Optimal Policy with General Signal Extraction

Esther Hauk
Andrea Lanteri
Albert Marcet

September 2016

Barcelona GSE Working Paper Series

Working Paper nº 932

Optimal Policy with General Signal Extraction*

Esther Hauk† Andrea Lanteri‡ Albert Marcet§

September 2016

Abstract

This paper studies optimal policy with partial information in a general setup where observed signals are endogenous to policy. In this case, signal extraction about the state of the economy cannot be separated from the determination of the optimal policy. We derive a non-standard first order condition of optimality from first principles and we use it to find numerical solutions. We show how previous results based on linear methods, where separation or certainty equivalence obtains, arise as special cases. We use as an example a model of fiscal policy and show that optimal taxes are often a very non-linear function of observed hours, calling for tax smoothing in normal times, but for a strong fiscal reaction to output when a recession is quite certain and the economy is near the top of the Laffer curve or near a debt limit.

JEL codes: C63, D82, E60, H60.
Keywords: Optimal Policy; Partial Information; Separation; Calculus of Variations; Fiscal Policy.

---

*The authors appreciate helpful comments from Wouter den Haan, Cosmin Ilut, Kyle Jurado, Jennifer La'O, Paul Levine, Martí Mestieri, Juan Passadore, Joe Pearlman, Aleh Tsyvinski, Robert Shimer, Jaume Ventura, Mike Woodford and other seminar participants at LSE Macro Work-in-progress, IAE-CSIC Macro work-in-progress, Midwest Macro Spring Meeting 2014, SED Annual Meeting 2014, University of Cambridge, Columbia University, University of Vienna, Duke University, Federal Reserve Board, Barcelona GSE Summer Forum 2016. Hauk acknowledges financial support from through CICYT project number ECO2015-67171-P and from the government of Catalonia (2014 SGR 1064). Hauk and Marcet have received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement no. 612796 and from the Spanish Ministry of Economy and Competitiveness through the Severo Ochoa Program for Centers of Excellence in R&D. Lanteri acknowledges support and hospitality from MOVE Barcelona and Institut d'Anàlisi Econòmica-CSIC. Marcet acknowledges support from the Axa Foundation, the Excellence Program of Banco de España and European Research Council under the EU 7th Framework Programme (FP/2007-2013) under grant agreement 324048-APMPAL.

†Institut d'Anàlisi Econòmica-CSIC, MOVE and Barcelona GSE; email: esther.hauk@iae.csic.es

‡Department of Economics, Duke University; email: andrea.lanteri@duke.edu

§Institut d'Anàlisi Econòmica-CSIC, ICREA, Barcelona GSE, UAB, MOVE; email: marcet.albert@gmail.com

> "In the policy world, there is a very strong notion that if we only knew the state of the economy today, it would be a simple matter to decide what the policy should be. The notion is that we do not know the state of the system today, and it is all very uncertain and very hazy whether the economy is improving or getting worse or what is happening. Because of that, the notion goes, we are not sure what the policy setting should be today. [...] In the research world, it is just the opposite. The typical presumption is that one knows the state of the system at a point in time. There is nothing hazy or difficult about inferring the state of the system in most models." (James Bullard, interview on *Review of Economic Dynamics*, November 2013)

1 Introduction

The opening quote states that inferring the underlying state of the economy is a key practical difficulty in setting macroeconomic policy. One could say that this is not such an insurmountable problem: policy-makers should choose the optimal policy taking into account the uncertainty ("haziness") about the underlying state, given the information available to them. However, a key difficulty of this problem is that the observed signals about the state of the economy are, in general, endogenous to policy decisions.

To our knowledge, all existing results in the literature on optimal policy with Partial Information assume either "separation", that is, that the filtering problem can be solved independently from the optimization problem, or "certainty equivalence", meaning that signal extraction and optimization can be done sequentially and uncertainty about the state does not affect the shape of the optimal policy. These cases constrain the application of Partial Information since the conditions for "separation" and "certainty equivalence" are not satisfied in most models currently used in macroeconomics. The main contribution of this paper is to provide optimality conditions that can be used to compute easily the solution in the general case, when the distribution of the fundamentals given the signal available to policy-makers is endogenous to policy. We then illustrate our results in a simple model of fiscal policy. We show how Partial Information can shed new light on the reasons for tax smoothing and the timing of fiscal adjustments in recessions. We also find that, in general, Partial Information introduces strong non-linearities in an otherwise linear setup.

To illustrate a case of endogenous signals in optimal policy, consider the fiscal policy response to the recent financial crisis. In 2008-2009 policy-makers observed a large fall in output and employment, but it was unclear whether the recession was due to a shock to productivity or to a demand shock or some combination of both. Nonetheless, policy-makers had to react to the recession implicitly making a guess about the nature of the shock. In that instance policy-makers opted for "stimulus packages" which, ex-post, given the depth of the recent recession, have generated larger deficits and have arguably worsened debt crises in some countries. Since the level of output and employment depend on whether an expansionary fiscal policy or austerity is adopted, the problem of signal extraction is endogenous to the choice of policy. Automatic stabilizers, e.g. income taxes and unemployment benefits, are leading examples of policies that respond to aggregate endogenous information such as output, while simultaneously influencing the level of output. This paper addresses the question of how to design such instruments optimally.

The existing literature on Ramsey-optimal policy has introduced models with non-linearities (such as debt limits, e.g., Aiyagari et al. 2002), but this literature almost always assumes Full Information: the government can observe the state of the economy and react to it contemporaneously. A small literature has produced important applications and technical contributions on models of optimal control with Partial Information. In particular, Svensson and Woodford (2003, 2004) study optimal policy in linear-quadratic models, where the optimal policy can be found under "certainty equivalence", in other words, the solution is found by applying the Full Information solution and evaluating the underlying shocks at their conditional expectation, so that a form of "separation" holds.[1]

However, "certainty equivalence" cannot be applied in general. In a non-linear model, no form of separation applies if the policy decision is taken simultaneously with the determination of the signal. More in general, in a dynamic model and under Rational Expectations, even if policy decisions are implemented with a delay, they affect the signal observed by the government currently. For example, tax rates are generally set for the following budget year, before the tax base (say, total income) is observed. Using common notation, taxes τ_{t+1} are decided in period t, before output y_{t+1} is observed, so it may seem that the tax and income are not determined simultaneously. But if y_t is observed when τ_{t+1} is chosen, the decision about τ_{t+1} simultaneously determines y_t under Rational Expectations. In fact, most models of optimal policy used in macroeconomics nowadays imply no separation between signals and policy.

We argue that Partial Information has important consequences for the design of optimal policy and for our understanding of real-world policy decisions. In particular, optimal policy is smooth when the economy is far from a crisis, but becomes highly responsive to signals when approaching a debt limit or the top of the Laffer curve. As a consequence, adjustments to shocks can be optimally delayed as a consequence of incomplete information and the policy function can be highly non-linear due to signal-extraction issues.

We first derive optimality conditions in a problem of optimal control with multi-dimensional uncertainty and an endogenous signal. The density of the signal depends on the policy variable and vice versa. We derive a first order condition (FOC) for the optimal policy relying on first principles. This optimality condition is different

[1] See Section 2 for more references and a detailed discussion of the literature.

from the standard FOC found in dynamic stochastic models: the FOC under Full Information needs to be weighted by a kernel that depends on the chosen policy. The FOC is derived for a general model so that our results can be widely applied. We show how “certainty equivalence” arises only in some special cases as an implication of our general theorem.

Our leading example is a two-period version of the standard fiscal policy model of Lucas and Stokey (1983). We introduce two shocks (to demand and to supply) and to make the issue of hidden information relevant we assume incomplete insurance markets as in Aiyagari et al. (2002). Then we solve for optimal Ramsey taxation under the assumption that the government does not observe the realizations of the shocks, but only some endogenous signal, such as output or hours worked.

This model yields interesting insights about the conduct of fiscal policy. First, hidden information can be a driver of tax smoothing, as the high frequency response of taxes to underlying shocks is small relative to the case of Full Information. However, we also find that at lower frequencies the tax response to shocks can be larger. The low short-term response of taxes is due to the fact that optimal policy averages all the possible unobserved contingencies. However, Partial Information can lead to accumulation of large debt over time, eventually leading to large fiscal adjustments and high long run volatility. This identifies a reason for tax smoothing that is quite different from the standard one in Ramsey policy under market completeness and Full Information (Lucas and Stokey, 1983).

A second policy implication is that Partial Information often leads to a very non-linear response of taxes to output. A government may go very quickly from not reacting to low observed output to increasing taxes very strongly as output falls further. This arises in particular with very high government spending, when future taxes could be close to the maximum of the Laffer curve, as was arguably the case in some European economies in the recent crisis. This sudden adjustment occurs once the signal is “bad enough” for policy-makers to be quite sure that a very low tax revenue is very likely. Importantly this inference is endogenous to the output observed and the policy chosen. Several European governments in the Great Recession made large fiscal adjustments a few years after the start of the downturn, leading to a large debt accumulation which eventually amplified the fiscal adjustments needed. Our model provides a framework where this delay in increasing taxes could be a feature of optimal decision making and not necessarily of irresponsible politicians.

Finally, we build an infinite-horizon model that confirms the results on tax smoothing and non-linear response in a dynamic setup. To the extent that our model contemplates that a delayed adjustment of fiscal policy is optimal it may be useful to detect when governments are optimally postponing tax adjustments or when they are simply behaving irresponsibly.

It should be noted that the technique we develop can be applied to many interesting open questions in optimal control under Partial Information. For example one can analyze macro policy when the government has incomplete information about agents' types, about the distribution of wealth or about agents' expectations. The government may have incomplete (or even wrong) information about the structure of the economy and it treats this uncertainty consistently in a Bayesian way. One can also apply our results to many other fields in economics.

The remainder of the paper is organized as follows. The related literature is discussed in Section 2. Section 3 introduces our two-period optimal fiscal policy model with incomplete markets and Partial Information. Section 4 contains the main theoretical contribution. It provides the first order condition for a general static model, it shows how this can be used to compute optimal solutions, it compares the Full Information solution with the Partial Information solution and discusses a case of "invertibility" when the two solutions coincide. We also study when "certainty equivalence"holds. In Section 5 we apply these results to our Ramsey fiscal policy problem. Section 6 presents the infinite-horizon model. Section 7 concludes.

2 Related literature

Most of the literature on optimal policy in dynamic models in the last thirty years has disregarded the issue of endogenous signal extraction. However, Partial Information and signal extraction were often present in the early papers on dynamic models with Rational Expectations.

Signal extraction with an *exogenous* signal is well understood; it goes as far back as Muth (1960). Typically, it just requires a routine application of the Kalman filter. Because the signal extraction problem is solved independently of policy choices, it is said that a "separation principle" between signal extraction and optimization applies.

Few papers have studied optimal policy when signals are endogenous. Pearlman (1992) and Svensson and Woodford (2003) consider linear Gaussian models where the policy-maker and the private sector have the same information set. In other words information is partial but symmetric. In this case, they show that the "separation principle" continues to hold. Baxter et al. (2007, 2011) derive an "endogenous Kalman filter" for all these cases which is equivalent to the solution of a standard Kalman filter of a parallel problem where all the states and signals are fully exogenous.

Closest to our work is Svensson and Woodford (2004). They consider optimal policy in a non-microfounded linear Rational Expectations model, where the government's information set is a subset of the private sector's information set. They show that, even though the "separation principle" fails because of asymmetric information, there is a suitable modification of the standard Kalman filter that works in the case of linearity and additively separable shocks. Moreover optimal policy has the "certainty equivalence" property: under Partial Information the government applies the Full Information policy to its best estimate of the state. Aoki (2003) applies these results to optimal monetary policy with noisy indicators on output and inflation. Nimark (2008) applies them to a problem of monetary policy where the central bank uses data from the yield curve while at the same time understanding that it is affecting them.

Our contribution is to consider a fully microfounded optimal policy model and to provide a solution to the general signal extraction and optimization problem when the government (or, more generally, a Stackelberg leader) conditions on a signal simultaneously determined with policy. In the general case separation does not apply. We find how "certainty equivalence" arises in the linear case using our generic theorem and we show a case where the correct solution is highly non-linear in nature and, therefore, a linear approximation can be misleading.[2]

The effect of policy choices on information extraction about unobserved variables is also considered in the armed-bandit problems of decision theory. For some applications to dynamic macro policy see Kiefer and Nyarko (1989), Wieland (2000a, 2000b) and Ellison and Valla (2001). For an application to monopolist behavior see Mirmann, Samuelson and Urbano (1993). Van Nieuwerburgh and Veldkamp (2006) use a similar learning framework to explain business-cycle asymmetries. In these papers the planner's decision influences the probability of next period's signal, but the current signal is unaffected by the current policy decision so that separation holds given the past policy decision. It would be interesting to blend the issue of influencing the distribution of future variables considered in these papers with the one studied in the current paper.

Another related strand of literature is that on robust control. Hansen and Sargent (2012) study Ramsey-optimal policy with ambiguity aversion and find that this leads to violations of "certainty equivalence" even in linear-quadratic setups. In their setup the optimal solution is found as if the state of the economy was known, for each possible value of the state and then the policy is chosen assuming the worst possible state. Therefore the optimal policy can be found without having to solve a filtering problem and the endogenous signal extraction issue that we address does not arise. In our setup is useful if one departs from the assumption that the planner has min-max preferences and has expected utility preferences, putting proper probabilities on the unobserved variables. Adam (2004) shows that the min-max criterion arises from a sequence of more and more risk-averse planners, so our setup is available for any level of risk aversion of the planner.

A wide literature considers models where competitive agents use prices as signals of unknown information.[3] Since prices are taken as given by competitive agents the

[2]Optimal non-linear policies have been found in the literature but for totally different reasons. Swanson (2006) obtains a non-linear policy when he relaxes the assumption of normality in the linear model with separable shocks. He considers a model where the "separation principle" applies. The non-linearity results entirely from Bayesian updating on the a priori non-Gaussian shocks.

[3]Lucas's (1972) seminal paper, and Guerrieri and Shimer (2013) analyze a competitive market in this setup. For an optimal policy problem see Angeletos and Pavan (2010) where those solving a signal extraction problem are the agents (not the government). The whole controversy about whether asymmetric information Rational Expectations equilibria could be reached as discussed in Townsend (1983) is also within this framework.

signal extraction problem can be solved with standard filtering techniques and the issue we address does not arise in this literature.

The literature on optimal contracts under private information and incentive compatibility constraints (or the "New Dynamic Public Finance" as in Kocherlakota, 2010) is perhaps less directly related to our work. This literature usually assumes revelation of the private information conditional on the equilibrium actions (the "invertibility" case) where, as we argue below in detail, Partial Information is not relevant and the filtering issue that we address does not arise. On the other hand, this literature assumes that agents react strategically to the optimal contingent policy set up by the principal, an issue that we abstract from. In our setup the government conditions on aggregate variables and, since agents are atomistic, the policy function ($\mathcal{R}$ in the text) does not affect agents' decision while the government action (the tax rate τ in our main example) does and it is taken as given by the agents. On the other hand, there is an interaction between the signal extraction problem and the optimal policy decision that the literature on incentive compatibility constraints often ignores. Blending the two issues would be of interest but it is left for future research.

3 A simple model of optimal fiscal policy

We first present a simple model of optimal policy. This will serve to illustrate the issue of general signal extraction and it will be of interest on its own. It is a simple two-period version of Lucas and Stokey (1983). We introduce incomplete insurance markets to be consistent with the Partial Information story.

3.1 Preferences and technology

The economy lasts two periods $t = 1, 2$. A government needs to finance an exogenous and deterministic stream of expenditure (g_1, g_2), where subscripts indicate time periods. The government levies distortionary income taxes (τ_1, τ_2) and issues bonds b^g in the first period that promise a repayment in second period consumption units with certainty.

The economy is populated by a continuum of agents. Each agent $i \in [0, 1]$ has utility function

$$E\left[U\left(c_1^i, l_1^i, c_2^i, l_2^i; \gamma\right)\right] \tag{1}$$

where

$$U\left(c_1^i, l_1^i, c_2^i, l_2^i; \gamma\right) = \gamma u\left(c_1^i\right) - v\left(l_1^i\right) + \beta\left[u\left(c_2^i\right) - v\left(l_2^i\right)\right]$$

where c_t^i and l_t^i for $t = 1, 2$ are consumption and hours worked respectively, with $u' > 0$, $u'' < 0$, $v' > 0$, $v'' > 0$.

We refer to γ, a random variable with distribution F_γ, as a "demand shock". When γ is high, agents like first period consumption relatively more than other goods.

Hence, a high value of γ makes them willing to work more in their intratemporal labor-consumption decision and also more impatient in their intertemporal allocation of consumption. Given that agents are identical in the remainder we drop the subscripts i for notational convenience.

The production function in each period is linear in labor and output is given by

$$y_t = \theta_t l_t \; for \; t = 1, 2. \tag{2}$$

The random variable $\theta_1 = \theta$ has distribution F_θ and we will refer to it as the "productivity shock". γ and θ are assumed to be independent. As far as θ_2 is concerned, we assume that both agents and government know with certainty that $\theta_2 = E\theta$, that is, the second period productivity is a known constant, equal to the mean of the first period shock.

To summarize, the state of the economy is fully described by a realization of the random variables $A \equiv (\gamma, \theta)$. These variables are observed at the beginning of period $t = 1$ by consumers and firms, but not by the government. The distributions F_γ and F_θ represent the government's perceived distribution of the exogenous shocks, which may or may not be equal to the true distribution of these variables. Thus this formulation encompasses the case of "true" uncertainty as well as the government's ignorance about the structure of the economy.

We consider agents that have Rational Expectations. To this end, denoting by Φ the space of possible values of A, we assume that agents know that fiscal policy is given by a triplet of functions $(\widetilde{\tau}_1, \widetilde{\tau}_2, \widetilde{b^g}) : \Phi \to R^3$ and these are actually the equilibrium values of taxes and government bonds for each A.

Consumers' choices and prices are contingent on the state $A \equiv (\gamma, \theta)$ observed by agents in period $t = 1$. Agents choose $(c_1, c_2, l_1, l_2, b) : \Phi \to R^5$ knowing the fiscal policy and the bond price function $q : \Phi \to R$. Obviously, the solution of the agents' problem in this setup coincides with the non-stochastic model where A is known. Uncertainty will only play a role in the government's problem, to be specified later.

Firms also observe θ at $t = 1$. Profit maximization implies that agents receive a wage equal to θ_t, observed by agents, so that the period budget constraints of the representative agent are

$$c_1 + qb = \theta l_1 (1 - \widetilde{\tau}_1) \tag{3}$$

$$c_2 = \theta_2 l_2 (1 - \widetilde{\tau}_2) + b \tag{4}$$

where q is the price of the government bond b. The above budget constraints have to hold for all realizations of A.

The government's budget constraints are analogous, they restrict the choice of the policy $(\widetilde{\tau}_1, \widetilde{\tau}_2, \widetilde{b^g})$.

3.2 Competitive equilibrium

Here we provide a definition of competitive equilibrium. The definition is standard in the literature, it is common to both the Full Information (FI) and the Partial

Information (PI) equilibria that we analyze.

Definition 1 *A* ***competitive equilibrium*** *is a fiscal policy* $(\widetilde{\tau}_1, \widetilde{\tau}_2, \widetilde{b^g})$*, price* q *and allocations* (c_1, c_2, l_1, l_2, b) *such that when agents take* $(\widetilde{\tau}_1, \widetilde{\tau}_2, q)$ *as given the allocations maximize the agents' utility (1) subject to (3) and (4). In addition, bonds and goods markets clear, so that* $b^g = b$ *and*

$$c_t + g_t = \theta l_t \; for \; t = 1, 2. \tag{5}$$

This definition embeds competitive equilibrium relations insuring that wages are set in equilibrium and that the budget constraint of the government holds in all periods due to Walras' law.

Utility maximization implies for all A

$$\frac{v'(l_1)}{u'(c_1)} = \theta\gamma(1 - \widetilde{\tau}_1) \tag{6}$$

$$\frac{v'(l_2)}{u'(c_2)} = \theta_2(1 - \widetilde{\tau}_2) \tag{7}$$

$$q = \beta \frac{u'(c_2)}{\gamma u'(c_1)} \tag{8}$$

As anticipated, the demand shock enters the first period labor supply decision described by (6) as well as the bond pricing equation (8). A competitive equilibrium is fully characterized by equations (3) to (8).

3.3 Ramsey equilibrium

To describe government behavior we now provide a definition of Ramsey equilibrium. As is standard we assume the government has full commitment, perfect knowledge about how taxes map into allocations for a given value of the underlying shocks A and that it chooses the best policy for households.

We first give the standard definition when both government and consumers observe the realization of A.[4]

Definition 2 *A* ***Full Information (FI-) Ramsey equilibrium*** *is a fiscal policy* (τ_1, τ_2, b^g) *that achieves the highest utility (1) when allocations are determined in a competitive equilibrium.*

Our interest is in studying optimal taxes under PI. More precisely, we assume that taxes in the first period have to be set before the shock A is known but after observing a **signal** s that potentially depends on aggregate outcomes observed in period 1, $s = G(c_1, l_1, q, A)$ for a given G.

[4] In these definitions we take for granted that we only consider tax policies for which a competitive equilibrium exists and is unique.

Definition 3 *A **Partial Information (PI-) Ramsey equilibrium** when government observes a signal s is a FI-Ramsey equilibrium satisfying*

1. *τ_1 is measurable with respect to s*
2. *fiscal policy (τ_1, τ_2, b^g) achieves the highest utility from among all equilibria satisfying 1.*

Restriction 1 can be expressed as the PI-Ramsey equilibrium having to satisfy

$$\tau_1 = \mathcal{R}(s) \; for \; all \; A \in \Phi, \; for \; some \; function \, \mathcal{R} : R \to R \tag{9}$$

We are interested in the case when this restriction prevents the PI-Ramsey equilibrium from achieving the FI version.

Note that consumers may or may not know that (9) holds, in any case they take as given the tax rate that arises from this equation and equilibrium. Even if they knew (9), they would not be able to exploit this knowledge in their optimization problem as we consider atomistic agents that cannot affect the aggregate signal and, hence, the tax rate.[5] In this model, as is standard in Ramsey equilibria, the tax level τ and the equilibrium allocations (and therefore s) are determined simultaneously as a consequence of the government's choice for $\mathcal{R}$.

FI- Ramsey equilibria

Using the so-called "primal approach" and standard arguments it is easy to show that an allocation is a competitive equilibrium if and only if, in addition to resource constraints (5), the following implementability condition holds

$$\gamma u'(c_1) c_1 - v'(l_1) l_1 + \beta \left[u'(c_2) c_2 - v'(l_2) l_2 \right] = 0. \tag{10}$$

The standard approach to find Ramsey policy under FI is to maximize (1) subject to (10) and the resource constraints. We now slightly deviate from this traditional approach in order to obtain a formulation of the FI problem that is as close as possible to the PI problem.

Implicit in the standard definition of FI Ramsey equilibrium is the assumption that the government knows how the economy reacts to a given tax policy given each value of A. We find it convenient to write out this reaction function explicitly. Using (5) for $t = 1$ to substitute out consumption in (6), we get

$$\frac{v'(l_1)}{u'(\theta l_1 - g_1)} = \theta \gamma (1 - \tau_1), \tag{11}$$

Letting h be the function that gives the l_1 that solves this equation for given τ_1, θ, γ we can rewrite the above equilibrium condition as

$$l = h(\tau, \theta, \gamma) \tag{12}$$

[5]This differs from the situation in the New Dynamic Public Finance, where consumers optimize given the policy function $\mathcal{R}$ which is a function of individual choices.

where we have again suppressed the time subscript from first period labor and tax rate. This shows how the allocation reacts to a tax choice.

Now we write the equilibrium utility function as a function of first period equilibrium allocations and shocks only. Using the resource constraints (5) to substitute out c_t in (10) gives one equation that, for each A, involves only the unknowns l_1, l_2. This defines implicitly a function that maps an equilibrium l_1 into a second period equilibrium labor l_2, call this map $L_2^{imp} : \Phi \times [0,1] \to R$,[6] so that

$$L_2^{imp}(A, l_1) \; for \; all \; A \in \Phi \tag{13}$$

solves (10). The welfare of the planner for each A can be written as

$$\begin{aligned} W(l; A) &\equiv U(\theta l - g_1, l, \theta_2 L_2^{imp}(l, A) - g_2, L_2^{imp}(l, A); \gamma) \\ &= U(c_1, l_1, c_2, l_2; \gamma) \end{aligned} \tag{14}$$

Note that the only arguments in W are the observed variable l and the shocks θ, γ. The functions U and L_2^{imp} are embedded in the definition of competitive equilibrium and known by the government.

Given the above discussion, the FI Ramsey Equilibrium reduces to solving

$$\begin{aligned} &\max_{\tau:\Phi\to R} E[W(l; A)] \\ &s.t.(12) \end{aligned} \tag{15}$$

Obviously the result is the same as to maximize (1) subject to (10) given A.

PI- Ramsey equilibria

We focus on the case when the signal is just labor so $s = l_1 = l$. The only difference under PI is that the additional constraint (9) appears and that the choice is over a tax contingent on the signal. Hence a **PI-Ramsey equilibrium** (given a signal l) solves

$$\begin{aligned} &\max_{\mathcal{R}:R\to R} E[W(l; A)] \\ &s.t.(12) \\ &\tau = \mathcal{R}(l) \end{aligned} \tag{16}$$

This gives rise to a non-standard maximization problem. We solve this problem in Section 4.

3.4 The economic consequences of PI for taxation policy

Before giving a mathematical solution it is worthwhile discussing the economic issues raised by limited information in the fiscal policy example we use. As is well known

[6] Under some specific assumptions on u and v, it will be possible to solve for l_2 as a function of l_1 in closed form. In general, the marginal effect of l_1 on l_2 is easily found by applying the implicit function theorem to (10). See our examples below.

the optimal FI policy is one of tax smoothing over time as the government wants to spread the distortions equally in the two periods. In the case of CRRA preferences

$$u(c) = \frac{c^{1+\alpha_c}}{1+\alpha_c},\ v(l) = B\frac{l^{1+\alpha_l}}{1+\alpha_l}$$

for $\alpha_c \leq 0$, $\alpha_l, B > 0$, tax smoothing will be perfect and Ramsey policy under FI involves setting a constant tax rate $\tau = \tau_1 = \tau_2$ that solves the intertemporal budget constraint

$$\tau\theta l_1 - g_1 + \beta\frac{u'(c_2)}{\gamma u'(c_1)}(\tau\theta_2 l_2 - g_2) = 0. \tag{17}$$

It is clear from (17) that the government needs to know the realization of both productivity and demand shock in order to implement this policy under FI. In particular, the realization of $\theta = \theta_1$ is a crucial piece of information, as it determines the revenue that a given tax rate, together with an observed level of hours worked, is going to raise. The demand shock γ also matters as it affects both the objective function and the interest rate that the government will have to pay on its debt. Furthermore, both shocks clearly contribute to the determination of an allocation (c_1, c_2, l_1, l_2).

Under PI the government can only condition its policy on l, without knowing what combination of the shocks gives rise to a given observation. Depending on the realizations, the government would like to set different tax rates and under some preference assumptions (e.g. log-quadratic, discussed in Section 5) it may even be the case that a certain increase in hours would call for a tax cut if driven by a high realization of γ, but it would call for a tax hike if driven instead by low θ. Since the government does not observe these shocks, this model gives a framework where private information matters.

Clearly, under PI the choice of constant taxes is not feasible. The government has to fix τ_1 while it is still uncertain about the revenue that this tax rate will generate and it will enter period 2 with an uncertain amount of debt. Once θ and γ are known in period 2, the government will have to set τ_2 so as to balance the budget in the second period in order to avoid default, so to the government τ_2 is unavoidably a random variable at the time of choosing τ_1.

Arguably, uncertain tax revenue is a crucial feature of actual fiscal policy decision, and tax rates are decided based on information from equilibrium outcomes that are observed frequently. In this sense, one can interpret this model as a simple model of optimal automatic stabilizers, as these are fiscal instruments that are designed to respond to endogenous outcomes, such as income or unemployment, independently of the source of fluctuations in these variables. The optimal design of these instruments requires a simultaneous determination of the density of taxable income and the policy. The next section studies a generic problem that allows the determination of taxes under limited endogenous information.

4 Optimal Control under General Signal Extraction (GSE)

This section (together with Appendix A) provides a formal treatment of Optimal Control under General Signal Extraction (GSE), including mathematical results concerning existence of a solution, a first order optimality condition that can be used to compute the equilibrium easily and second order optimality conditions.

Maximization problems such as (16) can be characterized as "Optimal Control under GSE". The key difficulty in (16) is that the policy choice $\mathcal{R}$ affects both the policy action τ as well as the distribution of the signal $s = l$. The signal provides information about the two unobserved exogenous shocks (γ, θ), so that the optimal policy depends on the conditional density $f_{(\gamma,\theta)|l}$, but this density depends itself on the tax policy $\mathcal{R}$. Therefore the choice of an optimal $\mathcal{R}$ must be consistent with the implied density $f_{(\gamma,\theta)|l}$.

Many problems in economics have this form. It can be thought of as a Stackelberg-reaction-function game where the "leader" (in Section 3, the government) chooses its policy (or reaction) function $\mathcal{R}$ optimally given the reaction of the "follower" h, when h is given and independent of the choice of $\mathcal{R}$. Unlike standard Stackelberg games the actions τ and s are determined jointly in this setup, there is a hierarchy only in the way leader and follower choose the reaction functions $\mathcal{R}, h$. Simultaneity is the standard assumption in Ramsey equilibria, where the equilibrium allocations are influenced only by the policy action τ, not by the whole policy function. This is a natural assumption when, as is standard in Ramsey taxation, followers are atomistic and the signal is an aggregate variable.[7]

Issues of GSE are pervasive whenever Partial Information is introduced in most currently used models. Notice that for the endogeneity of the signal to be an important issue, it is not necessary that the signal and the policy are realized in the same period: GSE arises even if tax rates are decided before the period they are actually levied (as it is arguably the case in practice). What matters is that taxes and the signal are jointly decided upon, even though they may be realized in different periods. To be specific, consider the following modifications of our fiscal policy example: first, assume now that at $t = 1$ the government makes an announcement about τ_2 contingent on hours in period 1, l_1; second, assume that there is a state variable that is determined at $t = 1$ that influences output next period, for example there is productive capital. Then, l_1 will respond to the announcement of τ_2, this will influence capital accumulation in period 1 and tax revenue $\tau_2 l_2$. In this situation since τ_2 is jointly determined with l_1 and they both influence tax revenue there is no separation between signal extraction and optimal policy. Therefore, under Rational Expectations issues of GSE arise generically even if tax rates are decided one period

[7]Simultaneity also occurs in the literature on supply function equilibria (Klemperer and Meyer, 1989). Here firms simultaneously choose a supply function.

in advance.

4.1 Optimization under GSE

We now present a generic problem of optimal control under GSE. This generalizes the PI- Ramsey Equilibrium Problem (16) without adding any difficulty to the proof. The generalization may be useful in other applications.

Consider a planner/government that chooses a policy variable $\tau \in \mathcal{T} \subset R$, observes endogenous signal $s \in \mathcal{S} \subset R$ at the time of choosing τ, when random variables $A \in \Phi \subset R^k$ have a given distribution F_A.[8]

The planner's objective is to maximize $E\left[W\left(\tau, s, A\right)\right]$ for a payoff function W. This nests the case when other endogenous variables enter the objective function, as these can be embedded in W. We did this in Section 3 through the use of the L_2^{imp} function defined in (13) to substitute out l_2 in the objective function.

The planner knows that a value for the policy variable τ maps into endogenous signals through the following equation

$$s = h(\tau, A) \quad a.s. \tag{18}$$

where a.s. statement is in the distribution of A.

The government knows W, h, F_A, $\mathcal{T}$, $\mathcal{S}$. It has to choose τ given an observation on s but it does not observe the value of A.

Optimal behavior under uncertainty implies that the government chooses a policy contingent on the observed variable s. Therefore, the government's problem is to choose a policy function $\mathcal{R} : \mathcal{S} \to \mathcal{T}$ setting policy actions equal to

$$\tau = \mathcal{R}(s). \tag{19}$$

To summarize, we wish to solve the following model of **Optimal Control with General Signal Extraction:**

$$\begin{aligned} &\max_{\{\mathcal{R}:\mathcal{S}\to\mathcal{T}\}} E\left[W\left(\tau, s, A\right)\right] \\ &s.t. \qquad (18), (19) \end{aligned} \tag{20}$$

We will derive optimality conditions that are non-standard and that can be easily used to provide a solution for the policy function $\mathcal{R}$. To see why a standard first order condition does not apply, we can rewrite the objective function as follows

$$\int E\left[W\left(\tau, s, A\right)\middle| s\right] f_s(s) ds \tag{21}$$

[8]It is possible to generalize the problem to the case of multidimensional policy instruments and signals. However, notation becomes cumbersome, hence we only refer to the univariate case in the following.

Taking derivatives with respect to τ and *if* $f_s(s)$ *could be taken as given* we would find the following optimality condition

$$E\left[W_\tau + W_s h_\tau | s\right] = 0 \quad for\ all\ s \tag{22}$$

This would be found by following that standard "recipe" for discrete-time first order conditions in dynamic stochastic models, namely, take the first order condition of the objective function under certainty, then condition on the information available at the time of the decision.

In most applications in dynamic models this would be correct, but it is not the correct FOC in our case because, in general, $\mathcal{R}$ determines the density f_s. To see this, notice that since s is determined implicitly by

$$s = h(\mathcal{R}(s), A),$$

hence $f_{s|A}$ depends on $\mathcal{R}$. Therefore $f_s = \int_A f_{s|A}\ f_A$ is also endogenous to $\mathcal{R}$. The derivative of f_s with respect to $\mathcal{R}$ should be taken into account in deriving optimality conditions, as we do below.

Let $S(\mathcal{R}, A)$ be the observable value of s induced by the shock A and a policy $\mathcal{R}$. Formally, $S(\mathcal{R}, A)$ is defined as follows: define H as

$$H(s, A; \mathcal{R}) \equiv s - h(\mathcal{R}(s), A), \tag{23}$$

then $S(\mathcal{R}, A)$ satisfies $H(S(\mathcal{R}, A), A; \mathcal{R}) = 0$ for all A given $\mathcal{R}$.

The policy variable that is realized for each value of the shocks A and for a given policy function $\mathcal{R}$ is then given by

$$T(\mathcal{R}, A) = \mathcal{R}\left(S(\mathcal{R}, A)\right)$$

Notice the following distinction between the objects S, T and $\mathcal{R}$: the latter is a function of s while S and T are functionals mapping $\mathcal{R}$ and the realizations of the shocks into $\mathcal{S}$, $\mathcal{T}$.

4.1.1 Reformulating the problem: relation with Calculus of Variations

It is easy to see that problem (20) can be reformulated as

$$\max_{\mathcal{R}} \int_\Phi W\left(\mathcal{R}(\mathcal{S}(\mathcal{R}, A)), S(\mathcal{R}, A), A\right) dF_A, \tag{24}$$

This formulation will be useful for actually finding optimality conditions and for comparison with the literature on calculus of variations.

"The basic problem in the subject that is referred to as *calculus of variations* consists of minimizing an integral functional of the type

$$\max_x \int \Lambda(x(t), x'(t), t) dt. \tag{25}$$

over a class of functions x" (p. 287 Clarke, 2013). Optimality conditions for this problem were first formulated by Euler in 1744. They have been applied in various economic models including, for example, the neoclassical growth model in continuous time.

However, there are two main differences between our (24) and the standard case (25). First, the objective function in the standard case involves the derivative x', while in our case $\mathcal{R}'$ is not an argument of W. Second, the object to be chosen in our case (namely $\mathcal{R}$) is *not* a function of the variable of integration (in our case A), while in the standard case x is a function of the variable of integration t. While the first difference is just a simplification, the second difference means that we can not solve (24) by appealing to the the standard Euler equation and, to our knowledge, this has not treated in the literature. In other words, in our case one has to choose the measurability conditions relating observables s and the underlying uncertainty, while this is not an issue in standard problems. The focus in this section is to formulate the optimality condition for (24)

4.2 Assumptions and restrictions on the policy function for existence of an equilibrium

Here we state some assumptions on the fundamentals of the problem and some restrictions on the policy function $\mathcal{R}$. These will be needed to ensure that the problem is well defined and to state the optimality conditions that are the focus of this section. The "Assumptions" can be readily checked from the fundamentals of the problem. We will comment on the role of each "Conditions" on $\mathcal{R}$. In subsection 4.10 we show how to impose these conditions to guarantee that the optimality condition can be applied and in our applications we show that these additional constraints are not binding. All the proofs can be found in Appendix A.

We maintain the assumption of univariate s and τ as this facilitates some of the proofs and the statement of conditions to ensure that the problem is well defined.

The following assumptions are on the fundamentals of the problem and are maintained in all the results below.

Assumption 1 *$W(\cdot, A)$ is continuous for all $A \in \Phi$ and $h(\cdot; A)$ is continuously differentiable for all $A \in \Phi$.*

Assumption 2 *The sets $\mathcal{T}, \mathcal{S}$ are bounded intervals, $\mathcal{T} = [\underline{\tau}, \overline{\tau}]\,, \mathcal{S} = [\underline{s}, \overline{s}]$, and Φ is compact.*

4.2.1 Conditions on the policy function $\mathcal{R}$

We add the following restriction to (20):

$$\mathcal{R} \in \mathcal{E}$$

for a given set $\mathcal{E} \subset \{functions\mathcal{R} : \mathcal{S} \to \mathcal{T}\}$. The set $\mathcal{E}$ will be a fundamental of the maximization problem.

We impose at the outset that the planner is constrained to choose policies that guarantee uniqueness. This can be justified by appealing to a principle that "good policy" the planner considers that an $\mathcal{R}$ giving rise to multiple equilibria is uniformly less preferable than any policy giving rise to a single equilibrium. Multiplicity of equilibria could be dealt with at the cost of having to introduce selection criteria or randomization in the model, we leave this for future research. So, we just impose

Condition 1 *$\mathcal{R}$ is such that for all values of $A \in \Phi$ the set $S(\mathcal{R}, A)$ is a singleton.*

In other words, there is a unique value of s that solves $H(s, A; \mathcal{R}) \equiv s - h(\mathcal{R}(s), A) = 0$. So we will require that all elements of $\mathcal{E}$ satisfy Condition 1.

Conditions 2-3 are needed to hold at the solution in order to obtain the optimality condition. Condition 2 will be naturally satisfied in the optimum given Assumption 1.

Condition 2 *$\mathcal{R}$ is continuous everywhere and differentiable almost everywhere in $\mathcal{S}$, where a.e. is with respect to the Lebesgue measure.*

Existence of a solution to $H(\cdot, A; \mathcal{R}) = 0$ will follow immediately under Assumption 1 for any $\mathcal{R}$ satisfying Condition 1: since $h(\mathcal{R}(\cdot), A)$ is continuous function mapping the compact set $\mathcal{S}$ into itself, Brower's fixed point theorem insures existence of equilibrium and that $S(\mathcal{R}, A)$ is non-empty for all A.

The system of equations $H(\cdot, A; \mathcal{R}) = 0$ is said to be well-conditioned if $\frac{\partial H}{\partial s} \neq 0$ at the solution or, equivalently, $h_\tau(\mathcal{R}(s), A)\mathcal{R}'(s) \neq 1$, for $s = S(\mathcal{R}, A)$. Well-conditioning given A is a slight requirement once a solution exists, but since we consider variations (or approximations) $\mathcal{R}+\alpha\delta$ for scalar α and functions δ, we need to guarantee that these variations $\mathcal{R} + \alpha\delta$ satisfy Condition 2 for small α. Therefore we need to guarantee that, $h_\tau(\mathcal{R}(s), A)\mathcal{R}'(s)$ is uniformly bounded away from 1 at the solution.[9] To ensure this we need the following

Condition 3 *$\mathcal{R}$ is such that the system of equations $s = h(\mathcal{R}(s), A)$ is* ***well conditioned uniformly*** *in A, namely, there is an $\varepsilon > 0$ and a $\eta > 0$ such that*

$$\left| \frac{h(\mathcal{R}(s), A) - h(\mathcal{R}(s'), A)}{s - s'} - 1 \right| > \epsilon,$$

for all $A \in \Phi$ and all $s, s' \in \mathcal{S}$ such that $|s - \mathcal{S}^*|, |s' - \mathcal{S}^*| < \eta$

Essentially this requires that H' is bounded away from 0 near the solutions for s. Since we allow for non-differentiabilities in $\mathcal{R}$, the restriction is imposed on the ratio $\frac{h(\mathcal{R}(s),A)-h(\mathcal{R}(s'),A)}{s-s'}$ instead of derivatives.

[9]The importance of this condition for our result is discussed in detail in Appendix A 8.3.1.

4.3 General formulation of the problem

Let $\mathcal{F}$ be the value of the objective function for a given $\mathcal{R}$.[10]

$$\mathcal{F}(\mathcal{R}) \equiv E\left[W\left(T\left(\mathcal{R}, A\right), S\left(\mathcal{R}; A\right), A\right)\right] \tag{26}$$

We can now re-define the Optimal Control with GSE problem as

$$\max_{\{\mathcal{R}:\mathcal{S}\to\mathcal{T}\}} \mathcal{F}(\mathcal{R}) \tag{27}$$
$$s.t. \mathcal{R} \in \mathcal{E}$$

and denote its solution by $\mathcal{R}^*$.

4.4 Apparent Partial Information: Invertibility

In some cases the government can still implement the FI policy even if it does not observe the shocks. This occurs whenever the information set of the government is invertible, allowing it to learn the true state of the economy A from observing the signal s.[11]

To formally define Invertibility, consider the set of all possible values (τ, s) in the FI case, namely

$$M^* \equiv \left\{(\tau, s) \in R^2 : \tau = \mathcal{R}^*_{FI}\left(A'\right) \ and \ s = h\left(\mathcal{R}^*_{FI}\left(A'\right), A'\right) \ for \ some \ A' \in \Phi\right\} \tag{28}$$

Let M^*_τ (M^*_y) denote the projection of M^* on $\mathcal{T}$ $(\mathcal{S})$

Definition 4 *Invertibility holds if for any $s \in M^*_s$ there exists a unique $\tau \in M^*_\tau$*

Clearly, Invertibility is satisfied if $h\left(\mathcal{R}^{FI*}\left(A\right), A\right) = s$ defines implicitly a unique value for A for all s.

Invertibility will often occur when the dimension of τ is the same as the dimension of A. Even if A is high-dimensional, Invertibility obtains if Φ is a finite set: in this case we can expect to be able to map an equilibrium into the shock since there are finitely many realizations, only by coincidence would the same equilibrium point (τ, l) occur for two different realizations of A.

[10] Notice that $\mathcal{F}$ maps the space of functions into R. The expectation operator integrates over realizations of A using the government's perceived distribution of A, so that the above objective function is mathematically well defined given the above definitions for $\mathcal{T}$, $\mathcal{S}$ and Assumptions 1-3.

[11] In the literature of optimal contracts under private information and incentive compatibility constraints this is the standard assumption, which amounts to assuming full revelation, but those papers concentrate on the difficulties raised by the reaction of the agents' to the choice of $\mathcal{R}$, this reaction is the reason that the FI solution is not reached in those papers.

Invertibility also holds in the supply function literature (Klemperer and Meyer, 1989) because uncertainty is onedimensional.

Proposition 1 *Under Invertibility* $\mathcal{R}^* = \mathcal{R}^{FI*}$

This follows from the fact that the PI case is a constrained FI problem, therefore the value in the PI case is less than or equal to the FI case, and the value of the FI case is achievable under Invertibility.

To illustrate a case of Invertibility, consider the example of Section 3, assume that $\gamma = 1$ with certainty. The government does not observe the random value of θ and it has to choose taxes observing l. This is only apparently a PI problem, because the government can infer θ from observing the labor choice, hence the government can implement the FI policy.

4.5 First order conditions for the general case of Partial Information

The case of interest in this paper arises when knowledge of (τ, s) is not sufficient to back out the actual realizations of the shocks from (18). Observe that Invertibility is violated if A has higher dimension than s and some element of A has a continuous distribution. In the rest of the paper we will focus on this case.[12] In terms of the fiscal policy example of the previous section, when the signal $s = l$ is observed, the values of θ, γ remain hidden even after the choice of τ has been made for a given observed labor l.

Without Invertibility, the solution (20) should take into account that the distribution is endogenous to policy while, at the same time, the policy depends on the optimal filtering of the fundamental shocks A given the observed signal s.

In order to proceed and derive optimality conditions, we now need to strengthen Assumption 1 as follows

Assumption 3 *W is differentiable everywhere with respect to τ and s. The derivatives $|W_\tau|, |W_s|, |h_\tau| < Q$ uniformly on (s, τ, A) for a constant $Q < \infty$ and h_τ is Lipschitz continuous for almost all A.*

The general first order condition (29) we derive below allows for a non-differentiable solution $\mathcal{R}^*$ but it needs to ensure that these non-differentiabilities happen with probability zero in equilibrium. Formally, consider the set of realizations that give rise to a signal s where $\mathcal{R}$ is non-differentiable.

Define $\Phi_{ND}(\mathcal{R}) \equiv \{A : \mathcal{R}\ is\ non-differentiable\ at\ s = S(\mathcal{R}, A)\}$. The proof of the general condition requires that $Prob(A \in \Phi_{ND}(\mathcal{R}^*)) = 0$.

That the solution $\mathcal{R}^*$ satisfies this condition can be guaranteed in various ways. In the case of Invertibility discussed in Part 1 of Corollary 1 below it can be checked that $\mathcal{R}^*$ is differentiable so the solution satisfies this condition. But we are interested

[12] Invertibility is also violated if $h\left(\mathcal{R}^{FI*}(\cdot), \cdot\right)$ is non-monotonic. We leave an exploration of this case for future work.

in the case where there is "true" PI so that the government's action can not reveal the shocks that occurred and $\mathcal{R}^*$ may be non-differentiable. For this purpose we make the following

Assumption 4 *Given any pair* $(s, \tau) \in \mathcal{S} \times \mathcal{T}$*, the set of realizations* $\{A : s = h(\tau, A)\}$ *has probability zero.*

This is easy to ensure, in general, but having at least one shock in A with a continuous density.

We can now state our main result, namely the first order optimality condition. Let $\mathcal{S}^*$ denote the support of $S(\mathcal{R}^*, A)$.

Proposition 2 *Assume Assumptions 1-4. Assume that a solution* $\mathcal{R}^*$ *exists and satisfies Conditions 1,2,3. Assume the solution is interior, more precisely* $\mathcal{S}^* \subset int(\mathcal{S})$*,* $\mathcal{S}^*$ *compact,* $\mathcal{R}^*(\mathcal{S}^*) \subset int(\mathcal{T})$ *and* $\mathcal{R}^* \in int(\mathcal{E})$*.*[13] *Then the solution* $\mathcal{R}^*$ *satisfies the following necessary first order condition*

$$E\left(\frac{W_\tau^* + W_s^* h_\tau^*}{1 - h_\tau^* \mathcal{R}^{*\prime}} \middle| S(\mathcal{R}^*, A) = s \right) = 0 \tag{29}$$

for all $s \in \mathcal{S}^*$ *where* $\mathcal{R}^{*\prime}$ *exists.*

It is understood tha $W_\tau^*, W_s^*, h_\tau^*$ denote W_τ, W_s, h_τ evaluated at s and $\tau = \mathcal{R}^*(s)$.

To find the optimality condition in Proposition 2 we use a variational argument similar to the one used often in calculus of variations. In the following paragraphs we show some of the key steps in the proof. For a detailed proof see Appendix A 8.3.2.

Take any function (a variation) $\delta : \mathcal{S} \to R$ and a constant $\alpha \in R$. Now consider reaction functions of the form $\mathcal{R}^* + \alpha\delta$. For a given δ consider solving the (one-dimensional) maximization problem

$$\max_{\alpha \in \Re} \mathcal{F}(\mathcal{R}^* + \alpha\delta) \tag{30}$$

so that now we maximize over small deviations of the optimal reaction function in the direction determined by δ. It is clear that

$$0 \in \arg\max_{\alpha \in \Re} \mathcal{F}(\mathcal{R}^* + \alpha\delta) \tag{31}$$

In Appendix A 8.3.2 we show that $\frac{\partial \mathcal{F}(\mathcal{R}^* + \alpha\delta)}{\partial \alpha}$ evaluated at $\alpha = 0$ is

$$E\left(\frac{W_\tau^* + W_s^* h_\tau^*}{1 - h_\tau^* \mathcal{R}^{*\prime}} \delta(S(\mathcal{R}^*, A)) \right) = 0. \tag{32}$$

[13]Interiority is with respect to the standard Eucledian distance for $\mathcal{S}$, $\mathcal{T}$ and with respect to the sup norm for $\mathcal{E}$.

The general characterization of conditional expectation as a function defining an orthogonal error implies (29).[14]

Notice that, as we anticipated, the first order condition (29) does not coincide with the standard FOC (22). The term $\frac{1}{1-h_\tau^* \mathcal{R}^{*\prime}}$ acts as a kernel, or measure change. This is the new term relative to the standard case when $f_{A|s}$ does not depend on $\mathcal{R}$. The term $\frac{1}{1-h_\tau^* \mathcal{R}^{*\prime}}$ captures the effect of the choice of $\mathcal{R}$ on the density f_s.

Furthermore, since under these Conditions $\mathcal{R}^*$ is continuous and differentiable a.e., we have that (29) characterizes the whole function $\mathcal{R}^*$.

4.6 A convenient optimality condition

Condition (29) is useful for comparability with the standard FOC and to point out the role of the kernel $\frac{1}{1-h_\tau R^{*\prime}}$, but it turns out to be less convenient for computations as it involves the derivative of the policy function $\mathcal{R}^{*\prime}$. An algorithm trying to approximate $\mathcal{R}^*$ numerically will have to ensure that not only $\mathcal{R}^*$ is well approximated but that its derivative is well approximated as well along the iterations. We now use an envelope condition to formulate the optimality condition where $\mathcal{R}^{*\prime}$ is not present.

Note that (29) conditions on s. If in addition we condition on some additional variables in A the remaining variables can be taken as given. More precisely, let us partition $A = (A_1, A_2)$ where $A_1 \in R$. Let $\Theta_2(s,\tau) = \{A_2 : s = h(\tau, A_1, A_2) \; for \; some \; (A_1, A_2) \in \Phi\}$ be the support of A_2's conditional on observing s, τ.

Condition 4 *Given $s{\in}\mathcal{S}^*$, for each $A_2 \in \Theta_2(s, \mathcal{R}^*(s))$ there is a unique value of A_1 that satisfies*

$$s = h(\mathcal{R}^*(s), A_1, A_2), \tag{33}$$

This defines an implicit function $A_1 = \mathcal{A}^*(s, A_2)$ that, given the information on s, maps a realization of A_2 into the corresponding A_1 for the optimal policy. Furthermore the partial derivative $h_{A_1}(\mathcal{R}^*(s), A)$ is bounded away from zero and either positive or negative for all $s{\in}\mathcal{S}^*$ and $A_2 \in \Theta_2(s, \mathcal{R}^*(s))$.

Obviously the following assumption is sufficient for Condition 6:

Assumption 5 *The partial derivative h_{A_1} exists, it is bounded away from zero and of the same sign for all (s, τ, A).*

This assumption is stronger than Condition 6 but can be checked without knowledge of the optimal policy.

In the following proposition we denote a random variable with upper case, say X with a density f_X, and we use lower case for the possible values, say x.

[14] More precisely, take random variables X, Y. We know the conditional expectation is a function $f(Y)$ with the property $E[(X - f(Y))g(Y)]$ for all g. Here $\frac{W_\tau^* + W_s^* h_\tau^*}{1-h_\tau^* \mathcal{R}^{*\prime}}$ plays the role of X, s plays the role of Y, 0 plays the role of $f(Y)$ and δ plays the role of g.

Proposition 3 *In addition to the assumptions and conditions needed for Proposition 2 assume Condition 4 and that A has a density. The optimality condition (29) is equivalent to*

$$\int_{\Theta_2(s,\mathcal{R}^*(s))} \frac{W_\tau^* + W_s^* h_\tau^*}{h_{A_1}^*} f_A(\mathcal{A}^*(s,a_2),a_2)da_2 = 0 \ for \ all \ s \in \mathcal{S}^* \tag{34}$$

Proof in Appendix A 8.4. As in Proposition 2 $\frac{W_\tau^*+W_s^*h_\tau^*}{h_{A_1}^*}$ is evaluated at equilibrium values, with the addition that we now take $A_1 = \mathcal{A}^*(s,a_2)$.

This can be used for computation of the optimal policy, since (34) gives one equation to solve for the unknown $\tau^* = \mathcal{R}^*(s)$. Given a candidate τ^* one can find $\Theta_2(s,\tau^*)$, evaluate all the functions involved in the integrand (as these functions are all known beforehand, τ^* only enters as an argument), compute the integral and if it is not zero iterate on τ^* until (34) holds. In particular, note that $\mathcal{R}^{*\prime}$ is not involved in this calculation.

In addition to simplifying computations this proposition honors the title of the paper. The proof highlights that $f_{A|S(\mathcal{R}^*,A)}$ depends on $\mathcal{R}^{*\prime}$ as it appears in the Jacobian in (71). Bayes' rule implies that the filter $f_{A_2|s}$ depends on the optimal policy $\mathcal{R}^*$.

4.7 Relation to previous results

With Proposition 2 in hand we can derive previous results available in the literature on optimal policy under PI as special cases. The following corollary shows some cases that have been considered in the literature and that gave rise to separation.

Corollary 1 *Assume for some $s \in \mathcal{S}^*$ one of the following hold*

1. *(Invertibility) there is a unique $A \in \Phi$ such that $S(\mathcal{R}^*,A) = s$ and $\mathcal{R}^*$ is differentiable.*
2. *(linearity of the reaction function) h_τ is a given constant for all $A \in \Phi$ such that $S(\mathcal{R}^*,A) = s$.*
3. *(exogenous signals) the welfare-relevant allocation is a function of τ, but the signal s is independent of τ, and given by $s = h(A)$, hence $h_\tau = 0$.*

Then the general FOC (29) reduces to the FOC (22).

The proof for Case 1 and 2 is trivial: in both cases $\frac{1}{1-h_\tau^*\mathcal{R}^{*\prime}}$ is known given s so that this terms goes out of the conditional expectation in (29) and it cancels out. Case 1 obtains a generalization of the Invertibility theorem we found before.

Case 2 highlights that for the kernel to be relevant, it has to be the case that it assumes different values for different possible realizations of the shocks, conditional on the signal. Hence, in models with a linear reaction function, where h_τ is a constant, the FOC simplifies to the standard one (22).

To discuss Case 3, it is helpful to distinguish two different roles played by the function h in our setup. First, h maps the shocks into a part of the allocation that enters the objective function W (in our fiscal policy example, hours). Second, it is what the PI literature calls the "measurement equation", mapping the shocks (and the policy) into an observed signal. In the case of exogenous signals, the signal is independent of the (endogenous) allocation, and only depends on exogenous shocks, including potentially a "noise shock".

To make a concrete example related to our fiscal policy application, a case of exogenous signals would be as follows: $l = f(\tau, A)$ is implicitly defined by equation (11); the signal is given by $s = h(A)$. For instance, as in the Kalman filter literature, we could have $s = \theta + \psi$, where ψ is noise and the state is $A = (\theta, \gamma, \psi)$. Then, the problem would be to choose a function $\mathcal{R}$ such that $\tau = \mathcal{R}(s)$ to maximize $EW(l)$ subject to $l = f(\tau, A)$. The associated first order condition (29) would simplify to

$$E\left[W_l f_\tau | s\right] = 0$$

as there would be no kernel related to the endogeneity of the distribution of the signal.

We now show that in the case of quadratic objective function and linear constraints, we can recover the result of "certainty equivalence" (Svensson and Woodford, 2004) as a special case of the FOC (29). This is stated in the following corollary.

Corollary 2 *If W is a quadratic function of τ and s and h is linear, then optimal policy has the "certainty equivalence" property, that is, optimal policy under PI calls for applying the FI policy function to the conditional expectation of the shocks.*

For simplicity of notation, consider the case where the objective function involves only quadratic terms and the linear terms are zero:

$$W(\tau, l) = -\frac{\omega_\tau}{2}\tau^2 - \frac{\omega_l l^2}{2} \tag{35}$$

and the reaction function is

$$l = h + h_\tau \tau + h_\theta \theta + h_\gamma \gamma \tag{36}$$

where the ω's are positive coefficients and the h's are constants.

Using Svensson and Woodford's (2004) notation, let $X \equiv (1, \theta, \gamma)'$. Then the FI optimal policy is

$$\tau = FX$$

where

$$F = \left(-\frac{\omega_l h_\tau h}{\omega_\tau + \omega_l h_\tau^2}, -\frac{\omega_l h_\tau h_\theta}{\omega_\tau + \omega_l h_\tau^2}, -\frac{\omega_l h_\tau h_\gamma}{\omega_\tau + \omega_l h_\tau^2}\right).$$

The FOC under PI (29) becomes

$$E\left[\omega_{\tau}\tau+\omega_{l}h_{\tau}\left(h+h_{\tau}\tau+h_{\theta}\theta+h_{\gamma}\gamma\right)|l\right]=0 \tag{37}$$

which can be rewritten as

$$\left(\omega_{\tau}+\omega_{l}h_{\tau}^{2}\right)\tau\omega_{l}h_{\tau}h+\omega_{l}h_{\tau}h_{\theta}E\left[\theta|l\right]+\omega_{l}h_{\tau}h_{\gamma}E\left[\gamma|l\right]=0 \tag{38}$$

which implies that optimal policy is given by

$$\tau=FE\left[X|l\right] \tag{39}$$

where F is the same vector of coefficients found under FI, that is, independently of the information structure and distribution of the shocks. This property of optimal policy with PI in linear models is called "certainty equivalence", as the government forms the best estimate of the state and behaves as if this estimate was certainty, or FI.

However, as in Svensson and Woodford (2004), notice that the "separation principle" does not hold, because one cannot compute the expectation of the state $E\left[X|l\right]$ without knowledge of the policy. To see this, consider for example $E\left[\theta|l\right]$, which can be rewritten as

$$E\left[\theta|l\right]=E\left[X|h+h_{\tau}FE\left[X|l\right]+h_{\theta}\theta+h_{\gamma}\gamma=l\right] \tag{40}$$

This shows that optimal policy (39) and signal extraction (40) have to be solved jointly as a system.

4.8 Second order condition

The following second condition further restricts the solution. It is stated for the case where W does not depend on τ, the case when τ is an argument can be treated similarly to obtain a more involved expression.

Proposition 4 *In addition to the assumptions and conditions of Proposition 2, assume second derivatives W_{ss} and $h_{\tau\tau}$ exist for all A and that these are uniformly bounded. Then for all $s\in\mathcal{S}^{*}$ where $\mathcal{R}^{*}$ is differentiable*

$$\mathbb{E}\left\{W_{ss}^{*}\left(\frac{h_{\tau}^{*}}{1-h_{\tau}^{*}\mathcal{R}^{*\prime}}\right)^{2}+W_{s}^{*}\frac{h_{\tau\tau}^{*}(1-h_{\tau}^{*}\mathcal{R}^{*\prime})+h_{\tau}^{*}h_{\tau\tau}^{*}+h_{\tau}^{*^{3}}\mathcal{R}^{*\prime\prime}}{(1-h_{\tau}^{*}\mathcal{R}^{*\prime})^{3}}\middle|s\right\}<0. \tag{41}$$

The proof is in Appendix A.

It should be noted that this condition is necessary and holds at all points of differentiability of $\mathcal{R}^{*}$.

4.9 Algorithm

Condition (34) gives the following algorithm to calculate the PI solution. Fix a value for s. We must be able to compute $\mathcal{A}^*(s, A_2)$ for a given candidate of $\tau = \mathcal{R}^*(s)$. Then at a possible candidate τ, we can evaluate the integrand in (34) for each possible A_2 since we have the corresponding $A_1 = \mathcal{A}^*(s, A_2)$. We compute the integral by running A_2 from the lowest to highest possible value of A_2 given the candidate τ. Note that these limits to the possible values of A_2, defining $\Theta_2(s, \mathcal{R}^*)$, are endogenous to τ and they have to keep A_1 within the admissible limits of the support of A_1. This operation maps a value of a candidate τ to the left side of (34), the optimal $\tau = \mathcal{R}^*(s)$ is found by solving this non-linear equation, making this integral as close as possible to zero.

4.10 Existence of a maximum

The necessary conditions that we found apply if a maximum exists. It is well known that such conditions may be misleading as for some problems a maximum $\mathcal{R}^*$ may not exist. Here we show how the set $\mathcal{E}$ can be restricted to insure existence of a maximum satisfying Conditions 1-3.[15] The optimality conditions will then apply to interior solutions.

Lemma 1 *Assume $\mathcal{R}$ satisfies Condition 2. Then $\mathcal{R}$ satisfies*

$$\left| \frac{h(\mathcal{R}(s), A) - h(\mathcal{R}(s'), A)}{s - s'} - 1 \right| > \epsilon, \tag{42}$$

for all $A \in \Phi$ and all $s, s' \in \mathcal{S}$ if and only if

$$\frac{h(\mathcal{R}(s), A) - h(\mathcal{R}(s'), A)}{s - s'} \leq 1 - \varepsilon \tag{43}$$

for all $A \in \Phi$ and all $s, s' \in \mathcal{S}$.

Proof in Appendix A 8.1. Notice that (42) is a strong form of Condition 3, where the boundedness of the derivative is required to hold for all values of s, s'. The lemma says that for this strong condition to hold it is enough to check (43). Using this result we can impose the latter inequality to ensure that Condition 3 is satisfied.

Define the bound $\underline{L}$ as

$$\underline{L} = (1 - \varepsilon) / \min_{(\tau, A) \in \mathcal{T} \times e308} h_\tau(\tau, A))$$

for some $1 > \varepsilon > 0$. Let $\mathbf{L}$ be a large positive constant.

Define the set

$$\mathcal{E} = \left\{ \mathcal{R} : \ for\ all\ x, y \in \mathcal{S}\ \tfrac{\mathcal{R}(x) - \mathcal{R}(y)}{x - y} \in [\underline{L}, \mathbf{L}] \ \right\} \tag{44}$$

[15] We leave for future work to research if Conditions 1 to 4 can be shown to hold for the "unrestricted" optimal solution $\mathcal{R}^*$.

Lemma 2 *If $\mathcal{E}$ is defined as in (44) then $\mathcal{R} \in \mathcal{E}$ satisfies Conditions 1-3 and $S(\mathcal{R}, A)$ is a singleton. Furthermore, $\mathcal{E}$ is a compact set in the sup norm.*

Proof in Appendix A 8.2. Although (44) is not the most general formulation of $\mathcal{E}$ needed for a well defined problem, this assumption is very general and it is easy to check given a candidate solution to the optimality conditions. The solutions that we find are well in the interior of $\mathcal{E}$, since they are not extremely steep anywhere and the equilibria are clearly uniformly well-conditioned. Therefore it is possible to check that the constraint $\mathcal{R} \in \mathcal{E}$ is not binding at the solutions we find and this is sufficient to guarantee that the optimality conditions above give the maximum.

Proposition 5 *The problem (20) subject to $\mathcal{R} \in \mathcal{E}$ for (44) has a solution $\mathcal{R}^*$ attaining the maximum.*

The proof follows from compactness of $\mathcal{E}$, continuity of the objective function and boundedness of $\mathcal{T}$, $\mathcal{S}$.

5 Solution of fiscal policy example with GSE

We now illustrate our computations of optimal policy in the model introduced in Section 3. In this section we restrict our analysis to the case of transitory shocks (hitting only in period 1) and observed labor. In Appendix D we also consider permanent shocks to productivity and observable output. In Section 6 we extend the analysis to an infinite-horizon model with shocks that hit in every period.

5.1 Low g and optimal tax smoothing

For simplicity we consider preferences that allow for an analytical reaction function h. Assume $u(c) = \log(c)$ and $v(l) = \frac{B}{2}l^2$. We show numerical examples with the following parameter values and distributional assumptions: let θ be uniformly distributed on a support $[\theta_{min}, \theta_{max}]$, γ uniformly distributed on $[\gamma_{min}, \gamma_{max}]$ and assume $\beta = .96$, B and the mean of θ are set to normalize average output to 1 and average hours to a third. Government expenditure is constant and equal to 25% of average output. The mean of γ is 1. The supports of both shocks imply a range of $\pm 10\%$ from the mean.

We first present the FI solution in order to illustrate the optimal response of taxes and allocations to the two different shocks we consider. Notice that the equilibrium condition (6) becomes

$$Bl_1 c_1 = \gamma\theta_1(1 - \tau_1) \tag{45}$$

and after substituting out consumption using the resource constraint, we obtain that labor supply is the positive root of a quadratic equation, so that the reaction function (12) specializes to

$$l = h(\tau, \theta, \gamma) = \frac{Bg_1\theta^{-1} + \sqrt{(Bg_1)^2\theta^{-2} + 4B\gamma(1 - \tau)}}{2B}. \tag{46}$$

While it is clear that hours are unambiguously increasing in the demand shock γ, it is important to note that the productivity shock θ has two opposing effects on l: the substitution effect between leisure and consumption and the wealth effect, that acts in the opposite direction. The extent to which hours are an informative signal about productivity depend crucially on the level of government spending, through its effect on marginal utility. If $g = 0$ then substitution and wealth effect exactly offset each other and hours are independent of θ. In presence of positive government spending, the wealth effect dominates the substitution effect and hence high realizations of θ will lead to low labor, ceteris paribus. The higher g, the higher the marginal utility of consumption, the stronger the wealth effect on labor supply, and the more informative hours become about productivity.

Furthermore, by setting the tax rate, the government can also affect the informativeness of signals. We will illustrate this interesting non-linear effect of g on the signal extraction by discussing both a case with low g (in this subsection) and a case with high g (in the following subsection).

In Figure 1 we show how hours and taxes move with the two different shocks under FI. On the left side of the figure, we keep γ constant and equal to its mean and we show that both hours and taxes are decreasing in the productivity shock. On the right side, we keep θ constant and equal to its mean and show that labor is increasing in γ, while taxes are decreasing. This shows that when we introduce PI with only hours being observed, if the government sees an increase in l, it would want to react in opposite directions depending on the source of the shock: this would call for a tax increase, if driven by low θ, or a tax cut if driven by high γ. Hence this model is particularly interesting to analyze optimal policy with endogenous PI since by observing a certain value l and imposing a tax rate τ, the government cannot infer the value of the shocks.

Under PI, for an intermediate value of l, the government is uncertain whether say both θ and γ are high, or vice versa, and in general there is a continuum of realizations (θ, γ) consistent with the observation of l and a policy $\mathcal{R}$. Therefore it cannot choose the policy under FI (constant taxes) since the realizations of γ and θ enter separately in (17).

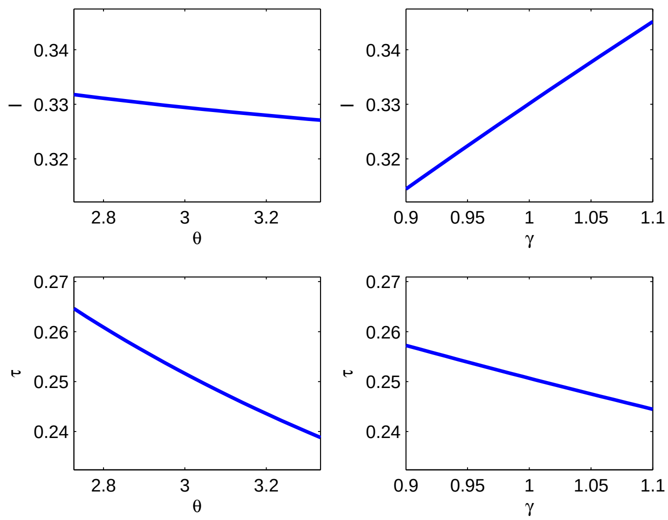
The partial derivatives h_τ and h_γ are easily obtained analytically. In particular

$$h_\tau(\tau, \theta, \gamma) = \frac{-1}{\sqrt{(Bg_1)^2 \left(\theta\gamma\right)^{-2} + \gamma^{-1} 4B(1-\tau)}}. \tag{47}$$

It is clear that both the productivity shock and the demand shock affect this slope, therefore endogenous signal extraction is an issue. Hence we proceed to find a $\mathcal{R}^*$ that satisfies (34) using the algorithm described in Subsection 4.9

Figure 2 illustrates the optimal policy for this case, plotting the tax rate against observed labor. The red line is $\mathcal{R}^*$, while the yellow region is the set of all equilibrium pairs (l^{FI}, τ^{FI}) that could have been realized under FI.

Figure 1: Hours and taxes with Full Information



Left panels: top, hours under FI as a function of the productivity shock; bottom, tax rate under FI as a function of the productivity shock. Right panels: top, hours under FI as a function of the demand shock; bottom, tax rate under FI as a function of the demand shock.

The limits of l are easy to find ex-ante by exploiting the fact that the extreme values for l coincide with the FI case. As l is increasing in γ and decreasing in θ, letting $l_{\min}$ and $l_{\max}$ be the extreme values of l in the PI solution, we have

$$\begin{aligned} l_{\min} &= L(\mathcal{R}^*; \theta_{max}, \gamma_{min}) = L^{FI}(\theta_{max}, \gamma_{min}) \\ l_{\max} &= L(\mathcal{R}^*; \theta_{\min}, \gamma_{\max}) = L^{FI}(\theta_{\min}, \gamma_{\max}) \end{aligned}$$

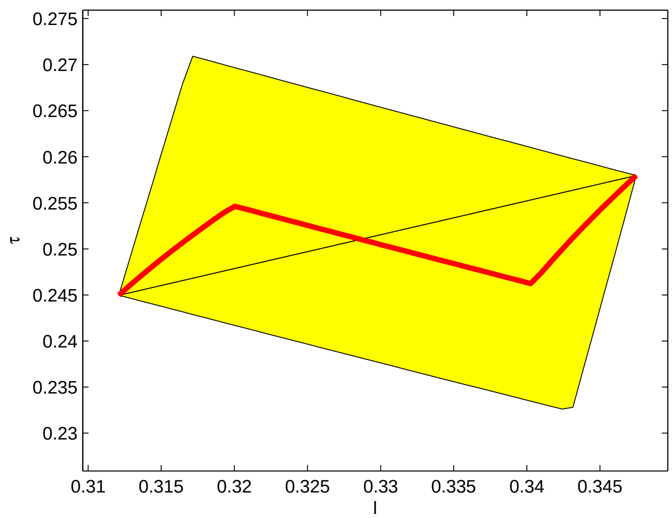
and the PI solution is in the interval $[l_{\min}, l_{\max}]$.

For these extremes values of the signal, there is full revelation, but anywhere between these two extremes the government has to choose a policy without knowing the values of γ, θ that give rise to equilibrium taxes or labor. It can be seen that the optimal policy calls for a tax rate in between the minimum and the maximum FI policies for each observation.

For low labor, the government learns that productivity must be high, so the tax rate can be rather low. The lowest labor realization leads to the FI equilibrium for $(\theta_{max}, \gamma_{min})$. Then taxes start to increase: higher l's signal lower expected productivity and hence revenue, as the set of admissible θ's is gradually including lower and lower realizations. This goes on up to a point where the set of admissible θ's conditional on l is the whole set $[\theta_{min}, \theta_{max}]$. From that point on, the tax rate changes slope and becomes decreasing with respect to l. This is because now, with any θ

being possible, increasing l signals an increasing expected revenue, hence allowing lower tax rates on average, up to the point where the highest θ's start being ruled out, at which point the policy becomes increasing again, up the full revelation point $l_{\max} = L^{FI}(\theta_{min}, \gamma_{max})$.

Figure 2: Optimal policy with low g

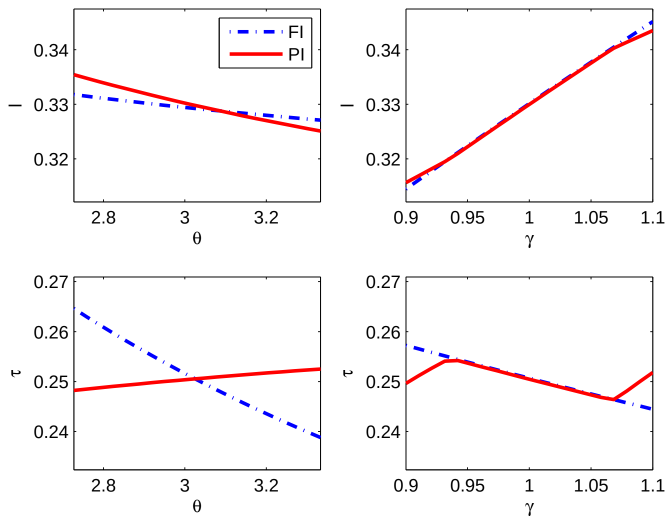


Optimal tax rate as function of hours under PI. Red line: $\mathcal{R}^*$; yellow region: set of FI pairs (l, τ) for all possible realizations of (θ, γ); black line: linear policy connecting the two full revelation points.

To gain further understanding on the implications of PI for the properties of the model, we plot again hours and taxes as functions of each shock individually in Figure 3. In all four panels, we reproduce the FI outcomes shown in Figure (FI) (blue dashed-dotted lines). The red lines represent the PI outcomes. For instance in the left panels we keep γ equal to its mean and we plot hours and taxes as functions of θ. Of course the government does not observe the values of θ and γ, but only hours. Interestingly, it can be seen that hours become more volatile in response to productivity shocks under PI, while taxes become smoother and change the sign of their response to θ. This is because under this parametrization the government learns little about the realizations of θ and hence optimally chooses to cut taxes as hours increase.[16] On the right-hand side, we plot again hours and taxes as functions on γ, keeping θ equal to the mean. For intermediate values of γ, the government is relatively confident about the realization of the demand shock, hence the policies under FI and PI are very close. However for extreme realizations the government is uncertain about which shock is

[16]We will see in the next subsection that this property of the solution will change with higher government expenditure.

Figure 3: Hours and taxes with Partial Information and Full Information



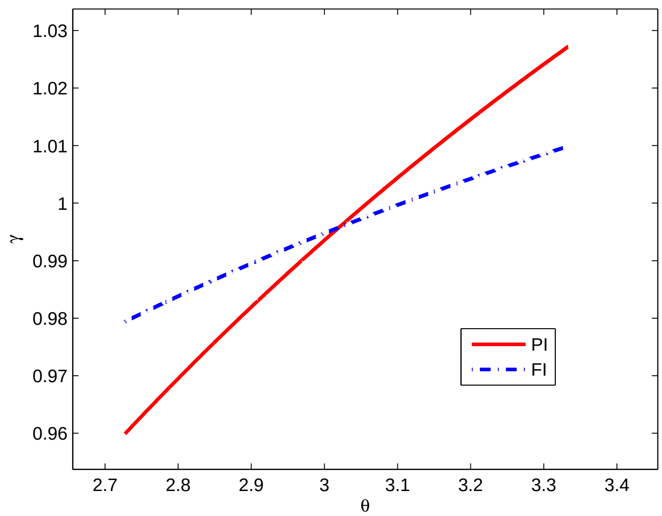
Left panels: top, hours as a function of the productivity shock (solid red line for PI, blue dashed-dotted line for FI); bottom, tax rate as a function of the productivity shock. Right panels: top, hours as a function of the demand shock; bottom, tax rate as a function of the demand shock.

driving hours, hence it cuts taxes for very low γ's and increases taxes for very high γ's, believing that changes in productivity are responsible for the observed behavior of hours.

We also plot the locus of admissible realizations of shocks for $l = .33$ in Figure 4. The wealth effect of productivity makes it an increasing function in the (θ, γ) space. As hours are increading in γ and decreasing in θ, a given level of hours could be due to combinations of high demand and high productivity, or low demand and low productivity. It should be noted that the locus of shocks realizations conditional on l is endogenous to policy. Importantly, τ affects the slope of this locus, implying that the government can choose to some extent on what shocks the signal extraction will be more precise. To see this, observe that a horizontal locus would imply revelation of the value of γ, while a vertical locus would imply revelation of the value of θ.

Optimal policy with PI calls for a substantial smoothing of taxes across states. This can be seen in Figure 5, where the equilibrium cumulative distribution function of tax rates under PI (red line) is contrasted with the one obtained under FI (blue dotted line). This result is rather intuitive and it carries a general lesson for optimal fiscal policy decisions under uncertainty: when the government is not sure about what type of disturbance is hitting the economy, it seems sensible to choose a policy that is not too aggressive in any direction and just aims at keeping the budget under control

Figure 4: Set of admissible shocks consistent with $l = .33$



Set of combinations of (θ, γ) that have positive density in equilibrium, conditional on a particular realization of the signal, namely $l = .33$. Solid red line for PI, blue dashed-dotted line for FI.

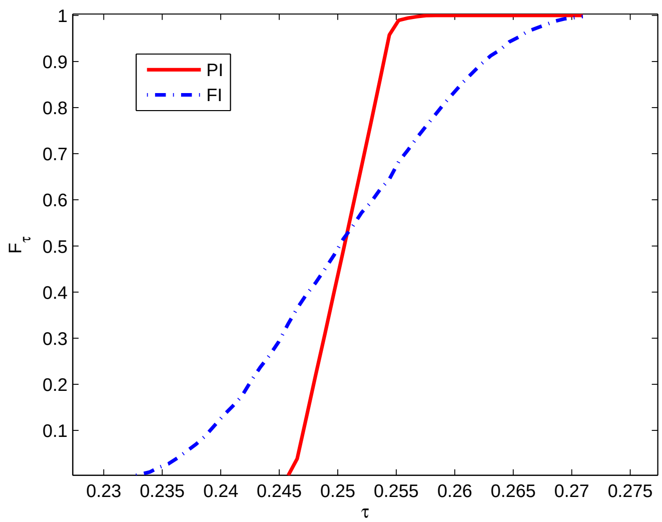
on average.

In our model, this smoothing of taxes across states will imply a larger variance of tax rates in the second period with respect to the FI policy. In the second period, all the uncertainty is resolved and the tax rate will be whatever is needed to balance the budget constraint. This is of course taken into account at the time of choosing a policy under uncertainty, so that we could say that optimal policy is very prudent while the source of the observed aggregate variables is not known and then responsive after uncertainty has been resolved.

This result is related to the question on whether taxes should be smooth across states or over time, depending on the completeness or incompleteness of financial markets. With complete markets and FI, tax smoothing happens across states (Lucas and Stokey, 1983). When markets are incomplete, the FI government substitutes tax smoothing across states with tax smoothing over time (Aiyagari et al., 2002). In our model, with incomplete markets and PI, we find that taxes are smoother across states than over time. This suggests that tax smoothing across states may not necessarily be an indication of market completeness and full insurance on the part of the government, but simply a sign of incomplete information about the state of the economy.

Because of this property, our model can rationalize the slow reaction of some governments to big shocks like the Great Recession. The Spanish example in the latest recession is a case in point. In 2008, it was far from clear how persistent the

Figure 5: Equilibrium CDF of tax rates



Cumulative Distribution Function of τ_1 under PI (solid red line) and FI (blue dashed-dotted line).

downturn would be and also whether is was demand-driven or productivity-driven and the government did not adjust its fiscal stance quickly, only to make large adjustments in the subsequent years. We will discuss this further in the paper.

5.2 Close to the top of the Laffer curve

Let us now look at the case where government expenditure is very high, equal to 60% of average output in both periods.[17] We will see that this leads to a very non-linear optimal policy and to an exception to tax-smoothing across states. This example is of interest for several reasons. From an economic point of view, Partial Information is of higher importance here: since the government needs to balance the budget in the second period it is thus now very concerned about the possibility of a very low level of productivity θ, as in this case tax revenue is low in the first period and a large amount of debt will need to be issued in this case. A high debt, combined with high future expenditure, may call for very high taxes in the future, it could even mean getting the economy closer to the top of the Laffer curve, where taxation is most distortionary and hence consumption is very low.

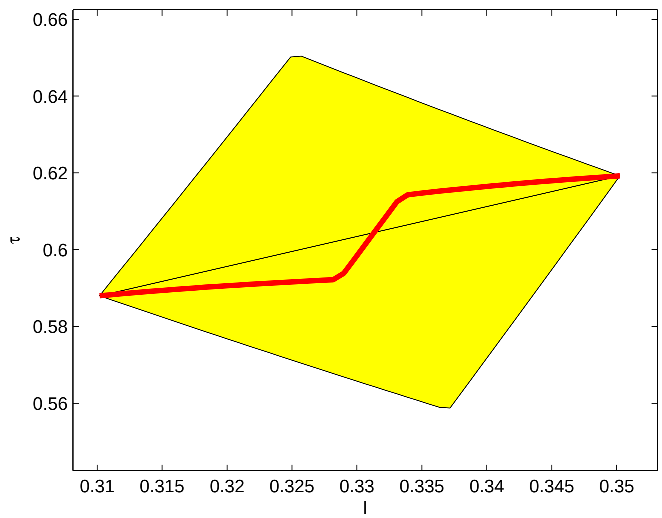
This example will also be of interest because the PI solution has some very different features from the FI outcome. High government spending makes the wealth effect of a productivity shock larger, with the consequence that hours worked become a stronger

[17]All other assumptions on preferences and shocks are the same as in the previous section.

signal about θ. The government will optimally exploit this in the signal extraction.

Figure 6 shows optimal policy for this case (red line), again contrasted with the set of tax-labor outcomes under FI (yellow region).

Figure 6: Optimal policy with high g

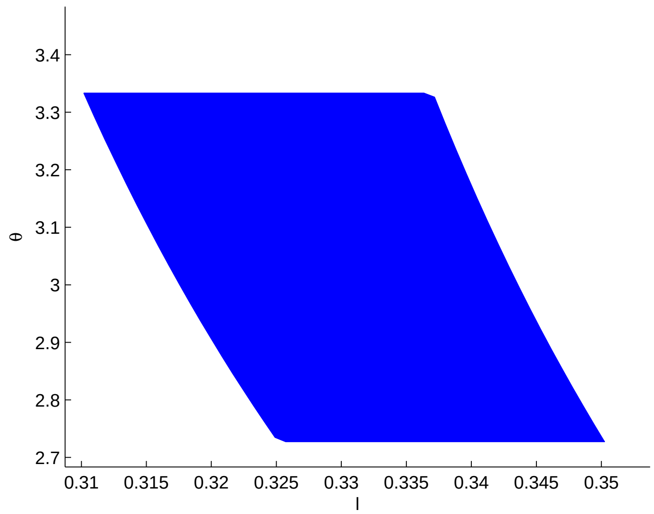


Optimal tax rate as function of hours under PI. Red line: $\mathcal{R}^*$; yellow region: set of FI pairs (l, τ) for all possible realizations of (θ, γ); black line: linear policy connecting the two full revelation points.

The figure shows that the optimal solution is highly non-linear. The derivative $\mathcal{R}^{*\prime}$ is positive and relatively high in a middle range of levels of l, but both to the left and to the right of this middle range $\mathcal{R}^*$ it is much flatter. Notice that this is the opposite of what happens with a low level of g in the previous subsection. When government expenditure is sufficiently low, the government is very uncertain about the true realization of θ. Hence higher labor does not allow a more precise signal extraction about productivity. On the other hand, when g is sufficiently high, there is an intermediate region of observables where the government becomes confident about low realizations of θ. In Appendix B we prove this result by illustrating how g affects the slope of the loci of realizations of the shocks through its impact on the wealth effect.

To illustrate how the PI policy involves a relatively precise signal extraction on θ with high government expenditure, consider the sets of possible realizations of θ under FI in Figure 7 and under PI in Figure 8. Consider Figure 7 first. It can be seen that under FI any realization of θ is consistent with an intermediate realization of l, but each of these θ's would call for a different tax rate. However, under PI, there can only be one tax rate for each observed l and the government uses this policy to extract information on θ.

Figure 7: Set of admissible θ's with FI

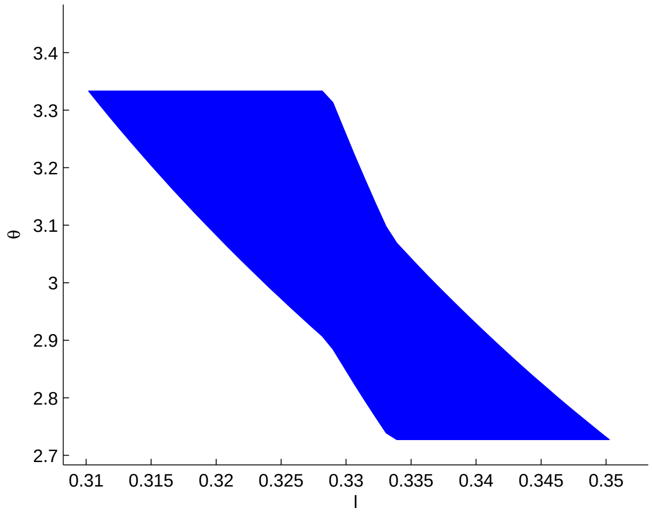


The blue region indicates, for each possible value of l, the corresponding set of realizations of θ with positive density in equilibrium under FI.

To see this, consider now Figure 8. The minimum value of l is only consistent with the highest possible θ (and lowest possible γ) because the wealth effect dominates. Under PI, increasing l from this point, the government becomes uncertain and lower realizations of productivity become consistent with the observations. At first, uncertainty is rising with l, but in the intermediate region of l's the government becomes more and more confident about low realizations of productivity. This leads to the sharp increase in the tax rate, which in turn gives rise to feedback effect on the set of possible θ's: high taxes discourage work effort, so higher labor now is an even stronger signal of low θ (high marginal utility from consumption). In this way an optimal policy and a conditional distribution of shocks consistent with it confirm each other in equilibrium.

Consistently with this analysis of the signal extraction, we also plot hours and taxes as functions of each shock individually, and we contrast the PI outcomes with the FI solution in Figure 9. On the left-hand side we consider productivity shocks only. As illustrated above, in the intermediate region of l's the government has a precise signal about θ, hence PI and FI policies and allocations are very close to each other. However for extreme realizations of θ the government is fooled about the source of the fluctuations and does hardly respond to productivity. On the right-hand side we consider only demand shocks. It can be seen that the PI government has very

Figure 8: Set of admissible θ's with PI

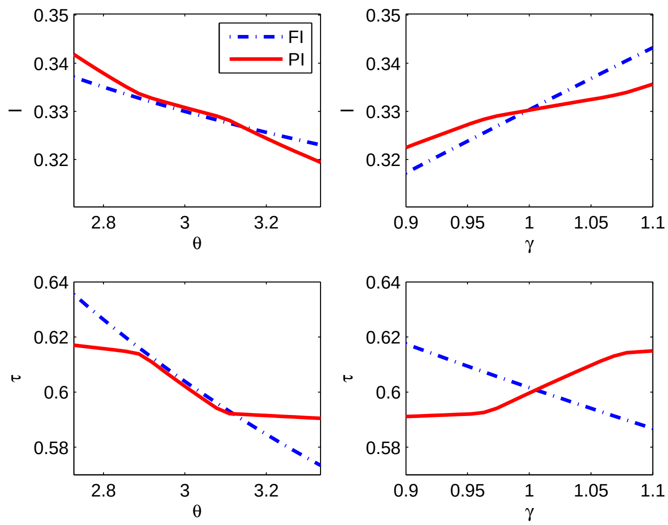


The blue region indicates, for each possible value of l, the corresponding set of realizations of θ with positive density in equilibrium under PI.

imprecise information about γ. Hence it responds to these shocks with the opposite slope with respect to the FI government.

The case of high government expenditure shows that optimal policy with PI can be very non-linear in order to avoid the worst outcomes, e.g. in the model hitting the top of the Laffer curve or in the real world a debt crisis. As shown in the previous subsection, when expenditure is low and there are no concerns related to the government budget constraint, policy has to be smooth, but when there are contingencies that are particularly dangerous for agents, then optimal policy calls for being very reactive to observables in order to prevent those cases to materialize. This is exemplified by the optimality of increasing taxes steeply in the first period to avoid having to distort the economy too heavily in the second period if realized productivity turn out to be low (and hence the fiscal deficit turns out to be high). This lesson seems relevant for the understanding of the fiscal policy reaction to the financial crisis in 2008 and afterwards, especially in countries like Spain and Italy, that arguably where in danger of getting close to the top of the Laffer curve, as testified by the fact that significant increases in taxes after 2009 did not raise the amount of revenue as much as it was desired by these governments.

Figure 9: Hours and taxes with PI and FI: high g



Left panels: top, hours as a function of the productivity shock (solid red line for PI, blue dashed-dotted line for FI); bottom, tax rate as a function of the productivity shock. Right panels: top, hours as a function of the demand shock; bottom, tax rate as a function of the demand shock.

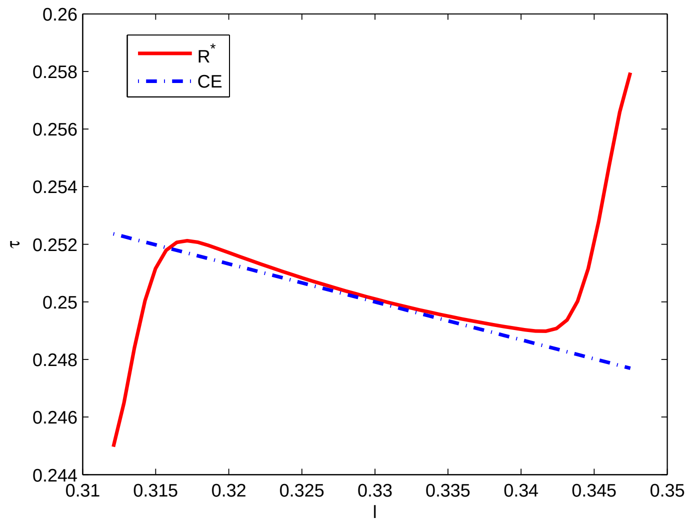
5.3 Linear-quadratic approximation

We now compare our solution to existing methods based on linear-quadratic optimization (Svensson and Woodford, 2004). In order to do so, we modify our distributional assumption and we assume that both θ and γ are normally distributed. We then truncate these distributions at three standard deviations from the mean in order to have a bounded support for the shocks in our solution method. The standard deviation of each of the shocks is assumed to be 3% of the mean.

In order to compute the linear approximation, we take a second-order approximation of the objective function and a first-order approximation of the reaction function h around the allocation and policy that arises under FI when the shocks take their mean value. Then, we compute the "certainty equivalent" policy as described in Section 4.7. Importantly, this policy can be found under FI and then applied to the PI case by simply computing the conditional mean of the shocks for each value of l. Figures 10 and 11 compare the optimal policy and the linear approximation in the case of low g and high g respectively. It can be seen that the approximation is quite accurate for intermediate realizations of labor, but less so for extreme values. This suggests that linear approximations can be misleading when there is endogenous PI and the economy is hit by large shocks.

An important difference between the non-linear solution and the linear approximation is that in the linearized model the government cannot affect the slope of the loci of shocks realizations conditional on the signal. In the Svensson and Woodford (2004) environment, this slope is constant and exogenous to policy, hence there is no scope for a choice of better signal extraction about one shock or the other.

Figure 10: Linear approximation, low g



Optimal tax rate as a function of the signal l, with $g = .25$. Red line: $\mathcal{R}^*$, blue dashed-dotted line: "certainty equivalent" linear approximation.

6 An infinite-horizon model with debt

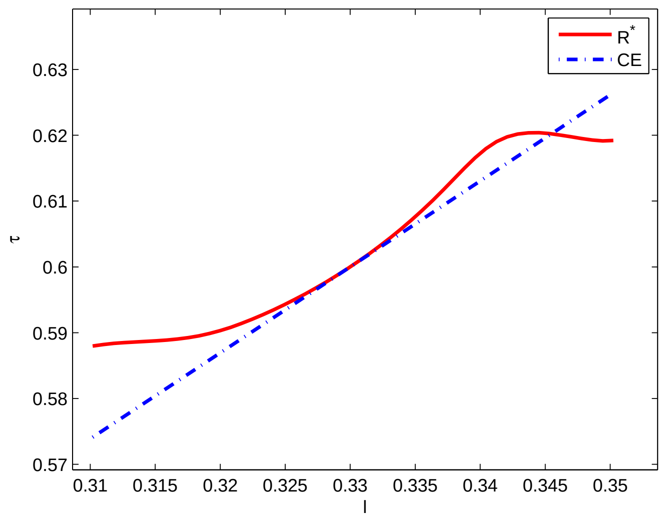
In this section we present an infinite horizon version of the optimal fiscal policy model we have considered. We will see that some key intuitions developed in the two-period model are still present and they lead to interesting dynamics. In particular, under PI the government sometimes reacts slowly to recessions and as a consequence needs to raise taxes by more and for a long time.

As is well known, in the case of non-linear utility, the current bond price depends on future taxes, hence the optimal policy under full commitment would be time inconsistent, leading to some complications in the solution of optimal policy.[18] In order to avoid these difficulties we assume linear utility from consumption.

We also assume that the shocks are i.i.d. over time as this reduces the number of state variables and allows us to abstract from the issues of government experimentation that have been studied in the armed-bandit literature of optimal policy that

[18]Under FI, this issue was first addressed in Aiyagari et al (2002).

Figure 11: Linear approximation, high g



Optimal tax rate as a function of the signal l, with $g = .6$. Red line: $\mathcal{R}^*$, blue dashed-dotted line: "certainty equivalent" linear approximation.

we discussed in Section 2.[19] In this way we are left with the simplest infinite horizon model of fiscal policy where endogenous signal extraction plays a role.[20]

6.1 Full Information

Our model under FI is a small variation of Example 2 of Aiyagari et al. (2002), with linear utility from consumption and standard convex disutility from labor effort. Preferences of the representative agent are given by:

$$E_0 \sum_{t=0}^{\infty} \beta^t \left[\gamma_t c_t - v(l_t) \right] \tag{48}$$

where γ_t is a demand shock, i.i.d. over time.

The period t budget constraint of the representative agent is

$$c_t + q_t b_t = \theta_t l_t (1 - \tau_t) + b_{t-1} \tag{49}$$

where θ_t is an i.i.d. productivity shock. Note that the government can only issue real riskless bonds b_t.

[19]For example Wieland (2000a, 2000b), Kiefer and Nyarko (1989), Ellison and Valla (2001)

[20]Combining this with issues of commitment and optimal experimentation is of interest but we leave it for future work.

The standard first order conditions for utility maximization are

$$\frac{v'(l_t)}{\gamma_t} = \theta_t(1 - \tau_t) \tag{50}$$

and

$$q_t = \beta \frac{\bar{\gamma}}{\gamma_t}. \tag{51}$$

where $\bar{\gamma}$ is the unconditional expectation of the demand shock γ.

The Ramsey government finances a constant stream of expenditure $g_t = g \; \forall t$ and chooses taxes and non-contingent one-period debt in order to maximize utility of the agent subject to the above competitive equilibrium conditions as well as the resource constraint $c_t + g = \theta_t l_t$. Under FI, the government can choose a sequence of taxes conditional on a sequence of shocks $A^t = (A_t, A_{t-1}, ..., A_0)$, where $A_t = (\theta_t, \gamma_t)$.

The period-t implementability constraint is

$$b_{t-1} = c_t - \frac{v'(l_t)}{\gamma_t} l_t + \beta \frac{\bar{\gamma}}{\gamma_t} b_t. \tag{52}$$

We now introduce an upper bound on debt, b_{max}. We will assume that whenever debt goes above this threshold, the government pays a quadratic utility cost $\beta\frac{\chi}{2}(b_t - b^{max})^2$ and we will set the parameter χ to be a large number in order to mimic a model with an occasionally binding borrowing constraint while still retaining differentiability of the problem.

The first order conditions for Ramsey allocations with respect to hours and debt are:

$$\gamma_t\theta_t - v'(l_t) + \lambda_t\theta_t - \frac{\lambda_t}{\gamma_t}\left[v'(l_t) + v''(l_t)l_t\right] = 0 \tag{53}$$

and

$$\lambda_t \frac{\bar{\gamma}}{\gamma_t} = E_t\lambda_{t+1} + \chi(b_t - b^{max}) I_{[b^{\max},\infty)}(b_t). \tag{54}$$

where λ_t is the Lagrange multiplier of constraint (49) and we denote by $I_{[b^{\max},\infty)}(b)$ the indicator function for the event $b > b^{max}$.

Thanks to the assumption of linear utility from consumption, the Ramsey policy is time-consistent and allocations satisfy a Bellman equation that defines a value function $V^{FI}(b_{t-1}, A_t)$. Thus optimal taxes are given by a time-invariant policy function $\tau_t = \mathcal{R}^{FI}(b_{t-1}, A_t)$

6.2 Partial Information

We start the description of the PI problem by specifying its timing. At the beginning of each period t, the Ramsey government observes the realization of the exogenous shocks of last period A_{t-1}, the value of its outstanding debt b_{t-1} and the realization of current labor l_t. Based on this information, but before knowing the value of A_t, it sets the tax rate τ_t. Formally, the choice of taxes at time t is contingent on - i.e. a function of - (A^{t-1}, l_t).

Note that because of the i.i.d. assumption on the shocks A_t, information about outstanding debt summarizes all the information about past realizations that is relevant in terms of the objective function and the constraints of the Ramsey problem. In other words, the government cares about past realizations of the exogenous shocks only to the extent that they affect the level of current outstanding debt. As a consequence, debt is a sufficient state variable in addition to the current observed signal l_t. Hence the optimal policy has a recursive structure and taxes are given by a policy function $\tau_t = \mathcal{R}(b_{t-1}, l_t)$.

Define V as the value of the utility (48) at the optimal choice for given initial debt, before seeing the realization of l_0. By a standard argument, the choice from period 1 onwards is feasible from period 0 onwards given the same level of debt. Therefore V satisfies the following Bellman equation

$$V(b) = \max_{\mathcal{R}:\, \Re^2 \to \Re_+} E\left[\gamma(\theta l - g) - v(l) + \beta V\left(\frac{(b + g - \theta l + \frac{v'(l)l}{\gamma})\gamma}{\beta\bar{\gamma}}\right) + \right.$$
$$\left. - \beta\frac{\chi}{2}\left(\frac{(b + g - \theta l + \frac{v'(l)l}{\gamma})\gamma}{\beta\bar{\gamma}} - b^{max}\right)^2\right] \tag{55}$$

where l satisfies $l = h(\mathcal{R}(b, l), \theta, \gamma)$ where $h(\tau, \theta, \gamma)$ is the labor that satisfies (50).

The only difference with respect to the reaction function in the two-period model is that now the government should recognize that debt affects labor indirectly through the tax rate.

With linear utility from consumption, the reaction function is given by

$$l = (v'^{-1}(\theta\gamma(1 - \tau)), \tag{56}$$

hence the locus of shocks realization conditional on l is increasing, that is, a certain level of hours can be generated by combinations of high productivity and low demand or vice versa. High θ_t and low γ is good news for the government for two reasons: revenue is high and the interest rate is low. On the other hand, low θ and high γ put pressure on the government budget constraint both by inducing low revenue and a high interest rate on the newly issued debt.

At this point it is important to pause the maths for a second and discuss how shocks and PI influence the optimal choice of taxes. Under incomplete markets a sequence of adverse shocks (low θ) will lead to an increase in debt. This will be more so under PI than under FI, because under PI the government only learns that a low θ and, therefore, a low tax revenue, occurred with a delay. The reason that b is an argument in $\mathcal{R}$ is that in the presence of incomplete markets, debt piles up after a few bad shocks, much more so than under FI, therefore the government will have to increase the level of taxes for a given l_t to avoid debt from becoming unsustainable.

Note that in (55) we have substituted future debt using the budget constraint (52). It is important to highlight a key difference with respect to the FI problem: while in that case a choice of τ_t implied a choice of b_t, now, b_t is a random variable even for a given choice of τ_t. In other words, just like in the two-period model, the government is uncertain about how much debt will need to be issued and in particular must take into account that bad realizations of productivity may lead to a debt level above b^{max}, if taxes are not sufficiently high.

In order to solve the model, we exploit its recursive structure, by solving for the PI first order condition at each point on a grid for debt and iterating on the value function of the problem. To see how this works, consider the objective function defined by the right-hand side of (55).

For a given guess for the value function, this is just a function of observed labor to which we can apply the main theorem of the paper (Proposition 3) and obtain the general first order condition with PI.[21]

This first order condition involves the derivative $V'(b)$. In standard dynamic programming it is well known that an envelope condition applies that allows the simplification of the derivative of the value function. In Appendix C we show that an analogous envelope condition holds under our PI model so that

$$V'(b) = E\left[\frac{\gamma}{\bar{\gamma}}V'(b')\right] - \chi\frac{\gamma}{\bar{\gamma}}(b' - b^{max})I_{[b^{\max},\infty)}(b'). \tag{57}$$

Hence by solving the first order condition using (57) and iterating on the Bellman equation (55), we can approximate the optimal policy. In the next subsection, we show some numerical results obtained after parametrizing the economy. While the model is not meant to be a quantitative model of fiscal policy, it is helpful to illustrate the key properties of optimal policy with endogenous signals, when the government faces a debt limit.

6.3 Numerical results: non-linearities and delayed adjustments

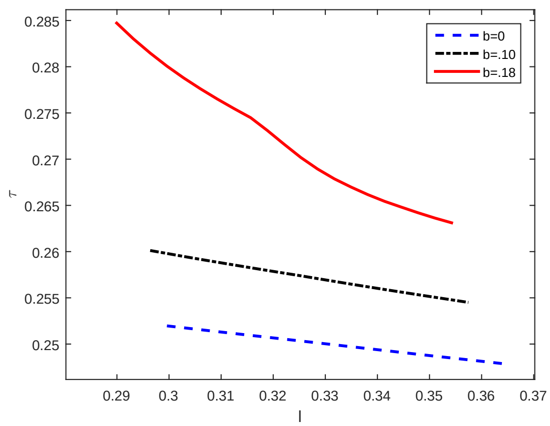
In order to parametrize the economy we assume quadratic disutility from labor, and the other parameters are as in the two-period model. The shocks are uniformly distributed on a support of ± 5% from their means, implying a volatility of 2.89%. The debt limit is set at 20% of mean output.

We now illustrate two interesting properties of optimal policy in this model. First, the response of taxes to endogenous signals is quite non-linear and it depends on the level of debt. When debt is low, optimal policy calls for smooth taxes regardless of the realization of labor. When debt is close to the limit, taxes become highly responsive to the signal, in particular in the central region of the signal, where there

[21]The FOC is explicitly shown in Appendix C.

is higher uncertainty on the state of the economy. Figure 12 illustrates this property by plotting taxes against labor, as given by the function $\mathcal{R}^*(b, \cdot)$ evaluated at different values of debt. Notice that taxes are in general decreasing in labor, because, with linear-quadratic utility, hours are increasing in θ, since the only effect of this shock is the substitution effect. It can be seen that for lower levels of debt (namely for $b = 0$ and $b = .1$), taxes are relatively flat with respect to the signal. However, close to the debt limit (b=.18), they become highly responsive, with a higher slope in the middle region of realizations of hours.

Figure 12: Tax policy for different levels of debt



Optimal tax rate as a function of the signal l, for different values of outstanding debt b. Red solid line: $b = .18$; black dashed-dotted line: $b = .10$; blue dashed line: $b = 0$.

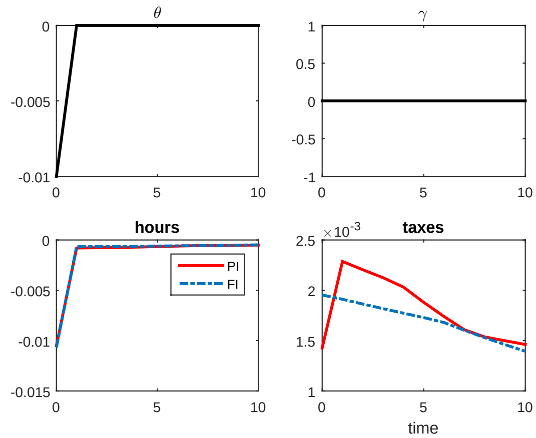
Second, optimal policy with PI can rationalize delayed fiscal adjustments following negative productivity shocks. While optimal policy with FI can respond contemporaneously to such shocks, under PI the government can only respond with a lag, after observing a higher than expected level of debt, leading to a delayed, persistent tax adjustment. This feature of the model is illustrated in the Impulse Response Function in Figure 13.[22]

We initialize both the FI and the PI economy at a level of debt equal to .12. When the productivity shock hits, the FI government is successful at making a tax correction to keep the economy sufficiently far from the borrowing constraint. On the other hand, the initial tax increase under PI is small, and the PI government imposes a fiscal adjustment only with a delay, after it learns that debt had increased. As a consequence of this delay, the postponed tax hike needs to be larger. Notice that because of market incompleteness, both the tax change under FI and PI are highly

[22]These non-linear impulse response functions are computed as percentage deviations from the path that would arise absent all shocks.

persistent, even though the shock is transitory as the government smooths taxes over time (Aiyagari et al. 2002).

Figure 13: Impulse response function: θ shock



Impulse response function to a negative 1% productivity shock. Top left panel: productivity shock; top right panel: demand shock; bottom left panel: hours; bottom right panel: tax rate. Solid red line: PI; blue dashed-dotted line: FI.

7 Conclusion

We derive a method to solve models of optimal policy with Partial Information generically. In our setup there is no separation between the optimization and signal extraction problem, the optimal decision influences the distribution of the shocks conditional on the observed endogenous signal, therefore the signal extraction and optimization problem are solved consistently and simultaneously. The method works in general and we show algorithms to find a solution. We also show that Partial Information on endogenous variables matters as some revealing non-linearities appear in very simple models. These non-linearities are due to the fact that in different regions of the observed signal the information revealed about the underlying state changes in a non-linear fashion.

Optimal fiscal policy under General Signal Extraction calls for smooth tax rates across states when the government budget is under control, and for regions of large response to aggregate data when the economy is close to the top of the Laffer curve or to a borrowing limit. Uncertainty about the state of the economy helps to understand the slow reaction of some European governments to the Great Recession, followed by sharp fiscal adjustments and prolonged downturns: as the signal worsens and it becomes more consistent with a slump in productivity the government becomes more certain about lower productivity. The tax choice can be highly non-linear as a function of the signal even in setups that are essentially linear.

Our methodology allows to study the effects of Partial Information about the state of the economy in models that feature important non-linearities. While we have illustrated the technique in a simple model of optimal fiscal policy, the methodology can be easily extended to address many other questions in optimal policy, such as models with heterogeneity and unobserved idiosyncratic shocks and models with unobserved beliefs fluctuations.

References

[1] Adam, K. (2004), "On the Relation between Bayesian and Robust Decision Making", *Journal of Economic Dynamics and Control*, vol. 28(10) 2105-2117

[2] Aiyagari, S.R., A. Marcet, T.J. Sargent and J. Seppala (2002) "Optimal Taxation without State-Contingent Debt", *Journal of Political Economy*, University of Chicago Press, vol. 110(6), pages 1220-1254, December.

[3] Angeletos, George-Marios and Alessandro Pavan (2009), "Policy with Dispersed Information", *Journal of the European Economic Association* 7(1):11–60

[4] Aoki, K. (2003), "On the Optimal Monetary Policy Response to Noisy Indicators" *Journal of Monetary Economics*, Elsevier, vol. 50(3), pages 501-523, April.

[5] Baxter, B., G. Liam and S. Wright (2007), "The Endogenous Kalman Filter", Birkbeck School of Economics, Mathematics and Statistics working paper BWPEF 0719

[6] Baxter, B., G. Liam and S. Wright (2011), "Invertible and Non-invertible Information Sets in Linear Rational Expectations Models", *Journal of Economic Dynamics & Control* 35: 295–311

[7] Clarke, F. (2013), *Functional Analysis, Calculus of Variations and Optimal Control*, Springer Graduate Texts in Mathematics

[8] Ellison, M. and N. Valla (2001), "Learning, Uncertainty and Central Bank Activism in an Economy with Strategic Interactions", *Journal of Monetary Economics*, Elsevier, vol. 48(1), pages 153-171, August.

[9] Guerrieri, V. and R. Shimer (2013) "Markets with Multidimensional Private Information", Society for Economic Dynamics Meeting Papers series 2013 n. 210.

[10] Hansen, L.P. and T.J. Sargent (2012), Three types of ambiguity, *Journal of Monetary Economics*, Elsevier, 59(5), 422-445.

[11] Kiefer, N. and Y. Nyarko, (1989), "Optimal Control of an Unknown Linear Process with Learning", *International Economic Review*, Dept. of Economics, U. of Pennsylvania and Osaka U. ISERA, vol. 30(3), pages 571-86, August.

[12] Klemperer, P.D. and M.A. Meyer (1989), "Supply Function Equilibria in Oligopoly under Uncertainty", Econometrica, Vol. 57, No. 6, pp. 1243-1277

[13] Kocherlakota, N.J. (2010), *The New Dynamic Public Finance*, Princeton University Press.

[14] Lucas, R.E., Jr. (1972), "Expectations and the Neutrality of Money", *Journal of Economic Theory*, 4(2), 103-124.

[15] Lucas, R.E., Jr. and N.L. Stokey (1983), "Optimal Fiscal and Monetary Policy in an Economy without Capital", *Journal of Monetary Economics* , 12, 55-93

[16] Mehra, R. and E.C. Prescott (1980), "Recursive Competitive Equilibrium: the Case of Homogeneous Households", *Econometrica* 48(6),1365–1379

[17] Mirman, L. J., Samuelson, L., and Urbano, A. (1993), "Monopoly Experimentation", *International Economic Review*, 549-563.

[18] Muth, J. F. (1960), "Optimal Properties of Exponentially Weighted Forecasts", *Journal of the American Statistical Association*, Volume 55, Issue 290

[19] Nimark, K. (2008), "Monetary Policy with Signal Extraction from the Bond Market", *Journal of Monetary Economics* 55, 1389–1400

[20] Pearlman, J., D. Currie and P. Levine (1986), "Rational Expectation Models with Private Information", *Economic Modelling* 3(2): 90-105

[21] Pearlman, J. (1992), "Reputational and Nonreputational Policies under Partial Information", *Journal of Economic Dynamics and Control* 16(2): 339-357

[22] Svensson, L.E.O. and M. Woodford (2003), "Indicator Variables for Optimal Policy", *Journal of Monetary Economics* 50, 691–720

[23] Svensson, L.E.O. and M. Woodford (2004), "Indicator Variables for Optimal Policy under Asymmetric Information", *Journal of Economic Dynamics & Control* 28, 661 – 690

[24] Swanson, E.T. (2006), "Optimal Nonlinear Policy: Signal Extraction with a Non-normal Prior", *Journal of Economic Dynamics and Control* 30: 185-203

[25] Townsend, R.M. (1983), "Forecasting the Forecasts of Others", *Journal of Political Economy*, 91(4), 546-588

[26] Van Nieuwerburgh, S., Veldkamp, L. (2006), "Learning Asymmetries in Real Business Cycles" *Journal of Monetary Economics*, Elsevier, vol. 53(4), pages 753-772, May.

[27] Volker, W. (2000a), "Monetary Policy, Parameter Uncertainty and Optimal Learning", *Journal of Monetary Economics* 46: 199-228

[28] Volker, W. (2000b), "Learning by Doing and the Value of Optimal Experimentation", *Journal of Economic Dynamics & Control* 24: 501-534

Appendix A

8 Proofs

8.1 Proof of Lemma 1

The proof uses the fact that for any continuous mapping $F : \mathcal{S} \to \mathcal{S}$, if there is a pair $s, s' \in \mathcal{S}$ such that $\frac{F(s)-F(s')}{s-s'} > 1$, then there exist $s'', s''' \in \mathcal{S}$ such that $\frac{F(s'')-F(s''')}{s''-s'''} = 1$. To obtain intuition for this the reader can draw such an F map and find that there are always such s'', s'''. The lemma follows if we take $F = h(\mathcal{R}(\,\cdot\,), A)$.

Assume $\mathcal{R}$ satisfies Condition 1 and there is an A and a pair $s > s'$ such that

$$\frac{h(\mathcal{R}(s), A) - h(\mathcal{R}(s'), A)}{s - s'} > 1 \tag{58}$$

Consider the cases

- *Case 1* $h(\mathcal{R}(s'), A) \geq s'$.

  Define the function $F : [s, \overline{s}] \to R$ as

  $$F(\mathbf{s}) = \frac{h(\mathcal{R}(\mathbf{s}), A) - h(\mathcal{R}(s'), A)}{\mathbf{s} - s'}$$

  where $\overline{s}$ denotes the upper bound of $\mathcal{S}$. Clearly F is continuous and (58) implies $F(s) > 1$. We have

  $$F(\overline{s}) = \frac{h(\mathcal{R}(\overline{s}), A) - h(\mathcal{R}(s'), A)}{\overline{s} - s'} \leq \frac{h(\mathcal{R}(\overline{s}), A) - s'}{\overline{s} - s'} \leq 1$$

  where the first inequality uses the Case and the last inequality the fact that $h(\mathcal{R}(\overline{s}), A) \in \mathcal{S}$ hence $h(\mathcal{R}(\overline{s}), A) \leq \overline{s}$. Therefore $F(\overline{s}) \leq 1$.

  Therefore by the intermediate value theorem there is an $s'' \in [s, \overline{s}]$ such that $F(s'') = 1$ hence

  $$\frac{h(\mathcal{R}(s''), A) - h(\mathcal{R}(s'), A)}{s'' - s'} = 1,$$

  violating Condition 3.

- *Case 2* $h(\mathcal{R}(s), A) < s$.

  An analogous argument gives that the ratio $\frac{h(\mathcal{R}(\mathbf{s}),A)-h(\mathcal{R}(s),A)}{\mathbf{s}-s}$ goes from a value ≤ 1 at $\mathbf{s} = \underline{s}$ to a value > 1 at $\mathbf{s} = s$ and Condition 3 is violated.

- *Case 3* $h(\mathcal{R}(s'), A) < s'$ and $h(\mathcal{R}(s), A) \geq s$.

  The function $F(\mathbf{s}) = h(\mathcal{R}(\mathbf{s}), A) - \mathbf{s}$ is continuous. Given the case, $F(s') < 0$ and $F(s) > 0$. Hence the intermediate value theorem implies that there is a fixed point $s'_f = h(\mathcal{R}(s'_f), A) \in [s', s]$. Furthermore, $F(\underline{s}) \geq 0$ hence there is a fixed point $s''_f = h(\mathcal{R}(s''_f), A) \in [\underline{s}, s']$. Therefore

$$\frac{h(\mathcal{R}(s''_f), A) - h(\mathcal{R}(s'_f), A)}{s''_f - s'_f} = 1$$

violating Condition 3.

These cases cover all possibilities. Therefore (58) is impossible and (43) is necessary and sufficient for condition 3.■

8.2 Proof of Lemma 2

Consider $\mathcal{R} \in \mathcal{E}$ throughout the proof.

Conditions 1 and 4 are obvious.

We detail the proof assuming $h_\tau < 0$, the proof works for any sign of h_τ as well.[23]

Proving Condition 3

We show that Lemma 1 holds.

Take any $x, y \in \mathcal{S}$, $x \neq y$ and any $A' \in \Phi$.

- If $\mathcal{R}(x) = \mathcal{R}(y)$ then $\frac{h(\mathcal{R}(s),A)-h(\mathcal{R}(s'),A)}{s-s'} = 0$ and (43) holds since $\varepsilon < 1$.
- If $\mathcal{R}(x) \neq \mathcal{R}(y)$ the mean value theorem and Assumption 1 gives

$$\frac{h(\mathcal{R}(x), A') - h(\mathcal{R}(y), A')}{x - y} = h_\tau(\tau, A')\frac{\mathcal{R}(x) - \mathcal{R}(y)}{x - y}$$

  for some $\tau \in [\mathcal{R}(x), \mathcal{R}(y)]$.

  - If $\frac{\mathcal{R}(x)-\mathcal{R}(y)}{x-y} > 0$ we have $h_\tau(\tau, A')\frac{\mathcal{R}(x)-\mathcal{R}(y)}{x-y} < 0$ so that (43) holds since $\varepsilon < 1$.
  - If $\frac{\mathcal{R}(x)-\mathcal{R}(y)}{x-y} < 0$, since $\frac{\mathcal{R}(x)-\mathcal{R}(y)}{x-y} \geq \underline{L}$ and $h_\tau < 0$ then

$$h_\tau(\tau, A')\frac{\mathcal{R}(x) - \mathcal{R}(y)}{x - y} \leq h_\tau(\tau, A')\underline{L} \leq 1 + \varepsilon$$

This proves that Lemma 1 holds and therefore Condition 3 holds.

Proving Condition 2

[23] It is easy to verify that in our application, for any separable preferences in consumption and labor effort, $h_\tau < 0$.

Consider x, y that are solutions for a given A, namely they satisfy $x = h(\mathcal{R}(x), A)$ and $y = h(\mathcal{R}(y), A)$. We have

$$x - y = h_\tau(\tau', A)\left[\mathcal{R}(x) - \mathcal{R}(y)\right].$$

If $x \neq y$ this would imply

$$h_\tau(\tau', A)\frac{\mathcal{R}(x) - \mathcal{R}(y)}{x - y} - 1 = 0$$

violating Condition 3. Therefore $x = y$. This proves that Condition 2 holds since there is a unique equilibrium.

Condition 1 follows because $\mathcal{R}$ is Lipschitz continuous.

To prove that $\mathcal{E}$ is compact take a sequence $\mathcal{R}_k \in \mathcal{E}$ for $k = 1, 2, ...$ Since $\mathcal{R}_k$ are Lipschitz the Arzelà-Ascoli theorem guarantees that there is a subsequence $\mathcal{R}_{k_i}$ such that $\mathcal{R}_{k_i} \to \mathcal{R}^{\lim}$ uniformly in $\mathcal{S}$ as $i \to \infty$. All that is left to prove is that for each x, y we have $\left|\frac{\mathcal{R}^{\lim}(x) - \mathcal{R}^{\lim}(y)}{x-y}\right| \in \left[\overline{L}, \mathbf{L}\right]$. Uniform convergence gives that for each pair x, y

$$\frac{\mathcal{R}_{k_i}(x) - \mathcal{R}_{k_i}(y)}{x - y} \to \frac{\mathcal{R}^{\lim}(x) - \mathcal{R}^{\lim}(y)}{x - y} \equiv d_{x,y} \; as \; i \to \infty.$$

Since $\mathcal{R}_k \in \mathcal{E}$ we have that $\frac{\mathcal{R}_{k_i}(x) - \mathcal{R}_{k_i}(y)}{x-y}$ belong to the compact set $\left[\overline{L}, \mathbf{L}\right]$ for all i, therefore the limit $d_{x,y}$ also belongs to the compact set $\left[\overline{L}, \mathbf{L}\right]$

This proves that any sequence $\mathcal{R}_k \in \mathcal{E}$ has a subsequence converging uniformly to a limit $\mathcal{R}_k \in \mathcal{E}$ so that $\mathcal{E}$ is compact.■

8.3 Proposition 2

8.3.1 Uniformly well-conditioned systems of equations.

To arrive at Proposition 2 we first need to show that for any variation δ and small α the equilibrium $S(\mathcal{R}^* + \alpha\delta, A)$ is well defined with probability one. This will be stated in Lemma 4. To highlight the importance of uniform well-conditioned system, we first state a generic result on the existence and uniqueness of solutions.

We give a result for systems of equations. The issue is that we want to guarantee that if $f^k \to f$ the solutions to $f^k = 0$ converge to the solution of $f = 0$.

We state assumptions that are sufficient for well-conditioning and that are easy to check for the case of our interest.

On the sets $X \subset R$, $Y \subset R^n$ we consider functions $f, f^k : X \times Y \to R$ for $k = 1, 2, ...$and the sup norm

$$d(f, g) = \sup_{X \times Y} |f(x, y) - g(x, y)|$$

We make the following assumptions

- **Assumption L1:** X and Y are compact, $f(\cdot, y)$ and $f^k(\cdot, y)$ are absolutely continuous for each y.
- **Assumption L2:** f^k converge uniformly to f, i.e. $d(f, f^k) \to 0$ as $k \to \infty$
- **Assumption L3:** for each $y \in Y$ there is a unique solution to $f(\cdot, y) = 0$. This solution lies in the interior of X.

  Hence, there is a well-defined mapping $\chi : Y \to X$ with the property $f(\chi(y), y) = 0$ and $\chi(Y) \subset int(X)$.

  Let $X^d(y) \subset X$ be the set containing all points where the partial derivatives $f_x(\cdot, y)$ and $f_x^k(\cdot, y)$ exist for all k. Note that because of absolute continuity $f(\cdot, y)$ and $f^k(\cdot, y)$ the set $X^d(y)$ has Lebesgue measure zero.
- **Assumption L4:** The partial derivatives of f^k converge as follows

$$\sup_{x \in X^d(y), y \in Y} \left| f_x(x, y) - f_x^k(x, y) \right| \to 0$$

- **Assumption L5:** f_x is uniformly bounded away from zero near the zeroes. Formally, there is a constant $K > 0$ and an $\varepsilon > 0$ such that for all $y \in Y$ and all x such that $|x - \chi(y)| < \varepsilon$, then $x \in X$ and if $f_x(x, y)$ exists then

$$|f_x(x, y)| > K$$

  Furthermore, for all y, the sign of $f_x(x, y)$ is the same for any x such that $|x - \chi(y)| < \varepsilon$ where f_x exists.

Note that the last assumption is generically satisfied by functions satisfying Assumption L3, it just requires the partial f_x to be away from 0 near the zeroes. For example if $f(x, y) = \phi x + x^3$ it excludes $\phi = 0$ but any other $\phi > 0$ works. The sign restriction excludes, for example, $f(x, y) = |x|$ but $|x| + \phi$ violates L3 for all $\phi \neq 0$.

Lemma 3 ***(Existence and uniqueness of solutions)*** *Under Assumptions L1-L5 there exists a unique solution to $f^k(\cdot, y) = 0$ for all y for k is sufficiently high. Furthermore, letting $\chi^k(y)$ denote this solution, we have $d(\chi^k, \chi) \to 0$ as $k \to \infty$.*

Proof

Define the neighborhood of the zeros as the set

$$XY^\varepsilon \equiv \{(x, y) : for\ all\ y \in Y\ and\ x\ such\ that\ |x - \chi(y)| < \varepsilon\ for\ some\ y \in Y\}$$

where ε is as in assumption L5.

We first prove that $f^k(\cdot, y)$ has a solution for all y for sufficiently high k.

Since $f(\chi(y), y) = 0$, the fundamental theorem of calculus gives

$$\int_{\chi(y)}^{x} f_x(\overline{x}, y) d\overline{x} = f(x, y) \tag{59}$$

for all $x > \chi(y)$ and all y. Consider a y such that $f_x(\chi(y), y) > 0$. Using assumption L5, if $|x - \chi(y)| < \varepsilon$ we have

$$\int_{\chi(y)}^{x} f_x(\cdot, y) > K(x - \chi(y)) \tag{60}$$

Plugging $x = \chi(y) + \varepsilon/2$ in (59) we have

$$0 < \frac{\varepsilon}{2} K < f(\chi(y) + \frac{\varepsilon}{2}, y)$$

A similar argument gives

$$0 > -\frac{\varepsilon}{2} K > f(\chi(y) - \frac{\varepsilon}{2}, y)$$

This all implies that $f(\chi(y) \pm \frac{\varepsilon}{2}, y)$ are bounded away from zero, and of oposite signs for all y.

Note that here we have used $\chi(Y) \subset int(X)$ in guaranteeing that f is well defined at $x = \chi(y) \pm \frac{\varepsilon}{2}$.

Uniform convergence of f^k implies that for a k sufficiently high $d(f^k, f) < \frac{\varepsilon}{4} K$. Together with the last two equations this implies that

$$\begin{aligned} 0 &< \frac{\varepsilon}{4} K < f^k(\chi(y) + \frac{\varepsilon}{2}, y) \\ 0 &> -\frac{\varepsilon}{4} K > f^k(\chi(y) - \frac{\varepsilon}{2}, y) \end{aligned}$$

Therefore $f^k(\cdot, y)$ takes a positive and a negative value near. This implies by the intermediate value theorem that a solution to $f^k(\cdot, y) = 0$ exists for all y and for k high enough.

A similar argument gives the reverse inequalities for y's such that the sign of f_x is negative near $\chi(y)$ and existence follows in that case as well.

Denote the set of solutions to $f^k(\cdot, y) = 0$ as $\chi^k(y)$, thus far we know $\chi^k(y)$ is a non-empty set.

We now prove that $\chi^k(y) \subset XY^\varepsilon$. We first show that

$$|f(x, y)| \geq V \; for \; all (x, y) \notin XY^\varepsilon \tag{61}$$

for a constant $V > 0$. Consider the infimum of $|f|$ outside XY^ε

$$V \equiv \inf_{X \times Y - XY^\varepsilon} |f(x, y)|$$

Compactness of $X \times Y - XY^{\varepsilon}$ and continuity of f implies the infimum is attained at some $(x^*, y^*) \in X \times Y - XY^{\varepsilon}$ such that $f(x^*, y^*) = V$. If $V = 0$ this implies that $x^* = \chi(y^*)$ and it contradicts the fact that all the zeroes of f are contained in XY^{ε}. Therefore $V > 0$ and this shows (61).

This and uniform convergence implies

$$\left|f^k(x,y)\right| > V/2 > 0 \ \forall (x,y) \in X \times Y - XY^{\varepsilon} \ for\ k\ large\ enough$$

so that $f^k(x,y) = 0$ can only happen for $(x,y) \in XY^{\varepsilon}$ for k large.

This proves that all solutions of f^k (namely (x,y) where $x \in \chi^k(y)$) must lie in XY^{ε} for k large enough.

All that is left to show is that $\chi^k(y)$ has only one element for k large enough. Applying (59) and using assumption L5 in a similar way as we did above implies that given any y for which f_x is positive in assumption L5, for any elements $x', x'' \in \chi^k(y)$ with $x' \geq x''$ then

$$\int_{x''}^{x'} f_x(\overline{x}, y) d\overline{x} = f(x', y) - f(x'', y) = 0$$

which is impossible since

$$\int_{x''}^{x'} f_x(\overline{x}, y) d\overline{x} > K(x' - x'') > 0$$

Therefore there exists a unique solution of $f^k(\cdot, y)$ for k large enough for all y or, equivalently, $\chi^k(y)$ is a singleton for all y.

To prove convergence of the solutions to $\chi^k(y)$, fix y, take any convergence subsequence $\left\{\chi^{k_j}(y)\right\}_{j=1}^{\infty}$, denote the limit $\lim_{j\to\infty} \chi^{k_j}(y) = L(y)$. Uniform convergence of f^k implies that $f(L(y), y) = 0$, so that $L(y)$ is a zero of $f(\cdot, y) = 0$ so assumption L3 implies $\chi(y) = L(y)$. Therefore $\chi^k(y) \to \chi(y)$ as $k \to \infty$.■

We apply Lemma 3 to show that solutions $S(\mathcal{R}^* + \alpha\delta, A)$ exist and are unique for α small. Here, $\delta : \mathcal{S} \to R$ is a variation of the policy function.

Lemma 4 *Consider a variation δ that is bounded differentiable everywhere and with a uniformly bounded derivative. Under Assumptions 1-4 and if $\mathcal{R}^*$ satisfies Conditions 1-4, the solution $S(\mathcal{R}^* + \alpha\delta, A)$ exists and is unique for α small enough.*

Proof

To ensure that the equilibria are well defined we truncate the variations

$$\begin{aligned} \left(\mathcal{R}^* + \alpha\delta\right)(s) &= \overline{s}\ if\ \mathcal{R}^*(s) + \alpha\delta(s) > \overline{s} \\ &= \underline{s}\ if\ \mathcal{R}^*(s) + \alpha\delta(s) < \underline{s} \\ &= \mathcal{R}^*(s) + \alpha\delta(s)\ otherwise \end{aligned} \tag{62}$$

We apply Lemma 3 when

-s takes the role of x and A takes the role of y in Lemma 1.
-$f(\cdot) \equiv H(\cdot;\mathcal{R}^*)$
-$f^k(\cdot) \equiv H(\cdot;\mathcal{R}^* + \alpha_k\delta)$
for a sequence $\alpha_k \to 0$. We now have to check that assumptions L1-L5 hold.
Assumptions 1, 2, and Condition 1 imply Assumption L1.
Absolute continuity of h implies

$$|H(s,A;\mathcal{R}^*) - H(s,A;\mathcal{R}^* + \alpha\delta)| < Q\alpha\,|\delta(s)|$$

where $Q < \infty$ is the uniform bound on $|h_\tau|$. Since δ is bounded this implies that $H(\cdot;\mathcal{R}^* + \alpha_k\delta)$ converge uniformly to $H(\cdot;\mathcal{R}^*)$ as in assumption L2.

Assumption L3 and Condition 2 are equivalent.

Consider s, A where the derivatives $h_\tau(\cdot, A)$ and $\mathcal{R}^{*\prime}$ exist. We have

$$H_s(s,A;\mathcal{R}^* + \alpha_k\delta) = 1 - h_\tau((\mathcal{R}^* + \alpha\delta)(s), A)(\mathcal{R}^{*\prime} + \alpha\delta')(s)$$

Therefore

$$\begin{aligned} |H_s(s,A;\mathcal{R}^*) - H_s(s,A;\mathcal{R}^* + \alpha_k\delta)| &= \\ |h_\tau((\mathcal{R}^* + \alpha\delta)(s), A)(\mathcal{R}^{*\prime} + \alpha\delta')(s) - h_\tau(\mathcal{R}^*(s), A)\mathcal{R}^{*\prime}(s)| &= \\ |\ [h_\tau((\mathcal{R}^* + \alpha\delta)(s), A) - h_\tau(\mathcal{R}^*(s), A)]\,(\mathcal{R}^{*\prime} + \alpha\delta')(s) + \alpha\delta'(s)\ | \le & \\ Q^L\alpha K^\delta(K^R + \alpha K^{\delta'}) + \alpha K^{\delta'} & \end{aligned}$$

the second equality follows from adding and substracting $h_\tau(\mathcal{R}^*(s), A)(\mathcal{R}^{*\prime} + \alpha\delta')(s)$ and where $K^{\delta'}$ is the bound on δ', K^δ the bound on δ, and Q^L the Lipschitz constant for h_τ.

This guarantees Assumption L4.

Assumption L5 is given by Condition 4.

Therefore, Lemma 3 implies that $S(\mathcal{R}^*{+}\alpha\delta, A)$ is well defined for α small enough.■

8.3.2 Proof of Proposition 2

Fix a variation $\delta : \mathcal{S} \to R$ such that δ is continuous, differentiable, $|\delta(s)| \le K^\delta$ and $|\delta'(s)| \le K^{\delta\prime}$ for all s for some constants $K^\delta, K^{\delta'} < \infty$.

Let $\mathcal{R}^* + \alpha\delta$ be the truncated variation (62). Since $\mathcal{R}^*$ is continuous the set $\mathcal{R}^*(\mathcal{S}^*)$ is compact, thus $\underline{s} < \mathcal{R}^*(s) < \overline{s}$ implies that it is bounded away from the limits $\underline{s}$,$\overline{s}$. Therefore, for α small enough $(\mathcal{R}^* + \alpha\delta)(s)$ is also contained in the open interval $(\underline{s}$,$\overline{s})$ and the truncation in (62) is not operational.

Consider the problem defined by (30). For α small enough $S(\mathcal{R}^* + \alpha\delta, A)$ is a well defined function by Lemma 2, and so is $\mathcal{F}$. It is clear that since $\mathcal{R}^* + \alpha\delta$ is a feasible policy function in the PI problem the solution of (30) is attained at $\alpha = 0$ so that

$$\left.\frac{d\mathcal{F}(\mathcal{R}^* + \alpha\delta)}{d\alpha}\right|_{\alpha=0} = 0 \tag{63}$$

if this derivative exists. We now prove that this derivative exists and that (63) implies (29).

The assumptions on W and h imply

$$|W_\tau|, |W_s|, |h_\tau| < Q$$

for some finite constant Q whenever the derivatives exist.

Take any sequence $\alpha_k \to 0$. For each k we can write

$$\frac{\mathcal{F}(\mathcal{R}^* + \alpha_k\delta) - \mathcal{F}(\mathcal{R}^*)}{\alpha_k} = \int_\Phi M_k(A) dF_A(A)$$

for

$$M_k(A) \equiv \frac{W\left(T(\mathcal{R}^*{+}\alpha_k\delta, A), S(\mathcal{R}^*{+}\alpha_k\delta, A), A\right) - W\left(T(\mathcal{R}^*, A), S(\mathcal{R}^*, A), A\right)}{\alpha_k}$$

Since $\mathcal{R}^*, W, h$ are differentiable at $s = S(\mathcal{R}^*, A)$ we have that

$$\begin{aligned} M_k(A) \to & \left.\frac{dW\left(\left(\mathcal{R}^* + \alpha\delta\right)\left(S(\mathcal{R}^*{+}\alpha\delta, A)\right), S\left(\mathcal{R}^*{+}\alpha\delta, A\right), A\right)}{d\alpha}\right|_{\alpha=0} & (64) \\ = \quad & W_\tau^{*\prime}(A)\left\{\left[\mathcal{R}^{*\prime}(A)\right] S_\delta^{*\prime}(A) + \delta^*(A)\right\} + W_s^{*\prime}(A) S_\delta^{*\prime}(A) & (65) \end{aligned}$$

where

$$\begin{aligned} \mathcal{R}^{*\prime}(A) &= \mathcal{R}^{*\prime}\left(S(\mathcal{R}^*, A)\right) \\ \delta^{*\prime}(A) &= \delta'\left(S(\mathcal{R}^*, A)\right) \\ \delta^*(A) &= \delta\left(S\left(\mathcal{R}^*, A\right)\right) \qquad (66) \\ W_x^*(A) &= W_x\left(\mathcal{R}^*\left(S(\mathcal{R}^*, A)\right), S\left(\mathcal{R}^*, A\right), A\right) \;\; for x = \tau, s \\ S_\delta^{*\prime}(A) &= \left.\frac{dS(\mathcal{R}^*{+}\alpha\delta, A)}{d\alpha}\right|_{\alpha=0} \end{aligned}$$

(note there is a slight abuse of notation as we use the symbol $\mathcal{R}^{*\prime}$ both for the function of s as well as for the function of A).

The only non-obvious term is the derivative $S_\delta^{*\prime}$. This can be found by applying the implicit function theorem to

$$S(\mathcal{R}^*{+}\alpha\delta, A) = h((\mathcal{R}^*{+}\alpha\delta)(S(\mathcal{R}^*{+}\alpha\delta, A)), A)$$

Carefully differentiating with respect to α we have

$$\begin{aligned} \frac{dS(\mathcal{R}^*{+}\alpha\delta, A)}{d\alpha} \;=\; & h_\tau((\mathcal{R}^*{+}\alpha\delta)(S(\mathcal{R}^*{+}\alpha\delta, A)), A)[(\mathcal{R}^{*\prime}{+}\alpha\delta')(S(\mathcal{R}^*{+}\alpha\delta, A)) \cdot \\ & \cdot \frac{dS(\mathcal{R}^*{+}\alpha\delta, A)}{d\alpha} + \delta(S(\mathcal{R}^*{+}\alpha\delta, A))] \end{aligned}$$

So that

$$S_\delta^{*\prime}(A) \equiv \left.\frac{dS(\mathcal{R}^*+\alpha\delta, A)}{d\alpha}\right|_{\alpha=0} = \frac{h_\tau^*(A)\delta^*(A)}{1-h_\tau^*(A)\mathcal{R}^{*\prime}(A)}$$

for a definition of $h_\tau^*(A)$ analogous to the notation in (66). Notice that Condition 3 allows us to divide by $1-h_\tau^*(A)\mathcal{R}^{*\prime}(A)$.

Using the boundedness assumptions it is clear that

$$|M_k(A)| \leq Q\left(\left[L+\alpha K^{\delta'}\right]\frac{Q}{\epsilon}+K^\delta+\frac{Q}{\epsilon}\right)$$

where L is the Lipschitz constant for $\mathcal{R}$, uniformly on k for almost all A.

Condition 5 implies that the limit in (65) occurs with probability one in A. Since M_k is uniformly bounded Lebesgue dominated convergence implies

$$\frac{\mathcal{F}(\mathcal{R}^*+\alpha_k\delta)-\mathcal{F}(\mathcal{R}^*)}{\alpha_k} \to \int_\Phi \left(\left[W_\tau^*\mathcal{R}^{*\prime}+W_s^*\right]S_\delta^{*\prime}+W_\tau^*\delta^*\right)dF_A$$

Since this holds for any sequence $\alpha_k \to 0$ it proves that the derivative of $\mathcal{F}$ with respect to α exists for any variation δ and from (63) we have

$$\int_\Phi \left(\left[W_\tau^*\mathcal{R}^{*\prime}+W_s^*\right]S_\delta^{*\prime}+W_\tau^*\delta^*\right)dF_A = 0 \tag{67}$$

Using the formula for $S_\delta^{*\prime}(A)$ and rearranging, we conclude that for any variation δ

$$\int_\Phi \frac{W_\tau^*+W_s^*h_\tau^*}{1-h_\tau^*\mathcal{R}^{*\prime}}\delta(S(\mathcal{R}^*,A))dF_A = 0 \tag{68}$$

Since (68) holds for any bounded δ with bounded derivative it also holds when δ is any bounded function measurable with respect to s. Therefore, the general definition of conditional expectation implies (29).

8.4 Proof of Proposition 3

Take random variables X, Y, Z. Assume there is a function g such that $X = g(Y, Z)$ a.s.; clearly for any function ϕ

$$E(\phi(X,Y) \mid Z) = E(\phi(g(Y,Z),Y) \mid Z). \tag{69}$$

Letting A_1 play the role of X, $S(\mathcal{R}^*, A)$ the role of Z, A_2 the role of Y and $\mathcal{A}^*$ the role of g we have the second equality in

$$E\left(\left.\frac{W_\tau^*+W_s^*h_\tau^*}{1-h_\tau^*\mathcal{R}^{*\prime}}\right| S(\mathcal{R}^*,A)=s\right) =$$
$$E\left(\left.\frac{W_\tau+W_sh_\tau}{1-h_\tau\mathcal{R}^{*\prime}}(\tau^*,s,A_1,A_2)\right| S(\mathcal{R}^*,A)=s\right) =$$

$$
E\left(\frac{W_\tau + W_s h_\tau}{1-h_\tau \mathcal{R}^{*\prime}}(\tau^*, s, \mathcal{A}^*(s, A_2), A_2)\middle| S(\mathcal{R}^*, A) = s\right) =
$$

$$
\int_{\Theta_2(s,\mathcal{R}^*(s))} \frac{W_\tau + W_s h_\tau}{1-h_\tau \mathcal{R}^{*\prime}}(\tau^*, s, \mathcal{A}^*(s, a_2), a_2) f_{A_2|S(\mathcal{R}^*,A)}(a_2, s) da_2 \tag{70}
$$

We apply Bayes' formula to find $f_{A_2|S(\mathcal{R}^*,A)}$. First we find $f_{S(\mathcal{R}^*,A)|A_2}$. Since $\mathcal{A}^*(s, A_2)$ is the inverse function of $S(\mathcal{R}^*, \cdot, A_2)$, by the change of variable rule the density $f_{S(\mathcal{R}^*,A)|A_2}$ exists and is given by

$$
f_{S(\mathcal{R}^*,A)|A_2}(s, a_2) = f_{A_1|A_2}(\mathcal{A}^*(s, a_2), a_2)\, |\mathcal{A}^{*\prime}(s, a_2)| \tag{71}
$$

where $\mathcal{A}^{*\prime}(s, a_2)$ is the partial derivative of $\mathcal{A}^*$ with respect to s. In writing $f_{A_1|A_2}$ we are using the assumption that A has a density. To find this derivative we apply once again the implicit function theorem to H and get

$$
\mathcal{A}^{*\prime} = \frac{1-h_\tau^* \mathcal{R}^{*\prime}}{h_{A_1}^*}, \tag{72}
$$

given Condition 6 this derivative is well defined. Plugging (72) into (71) gives $f_{S(\mathcal{R}^*,A)|A_2}$. Applying Bayes' rule we have

$$
f_{A_2|S(\mathcal{R}^*,A)} = \frac{f_{S(\mathcal{R}^*,A)|A_2} f_{A_2}}{\int f_{S(\mathcal{R}^*,A)|A_2} f_{A_2} da_2} \tag{73}
$$

where we use again the assumption that the density f_{A_2} exists.

Plugging the above formula for $f_{A_2|S(\mathcal{R}^*,A)}$ we see the expression in (70) becomes

$$
\int_{\Theta_2(s,\mathcal{R}^*)} \frac{W_\tau^* + W_s^* h_\tau^*}{1-h_\tau^* \mathcal{R}^{*\prime}}(\tau^*, s, \mathcal{A}^*(s, a_2), a_2) f_{A_2|S(\mathcal{R}^*,A)}(a_2, s) da_2 =
$$

$$
\int_{\Theta_2(s,\mathcal{R}^*)} \frac{W_\tau^* + W_s^* h_\tau^*}{1-h_\tau^* \mathcal{R}^{*\prime}}(\tau^*, s, \mathcal{A}^*(s, a_2), a_2) \frac{f_{A_1|A_2}(\mathcal{A}^*(s, a_2), a_2)}{f_{S(\mathcal{R}^*,A)}(s)} \left| \frac{1-h_\tau^* \mathcal{R}^{*\prime}}{h_{A_1}^*}(\tau^*, s, \mathcal{A}^*(s, a_2), a_2) \right| da_2
$$

Since $1 - h_\tau^* \mathcal{R}^{*\prime} > 0$ a.s. by Lemma 1, using (29) to set this equal to zero, using the fact that the denominator $f_{S(\mathcal{R}^*,A)}(s)$ drops out and that under Condition 6 $h_{A_1}^*$ is always of the same sign this gives (34).

We can finally backtrack all the steps to show that (34) implies (29).■

8.5 Proposition 4: Second Order Conditions

Proof of Proposition 4

We start from the first derivative of the objective function with respect to α and in order to simplify notation we drop the dependence of all functions on the shocks A.

Consider a variation $\mathcal{R}^* + \alpha\delta$ as in the proof of proposition 2. Let the first derivative be

$$\frac{dW}{d\alpha} = W'(s)S_\alpha \tag{74}$$

where the derivative of the signal is

$$S_\alpha = \frac{\delta(s)h_\tau((\mathcal{R}^* + \alpha\delta)(s))}{1 - h_\tau((\mathcal{R}^* + \alpha\delta)(s))((\mathcal{R}^{*\prime} + \alpha\delta')(s))} \tag{75}$$

By differentiating (74) again with respect to α we obtain

$$\frac{d^2W}{d\alpha^2} = W''(s)(S_\alpha)^2 + W'(s)S_{\alpha\alpha} \tag{76}$$

where

$$S_{\alpha\alpha} = \frac{\left[\delta^2 h_{\tau\tau} + \delta h_{\tau\tau}(\mathcal{R}^{*'} + \alpha\delta')S_\alpha + \delta' h_\tau\right]\left[1 - h_\tau(\mathcal{R}^{*'} + \alpha\delta')\right]}{\left[1 - h_\tau(\mathcal{R}^{*'} + \alpha\delta')\right]^2} + \tag{77}$$

$$+\frac{h_\tau\delta\left[\delta h_{\tau\tau}(\mathcal{R}^{*'} + \alpha\delta') + h_{\tau\tau}(\mathcal{R}^{*'} + \alpha\delta'^2 S_\alpha + \delta' h_\tau + h_\tau(\mathcal{R}^{*''} + \alpha\delta'')S_\alpha\right]}{\left[1 - h_\tau(\mathcal{R}^{*'} + \alpha\delta')\right]^2}$$

Doing some algebra, simplifying this expression, using an indicator function as variation δ and evaluating the expression at $\alpha = 0$ we get

$$S_{\alpha\alpha} = \frac{h_{\tau\tau}(1 - h_\tau\mathcal{R}^{*'}) + h_\tau h_{\tau\tau} + h_\tau^3\mathcal{R}^{*''}}{(1 - h_\tau\mathcal{R}^{*'})^3} \tag{78}$$

Finally, plugging (78) into (76), we get

$$\frac{d^2W}{d\alpha^2} = W''(s)(\frac{h_\tau}{1 - h_\tau\mathcal{R}^{*'}})^2 + W'(s)\frac{h_{\tau\tau}(1 - h_\tau\mathcal{R}^{*'}) + h_\tau h_{\tau\tau} + h_\tau^3\mathcal{R}^{*''}}{(1 - h_\tau\mathcal{R}^{*'})^3} \tag{79}$$

This gives the proposition.■

Appendix B: The effect of government expenditure on the signal extraction

In this Appendix we show that in the log-quadratic case an increase in g can change the sign of optimal policy in the intermediate region of the observables.

First, for a given $\bar{l}$ consider the locus $\tilde{\gamma}(\mathcal{R}(\bar{l}), \theta, \bar{l})$ which is the inverse of the reaction function h with respect to γ, implicitly defined by

$$h(\mathcal{R}(\bar{l}), \theta, \tilde{\gamma}(\mathcal{R}(\bar{l}), \theta, \bar{l})) - \bar{l} = 0. \tag{80}$$

Let ζ be the derivative of this function with respect to θ. This can be found using the implicit function theorem, which gives the positive slope

$$\zeta = -\frac{h_\theta}{h_\gamma} = \frac{g\sqrt{(Bg)^2 + 4B\theta^2\gamma(1-\tau)} + Bg^2}{2\theta^3(1-\tau)} \tag{81}$$

Now we differentiate ζ with respect to g and get

$$\frac{d\zeta}{dg} = \frac{\partial\zeta}{\partial g} + \frac{\partial\zeta}{\partial\tau}\frac{\partial\tau}{\partial g} \tag{82}$$

where all these partial derivatives are positive. Hence higher government expenditure makes the loci of realizations of the shocks steeper.

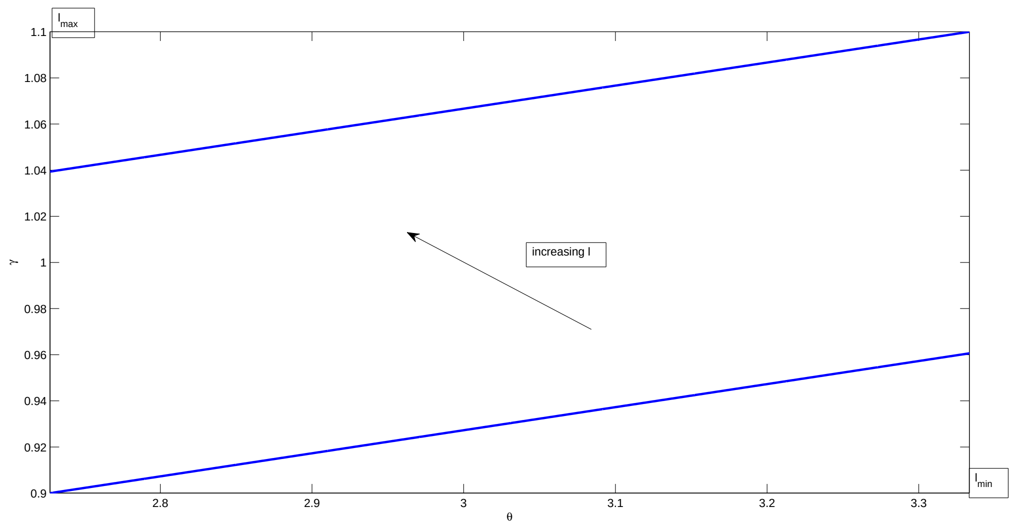
Now we illustrate how this effect changes the nature of the signal extraction on the shocks and hence the slope of optimal policy in the intermediate region of the observables. For this purpose we will take a first order approximation of the loci $\tilde{\gamma}$.[24]

Consider Figures 14 and 15. When g is sufficiently low, the map of loci (solid blue lines) moving in the direction of increasing l's looks like in Figure 14. Starting at l_{min} (bottom right corner) and increasing l the loci first hit the bottom-left corner, where the lowest θ becomes possible, and then the top-right corner, where the highest values for θ start to be inconsistent with the observed l's. Hence in the intermediate region of l's all θ's are possible, but clearly not all $\gamma's$. In this region, the government learns little about productivity. All the government learns is that the agent is working more as l increases so expected output is higher and taxes can be lower. This gives the negative slope of $\mathcal{R}^*$ with low government expenditure.

On the other hand, when g is sufficiently high, the slope of the loci becomes higher. Hence, as illustrated in Figure 15, in the intermediate region of l's the government learns that only a relatively small set of θ's is possible, whereas any γ is consistent with the observations. This leads to the positive slope of the optimal tax rate with high government expenditure.

[24] In our computed examples these loci are very close to linear.

Figure 14: Loci of shocks realizations with low g



X-axis: θ, Y-axis: γ. This figure illustrates how the set of (θ, γ) with positive density changes as the value of the signal l increases, starting from the bottom-right corner for l_{min} and moving up and left until reaching the top-left corner for l_{max}. With low g, these loci of shocks hit the bottom-left corner at a lower value of l than the top-right corner.

Appendix C: Derivation of the Envelope Condition (57)

In this Appendix we derive the Envelope Condition (57). First of all let us introduce the necessary notation. A tax policy is a function of debt and labor $\mathcal{R}(b,l)$ and labor is a function of a policy $\mathcal{R}$, outstanding debt and the exogenous shock, $L(\mathcal{R}; b, A)$ defined by the zero of

$$H(l, A, \mathcal{R}) \equiv l - h(\mathcal{R}(b,l), A), \tag{83}$$

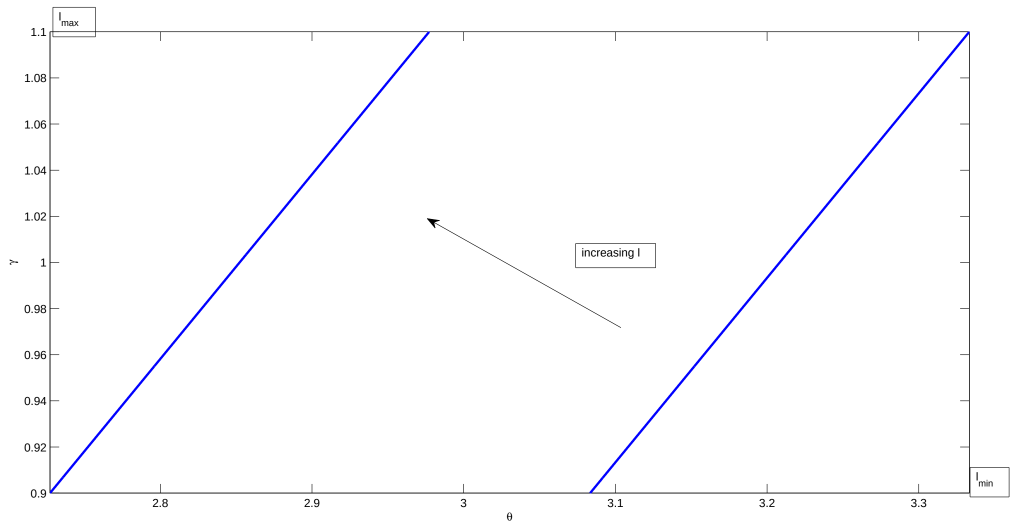
in analogy with the two-period model. By total differentiation of (83), the partial derivative of labor with respect to debt, L_b, is given by

$$L_b(\mathcal{R}, b, A) = \frac{\gamma\theta\mathcal{R}_b(b,l)}{v''(l) + \gamma\theta\mathcal{R}_L(b,l)}. \tag{84}$$

Now, for simplicity consider a case without borrowing penalty. In order to derive the envelope condition, we differentiate (55) with respect to b and get

$$V'(b) = E\left[\left(\gamma\theta - v'(l^*)\right)L_b^* + V'\left(b^{*'}\right)\left(\frac{\gamma}{\bar{\gamma}} + \beta b_L^{*'} L_b^*\right)\right]$$

Figure 15: Loci of shocks realizations with high g



X-axis: θ, Y-axis: γ. This figure illustrates how the set of (θ, γ) with positive density changes as the value of the signal l increases, starting from the bottom-right corner for l_{min} and moving up and left until reaching the top-left corner for l_{max}. With high g, these loci of shocks hit the top-right corner at a lower value of l than the bottom-left corner.

where

$$b_L^{*'} = \frac{-\theta\gamma + [v''(L(\mathcal{R}^*, A, b))L(\mathcal{R}^*, A, b) + v'(L(\mathcal{R}^*, A, b))]}{\beta\bar{\gamma}}$$

$$l^* = L(\mathcal{R}^*; b, A)$$

$$L_b^* = L_b(\mathcal{R}^*; b, A).$$

Using (84) we can write

$$V'(b) = E\left[\left(\gamma\theta - v'(l^*) + \beta V'\left(b^{*'}\right) b_L^{*'}\right) \frac{-\gamma\theta\mathcal{R}_b^*(l, b)}{v''(l) + \gamma\theta\mathcal{R}_L^*(l, b)} + V'\left(b^{*'}\right)\frac{\gamma}{\bar{\gamma}}\right]. \quad (85)$$

Using Proposition 3, the FOC of PI Ramsey problem is

$$E\left[\left(\theta\gamma - v'(l^*) + \beta V'(b'^*)b_L^{*'}\right)\frac{h_\tau^*}{1 - h_\tau^*\mathcal{R}_L^*}|\bar{l}\right] = 0 \quad (86)$$

for all $\bar{l}$. Furthermore, we have that the partial derivative of the reaction function h with respect to taxes is

$$h_\tau = \frac{-\gamma\theta}{v''(l)}.$$

So from (85) we get

$$V'(b) = E\left[\left(\gamma\theta - v'(l^*) + \beta V'\left(b^{*'}\right) b^{*'}_L\right) \frac{h^*_\tau \mathcal{R}^*_b(l,b)}{1 - h^*_\tau \mathcal{R}^*_L(L,b)} + V'\left(b^{*'}\right)\frac{\gamma}{\bar{\gamma}}\right].$$

Now, applying the law of iterated expectations, using the fact that $\mathcal{R}_b(l,b)$ is known given L, b and using (86), we obtain

$$\begin{aligned} V'(b) &= E\left[E\left(\left(\gamma\theta - v'(L^*) + \beta V'\left(b^{*'}\right) b^{*'}_L\right) \frac{h^*_\tau \mathcal{R}^*_b(L,b)}{1 - h^*_\tau \mathcal{R}^*_L(L,b)} \middle| L\right) + V'\left(b^{*'}\right)\frac{\gamma}{\bar{\gamma}}\right] && (87) \\ &= E\left[0 + V'\left(b^{*'}\right)\frac{\gamma}{\bar{\gamma}}\right] && (88) \end{aligned}$$

Finally, adding the marginal cost of excessive debt this becomes

$$V'(b) = E\frac{\gamma V'(b')}{\overline{\gamma}} - \frac{\gamma}{\bar{\gamma}}\chi(b' - b^{\max}) I_{[b^{\max},\infty)}(b').$$

Appendix D: Additional models

In this appendix we present two additional versions of our two-period model. In the first additional model, we assume that the productivity shock is permanent and we mantain labor as the signal. In the second model, we consider output as a signal, as opposed to labor.

Permanent productivity shock

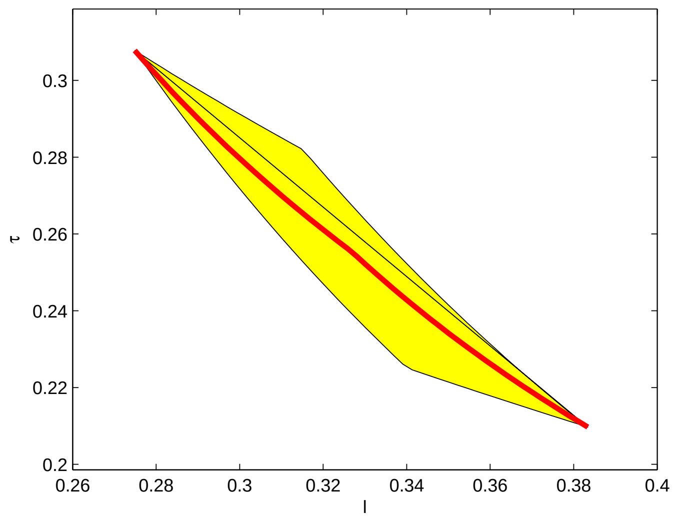
We now consider a permanent productivity shock, that is $\theta_2 = \theta_1 = \theta$, while we mantain the assumption of transitory demand shock γ. This setup would be uninteresting with log-quadratic preferences because hours would become independent of θ as wealth and substitution effects would cancel out. Therefore we assume linear utility from consumption (as in the infinite-horizon model) $u(c) = c$ and quadratic disutility from labor $v(l) = \frac{B}{2}l^2$, which give another case with an analytical solution for the reaction function h and its derivatives.

In particular, it is easy to see from the first order condition (6) that the reaction function (12) specializes to

$$l = h(\tau, \theta, \gamma) = \frac{\gamma\theta}{B}(1 - \tau), \qquad (89)$$

The optimal policy is illustrated in Figure 16 and is compared with the set of FI tax rates conditional on l. In general, $\mathcal{R}^*$ is decreasing as higher observed labor suggests higher conditional expectation for productivity, hence allowing to balance the intertemporal budget constraint with a lower distortionary tax. The figure also compares the optimal policy with a linear policy obtained connecting the two full revelation points with a straight line.

Figure 16: Optimal policy with permanent productivity shock



Optimal tax rate as a function of the signal l when the productivity shock is permanent. Red line: $\mathcal{R}^*$; yellow region: set of equilibrium (l, τ) under FI.

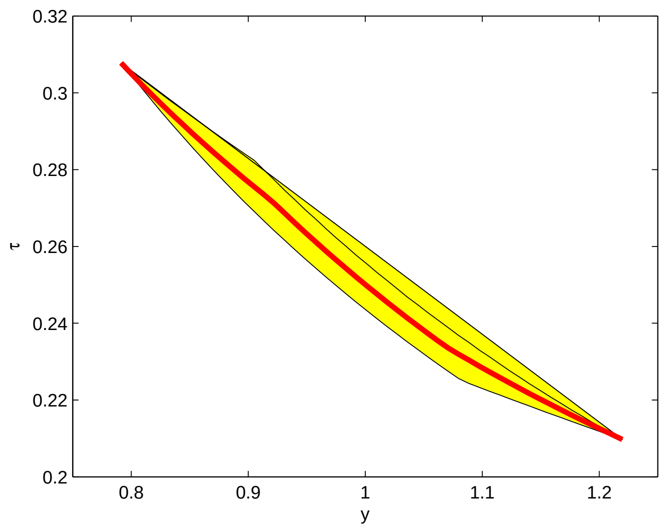
Observable output

Our benchmark example assumed that labor was observable. As a robustness check, we now consider the case when the signal observed by the government is output, so $s = y = \theta l$. The government cannot sort out the values of the θ and l independently, hence there still is a problem of signal extraction as the government can not back out the values of (θ, γ). In addition we mantain the assumptions of linear-quadratic utility and permanent productivity shock.

Figure 17 illustrates the optimal policy for this case. It can be seen that the result is remarkably similar to that obtained in the case with observable labor. However, in this case the government has a lot more information than in the previous case. This is because current revenue $\tau\theta l$ is known, and hence there is no uncertainty about the amount of debt that needs to be issued. The only uncertainty is about the amount of revenue that will be collected in the future, as the value of (permanent) productivity is unknown.[25] As we have seen, uncertainty about debt is key to get fiscal adjustments as in our benchmark example.

[25]Note that if we had assumed that θ_2 is not random the model would not be interesting, since in that case there would be no uncertainty about future revenue.

Figure 17: Optimal Policy with output observed



Optimal tax rate as a function of the signal y (output). Red line: $\mathcal{R}^*$; yellow region: set of equilibrium (l, τ) under FI.