution of housing documented in section 2.2.

### 2.5.3 Counterfactual exercise

Building upon the promise of the scarcity view in the previous section, I now use the multisector model to perform a counterfactual exercise, exploring the implications of alternative paths for $P_{s2}$ and $L_{s2}$.

The real price $P_{s2}$ of residential investment has risen in the last several decades in the US; furthermore, real output has grown substantially, putting pressure on the supply of residential land. I consider a counterfactual where these two forces are not present: where the real price of $P_{s2}$ is instead constant from the beginning of the sample period (1948) onward, and where the quantity of residential land $L_{s2}$ grows in tandem with real output from the beginning of the sample period onward.$^{30}$

In contrast to the exercises in section 2.5.2, which consider only local shocks to exogenous variables, this counterfactual involves large global changes. It requires additional, global assumptions to compute; for this purpose, I will assume that the

---

$^{30}$To make this modification, I assume that the quantity of land $L_{s2}$ was in reality constant, and then expand it in each year by a fraction equal to cumulative real GDP growth since 1948. Depending on the interpretation of $L_{s2}$, the assumption that it has been constant may or may not be appropriate.
production functions \((Z, F, G_1, G_2, H)\) each have a globally constant elasticity of substitution. I consider two choices of \(\{\sigma_i\}\): first, the benchmark \((\sigma_Z, \sigma_F, \sigma_{G_1}, \sigma_{G_2}, \sigma_H) = (0.6, 1.0, 0.75, 0.75, 1.0)\); and second, the alternative \((0.4, 0.5, 0.75, 0.75, 0.5)\) that sets \(\sigma_Z, \sigma_F, \) and \(\sigma_H\) (the \(\sigma\)s that govern the response to \(P_{z2}\) and \(L_2\), according to table 2.7) to the minimum values in the range \((2.21)\).

Figure 2-12 displays the results of this exercise, distinguishing between the housing and non-housing components of the net capital share. Consistent with table 2.8, there is little effect working through the non-housing component. Furthermore, the large initial increase and then decline in the housing component, through 1980, is left untouched by the counterfactuals. Much of the subsequent increase in the housing component, however, is eroded.

This is consistent with a role for rising residential investment costs, along with growing scarcity of residential land, in driving up housing's contribution to the net capital share: when these forces are reversed in a counterfactual, we see less of a rise. At the same time, figure 2-12 makes clear the limitations of this account. It does not explain the fall and rise in the non-housing component, nor can it explain all aspects of the housing time series. The scarcity view, therefore, is only a partial replacement for the accumulation view: it achieves better consistency with data and theory, but does not purport to explain more than a fragment of the evolving factor income distribution.

### 2.6 Conclusion

The aging Kaldor facts have retreated in the face of experience. Today, macroeconomists no longer claim that factor shares are constant—but what should replace the old consensus?

It is increasingly commonplace to believe that labor is ceding ground to capital. But a closer look at postwar experience reveals a murkier story, in which steady increase is limited to the gross capital share. The net share, by contrast, has fallen and then recovered; it consists of a large long-term increase in net capital income.
from housing, and a more volatile contribution from the rest of the economy with little cumulative movement in either direction.

Even more elusive than these facts is a cohesive explanation of them. The accumulation view, in both its Piketty (2014) and Karabarbounis and Neiman (2014b) variants, falters in multiple respects. It cannot explain the dominant role of housing, nor can it be readily reconciled with the evidence on elasticities of substitution. Outside of housing, there appears to be little correlation between the capital-income ratio and the net capital share.

The rise in housing's contribution to the capital share, by contrast, can be explained in part as the result of scarcity. The rising real cost of residential investment and the limited quantity of residential land have conspired to make housing more expensive, and given low elasticities of substitution this has meant a rise in housing's share of income.

With these trends in mind, policymakers concerned about the distribution of income should keep an eye on housing costs—many urban economists, including Glaeser, Gyourko and Saks (2005) and Quigley and Raphael (2005), have documented explicitly how restrictions on land use and residential construction inflate the cost of housing. Outside of housing, however, this paper raises more questions than it answers about the evolution of the net capital share: once the accumulation view has been discarded, there is no master narrative at hand that can explain the postwar fall and rise.

If anything, these results suggest that concern about inequality should be shifted away from the overall split between capital and labor, and toward other aspects of distribution, such as the within-labor distribution of income. Although the net capital share has at times seen dramatic shifts both up and down, away from housing its long-term movement has been quite small, and there is no compelling reason to suggest that this pattern will change going forward.

No doubt, however, the distribution between capital and labor will continue to be a salient issue: we surely have not seen the last of Ricardo (1821)'s principal problem of Political Economy.
2.7 Appendix

2.7.1 Figures

Figure 2-1: Average net capital share of private domestic value added for G7 countries.
Figure 2-2: Average gross capital share of private domestic value added for G7 countries.

Figure 2-3: Components of average net capital share of private domestic value added for G7 countries: housing (h) versus other (nh) sectors, weighted (w) and unweighted (uw).
Figure 2-4: Average net capital shares of corporate sector value added for G7 countries.

Figure 2-5: Net capital share of corporate sector value added in the US.
Figure 2-6: Decomposition of net capital share of corporate sector value added in the US: return on equipment, structures, land, and pure profits $\pi$.

Figure 2-7: Estimated constant, linear, and quadratic time trends for the corporate rate of return $r(t)$. 
Figure 2-8: Decomposition of net capital share of corporate sector value added in the US: return on equipment, structures, land, and pure profits \( \pi \), using quadratic trend for \( r(t) \).

Figure 2-9: Ratio of total market value to the recorded value of equipment, structures, and land ("book value"), US corporate sector.
Figure 2-10: Net capital share of private value added in the US.

Figure 2-11: Decomposition of net capital share of private domestic value added in the US: return on equipment (eq), non-residential structures (st-nh), non-residential land (l-nh), pure profits $\pi$, residential structures (st-h), and residential land (l-h).
Figure 2-12: Counterfactual paths—assuming no change in the real price of residential structures investment, and a constant ratio of the quantity of residential land to the quantity of real output—for the housing (lower) and non-housing (upper) components of the net capital share.
### 2.7.2 Tables

Table 2.2: Decadal averages for the net capital share of private domestic value added, broken into housing and non-housing ("other") components.

<table>
<thead>
<tr>
<th></th>
<th>1950s</th>
<th>1960s</th>
<th>1970s</th>
<th>1980s</th>
<th>1990s</th>
<th>2000s</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>USA</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Housing</td>
<td>5.3%</td>
<td>6.5%</td>
<td>5.7%</td>
<td>7.2%</td>
<td>8.4%</td>
<td>8.2%</td>
</tr>
<tr>
<td>Other</td>
<td>22.0%</td>
<td>21.7%</td>
<td>18.6%</td>
<td>18.4%</td>
<td>19.2%</td>
<td>19.4%</td>
</tr>
<tr>
<td>Total</td>
<td>27.3%</td>
<td>28.2%</td>
<td>24.2%</td>
<td>25.6%</td>
<td>27.5%</td>
<td>27.6%</td>
</tr>
<tr>
<td><strong>Japan</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Housing</td>
<td>4.2%</td>
<td>3.6%</td>
<td>4.1%</td>
<td>5.2%</td>
<td>7.0%</td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>31.2%</td>
<td>26.9%</td>
<td>25.7%</td>
<td>21.6%</td>
<td>20.1%</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>35.4%</td>
<td>30.5%</td>
<td>29.8%</td>
<td>26.9%</td>
<td>27.1%</td>
<td></td>
</tr>
<tr>
<td><strong>Germany</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Housing</td>
<td>3.6%</td>
<td>5.1%</td>
<td>5.9%</td>
<td>7.1%</td>
<td>9.8%</td>
<td>10.8%</td>
</tr>
<tr>
<td>Other</td>
<td>21.3%</td>
<td>19.8%</td>
<td>17.9%</td>
<td>16.6%</td>
<td>19.9%</td>
<td>18.0%</td>
</tr>
<tr>
<td>Total</td>
<td>24.9%</td>
<td>24.9%</td>
<td>23.8%</td>
<td>23.7%</td>
<td>29.7%</td>
<td>28.8%</td>
</tr>
<tr>
<td><strong>France</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Housing</td>
<td>1.2%</td>
<td>2.1%</td>
<td>3.8%</td>
<td>4.6%</td>
<td>5.8%</td>
<td>7.3%</td>
</tr>
<tr>
<td>Other</td>
<td>27.2%</td>
<td>23.9%</td>
<td>18.3%</td>
<td>21.6%</td>
<td>23.2%</td>
<td>23.4%</td>
</tr>
<tr>
<td>Total</td>
<td>28.4%</td>
<td>26.0%</td>
<td>22.1%</td>
<td>26.2%</td>
<td>29.0%</td>
<td>30.7%</td>
</tr>
<tr>
<td><strong>UK</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Housing</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4.3%</td>
<td>6.4%</td>
</tr>
<tr>
<td>Other</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>33.9%</td>
<td>32.5%</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>38.2%</td>
<td>38.9%</td>
</tr>
<tr>
<td><strong>Italy</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Housing</td>
<td>6.6%</td>
<td>6.6%</td>
<td>8.1%</td>
<td>10.4%</td>
<td>8.6%</td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>22.5%</td>
<td>24.0%</td>
<td>25.8%</td>
<td>21.2%</td>
<td>27.2%</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>29.1%</td>
<td>30.6%</td>
<td>33.8%</td>
<td>31.6%</td>
<td>35.8%</td>
<td></td>
</tr>
</tbody>
</table>

125
Table 2.4: Gross and net shares of factors and higher-level aggregates taken from 2013 decomposition in section 2.3.3, used to calibrate the multisector model.

<table>
<thead>
<tr>
<th></th>
<th>Gross aggregate share</th>
<th>Net aggregate share</th>
</tr>
</thead>
<tbody>
<tr>
<td>$N$</td>
<td>60%</td>
<td>68%</td>
</tr>
<tr>
<td>$K_e$</td>
<td>12%</td>
<td>7%</td>
</tr>
<tr>
<td>$K_{s1}$</td>
<td>12%</td>
<td>11%</td>
</tr>
<tr>
<td>$L_1$</td>
<td>3%</td>
<td>3%</td>
</tr>
<tr>
<td>$K_{s2}$</td>
<td>10%</td>
<td>8%</td>
</tr>
<tr>
<td>$L_2$</td>
<td>1%</td>
<td>1%</td>
</tr>
<tr>
<td>$\pi$</td>
<td>1%</td>
<td>2%</td>
</tr>
<tr>
<td>$H$</td>
<td>72%</td>
<td>76%</td>
</tr>
<tr>
<td>$G_1$</td>
<td>15%</td>
<td>14%</td>
</tr>
<tr>
<td>$G_2$</td>
<td>11%</td>
<td>9%</td>
</tr>
<tr>
<td>$F$</td>
<td>88%</td>
<td>90%</td>
</tr>
</tbody>
</table>

Table 2.3: Decadal averages for the net capital share of value added in the domestic corporate sector.

<table>
<thead>
<tr>
<th></th>
<th>1950s</th>
<th>1960s</th>
<th>1970s</th>
<th>1980s</th>
<th>1990s</th>
<th>2000s</th>
</tr>
</thead>
<tbody>
<tr>
<td>USA</td>
<td>23.2%</td>
<td>23.2%</td>
<td>19.7%</td>
<td>19.8%</td>
<td>20.9%</td>
<td>21.1%</td>
</tr>
<tr>
<td>Germany</td>
<td></td>
<td></td>
<td></td>
<td>24.2%</td>
<td>29.0%</td>
<td></td>
</tr>
<tr>
<td>France</td>
<td>22.1%</td>
<td>20.9%</td>
<td>19.0%</td>
<td>17.9%</td>
<td>22.1%</td>
<td>20.1%</td>
</tr>
<tr>
<td>UK</td>
<td>27.6%</td>
<td>24.4%</td>
<td>19.0%</td>
<td>22.7%</td>
<td>24.7%</td>
<td>25.3%</td>
</tr>
<tr>
<td>Italy</td>
<td></td>
<td></td>
<td></td>
<td>35.4%</td>
<td>34.6%</td>
<td></td>
</tr>
<tr>
<td>Canada</td>
<td>24.5%</td>
<td>26.1%</td>
<td>28.5%</td>
<td>24.3%</td>
<td>30.1%</td>
<td></td>
</tr>
</tbody>
</table>
Table 2.5: Gross shares of production within each higher-level aggregate in calibrated the multisector model, based on shares in table 2.4.

<table>
<thead>
<tr>
<th></th>
<th>Gross share</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>H</strong></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>83%</td>
</tr>
<tr>
<td>Ke</td>
<td>17%</td>
</tr>
<tr>
<td><strong>G1</strong></td>
<td></td>
</tr>
<tr>
<td>Ks1</td>
<td>82%</td>
</tr>
<tr>
<td>L1</td>
<td>18%</td>
</tr>
<tr>
<td><strong>G2</strong></td>
<td></td>
</tr>
<tr>
<td>Ks2</td>
<td>90%</td>
</tr>
<tr>
<td>L2</td>
<td>10%</td>
</tr>
<tr>
<td><strong>F</strong></td>
<td></td>
</tr>
<tr>
<td>H</td>
<td>83%</td>
</tr>
<tr>
<td>G1</td>
<td>17%</td>
</tr>
<tr>
<td><strong>Z</strong></td>
<td></td>
</tr>
<tr>
<td>F</td>
<td>89%</td>
</tr>
<tr>
<td>G2</td>
<td>11%</td>
</tr>
</tbody>
</table>
Table 2.6: Minimum and maximum elasticities (for choices of $\sigma_i$ within range) of net capital share with respect to shocks, in addition to elasticity for benchmark $\sigma_i$.

<table>
<thead>
<tr>
<th>Shock</th>
<th>Min</th>
<th>Max</th>
<th>Benchmark</th>
</tr>
</thead>
<tbody>
<tr>
<td>$r$</td>
<td>0.04</td>
<td>0.54</td>
<td>0.26</td>
</tr>
<tr>
<td>$P_e$</td>
<td>-0.18</td>
<td>0.15</td>
<td>0.00</td>
</tr>
<tr>
<td>$P_{e2}$</td>
<td>-0.00</td>
<td>0.16</td>
<td>0.07</td>
</tr>
<tr>
<td>$L_2$</td>
<td>-0.04</td>
<td>0.00</td>
<td>-0.01</td>
</tr>
</tbody>
</table>

Table 2.7: Sensitivity of the elasticity of net capital share to changes in each $\sigma_i$, starting at benchmark values.

<table>
<thead>
<tr>
<th>Sensitivity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shock</td>
</tr>
<tr>
<td>-------</td>
</tr>
<tr>
<td>$r$</td>
</tr>
<tr>
<td>$P_e$</td>
</tr>
<tr>
<td>$P_{e2}$</td>
</tr>
<tr>
<td>$L_2$</td>
</tr>
</tbody>
</table>

Table 2.8: Contribution to the aggregate elasticity of the net capital share, for benchmark $\sigma_i$.

<table>
<thead>
<tr>
<th>Contribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shock</td>
</tr>
<tr>
<td>-------</td>
</tr>
<tr>
<td>$r$</td>
</tr>
<tr>
<td>$P_e$</td>
</tr>
<tr>
<td>$P_{e2}$ (high elasticity case)</td>
</tr>
<tr>
<td>$P_{s2}$</td>
</tr>
<tr>
<td>$L_2$</td>
</tr>
</tbody>
</table>

128
2.7.3 Description of alternative procedure in section 2.3.2 for estimating path of $r$.

I sketch here the procedure in section 2.3.2 for estimating the effective required return $r(t)$ on capital for the US corporate sector. I specify the model in continuous time, and use superscripts to denote the time $t$ for economy of notation. I am also more explicit here about how the underlying continuous time flows are aggregated into the measured flow for a given time period.

Assumed stochastic processes

I assume two stochastic processes beyond what is already visible in the data:

- $\pi^t$, which is a stationary, ergodic process for the share of gross output that goes to profits.

- $\zeta^t$, which reflects stochastic pricing error for the total market value of corporations, with mean 1 (where 1 corresponds to no error).

Core relations

First relation: profit share of flows. We know that all non-profit income will be allocated between depreciation, labor, and the various types of capital. This is a flow relation

\[
(1 - \pi^t)Y^t = w^tL^t + \sum_i (\delta_i + r^t - g_P)P_i^tK_i^t
\]

which can be rewritten as

\[
\pi^t = 1 - \frac{w^tL^t}{Y^t} - \sum_i \frac{\delta_iP_i^tK_i^t}{Y^t} - \sum_i (r^t - g_P)\frac{P_i^tK_i^t}{Y^t}
\]  \hspace{1cm} (2.22)

or, if we don’t want to divide by $Y^t$, as

\[
\pi^tY^t = Y^t - w^tL^t - \sum_i \delta_iP_i^tK_i^t - \sum_i (r^t - g_P)P_i^tK_i^t
\]
Consolidating into an accumulated flow. Suppose that we write

\[ \int_t^{t+\Delta t} \pi^* s ds \]

\[ = \int_t^{t+\Delta t} (Y^* - W^* L^* - \sum_i \delta_i P_i^* K_i^* ds - \sum_i \int_t^{t+\Delta t} (r^* - g_P) P_i^* K_i^* ds \quad (2.23) \]

We can identify the first part as simply real net capital income during the period, while for the second term we must write

\[ \int_t^{t+\Delta t} (r^* - g_P) P_i^* K_i^* ds \approx \left((r^{t+\Delta t} - g_P) P_i^{t+\Delta t} K_i^{t+\Delta t} + (r^t - g_P) P_i^t K_i^t\right) \frac{\Delta t}{2} \quad (2.24) \]

Second relation: asset pricing. The expected discounted value of the profit stream from time \( t \) onward is (in real terms)

\[ Y^t \cdot \int_0^\infty e^{(g_Y - \delta_r)s - \int_t^{t+s} r^* du} E_t[\pi^{t+s}] \]

where \( \delta_r \) is the rate at which pure profits decay. (We can think of it as the rate at which a given company, for instance, on average loses the ability to make pure profits. There is no clear basis for picking \( \delta_r \), and I will choose \( \delta_r = .015 \), which implies a half-life of just below 50 years—within reason given the typical lifetimes of American corporations. Fortunately, the precise choice of \( \delta_r \) does not matter much for the results.)

This expected discounted value plus the value of capital itself is (again in real terms)

\[ Y^t \cdot \int_0^\infty e^{(g_Y - \delta_r)s - \int_t^{t+s} r^* du} E_t[\pi^{t+s}] ds + \sum_i P_i^t K_i^t \]

I assume that the market value of the corporate sector equals this overall value times \( \zeta^t \), the multiplicative stochastic pricing error that has mean 1, follows a stationary, ergodic process, and is drawn independently of \( \pi, \{P_i^t\}, \{K_i^t\}, \) and \( Y^t \):

\[ MV^t = \zeta^t \left( Y^t \cdot \int_0^\infty e^{(g_Y - \delta_r)s - \int_t^{t+s} r^* du} E_t[\pi^{t+s}] ds + \sum_i P_i^t K_i^t \right) \quad (2.25) \]
Define $OMV^t \equiv MV^t - \sum_i P_i^t K_i^t$, and rewrite (2.25) as

$$OMV^t = \zeta^t \left( Y^t \cdot \int_0^\infty e^{(\gamma - \delta)x} f_i^{t+s} r^s du E_t[\pi_t^{t+s}] ds \right) + (\zeta^t - 1) \sum_i P_i^t K_i^t \quad (2.26)$$

Second relation, part two: taking first differences. Now use (2.26) to compute

$$\frac{\phi(t)}{Y_{t-1,t}} \left( OMV^t - e^{-\delta \Delta t - f_i^{t+\Delta t} r^s du} \cdot OMV^{t+\Delta t} \right)$$

for some small $\Delta t$, dividing by $Y_{t-1,t}$ (to be defined later, but known at time $t$) and multiplying by any deterministic function $\phi(t)$ of $t$. Expanding the term inside the parentheses in (2.27), we obtain

$$OMV^t - e^{-\delta \Delta t - f_i^{t+\Delta t} r^s du} \cdot OMV^{t+\Delta t}$$

$$= \zeta^t \left( Y^t \int_0^\infty e^{(\gamma - \delta)x} f_i^{t+s} r^s du E_t[\pi_t^{t+s}] ds \right)$$

$$- \zeta^{t+\Delta t} \left( e^{-\gamma \Delta t} Y^{t+\Delta t} \int_0^\infty e^{(\gamma - \delta)x} f_i^{t+s} r^s du E_{t+\Delta t}[\pi_{t+s}] ds \right)$$

$$+ (\zeta^t - 1) Y^t \sum_i P_i^t K_i^t + e^{-\delta \Delta t - f_i^{t+\Delta t} r^s du} (\zeta^{t+\Delta t} - 1) Y^{t+\Delta t} \sum_i P_i^{t+\Delta t} K_i^{t+\Delta t} \quad (2.28)$$

Suppose now that we take the unconditional expectation of (2.27). Given the assumed independence of $\zeta^t, Y^t$, and $\pi^t$, (2.28) simplifies dramatically and we are left with

$$E \left[ \frac{\phi(t)}{Y_{t-1,t}} \left( OMV^t - e^{-\delta \Delta t - f_i^{t+\Delta t} r^s du} \cdot OMV^{t+\Delta t} \right) \right]$$

$$= E \left[ \frac{\phi(t)}{Y_{t-1,t}} Y^t \int_0^{\Delta t} e^{(\gamma - \delta)x} f_i^{t+s} r^s du \pi_t^{t+s} ds \right] \quad (2.29)$$
We can further manipulate (2.29), using the law of iterated expectations to obtain

\[
E \left[ \frac{\phi(t)}{Y^{t-1,t}} Y^t \int_0^{\Delta t} e^{(\gamma - \delta) s - \int_t^{t+s} r_u \pi_s^t + s \, ds} \, ds \right]
\]

(2.30)

\[
= E \left[ \frac{\phi(t)}{Y^{t-1,t}} E_t \left[ \int_0^{\Delta t} (e^{\gamma \pi^t Y^t}) e^{-\delta s - \int_t^{t+s} r_u \pi_s^t + s \, ds} \, ds \right] \right]
\]

(2.31)

\[
= E \left[ \frac{\phi(t)}{Y^{t-1,t}} \int_0^{\Delta t} e^{-\delta s - \int_t^{t+s} r_u \pi_s^t + s \, ds} Y^{t+s} \, ds \right]
\]

Assuming that \( \Delta t \) is small enough and \( \pi^{t+s} Y^{t+s} \) is sufficiently close to being continuous, we can approximate the integral inside (2.31) by

\[
\int_0^{\Delta t} e^{-\delta s - \int_t^{t+s} r_u \pi_s^t + s \, ds} \approx 1 + e^{-\delta s \Delta t - (r^t + r^{t+s})/2} \int_0^{\Delta t} \pi^{t+s} Y^{t+s} \, ds
\]

(2.32)

**Full estimation strategy.** We have shown that the unconditional expectation of (2.27), which we can approximate by

\[
E \left[ \frac{\phi(t)}{Y^{t-1,t}} \left( OMV^t - e^{-\delta s \Delta t - (r^t + r^{t+s})/2} OMV^{t+s} \right) \right]
\]

(2.33)

has unconditional expectation approximately equal to

\[
E \left[ \frac{\phi(t)}{Y^{t-1,t}} \cdot \frac{1 + e^{-\delta s \Delta t - (r^t + r^{t+s})/2}}{2} \cdot \int_0^{\Delta t} \pi^{t+s} Y^{t+s} \, ds \right]
\]

(2.34)

where according to (2.23) and (2.24), we can obtain the approximate flow of pure profits \( \int_0^{\Delta t} \pi^{t+s} Y^{t+s} \, ds \) in (2.34) as

\[
\int_0^{\Delta t} \pi^{t+s} Y^{t+s} \, ds \approx \int_t^{t+\Delta t} (Y^s - w^s L^s)

- \sum_i \delta_i P^i \pi^{t+s} K^s_i \, ds

- \sum_i \left( (r_i^t + \Delta t) P^t_i K_i^{t+\Delta t} + (r_i^t - g_i^t) P^t_i K_i^t \right) \frac{\Delta t}{2}
\]

(2.35)
where the first term is just the recorded net return on capital in the period \([t, t + \Delta t]\) as measured in the national accounts, while the second term can be derived from the nominal quantities \(P_iK_i\) of each type of capital.

The moments (2.33) and (2.34) are equal, and we can set the corresponding sample moments equal to each other. Generally I will look at an annual frequency, such that \(\Delta t = 1\). Given a functional form for \(\tau^t\) with \(n\) free parameters to be pinned down, we can choose \(n\) functions for \(\phi(t)\) to give us \(n\) sample moment conditions that determine those parameters.
Bibliography


Chapter 3

Unique Equilibrium in the Eaton-Gersovitz Model of Sovereign Debt

3.1 Introduction

A common view of sovereign debt markets is that they are prone to multiple equilibria: a market panic may inflate bond yields, deteriorate the sustainability of government debt and precipitate a default event, justifying investor fears. Indeed, Mario Draghi’s speech in July 2012, announcing that the ECB was “ready to do whatever it takes” to preserve the single currency, and the subsequent creation of the Outright Monetary Transactions (OMT) program, are widely seen as having moved Eurozone sovereign debt markets out of an adverse equilibrium: since then, bond spreads have experienced dramatic falls as fears of default have receded.

This chapter is coauthored with Adrien Auclert. We thank Iván Werning for inspiration, continued encouragement and many useful suggestions. We also thank Daron Acemoglu, Marios Angeles-tos, Fernando Broner, Daniel Green, Pooya Molavi, Juan Passadore, Jonathan Parker, Alp Simsek, Olivier Wang and Yu Xu for helpful comments. Remaining errors are our own. Adrien Auclert and Matthew Rognlie gratefully acknowledge financial support from the Macro-Financial Modeling group and NSF Graduate Research Fellowship program, respectively.
At the same time, in the last decade, a booming quantitative literature in the line of Eaton and Gersovitz (1981)—initiated by Arellano (2008) and Aguiar and Gopinath (2006), and summarized in Aguiar and Amador (2015)—has studied sovereign debt markets using an infinite-horizon incomplete markets model for which no result on equilibrium multiplicity was known. An example of how the literature viewed this issue is in Hatchondo, Martinez and Sapriza (2009):

Krusell and Smith (2003) show that, typically, there is a problem of indeterminacy of Markov-perfect equilibria in an infinite-horizon economy. In order to avoid this problem, we analyze the equilibrium that arises as the limit of the finite-horizon economy equilibrium.

In this paper, we show that equilibrium is unique in the benchmark infinite-horizon model with a Markov process for the exogenous driving state and exogenous value from default. Although we emphasize Markov perfect equilibrium—the usual equilibrium concept in the literature, and one for which our argument is especially direct—we prove that our core uniqueness result extends to subgame perfect equilibria more generally. We also extend our proof to several modifications of the benchmark model, as described below.

Why could multiplicity arise in the benchmark model we study? To build intuition, consider the simplest environment: one in which debt is restricted to be risk-free, as in Zhang (1997). A Markov perfect equilibrium of this model features a (constant) endogenous debt limit, which is the most that can be incurred today without the possibility that the government will want to default tomorrow. Suppose credit becomes tighter in the future—tomorrow’s debt limit falls. Since the government is less able to smooth consumption fluctuations, its perceived benefit from access to credit is now lower, and so its willingness to repay today’s debts falls. In response, investors lower today’s debt limit as well.

Through this process, an equilibrium with loose credit and high willingness to pay debts could turn into one with tight credit and low willingness to repay. Similarly, in the full Eaton and Gersovitz (1981) model with risky debt, investors’ pessimistic
expectations about the likelihood of default could translate into higher risk premia on debt—which, by making debt service more costly and continued access to credit markets less valuable, would encourage default and validate the original pessimism. This mechanism sounds appealing, and in our view it captures an important part of the common intuition for equilibrium multiplicity in sovereign debt markets.

Our results rule it out. The intuition remains simplest in the Zhang (1997) environment. If there are two equilibria with distinct debt limits, we consider two governments that are each at the limit in their respective equilibria. We argue that the government with less debt must have a strictly higher value: starting from that point, it can follow a strategy that parallels the strategy of the higher-debt government, maintaining its liabilities at a uniform distance and achieving higher consumption at every point by economizing on interest payments. But this contradicts the assumption that both governments start at their debt limits, where each must obtain the (constant) value of default. In short, once both governments have exhausted their debt capacity, the one with a strictly lower level of debt is strictly better off—meaning that this government should be able to borrow slightly more without running the risk of default, and cannot have exhausted its capacity after all.

Interestingly, this proof strategy by replication has echoes of that used by Bulow and Rogoff (1989) to rule out reputational equilibria in a similar class of models where sovereign governments retain the ability to save after defaulting. The original Bulow-Rogoff result is cast in a complete markets setting. In a second modification of the benchmark model, we specify the only punishment from default as the loss of ability to borrow. As an immediate corollary to our Eaton-Gersovitz uniqueness result, we then obtain the *incomplete markets Bulow-Rogoff result*: under this specification of default costs, the no-borrowing equilibrium is the unique equilibrium. Hence our general uniqueness result nests a key impossibility result for the sovereign debt literature. Here, our paper complements parallel and independent work by Bloise, Polemarchakis and Vailakis (2015), who explore the validity of the Bulow-Rogoff result in environments with general asset market structures.

We next explore the importance of the model’s assumptions by relaxing each of
them in turn, and then considering the robustness of our uniqueness result. We first consider a case where savings are exogenously bounded. We prove that uniqueness holds provided that the bound on savings is strictly positive. This provides continuity with the result of Passadore and Xandri (2014), who have found multiplicity when the bound is zero. Next, we consider a case where the value of default is no longer exogenous. When this endogeneity comes from the assumption that governments in default have a stochastic option to reenter markets (a typical assumption in the quantitative literature), we rule out multiplicity of the most widely suspected form—where bond prices in a favorable equilibrium dominate those in a self-fulfilling adverse one—and obtain complete uniqueness when shocks are independent and identically distributed. Finally, we explore the role of our timing and commitment assumptions by discussing the way in which they differ from those in other models from the literature that do feature multiple equilibria (Calvo, 1988, Cole and Kehoe, 2000 and Lorenzoni and Werning, 2014).

Our results are important because they show that the multiplicity intuition is not valid in a benchmark model that is accepted as a good description—both qualitative and quantitative—of sovereign debt markets. They provide an additional analytical result for a model about which few such results exist, making use of a powerful new proof technique along the way. And they show that alternative strategies to compute Markov perfect equilibria should all converge to the same solution. Our results are not directly applicable to all the extensions of the Eaton-Gersovitz model that the quantitative sovereign debt literature has considered, but they do suggest that multiplicity is unlikely in many of these cases as well, and therefore that the literature’s quantitative findings are probably not driven by a hidden equilibrium selection.

\footnote{In particular, under our assumptions, sunspots cannot influence equilibrium outcomes. Recently, Stangebye (2015) has explored the role of sunspots in two versions of the Eaton-Gersovitz model where our results do not apply—first, for short-term debt, when the domain of debt is exogenously restricted beyond what we consider in section 3.4.1, and second, for long-term debt.}

\footnote{While our focus is on sovereign debt, the benchmark model we study also constitutes the core of a literature that analyzes unsecured consumer credit (Chatterjee, Corbae, Nakajima and Rios-Rull, 2007), and we conjecture that equilibrium is also unique in many of the models used in that literature.}
Our objective is not to deny that sovereign debt markets can be prone to self-fulfilling crises, or that OMT may have ruled out a bad equilibrium. Instead, we hope that our results may help sharpen the literature’s understanding of the assumptions that are needed for such multiple equilibria to exist. Our replication-based proof strategy may also be of independent interest, as a general technique for proving uniqueness of equilibrium in infinite-horizon games.

Layout. The rest of the paper is organized as follows. Section 3.2 lays out the benchmark Eaton-Gersovitz model, and establishes uniqueness of Markov perfect equilibrium and uniqueness of subgame perfect equilibrium. Section 3.3 adapts our main proof to two related models. Section 3.3.1 proves uniqueness in the Zhang (1997) model, where debt is restricted to be risk-free, and section 3.3.2 derives the incomplete markets version of the Bulow and Rogoff (1989) result as a corollary of our main uniqueness result. Section 3.4 considers the robustness of our results as we relax various assumptions. Section 3.4.1 considers exogenous restrictions on savings. Section 3.4.2 considers the case where reentry is allowed after default. Section 3.4.3 discusses our timing and commitment assumptions, contrasting them with those in the related literature. Section 3.5 concludes.

3.2 Equilibrium uniqueness in the benchmark model

3.2.1 Model description

We now describe what we call the benchmark infinite-horizon model with Markov income (see Aguiar and Amador, 2015). We focus first on Markov perfect equilibria, in which the current states $b$ and $s$ encode all the relevant history. In section 3.2.3 we will show that this is without loss of generality, since one can specify this model as a game whose only subgame perfect equilibria are Markov perfect equilibria.

An exogenous state $s$ follows a discrete Markov chain with elements in $S$, $|S| = S \in \mathbb{N}$ and transition matrix $\pi(s'|s)$. Output $y(s)$ is a function of this underlying state.
At the beginning of each period, the government starts with some level of debt $b$. After observing the realization of $s$, it decides whether to repay $b$ or default. If it does not repay, it receives an exogenous value $V^d(s)$, which encodes all the consequences of default. For example, if default is punished by permanent autarky, with output also reduced by an exogenous cost $\tau(s) \in [0, y(s)]$, then $V^d$ is defined recursively by $V^d(s) = u(y(s) - \tau(s), s) + \beta \mathbb{E}_{s'|s}[V^d(s')]$. The assumption that $V^d(s)$ is exogenous and independent of $b$ follows the vast majority of the quantitative sovereign debt literature. This assumption is important for the current result. When the literature has considered endogenous $V^d(s)$, it has typically been by including a stochastic reentry option; we will consider this possibility in section 3.4.2.

If the government does not default, it receives $y(s)$ as endowment, pays $b$, and issues new bonds $b'$ that will be due next period, raising revenue $Q(b', s)$. Its (possibly state-dependent) flow utility from consumption is $u(c, s)$, so that the value $V$ from repayment is given by

$$V(b, s) = \max_{b'} u(c, s) + \beta \mathbb{E}_{s'|s}[V^o(b', s')]$$

s.t. $c + b = y(s) + Q(b', s)$ (3.1)

and the value $V^o$ including the option to default at the beginning of a period is given by

$$V^o(b, s) = \max_{p \in \{0, 1\}} pV(b, s) + (1 - p)V^d(s)$$

(3.2)

where $p = 1$ denotes the decision to repay and $p = 0$ denotes the decision to default.

Debt is purchased by risk-neutral international investors that demand an expected return of $R$. For convenience, we assume that when a government is indifferent between repayment and default, it chooses to repay: $p(b, s) = 1$ if and only if $V(b, s) \geq V^d(s)$. Since investors receive expected repayment $\mathbb{E}_{s'|s}[p(b', s')]$, if follows that the bond revenue schedule $Q$ is

$$Q(b', s) = \frac{b'}{R} \mathbb{E}_{s'|s}[p(b', s')] = \frac{b'}{R} \mathbb{P}_{s'|s}[V(b', s') \geq V^d(s')]$$

(3.3)
We are now ready to define Markov perfect equilibrium, which is the typical focus in the literature.

**Definition 1.** A Markov perfect equilibrium is a set of policy functions $p(b,s)$, $c(b,s)$, $b'(b,s)$ for repayment, consumption and next period borrowing, value functions $V(b,s)$ and $V^o(b,s)$, and a bond revenue schedule $Q(b',s)$ such that (3.1)-(3.3) are satisfied.

We first prove an existence result. For this we need a number of technical assumptions. We highlight the following, because of the crucial role it will play for our main uniqueness proof in section 3.2.2.

**Assumption 1.** For each $s \in S$, $u(c,s)$ is strictly increasing in $c$.

**Proposition 3.2.1.** Under assumption 1 and additional standard assumptions, a Markov perfect equilibrium exists. In any equilibrium, $V(b,s)$ is strictly decreasing in $b$ for each $s \in S$, and there exists a set of default thresholds $\{b^*(s)\}_{s \in S}$ such that the government repays in state $s$ if and only if $b < b^*(s)$. Both $V$ and $Q$ are uniquely determined by the thresholds $\{b^*(s)\}_{s \in S}$.

The proof, developed in appendix 3.6.1, is constructive and relies on a fixed-point procedure similar to the one used by the quantitative literature to search for an equilibrium. As highlighted by Aguiar and Amador (2015), this procedure involves iterating on a monotone and bounded operator in the space of default thresholds. These iterations converge to a fixed point, and our proof verifies that this fixed point defines an equilibrium. Our additional assumptions (including the continuity of $u$, a no-Ponzi restriction, and an upper bound on $R$) ensure that value functions exist, are continuous and finite-valued, and that default thresholds $b^*(s)$ are uniquely defined by the equalities

$$V(b^*(s),s) = V^d(s)$$

(3.4)

The set $\{b^*(s)\}_{s \in S}$ then characterizes the bond revenue schedule $Q$: following (3.3),

$$Q(b',s) = \frac{b'}{R} \mathbb{P}_{s'|s}[b' \leq b^*(s')] = \frac{b'}{R} \sum_{s':b' \leq b^*(s')} \pi(s'|s)$$

(3.5)
In the special case where \( u(c, s) = u(c) \), income is i.i.d, and \( V^d \) is the expected value of autarky, it is possible to show that \( b^*(s) \) is increasing in \( y(s) \) (see Arellano, 2008), but such monotonicity is not needed for our proof.

In this environment, it is natural to conjecture that multiple equilibria could be present. Starting from an equilibrium with default thresholds \( \{b^*(s)\}_{s \in S} \), lowering default thresholds increases the cost of borrowing a given amount \( b' \), through (3.5). This, in turn, lowers the value to the government of repaying any amount \( b \), shifting down the function \( V(\cdot, s) \) for every \( s \) and hence, through (3.4), lowering the levels of debt at which the government is tempted to default. We now show that, in the environment considered here, this mechanism is never powerful enough to generate another equilibrium.

### 3.2.2 Uniqueness of Markov perfect equilibrium

Suppose that we have two distinct revenue schedules \( Q \) and \( \tilde{Q} \), each derived via (3.5) from anticipated default thresholds \( \{b^*(s)\}_{s \in S} \) and \( \{\tilde{b}^*(s)\}_{s \in S} \). Let \( V \) and \( \tilde{V} \) be the value functions for a government facing these schedules. To prove uniqueness of equilibrium, we need to show that at most one of these value functions can be consistent with the default thresholds that generate it—in other words, that we cannot have both \( V(b^*(s), s) = V^d(s) \) and \( \tilde{V}(\tilde{b}^*(s), s) = V^d(s) \) for all \( s \).

The key observation of this paper is that we can derive a simple inequality for the two value functions \( V \) and \( \tilde{V} \), related to the maximum difference between the default thresholds. This inequality requires assumption 1 together with

**Assumption 2.** \( R > 1 \).

The basis of our inequality is a simple replication strategy we call *mimicking at a distance*. Suppose that \( b^*(s) \) exceeds \( \tilde{b}^*(s) \) by at most \( M > 0 \). Then we show that it is always weakly better to start with debt of \( b - M \) when facing prices \( \tilde{Q} \) than with debt of \( b \) when facing prices \( Q \), and indeed strictly better whenever \( V(b, s) \geq V^d(s) \).

This observation, formalized in lemma 3.2.2, will ultimately be the basis of the proof that distinct equilibria are impossible in proposition 3.2.3.
The argument is as follows. The government with debt $b - M$ facing prices $\tilde{Q}$ has the option to mimic the policy of the government with debt $b$ facing prices $Q$—always defaulting at the same points, and otherwise choosing the same level of debt for the next period minus $M$. Before it defaults, this government is better off because it pays less to service debt, allowing it to consume more.

Debt service, in turn, costs less for two reasons. First, the mimicking government is less likely to be above the default thresholds assumed by its revenue schedule. This is due to the choice of $M$: since $M$ is the maximum amount by which the default thresholds $b^*(s)$ exceed the thresholds $\tilde{b}^*(s)$, as long as the government facing $Q$ chooses debt of $M$ less than the government it is mimicking, it is weakly less likely to exceed $\tilde{b}^*(s)$ than the other government is to exceed $b^*(s)$. Second, the mimicking government has strictly less debt, meaning that the cost of providing an expected return of $R > 1$ on this debt is lower.

Following this policy, the mimicking government always consumes strictly more until default, implying strictly higher utility due to assumption 1. It thus obtains a weakly higher value, which is strictly higher as long as it does not default right away.

**Lemma 3.2.2** (Mimicking at a distance.). Let $Q$ and $\tilde{Q}$ be two distinct revenue schedules, with $Q$ reflecting expected default thresholds $\{b^*(s)\}_{s \in S}$ and $\tilde{Q}$ reflecting expected default thresholds $\{\tilde{b}^*(s)\}_{s \in S}$. Let $V$ and $\tilde{V}$ be the respective value functions for governments facing these revenue schedules. Define

$$M = \max_s b^*(s) - \tilde{b}^*(s)$$

and assume without loss of generality that $M > 0$. Then, for any $s$ and $b$,

$$\tilde{V}(b - M, s) \geq V(b, s)$$

with strict inequality whenever $V(b, s) \geq V^d(s)$.

147
Proof. First, note that for any \( b' \) and \( s \), applying (3.5) we have

\[
\tilde{Q}(b' - M, s) = \frac{(b' - M)}{R} \sum_{\{s': b' - M \leq \tilde{b}^*(s')\}} \pi(s'|s) \geq \frac{(b' - M)}{R} \sum_{\{s': b' \leq \tilde{b}^*(s')\}} \pi(s'|s)
\]

\[
> \left( \frac{b'}{R} \sum_{\{s': b' \leq \tilde{b}^*(s')\}} \pi(s'|s) \right) - M = Q(b', s) - M \tag{3.8}
\]

Thus the amount that a government with schedule \( \tilde{Q} \) can raise by issuing \( b' - M \) of debt is always strictly larger than the amount that a government with schedule \( Q \) can raise by issuing \( b' \) of debt, minus \( M \). The two intermediate inequalities in (3.8) reflect the two sources of this advantage. First, there are weakly more cases in which \( b' - M \leq \tilde{b}^*(s') \) than in which \( b' \leq \tilde{b}^*(s') \), and this higher chance of repayment makes it possible to raise more. Second, since assumption 2 requires \( R > 1 \), issuing \( M \) less debt costs strictly less than \( M \) in foregone revenue in the current period.

Now we can formally define the mimicking at a distance policy. For any states and debt levels \( s \) and \( b \), let the history \( s^0 \) be such that the state and debt owed at \( t = 0 \) are respectively \( s \) and \( b \). The optimal strategy for a government facing schedule \( Q \) induces an allocation \( \{c(s^t), b(s^{t+1}), p(s^t)\}_{s^0 \leq s^t} \) at all histories following \( s^0 \). We construct a policy for the government facing schedule \( \tilde{Q} \) in state \( s \) and debt level \( b - M \) as follows. For every history \( s^t \geq s^0 \), let

\[ \tilde{p}(s^t) = p(s^t) \]

and provided that \( p(s^t) = 1 \), choose consumption and next-period debt as

\[
\tilde{b}(s^t) = b(s^t) - M \\
\tilde{c}(s^t) = c(s^t) + \tilde{Q}(b(s^t) - M, s_t) - (Q(b(s^t), s_t) - M) \tag{3.9}
\]

This ensures that the budget constraint is satisfied at all histories \( s^t \) where repayment

---

\(^3\)\(b(s^t)\) is defined to be the amount of debt chosen at history \( s^t \) to be repaid in period \( t + 1 \).
takes place:

\[
\tilde{c}(s') + \tilde{b}(s'^{-1}) - \tilde{Q}(\tilde{b}(s'), s_t) = \tilde{c}(s') + b(s'^{-1}) - M - \tilde{Q}(b(s'), s_t) = c(s') + b(s'^{-1}) - Q(b(s'), s_t) = y(s_t)
\]

Furthermore, using (3.8) we see that \(\tilde{c}(s') > c(s')\): when there is repayment, the mimicking policy (3.9) sets consumption equal to consumption in the other equilibrium, plus a bonus \(\tilde{Q}(b(s') - M, s_t) - (Q(b(s'), s_t) - M) > 0\) from lower debt costs.

The mimicking policy, of course, need not be optimal; but since it is feasible, it serves as a lower bound for \(\tilde{V}(b - M, s)\):

\[
\tilde{V}(b - M, s) \geq \sum_{\tilde{p}(s') = 1} \beta^t \Pi(s') u(\tilde{c}(s'), s_t) + \sum_{\tilde{p}(s') = 0, \tilde{p}(s'^{-1}) = 1} \beta^t \Pi(s') V^d(s_t) \\
\geq \sum_{p(s') = 1} \beta^t \Pi(s') u(c(s'), s_t) + \sum_{p(s') = 0, p(s'^{-1}) = 1} \beta^t \Pi(s') V^d(s_t) = V(b, s)
\]

with strict inequality whenever \(p(s^0) = 1\) (or equivalently \(b \leq b^*(s)\)), since this implies \(\tilde{c}(s^0) > c(s^0)\) and \(u(c, s_0)\) is strictly increasing in \(c\) thanks to assumption 1. 

An illustration of the mimicking policy used in lemma 3.2.2 is given in figures 3-1 and 3-2, which depict time paths in a hypothetical two-state case. In this case, debt starts relatively high and the high-income state \(y(s_H)\) keeps recurring, leading the government to deleverage in anticipation of lower incomes in the future. Figure 3-1 shows the paths of \(b\) (filled circles) and the mimicking policy \(\tilde{b} = b - M\) (hollow circles), while figure 3-2 shows the paths of \(c\) (filled circles) and the consumption \(\bar{c} = c + \tilde{Q}(b - M, s) - (Q(b, s) - M)\) induced by the mimicking policy (hollow circles).

Although \(\bar{c}\) is always greater than \(c\) in figure 3-2, the gap \(\bar{c} - c\) differs across periods. This reflects fluctuations in the two sources of \(\bar{c} - c\): differences in default premia, and the lower cost of servicing \(\tilde{b} = b - M\) rather than \(b\). First, since both debt levels at \(t = 2\) are above the respective default thresholds for \(y(s_L)\), there is no difference
at \( t = 1 \) in the two default premia. At \( t = 3 \), however, the mimicking policy achieves a debt level below \( \tilde{b}^*(s_L) \), while the other policy has debt that remains above \( b^*(s_L) \). Thus the default premium disappears at \( t = 2 \) for the mimicking policy while still being paid for the other policy, leading to an expansion in the gap \( \tilde{c} - c \). From \( t = 4 \) onward both policies achieve debt levels below their \( s_L \) default thresholds, leading to the disappearance of all default premia. This causes the gap \( \tilde{c} - c \) to compress starting at \( t = 3 \).

The central observation is that if it starts with debt \( M = \max_s b^*(s) - \tilde{b}^*(s) \) below the other government, the mimicking government can keep itself at the fixed distance \( M \), achieving higher consumption along the way.

We now turn to the main result, which uses lemma 3.2.2 to rule out multiple equilibria \((V, Q)\) and \((\tilde{V}, \tilde{Q})\) altogether.

**Proposition 3.2.3.** In the benchmark model, Markov perfect equilibrium has a unique value function \( V(b, s) \) and debt price schedule \( Q(b, s) \).

**Proof.** Suppose to the contrary that there are distinct equilibria \((V, Q)\) and \((\tilde{V}, \tilde{Q})\). Proposition 3.2.1 shows that these are characterized by their default thresholds \( \{b^*(s)\}_{s \in S} \) and \( \{\tilde{b}^*(s)\}_{s \in S} \). Therefore, it suffices for us to show that the thresholds are unique.

Without loss of generality, assume that the maximal difference between \( b^* \) and \( \tilde{b}^* \)
is positive and is attained in a state $\bar{s} \in \mathcal{S}$:

$$\max_s b^*(s) - \tilde{b}^*(s) = b^*(\bar{s}) - \tilde{b}^*(\bar{s}) = M > 0$$

Applying lemma 3.2.2 for $s = \bar{s}$ and $b = b^*(\bar{s}) = \tilde{b}^*(\bar{s}) + M$, we know that

$$\tilde{V}(b^*(\bar{s}), \bar{s}) > V(b^*(\bar{s}), \bar{s})$$

But this contradicts the fact that $b^*(\bar{s})$ and $\tilde{b}^*(\bar{s})$ are default thresholds, which requires $\tilde{V}(b^*(\bar{s}), \bar{s}) = V(b^*(\bar{s}), \bar{s}) = V^d(\bar{s})$. Thus our premise of distinct equilibria cannot stand. \qed

The intuitive thrust of lemma 3.2.2 and proposition 3.2.3 is that distinct debt revenue schedules cannot both be self-sustaining. No two schedules $Q$ and $\tilde{Q}$ can simultaneously rationalize their corresponding default thresholds $b^*(\bar{s})$ and $\tilde{b}^*(\bar{s})$ in the state $\bar{s}$ where these thresholds differ most. Instead, the argument of lemma 3.2.2 shows that it is better for a government to start at the lower threshold $b^*(\bar{s})$ given schedule $Q$ than to start at the higher threshold $\tilde{b}^*(\bar{s})$ given schedule $\tilde{Q}$; at this point, any advantages of $\tilde{Q}$ over $Q$ are outweighed by the heavier debt burden, and the former government can use a simple mimicking strategy to guarantee itself strictly higher consumption than the latter. It follows that these cannot both be default thresholds, which by definition must be equally desirable, with common value equal to the default value $V^d(\bar{s})$.

### 3.2.3 Uniqueness of subgame perfect equilibrium

The arguments used to prove proposition 3.2.3 can be extended to show that this model admits a unique subgame perfect equilibrium. While the Markov perfect concept exogenously restricts equilibrium to depend on a limited set of states, subgame perfect equilibria allow an arbitrary dependence of strategies at time $t$ on the history $h^{t-1}$ of past states and actions. The following result shows that the current states $s$ and $b$ summarize this dependence, demonstrating that the Markov concept—which
has been the focus of much of the quantitative literature—is not restrictive. Proving this formally requires defining the game played by the government and international investors more precisely. Crucially, in this game, the value from government default is still exogenous—endogenizing the default option as part of the game is outside of the scope of this paper (see Kletzer and Wright, 2000, for such an exercise). Here we summarize our result, and relegate the description of the game and the proof to appendix 3.6.2. Let $V(h^{t-1}, s)$ be the value achieved by a government after history $h^{t-1}$, when the current exogenous state is $s$. Then the following result holds.

**Proposition 3.2.4.** Consider two subgame perfect equilibria $A$ and $B$. For any $(b, s)$, and any histories $(h_A, h_B)$ such that $b(h_A) = b(h_B) = b$, we have $V_A(h_A, s) = V_B(h_B, s)$.

The key to the proof of proposition 3.2.4 is to show that, conditional on the exogenous state $s$, a government with higher debt must have lower value, independently of the equilibrium that is played or the history of past actions. This in turn relies on another mimicking argument, whereby a government with lower debt can always choose a strategy that ensures it higher consumption and higher future value than its higher-debt counterpart.

### 3.3 Application to other models

The argument used to prove uniqueness of equilibrium in section 3.2 is very general and can be used in other contexts, as the following applications illustrate.

#### 3.3.1 Bewley models with endogenous debt limits

Consider a modification of the environment of section 3.2, in which lenders are restricted to offer a price of $\frac{1}{R}$ for every unit of debt that they buy. Borrowing must therefore be risk-free: this is the equilibrium defined in Zhang (1997). This restriction can be captured within the framework of the previous section by specifying that the
price of non-riskless debt is zero. Instead of (3.3), the bond revenue schedule becomes

\[ Q^z(b', s) = \frac{b'}{R} \mathbf{1}_{\{ V^z(b', s') \geq V^d(s') \ \forall s'|s \}} \]  

(3.10)

Define \( \phi(s) \) as the value that satisfies \( V^z(\phi(s), s) = V^d(s) \), and assume that for all \( s \) and \( s' \), \( \pi(s'|s) > 0 \). Then, writing \( \varphi \equiv \min_s \{ \phi(s) \} \), (3.10) becomes

\[ Q^z(b', s) = \frac{b'}{R} \mathbf{1}_{\{ \nu \leq \varphi \}} \]  

(3.11)

In other words, the model is a standard incomplete markets model in the tradition of Bewley (1977), with a debt limit \( \varphi \) determined endogenously by the requirement that the government should never prefer default.

We can immediately prove analogs of lemma 3.2.2 and proposition 3.2.3 in this new environment.

**Lemma 3.3.1.** Consider two distinct equilibria with value functions \( V \) and \( \widetilde{V} \) and debt limits \( \overline{\varphi} < \varphi \). Then, letting \( M = \varphi - \overline{\varphi} \), for any \( b \) and \( s \) we have

\[ \widetilde{V}(b - M, s) \geq V(b, s) \]  

(3.12)

with strict inequality whenever \( b \leq \varphi \).

**Proof.** Same as the proof of lemma 3.2.2, with (3.3) replaced by (3.11) and inequality (3.8) becoming

\[ \widetilde{Q}^z(b' - M, s) = \frac{b' - M}{R} \mathbf{1}_{\{ \nu - M \leq \overline{\varphi} \}} = \frac{b' - M}{R} \mathbf{1}_{\{ \nu \leq \varphi \}} > \frac{b'}{R} \mathbf{1}_{\{ \nu \leq \varphi \}} - M = Q^z(b', s) - M \]  

(3.13)

\[ \square \]

The intuition behind (3.12) and (3.13) is well known in this class of environment: an increase in the debt limit is equivalent to a translation of the value function, accompanied by a translation of the income process that reflects the interest costs.
of debt.⁴ Our earlier inequality (3.8) can be interpreted as a generalization of this result.

**Proposition 3.3.2.** In the model with riskless debt, Markov perfect equilibrium has a unique value function $V(b, s)$ and debt limit $\varphi$.

*Proof.* Same as the proof of proposition 3.2.3, but using lemma 3.3.1 rather than lemma 3.2.2.

As proposition 3.3.2 demonstrates, our approach for proving uniqueness is not limited to the standard Eaton-Gersovitz framework. The underlying replication argument can be adapted, with some modifications, to other environments common in the literature. As highlighted in the introduction, this particular application also illustrates the key intuition behind our main uniqueness result in section 3.2: a deterioration in the terms of borrowing cannot be self-sustaining in this class of models since, once governments have exhausted their debt capacity, those with less debt are always better off.

### 3.3.2 Bulow and Rogoff

Our proof is also related to that used by Bulow and Rogoff (1989) to rule out reputational equilibria in sovereign debt models where saving is allowed after default. As originally written, the Bulow-Rogoff result only applies directly to environments with complete markets, but a similar result also holds in the incomplete markets framework we study: if a government can save at a strictly positive net risk-free rate after defaulting, and there are no other exogenous penalties for default, then no debt can be sustained. Though this result has not—to our knowledge—been written formally until now, it has informally motivated the ingredients of modern variations on the Eaton-Gersovitz model, which all specify some exclusion from international markets after default, together with additional costs of default such as output losses.

⁴See, for example, Ljungqvist and Sargent (2012).
Define $V^{nb}(b, s)$ to be the value function for a government that can save at the risk-free rate but not borrow.

$$V^{nb}(b, s) = \max_{b'} u(c, s) + \beta \mathbb{E}_{s'}[V^{nb}(b', s')]$$

s.t. $c + b = y(s) + \frac{b'}{R}$, $b' \leq 0$ \hspace{1cm} (3.14)

Now, specify $V^d(s) \equiv V^{nb}(0, s)$, such that when the government defaults, its debt is reset at 0 and it can subsequently save but not borrow. We can now prove the incomplete markets analog of Bulow and Rogoff (1989), as a special case of proposition 3.2.3.

**Proposition 3.3.3** (Incomplete markets Bulow-Rogoff). *In the model with $V^d(s) = V^{nb}(0, s)$ (i.e. savings after default), no debt can be sustained: in the unique Markov perfect equilibrium, the default thresholds $b^*(s)$ equal 0 for all $s$, and $Q(b', s) = 0$ for all $b' \geq 0$. Hence $V(b, s) = V^{nb}(b, s)$.*

**Proof.** We first verify that when $V^d(s) = V^{nb}(0, s)$, there exists an equilibrium where the government will default for any positive amount of debt $b > 0$. This equilibrium is $(V^{nb}, Q^{nb})$, where $V^{nb}$ is given in (3.14) and the government faces

$$Q^{nb}(b', s) = \begin{cases} \frac{b'}{R} & b' \leq 0 \\ 0 & b' \geq 0 \end{cases}$$ \hspace{1cm} (3.15)

which is the revenue schedule induced by default thresholds identically equal to zero.

First, $Q^{nb}$ generates $V^{nb}$. The budget constraint in (3.14) is effectively the same as the constraint in (3.1) given prices (3.15); although (3.14) does not allow $b' > 0$ while (3.1) does, positive borrowing $b' > 0$ will never be optimal given prices (3.15) because it raises no revenue. Moreover, proposition 3.2.1 shows that the value function generated by the prices in (3.15) is decreasing in $b$; hence whenever $b \leq 0$, we have $V^{nb}(b, s) \geq V^{nb}(0, s) = V^d(s)$ for all $s$, so that default is never optimal.

Second, the default thresholds corresponding to $V^{nb}$ are identically equal to zero, thereby generating $Q^{nb}$. This also follows from the monotonicity of $V^{nb}$ in $b$ (propo-
sition 3.2.1): since

\[ V^{nb}(b, s) \geq V^d(s) = V^{nb}(0, s) \iff b \leq 0 \quad \forall s \]

we have \( b^*(s) = 0 \) for all \( s \).

Proposition 3.2.3 then implies that \( (V^{nb}, Q^{nb}) \) must be the unique Markov perfect equilibrium, and hence that there is no distinct equilibrium in which debt can be sustained. In particular, there is no equilibrium where the expectation of being able to borrow in the future is enough to discourage default and sustain some positive debt. This is the incomplete markets version of the Bulow and Rogoff (1989) result.

Going back to the proof of proposition 3.2.3, the intuition behind this result is that once a government has already borrowed the maximum amount that can obtain a nonzero price, access to debt markets offers no benefits beyond access to a market for savings. It is impossible to borrow more until some debt is repaid—and rather than repay and reborrow, it is cheaper to default and then run savings up and down in a parallel way, achieving higher consumption by avoiding the costs of debt service. No amount of debt is sustainable: whenever a government has borrowed the maximum, it will default with certainty, and in anticipation creditors will never allow any debt.

This resembles the logic behind the original Bulow and Rogoff (1989) result, which observed that for a reputational debt contract in complete markets, there must always be some state of nature in which a government can default and use the amount demanded for repayment as collateral for a sequence of state-contingent “cash in advance” contracts that deliver strictly higher consumption in every future date and state. The main idea behind their proof carries over to our incomplete markets environment, once the cash in advance contracts are replaced with a simple, parallel savings strategy. Our contribution here is to show that this result is a special case of a much broader equilibrium uniqueness result.\(^5\)

\(^5\)In parallel and independent work, Bloise et al. (2015) have established a sufficient condition under which the Bulow and Rogoff (1989) result survives in incomplete markets environments with a general asset market structure. This sufficient condition is a “high implied interest rates” condition, as in Alvarez and Jermann (2000). When the only available asset is a risk-free bond and the endowment process is bounded, this condition simplifies to \( R > 1 \) (our Assumption 2). Our result in this section
3.4 Extensions of the benchmark model

The benchmark Eaton-Gersovitz model covered in section 3.2 is often modified to achieve greater realism or more tractable computation. Recently, Passadore and Xandri (2014) have studied multiplicity that arises when the domain of debt values is bounded from below, restricting savings. Although this assumption has not been microfounded in the literature, it can be justified by an appeal to the institutional difficulties some governments face in maintaining net savings; it can also make computation more tractable, as argued by Chatterjee and Eyigungor (2012), who exclude savings on their grid of debt values. Meanwhile, both Aguiar and Gopinath (2006) and Arellano (2008) allow some probability of reentry each period for countries excluded from asset markets after default.

For bounded savings, we show in subsection 3.4.1 that a modified replication argument—using the concavity of $u$ in addition to its monotonicity—continues to deliver uniqueness, except in a limit case where Passadore and Xandri (2014) have demonstrated multiplicity. This robustness is perhaps surprising, since the mimicking strategy behind lemma 3.2.2 and proposition 3.2.3 can rely on the government’s ability to smooth consumption via savings rather than debt. For reentry, we show in subsection 3.4.2 that although our argument can no longer establish uniqueness, it still can rule out the most commonly hypothesized form of multiplicity, in which one equilibrium delivers uniformly better bond prices than another. It also implies uniqueness in the special case where states are independently and identically distributed. Finally, in subsection 3.4.3 we discuss other, more drastic modifications to the environment, in which our results no longer apply.

In sections 3.4.1 and 3.4.2, it will be necessary to impose more structure on the payout $V^d(s)$ from default. In particular, we assume that there is an exogenous flow utility $v^d(s)$ from each period spent in market exclusion following default, and therefore complements theirs, by exhibiting an explicit replication strategy with risk-free bonds, and reinterpreting the no-lending result as a result about equilibrium uniqueness.

6 In their numerical simulations, Chatterjee and Eyigungor (2012) find that this restriction does not bind.
that it is weakly less desirable than autarky: \( v^d(s) \leq u(y(s), s) \). This is common in the literature, which often specifies it in terms of an output cost \( \tau(s) \) such that \( v^d(s) = u(y(s) - \tau(s), s) \). When there is no reentry, \( V^d \) becomes the expected utility from receiving \( v^d \) in all subsequent periods:

\[
V^d(s_0) = E \left[ \sum_{t=0}^{\infty} \beta^t v^d(s_t) \mid s_0 \right]
\]  

(3.16)

3.4.1 Bound on savings

Here we consider the case of an exogenous bound on savings, which in our notation becomes a lower bound \( \bar{b} \leq 0 \) on debt. Section 3.2 already allowed for any upper bound \( \bar{b} \) on debt.

The proof of lemma 3.2.2 no longer directly applies when savings are bounded, because the mimicking strategy (3.9) may no longer be feasible. Indeed, lemma 3.2.2 is no longer always true: a government with \( \bar{b} \) debt, facing debt prices given by default thresholds \( \{b^*(s)\} \), may be better off than a government with \( \bar{b} - M \) debt, facing debt prices given by default thresholds \( \{b^*(s) - M\} \). When allowable savings are limited, the improved debt prices implied by higher default thresholds give the former government more ability to smooth consumption, and this benefit may outweigh the costs of a higher debt load—possibly justifying the higher default thresholds themselves. In this light, multiplicity seems plausible.

Nevertheless, it remains possible to demonstrate equilibrium uniqueness in nearly all cases. Whereas the approach in section 3.2 used only the strict monotonicity of \( u \), here we need an additional assumption on \( u \):

**Assumption 3.** For each \( s \in S \), \( u(c, s) \) is concave in \( c \).

With concave utility and bounded savings, higher debt prices are desirable because they make it cheaper for the government to smooth consumption using debt—but concavity also implies that the marginal gains from consumption smoothing decrease

---

7Imposing this structure involves some loss of generality, since we cannot longer make the value of defaulting depend on state \( s \) without also affecting the value of being excluded from markets in state \( s \) after originally defaulting in state \( s' \neq s \).
as the government does more of it. Initially, if the government faces very low debt prices, an improvement in these prices may dramatically increase the benefits of market access, so that the default probability falls by more than enough to sustain the price improvement. But eventually, as diminishing returns to consumption smoothing set in, price improvements are no longer self-sustaining. At some point, we reach an equilibrium—where for given bond prices, the benefits of market access produce a pattern of default exactly consistent with those prices. Lemma 3.4.1 and proposition 3.4.2, which are partial analogues of lemma 3.2.2 and proposition 3.2.3, demonstrate that this equilibrium is unique.

**Formal argument.** In many respects the formal argument echoes the replication argument from section 3.2, but there are also substantial differences. Rather than mimicking at a distance, which is no longer feasible, we use compressed mimicking. Given distinct revenue schedules \( Q \) and \( \tilde{Q} \) derived via (3.5) from default thresholds \( \{b^*(s)\} \) and \( \{b*(s)\} \), lemma 3.4.1 defines \( \lambda \) to be the minimum ratio between \( \tilde{b}^*(s) - b \) and \( b^*(s) - b \). For any \( s \) and \( \tilde{b} - b = \lambda(b - \tilde{b}) \), a government starting at \( (\tilde{b}, s) \) can compress by \( \lambda \) the optimal strategy for a government (which we call the target) starting at \( (b, s) \) whenever the target government repays and defaulting whenever the target government defaults.

As in section 3.2, the mimicking government by construction obtains weakly better prices than the target government for its debt. Unlike in section 3.2, the mimicking government need not achieve higher consumption than the target government. Instead, because it is compressing the target’s debt issuance plan by \( \lambda \), in each period it obtains consumption \( \tilde{c} \) that is weakly higher than the convex combination \( \lambda c + (1 - \lambda)y(s) \) of the target’s consumption \( c \) and state-\( s \) autarky income \( y(s) \). This inequality is strict when \( \tilde{b} < 0 \), where the mimicking government can consume extra due to forgone financing costs. Concavity of \( u \) then implies that \( u(\tilde{c}, s) \) is strictly greater than \( \lambda u(c, s) + (1 - \lambda)u(y(s), s) \). Summing the expected value across all periods, we obtain (3.18), the analog of (3.7); when \( \tilde{b} = 0 \), the strict inequality can also follow from \( v^d(s) < u(y(s), s) \).
Effectively, lemma 3.4.1 bounds the extent to which higher expected default thresholds, which induce higher debt prices and allow the government to smooth more easily, increase the value function relative to the value from default. If, relative to the minimum feasible debt $\tilde{b}$, we scale up the government’s level of debt by $\lambda^{-1}$, and we also scale up default thresholds by at most $\lambda^{-1}$, then (3.19) states that we cannot increase the value over default by more than $\lambda^{-1}$.

**Lemma 3.4.1.** Let $Q$ and $\tilde{Q}$ be two distinct revenue schedules, with $Q$ reflecting expected default thresholds $\{b^*(s)\}_{s \in S}$ and $\tilde{Q}$ reflecting expected default thresholds $\{\tilde{b}^*(s)\}_{s \in S}$. Let $V$ and $\tilde{V}$ be the respective value functions for governments facing these revenue schedules. Define

$$\lambda \equiv \min_s \frac{\tilde{b}^*(s) - b}{b^*(s) - \tilde{b}} \quad (3.17)$$

and assume, without loss of generality, that $0 \leq \lambda < 1$. Assume also that either $b < 0$, or $v^d(s) < u(y(s), s)$ for all $s$. Then for any $s$ and $b$ such that $V(b, s) \geq V^d(s)$, we have

$$\tilde{V}(\tilde{b}, s) > (1 - \lambda)V^d(s) + \lambda V(b, s) \quad (3.18)$$

where $\tilde{b} - b \equiv \lambda(b - \tilde{b})$. This can equivalently be written as

$$\tilde{V}(\tilde{b}, s) - V^d(s) > \lambda(V(b, s) - V^d(s)) \quad (3.19)$$

**Proof.** In appendix 3.7.1. \qed

In contrast to (3.7), inequality (3.18) in lemma 3.4.1 does not show that $\tilde{V}(\tilde{b}, s)$ is higher than $V(b, s)$. Fortunately, this is not needed to establish uniqueness in proposition 3.4.2. Instead, (3.19) suffices to obtain a contradiction. Inequality (3.19) shows that if a government facing $Q$ weakly prefers not to default at $(b, s)$ (so that $V(b, s) - V^d(s) \geq 0$), then a government facing $\tilde{Q}$ must strictly prefer not to default at $(\tilde{b}, s)$ (so that $\tilde{V}(\tilde{b}, s) - V^d(s) > 0$). It is therefore impossible for both $b$ and $\tilde{b}$ to be default thresholds for their respective value functions.
Proposition 3.4.2. If either \( b < 0 \) or \( v^d(s) < u(y(s), s) \) for all \( s \), Markov perfect equilibrium has a unique value function \( V(b, s) \) and debt price schedule \( Q(b, s) \).

Proof. If, to the contrary, we have distinct equilibria \((V, Q)\) and \((\tilde{V}, \tilde{Q})\) with default thresholds \( \{b^*(s)\} \) and \( \{\tilde{b}^*(s)\} \), define \( \lambda \) as in (3.17) and assume without loss of generality that \( 0 \leq \lambda < 1 \).

Let \( \bar{s} \) be the state where the minimum in (3.17) is obtained. Evaluating (3.19) at \( \tilde{b} = \tilde{b}^*(\bar{s}), b = b^*(\bar{s}) \), and \( s = \bar{s} \), we obtain

\[
0 = \tilde{V}(\tilde{b}^*(\bar{s}), \bar{s}) - V^d(\bar{s}) > \lambda (V(b^*(\bar{s}), \bar{s}) - V^d(\bar{s})) = 0
\]

which is a contradiction.

Proposition 3.4.2 clarifies the role of unlimited savings in the uniqueness result of proposition 3.2.3. Even if a government cannot save at all, equilibrium will be unique as long as there is some exogenous penalty for default, such that \( v^d(s) \) is less than the flow utility \( u(y(s), s) \) from consuming the endowment. This assumption is made in most of the quantitative literature (for example Arellano, 2008). Alternatively, if \( v^d(s) = u(y(s), s) \), then even an infinitesimal capacity to save \( b = -\epsilon < 0 \) is sufficient for uniqueness.

Combined with results elsewhere, proposition 3.4.2 offers a very sharp characterization of the conditions governing uniqueness. In the limit case excluded by the proposition,\(^8\) with both no savings \( (b = 0) \) and no default penalty aside from autarky \( (v^d(s) = u(y(s))) \), Passadore and Xandri (2014) demonstrate that multiplicity is possible. In this environment, autarky is always an equilibrium, since there is no incentive to repay debt when a zero price for future issuance is expected and there is no additional penalty for default. But, under certain conditions, incentives to repay emerge endogenously in an alternative equilibrium where positive debt prices are expected and the government repays to preserve its debt market access. Interestingly,

\(^{8}\)Recall that we do not consider cases where \( b > 0 \). Indeed, our proposition does not rule out multiplicity in such cases, where active governments must maintain some positive minimum debt level, and default frees the government from the cost of servicing this debt. Since \( b > 0 \) is rare in the literature and a positive minimum debt level can be difficult to interpret, we emphasize \( b \leq 0 \) instead.
this multiplicity embodies the intuition that propositions 3.2.3 and 3.4.2 reject whenever any level of government savings is allowed: in Passadore and Xandri (2014), expectations of high bond prices can be self-sustaining, encouraging the low default rates needed to justify high prices.

### 3.4.2 Stochastic market reentry

In the literature, a very common departure from the benchmark model of Section 3.2 is an assumption that market reaccess is possible after default. This makes the value of default depend on the equilibrium value of borrowing, implying that lemma 3.2.2 and proposition 3.2.3 do not directly apply. Nevertheless, we will be able to rule out the most commonly hypothesized form of multiplicity—the existence of distinct “favorable” and “adverse” equilibria, in which the favorable equilibrium offers uniformly better debt prices $Q$.

To be concrete, suppose that it is possible to re-access markets with zero debt after a stochastic period of exclusion, which has independent probability $1 - \lambda$ of ending in each period. That is, replace (3.16) by\(^9\)

$$V^d(s) = v^d(s) + \beta \lambda E_{s'} [V^d(s')] + \beta (1 - \lambda) E_{s'} [V^d(0, s')]$$

(3.20)

In this framework, we can now prove the following specialized analog of proposition 3.2.3.

**Proposition 3.4.3.** In the model with stochastic reentry, there do not exist two distinct equilibria $(V, Q)$ and $(\tilde{V}, \tilde{Q})$ such that $Q(b, s) \geq \tilde{Q}(b, s)$ for all $b$ and $s$.

**Proof.** In appendix 3.7.2. \(\square\)

In general, the endogeneity of $V^d(s)$ in (3.20) makes it difficult to analytically characterize equilibria. In the particular case examined by proposition 3.4.3, however, the proof strategy from proposition 3.2.3 still applies with some modification. The

---

\(^9\)This formulation is the one used by Arellano (2008) and Aguiar and Gopinath (2006). It does not encompass the possibility of recovery on defaulted debt or debt renegotiation (see for example Yue, 2010).
core insight is that if $Q \geq \tilde{Q}$, then $V^d \geq \tilde{V}^d$, because a government facing a uniformly better price schedule after reentry is better off. Furthermore, if $Q$ and $\tilde{Q}$ are distinct and $Q \geq \tilde{Q}$, there must be some $s$ for which $b^*(s) > \tilde{b}^*(s)$. We then can apply the argument from lemma 3.2.2 and proposition 3.2.3, having a government in the $(\tilde{V}, \tilde{Q})$ equilibrium mimic the strategy of a government in the $(V, Q)$ equilibrium. The fact that $\tilde{V}^d \leq V^d$ only helps our argument, since it is further reason why government in the $(\tilde{V}, \tilde{Q})$ equilibrium will prefer the mimicking strategy to default.

In short, when there is reentry, uniformly higher bond prices defeat themselves: they make default and eventual reentry more attractive, raising the probability of default and pushing bond prices back down.

Although we cannot prove uniqueness more generally, this result does rule out the popular hypothesis—as discussed in the introduction—that sovereign debt markets can vary between self-sustaining “favorable” and “adverse” equilibria. Instead, if multiplicity exists, we know that it must be a surprising kind of multiplicity: among any two equilibria, each must offer cheaper borrowing in some places and more expensive borrowing in others.

**Special case with iid exogenous state.** It is possible to demonstrate full uniqueness in one special case. Suppose now that $s$ follows an iid process with probability $\pi(s)$. It follows that the expected value from reentry $E_{s'}[V^o(0, s')]$ in (3.20) is independent of the states preceding $s'$, and we can denote this expectation by $V^{re}$. The iid assumption also implies that the debt price schedule $Q$ depends only on the debt amount $b'$, not the current state $s$, as (3.5) reduces to

$$Q(b') = \frac{b'}{R} \sum_{\{s': b' \leq b^*(s')\}} \pi(s')$$  \hspace{1cm} (3.21)

**Proposition 3.4.4.** In the model with iid states and stochastic market reentry, Markov perfect equilibrium has a unique value function $V(b, s)$ and debt price schedule $Q(b)$.

*Proof. In appendix 3.7.3.*

Proposition 3.4.4 follows for reasons similar to proposition 3.4.3. For any distinct
equilibria \((V, Q)\) and \((\tilde{V}, \tilde{Q})\), the only difference between the default value functions \(V^d\) and \(\tilde{V}^d\) arises from the expected reentry value, which is now just a scalar \(V^{\text{re}}\). Whichever equilibrium has the higher reentry value must have a more favorable bond price schedule, meaning that at least one of its default thresholds is higher. As with proposition 3.4.3, we can then invoke a mimicking argument to show that the equilibrium with a higher default value cannot also have a higher default threshold for some \(s\).

This result further emphasizes how subtle any multiplicity in the model with reentry, if it exists, must be: it must rely, in some way, on the transition probabilities of the Markov process being non-iid.

3.4.3 Other variations on the model and multiplicity results

We have showed that the benchmark Eaton-Gersovitz model of sovereign debt with default does not admit multiple equilibria, and that this uniqueness result partly extends to the more complex environments of subsections 3.4.1 and 3.4.2. Nevertheless, multiplicity arises in several other sovereign debt models in the literature. This section reviews the ways in which these models sidestep the uniqueness result present in the benchmark framework.

Markov perfect equilibrium in the model we studied includes a price function \(Q(b', s)\), which depends only on the current state \(s\) and the bond payment \(b'\) promised tomorrow. After observing \(s\), in each period the government can choose either to default or to repay and sell some quantity \(b'\) of bonds for next period. Once the government chooses to repay and selects some \(b'\), there is no uncertainty about the amount \(Q(b', s)\) that will be raised; no further choices are made until the next period, when the next state \(s'\) is realized and the process repeats itself. As presented in appendix 3.6.2, this process can be explicitly written as a game between governments and risk-neutral investors. It is possible to define subgame perfect equilibria in this game, and proposition 3.2.4 shows that uniqueness still holds for these equilibria in the benchmark model.

Our uniqueness result can disappear if the timing and action space of the game
are modified. For instance, in the model of Cole and Kehoe (2000), the government has the option to default after observing the outcome of the current period’s bond auction. If it defaults, it can keep the proceeds of the auction but avoid repayment on its maturing debt. Given enough risk aversion, this option is preferable when the current period’s auction yields little revenue, and the cost of repaying maturing debt out of current-period resources is prohibitively high. A coordination problem among creditors thus emerges, leading to multiple equilibria: they might either offer high prices, in which case the government will repay, or offer low prices, in which case the government will default and thereby justify the low prices. The literature sometimes refers to this phenomenon as “rollover multiplicity”. It is absent in the model we study, which excludes the option to default after revenue from the auction comes in; but it captures an important intuition, which is that rolling over large amounts of short-term debt can be a source of fragility.\footnote{Our results explain why the emerging quantitative literature evaluating the importance of nonfundamental forces in explaining the recent Eurozone crisis (Conesa and Kehoe (2015), Bocola and Dovis (2015)) has turned to a Cole-Kehoe formulation of the timing: even in more complex quantitative models, the Eaton-Gersovitz timing tends to generate uniqueness and is therefore poorly suited for such an exercise.}

In the model of Calvo (1988), multiplicity arises because of the way the bond revenue-raising process works. In the Calvo model, a government borrows an exogenous amount $b$ at date 0 and inherits a liability of $R_b b$ at date 1. It then uses a mix of distortionary taxation and debt repudiation to finance a given level of government spending. Since a higher interest rate $R_b$ tilts the balance towards more repudiation at date 1, and since investors need to break even when lending to the government, there exist two rational expectations equilibria: one with high $R_b$ and high repudiation, and one with low $R_b$ and low repudiation. This is sometimes called “Laffer curve multiplicity” in reference to the shape of the bond revenue curve that arises in this model (the function that gives bond revenue $b$ as a function of promised repayment $R_b b$ has an inverted-V shape). In the model we study, the government directly announces the amount it will owe tomorrow, allowing it to avoid the downward-sloping part of the bond revenue curve.\footnote{Interestingly, the setup of the original Eaton and Gersovitz (1981) model does not let the government choose on the bond revenue curve \textit{a priori}, although their analysis focuses on equilibria} Lorenzoni and Werning (2014) make a forceful
argument that such an assumption requires a form of commitment that governments are unlikely to have: in practice, if they raise less auction revenue than expected, they may auction additional debt rather than making the burdensome fiscal adjustments that are otherwise necessary.

In effect, both the rollover multiplicity of Cole and Kehoe (2000) and the Laffer curve multiplicity of Calvo (1988) emerge from a more elaborate game between governments and investors. They create self-fulfilling alternate equilibria by allowing governments to act in ways ruled out by the game-theoretic formulation of the benchmark Eaton-Gersovitz model: when auction revenue is insufficient, governments can either take the revenue and then default (as in Cole and Kehoe, 2000) or dilute investors by issuing more debt in the same period (as in Lorenzoni and Werning's interpretation of Calvo, 1988). Since the Eaton-Gersovitz model alone cannot produce multiplicity, these modifications to the game may prove important to interpreting any multiplicity we see in practice. More generally, they suggest that a detailed look at institutions, and the practical options available to sovereign debtors when they raise funds in debt markets, is necessary to understand when the Eaton-Gersovitz model succeeds and when it fails as a benchmark.

Finally, another important strand of the literature considers long-term debt, as in Hatchondo and Martinez (2009). Here, uniqueness of equilibrium remains uncertain: multiplicity has not been explicitly demonstrated, but our mimicking-based proof of uniqueness breaks down when bond prices are influenced by the likelihood of endogenous default in the arbitrarily distant future. In a related continuous time environment, Lorenzoni and Werning (2014) find multiple equilibria: in their model, an adverse shift in the bond price schedule forces the government onto a path of increasing debt, justifying the initial shift. Although their analysis does not adapt directly to the Hatchondo and Martinez (2009) model, it does suggest that multiple equilibria may be present.

---

in which it effectively does.
3.5 Conclusion

We prove that the Eaton-Gersovitz model and several of its variants have a unique equilibrium. Our results settle an important outstanding question in the literature, making use of a replication-based proof that may be applicable more generally. By showing that no changes in a government’s reputation for repayment can be self-sustaining, we rule out a widely suspected source of multiple equilibria in sovereign debt markets, and invite a renewed focus on alternative economic mechanisms that can generate multiplicity.

3.6 Appendix: Proofs for section 3.2

3.6.1 Existence of Markov perfect equilibrium

We prove existence of Markov perfect equilibrium constructively, following a fixed point procedure similar to the one typically used by the sovereign debt literature to find an equilibrium. Section 3.6.1 defines a functional $B(V)$ mapping value functions to default thresholds, and proves properties of this mapping. Section 3.6.1 defines a functional $V(B)$ mapping default thresholds to value functions, and proves properties of that mapping. Finally, section 3.6.1 shows that iterating on the operator $T = B \circ V$, starting from thresholds identically equal to zero, produces a limit set of default thresholds that constitute an equilibrium. Throughout we will need the following additional, technical assumptions.

Assumption 3.6.1. $u(c, s)$ is continuous in $c$ for every $s$

Assumption 3.6.2. $u(0, s) = -\infty$

Assumption 3.6.3. There exist $\gamma > 0$ and $\kappa > 0$ such that $u(c, s) \leq \gamma c^\kappa$ for all $c, s$; and $\beta R^\kappa < 1$

Assumption 3.6.4. There exists an upper bound $\bar{b}$ such that $b \leq \bar{b}$

Assumption 3.6.5. $-\infty < V^d(s) \leq V^{nb}(0, s)$, where $V^{nb}$ is defined in (3.14)
Assumptions 3.6.1-3.6.4 guarantee that the value function is well-defined and finite, and that default thresholds are uniquely defined. Assumption 3.6.5 is needed to ensure that governments with positive assets are not tempted to default.

**Default thresholds for given \( V : \mathcal{B}(V) \)**

Consider a set of \( S \) strictly decreasing, continuous functions \( V(b, s) \). For each state \( s \), define the threshold \( b^*(s) \) as \(-\infty\) if \( \sup_b V(b, s) < V^d(s) \), or \(+\infty\) if \( \inf_b V(b, s) > V^d(s) \). In other cases, let \( b^*(s) \) be equal to the unique solution to

\[
V(b^*(s), s) = V^d(s)
\]

This defines a functional \( \mathcal{B}(V) \). The following shows that this is a monotone mapping, and provides conditions on \( V \) under which \( \mathcal{B}(V) \) is positive and bounded.

**Lemma 3.6.1.** The following propositions hold for every \( s \).

1. If \( V(0, s) \geq V_{nb}(0, s) \), then \( b^*(s) \geq 0 \)

2. If \( V(b_s, s) = -\infty \), then \( b^*(s) < b_s \)

3. If \( V^A(b, s) \geq V^B(b, s) \) for all \( b \), then the respective default thresholds satisfy \( b^A(s) \geq b^B(s) \)

**Proof.** The proof follows because \( V \) is continuous and strictly decreasing. Assumption 3.6.5 guarantees that \( V(0, s) \geq V_{nb}(0, s) \geq V^d(s) = V(b^*(s), s) \), so 1 holds. Assumption 3.6.5 also guarantees that \( V^d(s) \) is finite, so \( V(b_s, s) < V(b^*(s), s) \), and 2 holds. Finally, \( V^B(b^B(s), s) = V^d(s) = V^A(b^A(s), s) \geq V^B(b^A(s), s) \), so 3 holds. \( \square \)
Value functions $V$ given default thresholds: $V(B)$

Consider now a set of positive default thresholds $B = \{b^*(s)\}$, $b^*(s) \geq 0$. Define $V$ as the solution to

$$
V(b,s;\{b^*(s')\}) \equiv \max_{b(s'),p(s') \in [0,1]} \left\{ \sum_{s'} \beta^t \Pi(s') u(c(s'),s_t) 1\{p(s')=1\} \right.
$$

$$
+ \sum_{s'} \beta^t \Pi(s') V^d(s_t) 1\{p(s')=0,p(s_t-1)=1\} \right\}
$$

s.t. $c(s') = y(s_t) + \frac{b(s'_t)}{R} \left( \sum_{s_{t+1}} \pi(s_{t+1}|s_t) \cdot 1\{b(s'_t) \leq b^*(s_{t+1})\} \right) - b(s'_t-1)$

(3.22)

where $p(s') \leq p(s'_t-1)$

$p(s_0) = 1$

$b(s^{-1}) = b$

$s_0 = s$

We now show that this generates a well-defined mapping $V(B)$, and prove properties of this mapping, including monotonicity and continuity.

**Lemma 3.6.2.** The following propositions hold for every $s$.

1. The maximum in (3.22) is attained for any $b$, and $V(b,s) < \infty$

2. $V(b,s)$ is strictly decreasing in $b$

3. $V(b,s)$ is continuous in $b$ for every $s$

4. $V(b,s;\{b^*(s')\})$ is increasing in $\{b^*(s')\}$

5. $V(0,s) \geq V^{nb}(0,s)$

6. $V\left(\frac{b}{R} + y(s),s\right) = -\infty$

7. $V(b,s;\{b^*(s')\})$ is continuous in $\{b^*(s')\}$
Proof. We prove each of the propositions in turn

1. We restrict ourselves to cases where such that $V(b, s) > -\infty$, otherwise the proposition is trivial. We prove that the maximum is attained by showing that the problem in (3.22) is the maximization of an upper semicontinuous function on a compact set, and exhibit an upper bound to show $V(b, s) < \infty$. First, assumption 3.6.2 guarantees that $c(s^t) > 0$, which (given that assets receive the risk-free rate) bounds the rate of growth of assets: there exists $D > 0$ such that $b(s^t) \geq -DR^{t+1}$. Together with assumption 3.6.4, this guarantees that $b(s^t)$ must be chosen on a compact interval $[-DR^{t+1}, \bar{b}]$, and hence that the set of all arguments $\{b(s^t), p(s^t)\}$ is compact. Second, these bounds on $b(s^t)$ place an upper bound on $c(s^t)$ which, together with assumption 3.6.3, yields a bound on flow utility, $\beta^t u(c(s^t), s_t) \leq (\beta R^c)^t \bar{u}$ where $\bar{u} < \infty$. Third, the presence of the default option implies that flow utility along the no-default path is bounded below in all periods, $\beta^t u(c(s^t), s_t) \geq \beta^t u$ for $\bar{u} > -\infty$. Summing up, we have bounds on flow utility:

$$\beta^t u \leq \beta^t u(c(s^t), s_t) \leq (\beta R^c)^t \bar{u} \quad (3.23)$$

Next, all partial sums in the maximand (3.22) are upper semicontinuous in the argument. This follows from the fact that they consist entirely of continuous functions except $1_{\{b(s^t) \leq b^*(s_{t+1})\}}$, which is upper semicontinuous. Inequality (3.23) together with $\beta < 1$ and $\beta R^c < 1$ allows one to apply the Weierstrass M-test to conclude that the sum converges uniformly, and hence that the limit is also upper semicontinuous in the argument. Hence the maximum in (3.22) is attained. Finally, (3.23) together with the fact that default values are finite guarantee that the objective in (3.22) is uniformly bounded from above, and hence the maximum $V(b, s) < \infty$ as well.

2. Fix $s$ and consider $\bar{b} > b$. Consider the optimal plan $\{\bar{b}(s^t), \bar{p}(s^t)\}$ starting at $(\bar{b}, s)$. Then the plan $\{\bar{b}(s^t), \bar{p}(s^t)\}$ is also feasible starting at $(b, s)$, so that,
letting \( Q = \frac{\tilde{u}(s^0)}{R} \sum \{ s' \mid \tilde{b}(s^0) \leq \tilde{r}(s') \} \pi(s'|s) \), we have

\[
V(b, s) - V(\tilde{b}, s) \geq u(y(s) + Q - b, s) - u(y(s) + Q - \tilde{b}, s)
\]

\[
> u\left( y(s) + Q - \tilde{b}, s \right) - u\left( y(s) + Q - \tilde{b}, s \right) = 0
\]

3. Fix \((b, s)\) and let \( \epsilon > 0 \). We show that (i) there exists \( \delta_1 > 0 \) such that for any \( b < \tilde{b} < b + \delta_1 \), \( V(\tilde{b}, s) > V(b, s) - \epsilon \), and (ii) there exists \( \delta_2 > 0 \) such that for any \( b > \tilde{b} > b - \delta_2 \), \( V(\tilde{b}, s) < V(b, s) + \epsilon \). Together with \( V \) being strictly decreasing, (i) and (ii) establish continuity.

For (i), consider the optimal plan \( \{b(s^t), p(s^t)\} \) starting at \((b, s)\). This plan is also feasible starting at \((\tilde{b}, s)\) and delivers the same consumption at every point except \( t = 0 \), where consumption is \( \tilde{b} - b \) lower. Hence letting \( c(s^0) \) be the \( t = 0 \) consumption level for the optimal plan starting at \((b, s)\), we know

\[
V(b, s) - V(\tilde{b}, s) = u(c(s^0)) - u(c(s^0) - \delta_1)
\]

will be \( < \epsilon \) as desired if \( \delta_1 > 0 \) is defined via continuity of \( u \) such that \( |u(c) - u(c(s^0))| < \epsilon \) for all \( |c - c(s^0)| < \delta_1 \).

For (ii), we must appeal to a uniform continuity argument to choose \( \delta_2 \). We first find a compact set \([c, \overline{c}]\) such that any optimal plan with \( \tilde{b} < b \) (and hence \( V(\tilde{b}, s) > V(b, s) \)) has first period consumption \( \overline{c}(s^0) \in [c, \overline{c}] \). To do this, recall from section 3.6.1 that the sum of all terms in (3.22) for \( t \geq 1 \) is bounded from above by an upper bound \( \overline{V} < \infty \). Hence the initial consumption level \( \overline{c}(s^0) \) associated with an optimum \( V(\tilde{b}, s) > V(b, s) \) must be such that

\[
u(\overline{c}(s^0), s_0) + \overline{V} \geq V(b, s)
\]

(3.24)

From assumption 3.6.2, \( u(\overline{c}(s^0), s_0) \to -\infty \) as \( \overline{c}(s^0) \to 0 \), and hence for (3.24) to be satisfied we must have \( \overline{c}(s^0) \geq \xi > 0 \) for some lower bound \( \xi \). We also know that \( \overline{c}(s^0) \leq \tilde{b}/R + \overline{y} - b(s^0) \equiv \overline{c} \), giving us an upper bound. Since \( u \) is
continuous and \([c, \bar{c}]\) is a compact interval, we can pick a single \(\delta_2 > 0\) such that 
\[|u(c_A, s_0) - u(c_B, s_0)| < \epsilon \]
for all \(c_A \in [c, \bar{c}]\) and \(|c_B - c_A| < \delta_2\).

Now consider the optimal plan \(\{\bar{b}(s'), \bar{r}(s')\}\) starting at \((\bar{b}, s)\). This plan is also feasible starting at \((b, s)\) and delivers the same consumption at every point except \(t = 0\), where consumption is \(b - \bar{b}\) lower. Hence we have

\[V(\bar{b}, s) - V(b, s) = u(\bar{c}(s^0), s_0) - u(\bar{c}(s^0) - \delta_2, s_0) < \epsilon\]
as desired.

4. Since \(b^*(s') \geq 0\), increasing \(b^*(s')\) always weakly increases

\[\frac{b(s')}{R} \left( \sum_{s_{t+1}} \pi(s_{t+1}|s_t) \cdot 1_{\{b(s') \leq b^*(s_{t+1})\}} \right)\]

when \(b(s') \geq 0\) and leaves it unchanged when \(b(s') \leq 0\), which completes the proof.

5. Follows from 4, since \(V^{nb}(0, s)\) is the value with default thresholds all equal to zero, as shown in section 3.3.2.

6. Assumption 3.6.4 ensures that for any \(b' > 0\), \(\frac{b'}{R} \sum_{s':b' \leq b^*(s')} \pi(s'|s) < \frac{k}{R}\). Hence feasible consumption at date 0 is \(c(s^0) < y(s) + \frac{\bar{b}}{R} - \left(\frac{k}{R} + y(s)\right) = 0\).

Given that the continuation value for any \(b'\) is finite, 6 follows from assumption 3.6.2.

7. Fix \(b\) and \(s^0\). Let \(\epsilon > 0\) and let \(\{b^*(s')\}\) be a set of default thresholds. In an argument similar to the proof of 3, we show that (i) there exists \(\delta_1\) such that, for any alternative set of default thresholds \(\{\hat{b}^*(s')\}\) such that \(|b^*(s') - \hat{b}^*(s')| < \delta_1\) for all \(s'\), we have \(V(b, s, \{\hat{b}^*(s')\}) > V(b, s, \{b^*(s')\}) - \epsilon\), and (ii) there exists \(\delta_2\) such that, for any \(\{\hat{b}^*(s')\}\) such that \(|b^*(s') - \hat{b}^*(s')| < \delta_2\) for all \(s'\), we have \(V(b, s, \{b^*(s')\}) > V(b, s, \{\hat{b}^*(s')\}) - \epsilon\). Combining (i) and (ii) then proves continuity. In both cases, we use the fact that a government facing debt thresholds that are lower by at most \(\delta\) can guarantee itself a consumption plan
that is only $\delta$ below that of a government with reference debt thresholds at date 0—and above at every other date—using a mimicking strategy, as embodied in the following claim.

**Claim.** Assume that $|b^*(s') - \tilde{b}^*(s')| < \delta$. Let $\{b(s'), p(s')\}$ be a plan that achieves consumption $c(s')$ subject to the default thresholds $\{b^*(s')\}$ starting from $(b, s')$. Then there is another plan $\{\tilde{b}(s'), p(s')\}$ that achieves consumption $\tilde{c}(s')$ subject to the default thresholds $\{\tilde{b}^*(s')\}$ such that $\tilde{c}(s') > c(s')$ for all $t \geq 1$ and $\tilde{c}(s^0) > c(s^0) - \delta$.

**Proof of claim.** Define $\tilde{b}(s') \equiv b(s') - \delta$ for all $t \geq 0$ and $\tilde{b}(s^{-1}) \equiv b(s^{-1}) = b$. Then compute

$$\tilde{c}(s^t) = y(t) + \frac{b(s^t) - \delta}{R} \left( \sum_{s_{t+1}} \pi(s_{t+1} | s_t) \cdot 1_{\{b(s^t) - M \leq \tilde{b}^*(s^t)\}} \right) - (b(s^t-1) - \delta)$$

$$\geq y(t) + \frac{b(s^t)}{R} \left( \sum_{s_{t+1}} \pi(s_{t+1} | s_t) \cdot 1_{\{b(s^t) \leq b^*(s^t)\}} \right) - b(s^t-1) + \left( 1 - \frac{1}{R} \right) \delta > c(s^t)$$

and

$$\tilde{c}(s^0) = y(0) + \frac{b(s^0) - \delta}{R} \left( \sum_{s_1} \pi(s_1 | s_0) \cdot 1_{\{b(s^0) - M \leq \tilde{b}^*(s_1)\}} \right) - b(s^{-1})$$

$$\geq y(0) + \frac{b(s^0)}{R} \left( \sum_{s_1} \pi(s_1 | s_0) \cdot 1_{\{b(s^0) \leq b^*(s_1)\}} \right) - b(s^{-1}) - \frac{\delta}{R} > c(s^0) - \delta$$

To prove (i), consider the plan $\{c(s^t), b(s^t), p(s^t)\}$ that achieves $V(b, s, \{b^*(s')\})$. Using the continuity of $u(c, s^0)$, let $\delta_1$ be such that $|u(c', s_0) - u(c(s^0), s_0)| < \epsilon$ for all $|c' - c(s^0)| < \delta_1$. Then, whenever the thresholds $\{\tilde{b}^*(s')\}$ are such that $|b^*(s') - \tilde{b}^*(s')| < \delta_1$ for all $s'$, it follows from the claim that there is a consumption plan $\{\tilde{b}(s^t), p(s^t)\}$ for these thresholds that achieves consumption above $c(s^0) - \delta$ in the first period and above $c(s^t)$ everywhere else, and hence value greater than $V(b, s, \{b^*(s')\}) - \epsilon$. 

173
To prove (ii), suppose for some \( \{b^*(s')\} \) that \( V(b, s, \{b^*(s')\}) \geq V(b, s, \{b^*(s')\}) \) (otherwise, the desired inequality is immediate), and let \( \{\tilde{b}(s'), p(s')\} \) be the plan attaining the optimum for \( V(b, s, \{b^*(s')\}) \). We can establish using the argument from the proof of 3 we can pick a single \( \delta_2 > 0 \) such that \( |u(c, s_0) - u(\bar{c}(s_0), s_0)| < \epsilon \) for \( |c - \bar{c}(s_0)| < \delta_2 \). It follows from the claim that there is a plan \( \{\tilde{b}(s'), p(s')\} \) that (subject to the default thresholds \( \{b^*(s')\} \)) achieves consumption \( \tilde{c}(s') \) that is strictly greater than \( \bar{c}(s') \) for all \( t \geq 1 \) and strictly greater than \( \bar{c}(s^0) - \delta_2 \) for \( t = 0 \). From the choice of \( \delta_2 \) we know that \( |u(\tilde{c}(s^0), s_0) - u(\bar{c}(s^0), s_0)| < \epsilon \), and hence that the proposed plan \( \{\tilde{b}(s'), p(s')\} \) gives value strictly greater than \( V(b, s, \{b^*(s')\}) - \epsilon \). It follows that \( V(b, s, \{b^*(s')\}) > V(b, s, \{b^*(s')\}) - \epsilon \) as desired.

\[ \square \]

Existence of equilibrium

Using the operators defined in section 3.6.1-3.6.1, we can define the operator \( T = B \circ V \).

**Lemma 3.6.3.** The operator \( T \) is monotone increasing and maps the set

\[ \prod_s \left[ 0, y(s) + \frac{\bar{b}}{R} \right] \]

onto itself.

**Proof.** Monotonicity follows by combining lemmas 3.6.13 and 3.6.24. By combining lemmas 3.6.11 and 3.6.25, we obtain that \( Tb^*(s) \geq 0 \) whenever \( b^*(s) \geq 0 \). By combining lemmas 3.6.12 and 3.6.26, we obtain that \( Tb^*(s) \leq y(s) + \frac{\bar{b}}{R} \) for each \( s \).

Let \( b^*(s) = 0 \) for every \( s \). For \( n \geq 1 \) define the sequence

\[ b^n = Tb^{n-1} \]

By lemma 3.6.3, the sequences \( b^n(s) \) are increasing and bounded for every \( s \). Hence
they converge to form a set of thresholds \( \{b^*\} \). Define \( V^n = V(b^n) \) and \( V^\infty = V(b^\infty) \). From Lemma 3.6.27 it follows that \( V^\infty(b, s) = \lim_{n \to \infty} V^n(b, s) \). Next, because \( V^n \) is a sequence of continuous bijective functions with continuous inverses, whose limit \( V^\infty \) is continuous and bijective, and since by definition \( B(V^n)(s) = (V^n)^{-1}(V^d(s), s) \), we have that

\[
B\left(\lim_{n \to \infty} V^n\right) = \lim_{n \to \infty} B(V^n)
\]

and therefore

\[
B(V^\infty) = \lim_{n \to \infty} T b^n = b^* \tag{3.25}
\]

So \((V^\infty, b^\infty)\) constitutes an equilibrium, as we set out to prove. To map these objects to those in the main text, define \( V = V^\infty \) and the bond price schedule \( Q \) as

\[
Q(b', s) = \frac{b'}{R} \mathbb{E}_{s'|s} [b' \leq b^* (s')] = \frac{b'}{R} \sum_{s': b' \leq b^* (s')} \pi (s'|s)
\]

then \((V, Q)\) is a Markov perfect equilibrium, since (3.1)-(3.2) is the recursive formulation of the problem in (3.22) for the schedule \( Q \) generated by the thresholds \( B(V) \), and (3.25) guarantees that (3.3) holds.

### 3.6.2 Uniqueness of subgame perfect equilibrium

This appendix proves uniqueness of the subgame perfect equilibrium in the game of section 3.2. In order to define the game explicitly, we assume that there exist overlapping generations of two-period lived international investors. The set of investors born at time \( t \) is denoted by \( I_t \). We assume that \( I_t \) is finite, that \( |I_t| \geq 2 \), and that all investors are risk-neutral with preferences given by

\[
-q_i a^t_{i+1} + \frac{1}{R} \mathbb{E}_t [a^t_{i+1} p_{i+1}^{t+1}]
\]

where \( R > 1 \). We next describe the sequence of actions.

Every period, with incoming history \( h^{t-1} \), after Nature realizes the exogenous state
\( s_t \), the government chooses repayment \( p_t \). If it chooses \( p_t = 0 \) (default), it obtains value \( V^d(s_t) \), investors receive zero, and the game ends.

If it chooses \( p_t = 1 \), the government receives income \( y(s_t) \geq 0 \) and chooses next period debt \( b_{t+1} \). Next, every investor \( i \) simultaneously bids a price \( q^i_t \geq 0 \) for the government’s debt. Given bids \( q^i_t \), an auctioneer allocates the bonds \( a^i_{t+1} \) according to the following rule:

\[
    a^i_{t+1} = \begin{cases} 
        \frac{b_{t+1}}{J} & \text{if } q^i_t = \max_{i'} q_{i'}^t \\
        0 & \text{otherwise}
    \end{cases}
\]

where \( J \) is the number of investors bidding the maximum price. History for period \( t \) is now \( h^t = (h^{t-1}, s_t, b_{t+1}, \{q^i_t\}) \).

The government receives \( Q_t = q_t b_{t+1} \) where \( q_t = \max_{i'} q^i_t \) and repays debt \( b_t \) to previous investors. Its consumption is then

\[
    c_t = y(s_t) - b_t + q_t b_{t+1}
\]

for which it receives flow utility \( u(c_t, s_t) \), and expected value

\[
    V(h^{t-1}, s_t) = \begin{cases} 
        u(c_t, s_t) + \beta \mathbb{E}_t [V(h^t, s_{t+1})] & \text{if } p_t = 1 \\
        V^d(s_t) & \text{if } p_t = 0
    \end{cases}
\]

(3.27)

**Definition 2.** A government strategy is \( p(h^{t-1}, s_t), b'(h^{t-1}, s_t) \) specifying the repayment and next period debt decision after each history \( h^{t-1} \) and state \( s_t \). A strategy of investor \( i \) born at time \( t \) is a price bid \( q^i(h^{t-1}, s_t, b_{t+1}) \).

Together, investor strategies imply a bond revenue function \( Q(h^{t-1}, s_t, b_{t+1}) \).

**Definition 3.** A subgame perfect equilibrium consists of strategies for the government and investors such that at each \( (h^{t-1}, s_t) \):

1. \( p(h^{t-1}, s_t), b'(h^{t-1}, s_t) \) maximize (3.27)

2. For all \( i \in \mathcal{I}_t, q^i(h^{t-1}, s_t, b_{t+1}) \) maximizes (3.26)
In any subgame perfect equilibrium, investor maximization leads to

\[
q(h^{t-1}, s_t, b_{t+1}) = \frac{1}{R} \mathbb{E}_t \left[ p(h^t, s_{t+1}) \right]
\]  

(3.28)

We retain the other assumptions from the model in section 3.2 on \( u, V^d \), and the no-Ponzi bound on debt \( \bar{b} \). These include assumption 1 and assumptions 3.6.1 through 3.6.5. Importantly, assumption 3.6.5 continues to imply that a government with debt \( b < 0 \) never finds it optimal to default, so \( q(h, s, b') = \frac{1}{R} \) for any \( b' < 0 \).

The following lemma is crucial to the proof of unique equilibrium. It shows that in equilibrium, regardless of the history of play, a government with a strictly lower level of debt can always achieve a weakly higher value than a government with more debt in the same state, and is also weakly more likely to repay. Like the proof of lemma 3.2.2, it uses a mimicking-based argument, although here the proof is written in a recursive setting and must deal with technical complications that arise from the more general notion of equilibrium.

**Lemma 3.6.4.** Consider two subgame perfect equilibria \( A \) and \( B \). For any \((h_A, h_B, s)\), if \( b(h_A) > b(h_B) \) then \( V_A(h_A, s) \leq V_B(h_B, s) \), and \( p_B(h_B, s) = 1 \) if \( p_A(h_A, s) = 1 \).

**Proof.** Define

\[
M \equiv \sup_{h_A, h_B, s} \{ b(h_A) - b(h_B) \} \text{ s.t. } V_A(h_A, s) \geq V_B(h_B, s) \text{ and } p_A(h_A, s) = 1 \}
\]

Assume \( M > 0.12 \) Let \( 0 < \epsilon < \frac{R-1}{R+1}M \), and let \((h_A, h_B, s)\) be such that \( V_A(h_A, s) \geq V_B(h_B, s) \), \( p_A(h_A, s) = 1 \) and \( b(h_A) > b(h_B) + M - \epsilon \). Define

\[
\tilde{b}'_B = b'_A(h_A, s) - M - \epsilon
\]

(3.29)

---

12One can rule out the case \( M = \infty \) through a more direct mimicking argument: whenever \( b(h_A) - b(h_B) > \bar{b} \), where \( \bar{b} \) is the upper bound on debt, then a government at \((h_B, s)\) can mimic at distance \( \bar{b} \) the strategy of a government at \((h_A, s)\), with weakly more favorable prices (and hence strictly higher consumption due to its lower \( b \)) guaranteed because it will never be in debt.

177
and continuation histories

\[ h_A' = (h_A, s, b_A'(h_A, s), \{q_A^i\}) \]

\[ \tilde{h}_B' = (h_B, s, \tilde{b}_B, \{\tilde{q}_B^i\}) \]

This is a feasible choice for the B government at \((h_B, s)\) because we assume that debt can be chosen at any level below some upper bound. We aim to prove that through this choice of \(\tilde{b}_B\), the government in the B equilibrium achieves expected utility strictly greater than \(V_A(h_A, s)\), thus establishing that \(V_B(h_B, s) > V_A(h_A, s)\), a contradiction. We first establish that continuation utility for B is weakly greater in each future state, and then that current consumption is strictly greater, than their corresponding values for A.

We have, for all \(s' \in S\),

\[ V_B(\tilde{h}_B', s') \geq V_A(h_A', s') \quad (3.30) \]

Indeed, if \(p_A(h_A', s') = 0\), then immediately \(V_B(\tilde{h}_B', s') \geq V^d(s') = V_A(h_A', s')\).

Moreover, if \(p_A(h_A', s') = 1\) then, since \(b(h_A') - b(\tilde{h}_B') > M\) by (3.29), we must have \(V_B(\tilde{h}_B', s') > V_A(h_A', s') \geq V^d(s')\).

This last observation also implies that \(p_B(\tilde{h}_B', s') = 1\) whenever \(p_A(h_A', s') = 1\). Hence, using the pricing condition (3.28), we also have

\[ q_B(h_B, s, \tilde{b}_B') \geq q_A(h_A, s, b_A') \quad (3.31) \]

Using (3.31), we now show that the consumption achieved by B from the choice of \(\tilde{b}_B\) is strictly greater than that achieved by A. Indeed, using the flow budget constraints of both governments, and dropping dependence on history for ease of notation:

\[ \tilde{c}_B = c_A + b_A - b_B + \tilde{q}_B \tilde{b}_B - q_A b_A' \]

\[ \geq c_A + M - \epsilon + (\tilde{q}_B - q_A) b_A' + \tilde{q}_B (\tilde{b}_B' - b_A') \quad (3.32) \]
where the inequality follows from the definition of $A$ and $B$.

Now if $b'_A < 0$ then, since $\tilde{b}_B \leq b'_A < 0$ as well we have $q_A = \tilde{q}_B = \frac{1}{R}$, and hence $(\tilde{q}_B - q_A) b'_A = 0$. If $b'_A \geq 0$ then using (3.31), $(\tilde{q}_B - q_A) b'_A \geq 0$.

Moreover, from (3.29), $\tilde{b}_B' - b'_A = -M - \epsilon$, and using $\tilde{q}_B \leq \frac{1}{R}$, $\tilde{q}_B (\tilde{b}_B' - b'_A) \geq -\frac{1}{R} (M + \epsilon)$. Using these inequalities in (3.32),

$$\tilde{c}_B \geq c_A + M - \epsilon - \frac{1}{R} (M + \epsilon)$$
$$\geq c_A + \left(1 - \frac{1}{R}\right) M - \epsilon \left(1 + \frac{1}{R}\right)$$
$$> c_A$$

(3.33)

where the last line follows from the choice of $\epsilon$.

Since the utility from choosing $\tilde{b}_B$ provides a lower bound on $V_B (h_B, s)$, we have

$$V_B (h_B, s) \geq u (\tilde{c}_B, s) + \beta \sum_{s'} V_B (\tilde{h}_B', s')$$
$$> u (c_A, s) + \beta \sum_{s'} V_A (h_A', s') = V_A (h_A, s)$$

where the second line follows from (3.30) and (3.33). This contradicts $M > 0$. Hence $M \leq 0$. We have proved that for $(h_A, h_B, s)$, if $V_A (h_A, s) \geq V_B (h_B, s)$ and $p_A (h_A, s) = 1$ then $b (h_A) \leq b (h_B)$.

So if $b (h_A) > b (h_B)$, either $p_A (h_A, s) = 0$ so that $V^d (s) = V_A (h_A, s) \leq V_B (h_B, s)$, or $p_A (h_A, s) = 1$ and $V_A (h_A, s) < V_B (h_B, s)$. The lemma is proved. \hfill $\square$

With lemma 3.6.4 in hand, the proof of proposition 3.2.4 requires only one additional step. We need to show that the value function is uniquely determined by $b$ and $s$. If two governments start with the same levels of $b$ and $s$, either one of them can mimic the other but choose $\epsilon$ less debt in the next period; lemma 3.6.4 implies that from this point forward, the mimicking government is weakly better off. The utility loss from paying down $\epsilon$ debt in the initial period can be made arbitrarily small by choosing arbitrarily small $\epsilon$, and hence the mimicking government's value must be weakly higher. Since this argument works in both directions, we conclude that the
value is indeed uniquely determined by $b$ and $s$.

**Proof of proposition 3.2.4.** Consider $(h_A, h_B)$ such that $b(h_A) = b(h_B) = b$. At $(h_B, s)$ consider the feasible choice

$$
\tilde{b}_B' = b'_A(h_A, s) - \epsilon
$$

for some $\epsilon > 0$. Define continuation histories

$$
\begin{align*}
\tilde{h}_A' &= (h_A, s, b'_A(h_A, s), \{q_A^i\}) \\
\tilde{h}_B' &= (h_B, s, \tilde{b}_B', \{q_B^i\})
\end{align*}
$$

From Lemma 3.6.4,

$$
V_B(\tilde{h}_B', s') \geq V_A(\tilde{h}_A', s') 
$$

(3.34)

and $B$ repays if $A$ repays. Hence $\tilde{q}_B \geq q_A$ by the pricing condition (3.28). Moreover,

$$
\begin{align*}
\tilde{c}_B &= c_A + \tilde{q}_B\tilde{b}_B' - q_Ab'_A \\
&= c_A + (\tilde{q}_B - q_A)b'_A + \tilde{q}_B(\tilde{b}_B' - b'_A) \\
&\geq c_A - \frac{1}{R} \epsilon
\end{align*}
$$

(3.35)

where the inequality follows as in (3.33) in the proof of Lemma 3.6.4.

Now,

$$
V_B(h_B, s) - V_A(h_A, s) \geq u(\tilde{c}_B) - u(c_A) + \beta \sum_{s'} \left( V_B(\tilde{h}_B', s') - V_A(\tilde{h}_A', s') \right)
$$

$$
\geq u\left( c_A - \frac{1}{R} \epsilon \right) - u(c_A)
$$

where inequality follows form (3.34) and (3.35). Taking the limit as $\epsilon \to 0$ and using continuity of $u$, we obtain $V_B(h_B, s) \geq V_A(h_A, s)$. The symmetric argument implies that $V_B(h_B, s) \leq V_A(h_A, s)$, which concludes the proof. \qed
3.7 Appendix: Proofs for section 3.4

3.7.1 Proof of lemma 3.4.1

Proof. First, note that for any \(x\) and \(s\), we have

\[
\bar{Q}(\lambda x + b, s) = \frac{(\lambda x + b)}{R} \sum_{\{s': x \leq b^*(s') - b\}} \pi(s'|s)
\]

\[
\geq (1 - \lambda)b + \frac{(\lambda x + b)}{R} \sum_{\{s': x \leq b^*(s') - b\}} \pi(s'|s) = \lambda Q(s, x + b) + (1 - \lambda)b
\]

where there is strict inequality if \(b < 0\).

Now we can formally define the mimicking at a distance policy. Suppose that at time 0 we have state \(s\) and debt level \(b\). The equilibrium \((V, Q)\) induces an allocation \(\{c(s^t), b(s^t-1), p(s^t)\}_{s^t \geq s^0}\) at all histories following \(s^0\). We construct a policy for the government in the equilibrium \((\tilde{V}, \tilde{Q})\) starting at \(s^0\) as follows. For every history \(s^t \geq s^0\), let \(\tilde{p}(s^t) = p(s^t)\), and whenever \(p(s^t) = 1\) define a plan for debt

\[
\bar{b}(s^t) = \lambda(b(s^t-1) - b)
\]

The resulting consumption path, again for \(p(s^t) = 1\), satisfies

\[
\bar{c}(s^t) = y(s_t) - \bar{b}(s^t-1) + \bar{Q}(\bar{b}(s^t), s_t)
\]

\[
= y(s_t) - \lambda b(s^t-1) - (1 - \lambda)b + Q(b(s^t) - b) + b, s_t)
\]

\[
\geq (1 - \lambda)y(s_t) + \lambda(y(s_t) - b(s^t-1) + Q(s_t, b(s^t)))
\]

Using the concavity of \(u\), whenever \(p(s^t) = 1\) we have

\[
u(\bar{c}(s^t), s_t) \geq (1 - \lambda)u(y(s_t), s_t) + \lambda u(c(s^t), s_t) \geq (1 - \lambda)v^d(s_t) + \lambda u(c(s^t), s_t)
\]

(3.37)
where the strict inequality from (3.36) persists in (3.37) whenever $b < 0$, and by assumption, $u(y(s_t), s_t) > v^d(s_t)$ gives strict inequality whenever $b = 0$. Furthermore, in states where $p(s') = 0$, we have

$$u(c(s'), s_t) = u(c(s'), s_t) = v^d(s_t) \quad (3.38)$$

Summing (3.37) and (3.38) across all times and states to obtain the expected discounted value, we obtain the result. \qed

### 3.7.2 Proof of proposition 3.4.3

**Proof.** Suppose to the contrary that there exist two distinct equilibria $(V, Q)$ and $(\tilde{V}, \tilde{Q})$, with associated default thresholds $\{b^*(s)\}_{s \in S}$ and $\{\tilde{b}^*(s)\}_{s \in S'}$, such that $Q(b, s) \geq \tilde{Q}(b, s)$ for all $b$ and $s$. It follows that $V(s, b) \geq \tilde{V}(s, b)$ for all $b$ and $s$ as well, since a government facing the weakly higher debt schedule $Q$ can always replicate the policy of the government facing $\tilde{Q}$, achieving weakly higher consumption in the process.\textsuperscript{13}

Since $Q$ and $\tilde{Q}$ are distinct, there exists some $s'$ such that $b^*(s') > \tilde{b}^*(s')$, and we define

$$M = \max_s b^*(s) - \tilde{b}^*(s) > 0 \quad (3.39)$$

We first seek to prove that, for any $s$ and $b \leq b^*(s)$

$$\tilde{V}(b - M, s) - \tilde{V}^d(s) > V(b, s) - V^d(s) \quad (3.40)$$

To do so, we use the same mimicking at a distance argument as in lemma 3.2.2, although the calculation becomes somewhat more complicated. Writing $s^0 \equiv s$, we continue to set $\tilde{b}(s') = b(s') - M$ and $\tilde{p}(s') = p(s')$, along with the consumption policy

---

\textsuperscript{13}Explicitly, it can set $b = \tilde{b}$, $p = \tilde{p}$, $c = \tilde{c} + Q(s, b) - \tilde{Q}(s, b) \geq 0$. 182
\( \widetilde{c}(s^t) \) in (3.9). This strategy places a lower bound on \( \tilde{V}(b - M, s) \):

\[
\tilde{V}(b - M, s) \geq \sum_{p(s^t) = 1} \beta^t \Pi(s^t) u(\widetilde{c}(s^t)) + \sum_{p(s^t) = 0, p(s^{t-1}) = 1} \beta^t \Pi(s^t) \tilde{V}^d(s_t) \quad (3.41)
\]

Subtracting the corresponding expression for \( V(b, s) \), and using \( \widetilde{c}(s^t) > c(s^t) \), we have

\[
\tilde{V}(b - M, s) - V(b, s) > \sum_{p(s^t) = 0, p(s^{t-1}) = 1} \beta^t \Pi(s^t) \left( \tilde{V}^d(s_t) - V^d(s_t) \right) \quad (3.42)
\]

Subtracting \( \tilde{V}^d(s_t) - V^d(s_t) \) from both sides we obtain

\[
\left( \tilde{V}(b - M, s) - \tilde{V}^d(s_t) \right) - (V(b, s) - V^d(s)) > - (\tilde{V}^d(s) - V^d(s)) + \sum_{p(s^t) = 0, p(s^{t-1}) = 1} \beta^t \Pi(s^t) \left( \tilde{V}^d(s_t) - V^d(s_t) \right) \quad (3.43)
\]

and to prove (3.40) it suffices to show that the right side of (3.43) is nonnegative.

Expanding \( \tilde{V}^d(s_t) - V^d(s_t) \) gives

\[
\tilde{V}^d(s_t) - V^d(s_t) = \sum_{s_T} \beta^t (1 - \lambda) \lambda^{t-t-1} \Pi(s_T|s_t) \left( \tilde{V}^o(0, s_T) - V^o(0, s_T) \right) \quad (3.44)
\]

Now, using (3.44), we can rewrite the right side of (3.42) as

\[- \sum_{s^t > s^0} \beta^t (1 - \lambda) \lambda^{t-t-1} \Pi(s^t) \left( \tilde{V}^o(0, s_T) - V^o(0, s_T) \right) + \sum_{p(s^t) = 0, p(s^{t-1}) = 1} \sum_{s^t > s^t} \beta^t (1 - \lambda) \lambda^{t-t-1} \Pi(s^t) \left( \tilde{V}^o(0, s_T) - V^o(0, s_T) \right) \]

which can be rearranged as

\[
\lambda^{-1} (1 - \lambda) \sum_{s^t > s^0} \beta^t \Pi(s^t) \left( V^o(0, s_T) - \tilde{V}^o(0, s_T) \right) \cdot \left( 1 - \sum_{s^t > s^t > s^0} \lambda^{t-t} \cdot 1_{p(s^t) = 0, p(s^{t-1}) = 1} \right) \quad (3.45)
\]

Since for any \( s^t \) there exists at most one \( s^t \) such that \( p(s^t) = 0 \) and \( p(s^{t-1}) = 1 \), the rightmost factor in parentheses is nonnegative. Since in addition \( V(0, s_T) \geq \tilde{V}(0, s_T) \),
the preceding factor is nonnegative as well, and hence (3.45) is nonnegative. (3.40) therefore follows.

Finally, suppose that the maximum in (3.39) is attained at \( \bar{s} \), so that \( b^*(\bar{s}) = \tilde{b}^*(\bar{s}) + M \). Applying (3.40), we have

\[
0 = \tilde{V}(\tilde{b}^*(\bar{s}), \bar{s}) - \tilde{V}^d(\bar{s}) > V(b^*(\bar{s}), \bar{s}) - V^d(\bar{s}) = 0
\]

which is a contradiction. \( \square \)

### 3.7.3 Proof of proposition 3.4.4

**Proof.** Write \( V^{re} = \mathbb{E}_0[V^o(0, s')] \), and similarly \( \tilde{V}^{re} \) for a conjectured alternative equilibrium. First, observe that if \( \tilde{V}^{re} = V^{re} \), then the two equilibria have the same expected value from default \( V^d \), and we can apply proposition 3.2.3 taking \( V^d \) as given to conclude that the two equilibria must be the same.

Otherwise, assume without loss of generality that \( V^{re} > \tilde{V}^{re} \). It cannot be that \( \bar{Q}(b') \geq Q(b') \) for all \( b' \), since in that case a government starting with zero debt and facing the weakly higher debt schedule \( \bar{Q} \) could always replicate the policy of the government facing \( Q \), achieving weakly higher consumption in the process. This would imply \( \tilde{V}^{re} \geq V^{re} \), a contradiction. Hence \( Q(b') > \bar{Q}(b') \) for some \( b' \). From this point on, the proof is the same as the proof for proposition 3.4.3 starting with the definition of \( M \) in (3.39), except that we can replace (3.44) with simply

\[
\tilde{V}^d(s_t) - V^d(s_t) = \sum_{r > t} \beta^{r-t}(1 - \lambda) \lambda^{r-t-1} (\tilde{V}^{re} - V^{re}) \tag{3.46}
\]

allowing us to replace (3.45) with

\[
\lambda^{-1}(1 - \lambda) \sum_{\tau > 0} \beta^{\tau}(V^{re} - \tilde{V}^{re}) \cdot \left(1 - \sum_{r > t > 0} \lambda^{r-t} \cdot 1_{\{p(s^t) = 0, p(s^{t-1}) = 1\}} \right) \tag{3.47}
\]

again concluding that the expression is nonnegative, from which a contradiction follows. \( \square \)
Bibliography


Conesa, Juan Carlos and Timothy J. Kehoe, “Gambling for Redemption and Self-Fulfilling Debt Crises,” Manuscript, Stony Brook University and University of Minnesota, January 2015.


