Endogenously (Non-)Ricardian Beliefs*

William Branch
University of California, Irvine

Emanuel Gasteiger
Freie Universität Berlin

April 14, 2018

This paper develops a theory of endogenously (non-)Ricardian beliefs. The foundation of the theory builds on the insights in Eusepi and Preston (2018a, Fiscal Foundations of Inflation: Imperfect Knowledge, American Economic Review, forthcoming), who show that when the private sector has imperfect knowledge about long-run monetary and fiscal policy, in particular whether the path of future taxes suffices to equate the government's intertemporal budget constraint, then Ricardian equivalence can fail even though the policy regime is Ricardian. The novelty here is a restricted perceptions viewpoint: agents, when facing a complex forecasting problem, prefer to forecast aggregate state variables with parsimonious models that are optimal within the restricted class, i.e., a restricted perceptions equilibrium (RPE). A misspecification equilibrium is a refinement of an RPE where the choice of restricted model is endogenous. Two natural forecasting rules arise in which for one rule Ricardian beliefs arise as a self-confirming equilibrium, while the other features an equilibrium with non-Ricardian beliefs. We show that (1.) there can exist misspecification equilibria where beliefs are endogenously (non-)Ricardian, (2.) multiple equilibria exist where the economy can coordinate on either a Ricardian or non-Ricardian equilibrium. The theory suggests a novel interpretation of U.S. inflation data as generated by endogenous belief-driven regime change. We explore this possibility quantitatively and provide evidence of time-varying (non-)Ricardian beliefs. Several counterfactual exercises illustrate a novel and nuanced trade-off in designing monetary policy rules.

*Gasteiger is also affiliated with the Business Research Unit (BRU-IUL) at Instituto Universitário de Lisboa (ISCTE-IUL) and gratefully acknowledges financial support from Fundação para a Ciência e a Tecnologia (UID/GES/00315/2013), the Berlin Economics Research Associates (BERA) program and the Fritz Thyssen Foundation. All remaining errors are the responsibility of the author.

1 Introduction

This paper proposes a theory of expectation formation where Ricardian equivalence, or its failure, arises endogenously as an equilibrium outcome. The theory builds on the imperfect knowledge environment in the seminal Eusepi and Preston (2018a) where individuals and firms have imperfect knowledge about the future path of government debt and how taxes will be adjusted accordingly. Agents hold subjective beliefs over the paths of government debt and the other endogenous state variables. The departure point in this paper is to endow agents with the choice of one of two forecasting models: the first nests Ricardian beliefs within a self-confirming equilibrium, while the other does not. The fact that Ricardian equivalence holds on, but not off, the self-confirming equilibrium path is the key observation in constructing equilibria where Ricardian equivalence fails. The equilibrium concept is a *misspecification equilibrium* where the choice of models is endogenous and agents only select the best performing statistically optimal models. When certain necessary and sufficient conditions are satisfied, beliefs are Ricardian and (self-confirming) Ricardian equivalence is sustained within a misspecification equilibrium. Critically, we demonstrate the possibility of multiple equilibria, where there exists simultaneously (non-)Ricardian beliefs. This latter possibility suggests an alternative interpretation of U.S. inflation data as arising from *belief-driven* regime-shifts.

A long important question in economics is when does Ricardian equivalence hold or fail? The answer to this question is important for the design of monetary and fiscal policy as it can be the difference between inflation being a monetary or fiscal phenomenon. An extensive literature has identified conditions under which there is a "fiscal theory of the price-level" that can arise when Ricardian equivalence fails. For example, Leeper (1991), and the ensuing literature, show that in a policy regime where fiscal policy is non-Ricardian and monetary policy is not committed to price stability, there exists a unique rational expectations equilibrium where government debt becomes an important state variable. Davig and Leeper (2006) provide evidence in favor of a model of inflation driven by regime-switching policy regimes. Recently, Bianchi and Ilut (2017) incorporates imperfect knowledge and learning about the policy regimes to show that mistaken private sector beliefs are critical for generating high inflation rates during the 1970's.[1]

The breakthrough paper by Eusepi and Preston (2018a) opened a new avenue for research into the implications for inflation in an environment where the private sector has imperfect knowledge about long-run fiscal and monetary policies.[2] In these models, individuals have imperfect knowledge about whether the paths for primary surpluses will adjust to satisfy the government's intertemporal budget constraint. Eusepi and Preston (2018a) develop their

[1] See Leeper and Leith (2016) for a recent overview of this literature.

[2] See also Evans et al. (2012), Eusepi and Preston (2012), and Woodford (2013).

insights within a New Keynesian model where individuals and firms have imperfect knowledge but form expectations from a well-specified forecasting model that nests the rational expectations equilibrium. These agents behave like econometricians by estimating the coefficients of their model in real-time. When these estimated coefficients depart from their rational expectations equilibrium values, Ricardian equivalence fails even though the policy regime is Ricardian. Eusepi and Preston (2018a) provide strong empirical evidence in favor of imperfect knowledge as an explanation for observed U.S. inflation and a central role played by non-Ricardian beliefs in high-debt economies. Relatedly, Woodford (2013) imparts to the private-sector a parsimonious, but misspecified, forecasting model for the economy that leads to a *restricted perceptions equilibrium* where Ricardian equivalence fails.

In the present study, we construct an economic environment where whether beliefs are Ricardian or not is determined endogenously within an equilibrium. The basic economic environment is New Keynesian where households in the economy are *ex-ante* identical and derive decision rules from an infinite horizon optimization problem given their subjective beliefs about payoff-relevant aggregate variables. With a Calvo (1983) nominal pricing friction gives rise to the usual New Keynesian IS and AS equations depending on subjective beliefs. Economic policy is given by a pair of linear feedback rules for nominal interest rates and lump-sum taxes. Fiscal policy is Ricardian in that taxes are adjusted to satisfy the government's long-run budget constraint and monetary policy is active and committed to price stability. In a "temporary equilibrium", without imposing private-sector Ricardian beliefs, the aggregate state variables depend on the existing stock of debt and the primary surplus. A forecasting model linear in these variables, as well as the other state variables, nests the rational expectations equilibrium.

We formalize our ideas by taking a step away from rational expectations and adopt a restricted perceptions viewpoint (see, Branch and McGough (2018), Woodford (2013)): individuals formulate expectations from one of two parsimonious forecasting models restricted to include a single fiscal variable as a predictor. In a *restricted perceptions equilibrium* agents' beliefs are optimal within the restricted class implying that, within the context of their model, the agents cannot detect their misspecification. We refine the set of restricted perceptions equilibria by endogenizing the predictor choice within a *misspecification equilibrium* where private sector beliefs come only from those misspecified models that forecast best in a statistical sense. The first model, which includes the existing stock of debt, naturally formalizes endogenous Ricardian beliefs as a self-confirming equilibrium, while the second model, which includes the primary surplus as a predictor, does not. In our theoretical analysis, we provide necessary and sufficient conditions for Ricardian beliefs to emerge as a self-confirming equilibrium and for non-Ricardian beliefs to be sustained in equilibrium.

Our main results show that (non-)Ricardian beliefs arise endogenously as an equilibrium outcome and the data predict that beliefs are heterogeneous, time-varying and non-

Ricardian. We begin by showing that the debt-based forecasting model leads to a restricted perceptions equilibrium where Ricardian equivalence is a self-confirming equilibrium: although, out of equilibrium, the debt forecasting model is misspecified, in equilibrium beliefs about the path for future debt is correct and real variables display a (weak) Ricardian equivalence. Conversely, the surplus-based model leads to a restricted perceptions equilibrium where Ricardian equivalence fails. Contrasting the equilibrium paths, the non-Ricardian path for output and consumption reacts less strongly on impact from a tax innovation but the effect is persistent.

We then turn to our main interest: providing necessary and sufficient conditions for the existence of endogenously (non-)Ricardian beliefs. We accomplish this by focusing on the properties of a *misspecification equilibrium* (Branch and Evans (2006a)) where individuals only select the best performing model. Because of the self-referential features of the model, this predictor choice is endogenous and depends on the distribution of agents across models. Depending on the coefficients in the policy rule and other structural parameters, it is possible for (non-)Ricardian beliefs to arise as the unique misspecification equilibrium. Most interestingly, under certain conditions – consistent with standard parameter estimates – multiple equilibria exist. That is, an economy can coordinate on a Ricardian self-confirming equilibrium or a non-Ricardian equilibrium. The existence of multiple equilibria implies the possibility for real-time learning dynamics that recurrently switch between the basins of attraction for each of the equilibria.

The latter result suggests an alternative, natural interpretation of U.S. inflation data. In particular, multiple equilibria imply that along a real-time learning path – where agents update their model coefficients and choose their models in real-time – the extent of non-Ricardian beliefs can evolve over time endogenously and display *regime-switching beliefs*. The model-predicted paths for the endogenous state variables, including the latent subjective belief states such as the distribution of agents across forecasting models, is estimated using the Extended Kalman Filter on U.S. data for the period 1960-2007:3. The estimates indicate that the data are consistent with a model that switches between mostly Ricardian and non-Ricardian agents. Like Eusepi and Preston (2018a) the model here holds the policy regime constant. The estimates of the latent states suggest that the late 1980's-1990's was a period of non-Ricardian equilibria.

The paper then turns to several counterfactual exercises that further exposit the mechanics of the model. One set of counterfactuals looks at the consequences for the economy if the monetary policy rule was chosen to take a more, or less, hawkish stance. There is a nuanced trade-off faced by policymakers that arises directly because of the theory of expectation formation proposed here. On the one hand, a more hawkish policy rule, all else equal, would have led to more frequent belief regime-switching which would have produced greater economic volatility, offsetting the usual stabilizing effect of a more hawkish policy.[3] On the other hand, a more dovish monetary policy rule would have coordinated the economy on the Ricardian self-confirming equilibrium more often, but with the less aggressive policy response economic volatility also would have been higher.

Is our specification for beliefs reasonable? We argue yes, for three reasons. First, to know that these models are misspecified requires the agents to know the form of the model consistent forecasting equation. However, model consistency here requires that agents hold a great deal of knowledge about the structural features of the economy such the beliefs, constraints, and decision rules of the other agents in the economy including the government. In a complex forecasting environment, especially with a large number of state variables and a potential degrees-of-freedom limitation, it is standard applied econometric practice to formulate parsimonious forecasting models. Here that parsimony takes the form of models that include a single fiscal variable, either government surplus or debt. Moreover, since the debt-model nests Ricardian beliefs as a self-confirming equilibrium, it is a perfectly reasonable predictor for agents to adopt. However, we do not impose this *a priori* and allow agents to select the surplus-model if it produces more accurate forecasts. Thus, the choice of model is an equilibrium outcome. Second, it is a natural way to formalize endogenously (non-)Ricardian beliefs. In the end, we give private-sector agents a choice between a self-confirming Ricardian equilibrium and a non-Ricardian equilibrium. Then, we let the data speak to what extent are beliefs (non-)Ricardian. Finally, the mean dynamics of a correctly specified model show that the non-Ricardian equilibrium can be an "escape" point, therefore, its existence has implications for the dynamics of an economy that does not impose restricted perceptions. The key assumption is that beliefs are potentially misspecified off the equilibrium path. Then learning dynamics can generate endogenous escapes from the self-confirming equilibrium. The paper proceeds as follows. The subsequent Section 2 discusses the related literature followed by Section 3's presentation of the model. Section 4 delivers the primary theoretical results, beginning first with the special analytic case considered in Woodford (2013). Section 5 presents the quantitative analysis, while section 6 concludes.

2 Related literature

This paper is related to a large literature that examines monetary policy design when rational expectations are replaced with an adaptive learning rule. Key contributions include Bullard and Mitra (2002), Evans and Honkapohja (2003), and Preston (2005). Typically models in this literature endow agents with correctly specified forecast models and focus on expectational stability of rational expectations equilibria as an equilibrium refinement and

[3] See Orphanides and Williams (2005) and Eusepi and Preston (2018b) for details on how monetary policy is able to anchor expectations in imperfect knowledge environments.

desirable outcome for monetary policy rules. There has also been research that characterizes fiscal and monetary policy interaction, e.g., Leeper (1991) under adaptive learning, c.f., Evans and Honkapohja (2007) and Branch et al. (2008). Gasteiger (2018) directly extends these frameworks to include heterogeneous expectations, while Eusepi and Preston (2011, 2012) study the implications in a sticky price model. Furthermore, Evans et al. (2012) examine the conditions under which Ricardian equivalence holds or fails when expectations are formed via adaptive learning.

The theory or restricted perceptions proposed here fits into a growing branch in the literature that equips agents with plausibly misspecified forecasting models and proposes equilibria in which beliefs are optimal within the restricted class, see Sargent (1999), Branch (2006), Branch and Evans (2006b), Sargent (2008), and Branch and McGough (2018). This paper builds on an insight from Woodford (2013) where an example of a restricted perceptions equilibrium is considered that leads to a failure of Ricardian equivalence, in particular when agents forecast with the surplus-model even though the policy regime is Ricardian. We extend Branch and Evans (2011) to an environment with fiscal and monetary policy interaction.

The theory is also closely related to Sargent (1999), Cho et al. (2002), and Williams (2018). These papers all study the escape dynamics from self-confirming equilibria. Much of the insight in this paper is related to the escape dynamics models. The dynamics in our model are also closely related to Cho and Kasa (2015) and Cho and Kasa (2017), which make innovations in applying large deviation theory to the problem of private sector model selection. In particular, Cho and Kasa (2017) develop a model of expectation formation where agents have available two forecasting models, one which is self-confirmed in an equilibrium and the other is misspecified on and off the equilibrium path. Rather than selecting a single model, each agent makes forecasts as a Bayesian average of the two forecasting models. They show that it is possible for an asset-pricing model to converge to the restricted perceptions equilibrium with full probability weight assigned to the misspecified model. Here we also have two models, one that can be self-confirmed along an equilibrium path and the other cannot. Our results show that an equilibrium can emerge where everyone has the misspecified beliefs that, in the context of the model presented here, imply non-Ricardian equivalence.

There is a long tradition of constructing equilibria with the property that inflation is (partly) driven by fiscal policy. In his original contribution, Leeper (1991) shows that an active fiscal policy, combined with a monetary policy not committed to price stability, will generate inflation driven by fiscal variables, i.e., the “fiscal theory of the price-level.” See also, Sims (1994), Cochrane (2001) and Woodford (2001). Recent related rersearch explain the *Great Inflation* and the *Volcker Disinflation* via recurrent change between non-Ricardian and Ricardian policy regimes. Examples include Davig and Leeper (2006), Sims (2011), and Bianchi and Ilut (2017). These papers also derive their results from an important role given to non-Ricardian beliefs. When agents assign a positive probability to a change from the Ricardian policy regime to the non-Ricardian policy regime, then the beliefs imply failure of Ricardian equivalence and inflation is, in part, a fiscal phenomenon. This has two implications. First, as discussed in Leeper and Leith (2016), there may be an observational equivalence between the Ricardian and non-Ricardian regimes that makes econometric identification of policy regimes elusive. Thus, it is open whether belief-driven regime-change of the type identified here is a plausible alternative. Finally, the results here do not show that policy regime change is not an important part of the inflation story. In fact more subtle changes, within the Ricardian policy regime, can generate belief-driven regime change.

The theory here is inspired by, and builds on, Eusepi and Preston (2018a) who show that replacing rational expectations with an adaptive learning rule produces temporary equilibrium dynamics that feature departures from Ricardian equivalence. They also estimate a quantitative version of their model and conduct counter-factual analyses that demonstrate that perceived net wealth may be an especially important factor in high debt economies.

3 Model

Woodford (2013) demonstrates that, following Eusepi and Preston (2018a), restricted perceptions about the government's intertemporal budget constraint can lead to a failure of Ricardian equivalence even in instances where policy would be Ricartdian under rational expectations, i.e., active monetary/passive fiscal in the Leeper (1991) sense. We extend the Woodford (2013) framework to see how, and whether, (non-)Ricardian beliefs arise in equilibrium.

The setting is a New Keynesian model with subjective expectations, in particular, a simplified version of Preston (2005), based on Woodford (2003, ch.4).[4] Households and firms have beliefs about payoff-relevant aggregate variables and choose consumption, leisure, and one-period government debt, the only asset available to households, to solve their intertemporal optimization problem given their subjective beliefs about future state variables.[5] In Woodford's framework, households turn over wage-setting and labor supply decisions to a

[4]In Eusepi and Preston (2018a) there are two assets, one period government bonds in zero net-supply and longer maturity bonds. Eusepi and Preston (2018a) demonstrate the important role that maturity structure, combined with imperfect knowledge and learning, can play in generating non-Ricardian wealth effects.

[5]It is typical in boundedly rational learning models to assume that agents optimize intertemporally with the *anticipated utility* approach (see, e.g., Kreps, 1998). This approach assumes that agents take their beliefs as given when solving their optimization problem and, so, if their beliefs evolve along a learning path the agent, when optimizing, is always dogmatic that the learning process has come to an end. In our main theoretical analysis, this assumption is unnecessary because we assume stationary beliefs within the restricted perceptions equilibrium. In the quantitative analysis, when we allow for real-time learning, then we implement anticipated utility.

union and are, therefore, obligated to supply labor to a firm on the union's terms. This is a stylized assumption that renders the consumption rule of the household analogous to the one in an economy where the household receives a stochastic endowment. However, because firms are monopolistically competitive, and face a Calvo (1983) nominal pricing friction, there is endogenous variation in hours and output. This formulation is particularly helpful in the presence of heterogeneous beliefs. Finally, the model is closed by specifying a Taylor-rule for the setting of nominal interest rates and a fiscal rule that passively adjusts the primary surplus to the stock of existing debt.

Households.

Woodford (2013) derives an individual's consumption function,

$$c_t^i = (1-\beta)b_t^i + \sum_{T=t}^{\infty} \beta^{T-t} E_t^i \{(1-\beta)(Y_T - \tau_T) - \beta\sigma(\beta i_T - \pi_{T+1}) + (1-\beta)s_b(\beta i_T - \pi_T) - \beta(\bar{c}_{T+1} - \bar{c}_T)\}, \tag{1}$$

where b_t^i is the individual's holdings of real government debt, Y_t, π_t, i_t, τ_t are aggregate output, the inflation rate, the nominal interest rate (i.e. the monetary policy instrument), and lump-sum taxes (the fiscal instrument). The government also an exogenous sequence G_t of its own private consumption of the good. The parameter $0 < \beta < 1$ is the discount rate, σ is the elasticity of intertemporal substitution, and $s_b \equiv \bar{b}/\bar{Y}$ is the steady-state debt-to-GDP ratio. The random variable $\bar{c}_t$ is a preference shock. All variables are in log deviation from the zero-inflation steady state. Following Woodford, we find it convenient to represent the fiscal policy instrument in terms of the surplus $s_t = \tau_t - G_t$.

The first two terms in (1) dictate how consumption responds to government bond holdings and disposable income, respectively. The first term is sometimes called a "wealth effect". The third term, parameterized by σ, captures an intertemporal substittution effect resulting from variations in the (perceived) *ex-ante* real interest rate. The fourth term, pre-multiplied by s_b, is the perceived real return on government bond holdings. Woodford (2013) describes this term as an "income effect." Note that from the final term that a positive preference shock, $\bar{c}_t$, implies a stronger desire for contemporaneous consumption.

Following the seminal Eusepi and Preston (2018a), Woodford (2013) derives equation (1) without assuming that individuals have structural knowledge about the government's intertemporal budget constraint. Even though fiscal policy is set passively, individuals do not necessarily know this as well as have imperfect knowledge about the structural form of the government's endogenously determined budget constraint. Instead, they form subjective beliefs over the evolution of aggregate variables. If they get those beliefs right then they will properly account for the evolution of debt and beliefs will be Ricardian. Otherwise, beliefs may be non-Ricardian. Eusepi and Preston (2018a) demonstrate that along an adaptive learning path, beliefs will temporarily diverge from Ricardian equivalence.

We are now ready to formally define Ricardian and non-Ricardian beliefs. Both Eusepi and Preston (2018a) and Woodford (2013) define Ricardian private sector beliefs as being consistent with the government's intertemporal budget constraint. In particular, Ricardian beliefs arise when the following condition on *beliefs* is satisfied

$$E_t^i \left\{ \sum_{T=t}^{\infty} \beta^{T-t} \left[s_T - s_b(\beta i_T - \pi_T) \right] \right\} = b_t.$$

Then imposing Ricardian beliefs onto the consumption rule (1) leads to a consumption function that satisfies Ricardian equivalence:

$$c_t^i = \sum_{T=t}^{\infty} \beta^{T-t} E_t^i \{ (1-\beta)(Y_T - g_T) - \beta\sigma(\beta i_T - \pi_{T+1})$$

where $g_t = G_t + \bar{c}_t$ is a composite consumption shock.

On the other hand, with non-Ricardian beliefs the path of future surpluses has a direct effect on consumption:

$$\begin{aligned} c_t^i = & \sum_{T=t}^{\infty} \beta^{T-t} E_t^i \{ (1-\beta)(Y_T - g_T) - \beta\sigma(\beta i_T - \pi_{T+1}) \} \\ & (1-\beta) b_t^i + \sum_{T=t}^{\infty} \beta^{T-t} E_t^i \{ (1-\beta) s_b(\beta i_T - \pi_T) - s_T \} \}, \end{aligned}$$

Evidently, non-Ricardian beliefs lead households to perceive holdings of government debt as real wealth and a change in the expected path for future surpluses can have a real effect on consumption. In our theory, we posit two forecasting models that, in equilibrium, will differ in whether beliefs are Ricardian or not.

One can rearrange terms in (1) so that

$$c_t^i = (1-\beta) b_t^i + (1-\beta)(Y_t - \tau_t) - \beta[\sigma - (1-\beta)s_b] i_t - (1-\beta) s_b \pi_t + \beta \bar{c}_t + \beta E_t^i v_{t+1}^i, \quad (2)$$

where the subjective composite variable v_t^i is defined as

$$v_t^i \equiv \sum_{T=t}^{\infty} \beta^{T-t} E_t^i \{ (1-\beta)(Y_T - \tau_T) - [\sigma - (1-\beta)s_b](\beta i_T - \pi_T) - (1-\beta)\bar{c}_T \}$$

This variable comprises all payoff-relevant aggregate variables over which a household formulates subjective beliefs.

Following Woodford (2013), express v_t^i recursively as

$$v_t^i = (1-\beta)(Y_t - \tau_t) - [\sigma - (1-\beta)s_b](\beta i_t - \pi_t) - (1-\beta)\bar{c}_t + \beta \widehat{E}_t^i v_{t+1}^i, \tag{3}$$

Rather than needing to specify beliefs about each of the aggregate variables that comprise v^i, the agent just needs to forecast this subjective continuation-value variable.[6]

Firms.

Firms are monopolistically competitive and face a nominal pricing friction based on Calvo (1983). An individual firm j produces a differentiated good, but can only optimally (re-)set its price in period t, if it belongs to the share $0 < \alpha < 1$ of firms that receives the idiosyncratic signal to reset the price.

The implied aggregate inflation dynamics are given by

$$\pi_t = (1-\alpha)p_t^*, \tag{4}$$

where $p_t^* \equiv \int p_t^*(j)dj$ is the average of the price changes compared to price level p_{t-1} by firms that can reset prices and $p_t^*(j)$ denotes the individual price change of firm j.

A firm j that can optimally reset price $p_t^*(j)$ will do so to satisfy the first-order condition

$$p_t^*(j) = (1-\alpha\beta)\sum_{T=t}^{\infty}(\alpha\beta)^{T-t}\left(E_t^j p_T^{\text{opt}} - p_{t-1}\right),$$

where $E_t^j p_T^{\text{opt}}$ is the perceived optimal price in period T. This condition can be written recursively:

$$p_t^*(j) = (1-\alpha\beta)\left(E_t^j p_t^{\text{opt}} - p_{t-1}\right) + (\alpha\beta)E_t^j p_{t+1}^*(j) + (\alpha\beta)\pi_t, \qquad \text{where} \tag{5}$$
$$E_t^j p_{t+1}^*(j) \equiv (1-\alpha\beta)\sum_{T=t}^{\infty}(\alpha\beta)^{T-t}\left(E_t^j p_{T+1}^{\text{opt}} - p_t\right).$$

[6] On the surface, formulating expectations over future v_t^i seems to be adopting the Euler equation approach of one-step ahead forecasting and decision-making. However, the derivation of the consumption function and v_t^i is based on the infinite-horizon approach where the household's consumption/savings decisions solve their entire sequence of Euler equations, flow budget constraints, and transversality condition given their subjective. We show below how these consumption rules can be aggregated with heterogeneous agents.

Policy.

Monetary policy is described by a Taylor (1993) rule,

$$i_t = \phi_\pi \pi_t + \phi_y y_t + w_t, \tag{6}$$

where the coefficients are assumed to be $\phi_\pi, \phi_y > 0$ and the monetary policy shock is $w_t \sim \text{iid}(0, \sigma_w^2)$. In the theoretical analysis, we assume white noise shocks. However, in the quantitative analysis we assume all exogenous shocks are stationary AR(1)'s.

Fiscal policy is characterized by a rule for the real primary surplus, $s_t \equiv \tau_t - G_t$, where G_t measures (exogenous) government purchases. Fiscal policy follows the rule

$$s_t = \phi_b b_t + z_t, \tag{7}$$

where the surplus shock is $z_t \sim \text{iid}(0, \sigma_z^2)$. This is a standard fiscal policy rule since Leeper (1991).

Finally, the government faces a flow budget constraint that, linearized around the steady-state, can be written as

$$b_{t+1} = \beta^{-1}[b_t - s_b \pi_t - s_t] + s_b i_t. \tag{8}$$

Notice that, as b_t is predetermined, this specification of monetary and fiscal policy nests the classic policy set-up of Leeper (1991). The parameter s_b plays an important role in the results presented below. When $s_b = 0$ the bond and primary surplus paths are exogenous while $s_b > 0$ implies that they are endogenous and affected, in part, by monetary policy.[7]

Throughout the paper, analysis is restricted to the *active* monetary (AM) and *passive* fiscal (PF) fiscal policy regime

$$\begin{gathered} 1 < \phi_\pi + \frac{1-\beta}{\kappa}\phi_y \\ (1-\beta) < \phi_b < 1. \end{gathered} \tag{9}$$

Under the benchmark *rational expectations hypothesis*, there is local determinacy (see, Leeper, 1991) implying that this locally unique *rational expectations equilibrium* displays Ricardian equivalence and is stable under least-squares learning (see, Evans and Honkapohja, 2007). We follow Eusepi and Preston (2012) in assuming a constant policy regime that is consistent with Ricardian equivalence under rational expectations as a stark example of the potentially

[7]This formulation arises in a cashless environment that allows us to abstract from the effect of monetary aggregates appearing in the consolidated budget constraint.

important role of endogenously time-varying Ricardian beliefs.

MARKET CLEARING AND EQUILIBRIUM.

Aggregate demand is given by the income- expenditure identity, i.e.,

$$Y_t = \int c_t^i di + G_t. \tag{10}$$

Applying (2) and (3) to (10) allows us to express aggregate demand as

$$Y_t = g_t + (1-\beta)b_t + v_t - \sigma\pi_t, \tag{11}$$

where a composite exogenous disturbance $g_t \equiv \bar{c}_t + G_t$, such that $g_t \sim \text{iid}(0, \sigma_g^2)$ and bond market clearing requires that the aggregate supply of one-period government bonds $b_t \equiv \int b_t^i di$ are defined accordingly.[8] Similarly, for the subjective state variable, i.e., $v_t = \int v_t^i di$.

Notice that we can use (8) and (11) to express the composite variable defined in (3) as

$$v_t^i = (1-\beta)v_t + (1-\beta)\beta(b_{t+1} - b_t) - \beta\sigma(i_t - \pi_t) + \beta E_t^i v_{t+1}^i. \tag{12}$$

Averaging over expectations in (12) and applying the result to (11) yields the "IS equation" without *a priori* imposing Ricardian beliefs:

$$Y_t = g_t - \sigma i_t + (1-\beta)b_{t+1} + \int E_t^i v_{t+1}^i di. \tag{13}$$

Next, as in Woodford (2013, Section 2.3), in equilibrium the optimal price in this model can be expressed as

$$p_t^{\text{opt}} = p_t + \xi\left(Y_t - Y_t^n\right) + \mu_t, \tag{14}$$

where $\xi > 0$ is a composite term of structural parameters measuring the output elasticity of a firm's optimal price.[9] The exogenous random variable Y_t^n is the natural level of output that captures exogenous demand shocks and μ_t represents disturbances to the desired markup over marginal cost.

As the firm's price is a decision variable, it is natural to impose that $E_t^j p_t^{\text{opt}} = p_t^{\text{opt}}$. It follows, then, from plugging (14) and (4) into (5) that

$$p_t^*(j) = (1-\alpha)p_t^* + (1-\alpha\beta)\left[\xi y_t + \mu_t\right] + \alpha\beta E_t^j p_{t+1}^*(j), \tag{15}$$

[8] Below, we discuss market clearing with heterogeneous beliefs in greater detail.
[9] The term is defined in Woodford (2003, ch.3-4).

Again averaging across firms, using (4), defining the output gap as $y_t \equiv Y_t - Y_t^n$, parameter $\kappa \equiv [(1-\alpha)(1-\alpha\beta)\xi]/\alpha$, and the cost-push supply shock as $u_t \equiv \{[(1-\alpha)(1-\alpha\beta)]/\alpha\}\mu_t$, yields the New Keynesian Phillips curve

$$\pi_t = (1-\alpha)\beta \int E_t^j p_{t+1}^*(j)dj + \kappa y_t + u_t, \tag{16}$$

Let us next define a *temporary equilibrium* (TE) following Woodford (2013).

DEFINITION 1. *Given exogenous processes* $\{g_t, u_t, w_t, z_t\}$*, a temporary equilibrium, given subjective expectations, is a stochastic process for the triple* $\{\pi_t, y_t, b_{t+1}\}$*, where the latter satisfy* (8)*,* (16)*,* (13)*,* $\{p_t^*(j), v_t^i\}$ *satisfy* (12)*-*(15)*, and a path for policy* $\{i_t, s_t\}$ *given by rules* (6) *and* (7)*.*

Below, we refine the equilibrium by a theory of subjective expectations that are determined within a *restricted perceptions equilibrium* and a further refinement, *misspecification equilibria*, that pins down the distribution of (potentially) heterogeneous beliefs.

Model misspecification.

A key insight from Eusepi and Preston (2018a) is that in environments where people have imperfect knowledge of the economic data generating process it is not reasonable to impose *a priori* that beliefs by individual i about future spending and taxes satisfies

$$E_t^i \left\{ \sum_{T=t}^{\infty} \beta^{T-t} \left[s_T - s_b(\beta i_T - \pi_T) \right] \right\} = b_t.$$

Theoretically, we construct equilibria, and provide conditions under which, beliefs along an equilibrium path may, or may not, satisfy this Ricardian belief condition.

Under *homogeneous* rational expectations, the equilibrium laws of motion are given by the equations:

$$\begin{aligned}
b_{t+1} &= \beta^{-1}(b_t - s_b \pi_t - s_t) + s_b i_t \\
y_t &= v_t - \sigma \pi_t + g_t + (1-\beta) b_t \\
v_t &= (1-\beta)(b_{t+1} - b_t) - \sigma(i_t - \pi_t) + E_t v_{t+1} \\
\pi_t &= \kappa y_t + (1-\alpha)\beta E_t p_{t+1}^* + u_t \\
s_t &= \phi_b b_t + w_t \\
i_t &= \phi_\pi \pi_t + \phi_y y_t + z_t
\end{aligned}$$

It is straightforward to verify, therefore, that under full information rational expectations that the equilibrium law of motion is of the form

$$\begin{bmatrix} \pi_t \\ v_t \\ y_t \end{bmatrix} = A \begin{bmatrix} b_t \\ s_t \end{bmatrix} + \eta_t$$

where η_t is a vector of composite disturbances and A is conformable. It follows that in order to formulate rational expectations they can adopt linear forecast rules that depend on both the stock of beginning of period debt, b_t, and the primary surplus, s_t. Alternatively, if agents formulated expectations with a set of rules that depend on end of period debt, b_{t+1}, the same (unique) rational expectations equilibrium would be achieved.

The bounded rationality assumption that we make in this paper, following in the footsteps of a burgeoning adaptive learning literature, is that the agents in this economy make econometric forecasts of payoff-relevant aggregate state variables whose perceived paths the agents take as given and beyond their control. Thus, when it comes to formulating expectations the individuals and firms seek to *optimize* their forecasts given their information and abilities. Our modeling strategy is to restrict that information and abilities of the agents so that they favor parsimonious forecasting models. Imagining the individuals and firms in our economy as an economic forecaster disciplined by rational expectations, they would formulate and estimate VAR forecasting models in b_t, s_t and any serially correlated shocks. However, knowing the model consistent forecasting model involves knowing the government's intertemporal constraint, as well as the expectations of the other agents. Following the adaptive learning literature, it is reasonable to replace rational expectations with subjective expectations formulated from a well-specified forecasting model. In many environments, however, forecasters prefer parsimonious models, face degrees of freedom limitations, and often use simpler models with a restricted set of regressors.

Following this logic, the key assumption in this paper is that agents will formulate their expectations from one of two restricted models, each of which includes a *single fiscal variable*: s_t or b_t. We could specify this parsimony in other ways, of course, but this approach is particularly interesting. As Woodford (2013) shows in a particular case (see below), when agents' restricted variable includes s_t, but not b_t, then Ricardian equivalence does not hold in the restricted perceptions equilibrium. When the restricted variable includes b_t and not s_t, as we show below, this leads to a weak form of Ricardian equivalence which, importantly, arises as a self-confirming equilibrium.

These two restricted forecasting models are natural ways to formalize endogenously (non-)Ricardian. We do this as follows. All individuals and firms make a choice about which fiscal variable to include in their forecasts. The coefficients of the restricted forecasting models are derived from the optimal linear projection of the aggregate state variables onto the restricted space of regressors, all of which is determined jointly in a *restricted perceptions equilibrium*

(RPE). In a *misspecification equilibrium* (ME), the distribution of the population across the two possible forecasting models is endogenous and determined by a discrete choice between models. If a misspecification equilibrium exists where this distribution is massed on the b_t-forecasting model then we say that there are *endogenously Ricardian beliefs*. On the other hand, if a misspecification equilibrium exists where all agents use the s_t-forecasting model then there are *endogenously non-Ricardian beliefs*. In the theoretical analysis, we explore these equilibria including the possibility of heterogeneous expectations and multiple equilibria. In the quantitative analysis, we explore whether real-time learning can lead to fluctuations in the proportion of the population with non-Ricardian beliefs and assess whether this can be a candidate explanation for recent U.S. macroeconomic data.

Specifically, expectations are formed from one of the following forecasting models, sometimes called perceived laws of motion (PLM):

$$\begin{aligned} PLM_s: \mathbf{Z}_t &= \boldsymbol{\psi}^s s_{t-1} + \eta_t \Rightarrow E_t^s \mathbf{Z}_{t+1} = \boldsymbol{\psi}^s s_t \\ PLM_b: \mathbf{Z}_t &= \boldsymbol{\psi}^b b_{t-1} + \eta_t \Rightarrow E_t^b \mathbf{Z}_{t+1} = \boldsymbol{\psi}^b b_t \end{aligned}$$

where $\mathbf{Z}_t' = (v_t, p_t^*, b_{t+1})$, η_t is a perceived noise, and the coefficient matrix, for $k = s, b$,

$$\boldsymbol{\psi}^k = \left(\psi^k, \Gamma^k\right)'$$

and $\psi^k = \left(\psi_v^k, \psi_{p^*}^k\right)'$. In a *restricted perceptions equilibrium* (RPE) the coefficients will satisfy the least-squares orthogonality condition:

$$E x_t^k \left(\mathbf{Z}_t - \boldsymbol{\psi}^k x_t^k\right) = 0$$

with $x_t^k \in \{s_t, b_t\}$. Beliefs, parameterized by $\boldsymbol{\psi}^k$, are derived from the optimal projection of the aggregate variables $\mathbf{Z}_t$ onto the restricted explanatory variable $\boldsymbol{\psi}^k$. It follows that

$$\boldsymbol{\psi}^k = \left[E\left(x_t^k\right)^2\right]^{-1} E\mathbf{Z}_t x_t^k \equiv S\left(\boldsymbol{\psi}^k\right)$$

DEFINITION 2. *A restricted perceptions equilibrium is a fixed point* $\boldsymbol{\psi}_*^k = S\left(\boldsymbol{\psi}_*^k\right)$.

We do not impose *a priori* which of the two PLM's that individuals and firms use to formulate expectations. Instead, we endow each agent with a discrete choice: they can forecast with the s_t-model or the b_t-model, and like the selection of model parameters, they will do so to minimize their forecast errors. We adopt the rationally heterogeneous expectations approach first pioneered by Brock and Hommes (1997). Agents make a predictor selection in a random-utility setting and in the limit of vanishingly small noise the agents will only select the best-performing statistical models. This approach pins down the distribution of agents in an equilibrium as follows.

Let n denote the fraction of agents who have selected model-s, leaving $1-n$ of the population forecasting with model-b.[10] They can rank these choices by calculating the relative mean square error (MSE):

$$EU^i = -E\left[(\mathbf{Z}_t - E_t^i[\mathbf{Z}_t^i])\right]' \times \mathbf{W} \times E\left[(\mathbf{Z}_t - E_t^i[\mathbf{Z}_t^i])\right], \tag{17}$$

where $\mathbf{W}$ is a weighting matrix.[11] Consequently, we define relative predictor performance $F(n) : [0,1] \to \mathbb{R}$ as $F(n) \equiv EU^1 - EU^2$.

Building on Brock and Hommes (1997), we assume that the distribution of agents across the two forecasting models, n, is pinned down according to the multinomial logit (MNL) map (see, e.g., Branch and Evans, 2011)

$$n = \frac{1}{2}\left\{\tanh\left[\frac{\omega}{2}F(n)\right] + 1\right\} \equiv T_\omega(n),$$

where ω denotes the "intensity of choice". The MNL map relates the fraction of agents adopting model-s, n, is an increasing function of its relative forecast accuracy, measured by the function $F(n)$.

DEFINITION 3. *A* misspecification equilibrium *is a fixed point* $n_* = T_\omega(n_*)$.

The neoclassical case $\omega \to \infty$ warrants special attention as, in this case, the agents only select the best-performing statistical models. The following proposition, an immediate consequence of the continuity of $T_\omega : [0,1] \to [0,1]$, details the range of *possible* equilibrium outcomes.

PROPOSITION 1. *Let* $N_*(\omega) = \{n_* | \ n_* = T_\omega(n_*)\}$ *denote the set of misspecification equilibria. As* $\omega \to \infty$, N_* *has one of the following properties:*

1. *If* $F(0) < 0$ *and* $F(1) < 0$ *then* $n_* = 0 \in N_*$.
2. *If* $F(0) > 0$ *and* $F(1) > 0$ *then* $n_* = 1 \in N_*$.
3. *If* $F(0) < 0$ *and* $F(1) > 0$ *then* $\{0, \hat{n}, 1\} \subset N_*$ *, where* $\hat{n} \in (0,1)$ *is such that* $F(\hat{n}) = 0$.
4. *If* $F(0) > 0$ *and* $F(1) < 0$ *then* $n_* = \hat{n} \in N_*$*, where* $\hat{n} \in (0,1)$ *is such that* $F(\hat{n}) = 0$.

[10] For simplicity, we assume that households and firms are distributed across models identically. This is a simplification that could be relaxed as follows. Instead, there could be a distribution n_h of households across models and a fraction n_f of firms. One can imagine that the simplification arises in an environment where households and managers of firms occupy the same residence and, thereby, share their expectations each morning, say, over breakfast.

[11] The main results do not depend heavily on the weights, so for simplicity we set $W = I$.

The first condition in Proposition 1 implies that the b-model forecasts best when all of the agents use model-b or if they all use model-s; $n_* = 0$ is evidently a misspecification equilibrium in such cases. Conversely, when $F(0) > 0$ and $F(1) > 0$ then $n_* = 1$ is a misspecification equilibrium. Outside of these polar cases, there is also the possibility of multiple misspecification equilibria, $n = 0, \hat{n}, 1$, for some $0 < \hat{n} < 1$. This case will make repeated appearances in various points of the remainder of this paper. As we will see, the $n = 0$ misspecification equilibrium can be thought of as a self-confirming equilibrium with *weakly Ricardian beliefs* and the $n = 1$ will correspond to *non-Ricardian beliefs*.[12] The multiple equilibria case is particularly interesting because it implies that a real-time learning version of the model may feature endogenous regime-switching in and out of Ricardian equilbria. A possibility we establish below and explore in the quantitative analysis.

Temporary equilibrium with heterogeneous beliefs.

Having introduced the temporary equilibrium equations and the nature of heterogeneous beliefs, we turn now to the aggregation of the temporary equilibrium with heterogeneous beliefs. Begin with the consumption function for a household of type i:

$$c_t^i = (1-\beta)b_t^i + \bar{c}_t - \sigma\pi_t + v_t^i$$

where

$$v_t^i = E_t^i \sum_{T \geq t} \beta^{T-t} \left\{(1-\beta)(Y_T - \tau_T) - [\sigma - (1-\beta)s_b]\,(\beta i_T - \pi_T) - (1-\beta)\bar{c}_T\right\}$$

It is useful to define the recursion

$$\hat{v}_t = (1-\beta)(Y_t - \tau_t) - [\sigma - (1-\beta)s_b]\,(\beta i_t - \pi_t) - (1-\beta)\bar{c}_t + \beta\hat{v}_{t+1}$$

Therefore, after simplying and imposing the government budget constraint, we have

$$v_t^i = E_t^i \hat{v}_t = (1-\beta)(Y_t - \tau_t) - [\sigma - (1-\beta)s_b]\,(\beta i_t - \pi_t) - (1-\beta)\bar{c}_t + \beta E_t^i \hat{v}_{t+1}$$

Now we can derive the recursion for the aggregate $v_t \equiv \int v_t^i di$. With two different expectations-types, a fraction n of "type-1" and $1-n$ of "type-2" It follows that

$$\begin{aligned} v_t &= n v_t^1 + (1-n)v_t^2 \\ &= (1-\beta)v_t + (1-\beta)\beta(b_{t+1} - b_t) - \beta\sigma(i_t - \pi_t) + \beta n E_t^1 \hat{v}_{t+1} + (1-n)E_t^2 \hat{v}_{t+1} \\ &= (1-\beta)(b_{t+1} - b_t) - \sigma(i_t - \pi_t) + \hat{E}_t \hat{v}_{t+1} \end{aligned}$$

[12]For discussion of self-confirming equilibria see Sargent (1999). A self-confirming equilibrium is a stronger concept than RPE as it requires that agents' beliefs are correct in equilibrium, though they may be misspecified off the equilibrium path.

where $\hat{E} = nE^1 + (1-n)E^2$ is an aggregate expectations operator. Notice, though, that

$$\begin{aligned}\hat{E}_t \hat{v}_{t+1} &= \hat{E}_t \left[(1-\beta)v_{t+1} + (1-\beta)\beta(b_{t+2} - b_{t+1}) - \beta\sigma(i_t - \pi_t) + \beta\hat{v}_{t+2}\right] \\ &= \hat{E}_t v_{t+1}\end{aligned}$$

where the last equality depends on the law of iterated expectations holding at the aggregate level. This needs to be a part of the assumptions on beliefs. There is, therefore, a natural recursion for aggregate v:

$$v_t = (1-\beta)(b_{t+1} - b_t) - \sigma(i_t - \pi_t) + \hat{E}_t v_{t+1}$$

A similar argument shows that $E_t^1 \hat{v}_{t+1} = E_t^1 v_{t+1}$. So that we can write

$$v_t^i = (1-\beta)v_t + (1-\beta)\beta(b_{t+1} - b_t) - \beta\sigma(i_t - \pi_t) + \beta E_t^i v_{t+1}$$

DEFINITION 4. *Given beliefs $E_t^i v_{t+1}, E_t^i p_{t+1}^{*i}$ a temporary equilibrium is a triple (b_{t+1}, π_t, y_t) and a policy (s_t, i_t) so that the bond and goods markets clear and the government budget constraint is satisfied. In particular, the following equations are satisfied*

$$\begin{aligned}b_{t+1} &= \beta^{-1}\left[b_t - s_b\pi_t - s_t\right] + s_b i_t \\ \pi_t &= (1-\alpha)\beta\hat{E}_t p_{t+1}^{*i} + \kappa y_t + u_t \\ y_t &= g_t - \sigma i_t + (1-\beta)b_{t+1} + \hat{E}_t v_{t+1}\end{aligned}$$

where

$$v_t = (1-\beta)(b_{t+1} - b_t) - \sigma(i_t - \pi_t) + \hat{E}_t v_{t+1}$$

4 Theoretical results

4.1 A simple example

We begin with a special case, first exposited by Woodford (2013), which reduces the fiscal variables, b_t, s_t, to follow exogenous processes, $\pi_t = 0$ for all t, and households have a simple permanent income problem to solve. The case under consideration here sets $\phi_y = s_b = \kappa = 0$ and $\alpha = 1$. We further shut down all of the exogenous disturbances except for the fiscal shock z_t. In the next subsection, we relax all of these parameter restrictions except $s_b = 0$. The case where $s_b > 0$ requires numerical analysis.

4.1.1 Restricted perceptions equilibria

We begin by characterizing the restricted perceptions equilibria with an exogenous distribution n. In the sequel, we endogenize n within a misspecification equilibrium. In this special case of the model, households need only forecast the continuation variable v_{t+1}. Consequently, agents' forecasts are projections from one of the following two regression models

$$\begin{bmatrix} v_t \\ b_t \end{bmatrix} = \begin{bmatrix} \psi^s \\ \Gamma_b^s \end{bmatrix} s_{t-1} + \eta_t^s \Rightarrow \begin{bmatrix} E_t^1 v_{t+1} \\ E_t^1 b_{t+1} \end{bmatrix} = \begin{bmatrix} \psi^s \\ \Gamma_b^s \end{bmatrix} s_t \tag{18}$$

$$\begin{bmatrix} v_t \\ b_t \end{bmatrix} = \begin{bmatrix} \psi^b \\ \Gamma_b^b \end{bmatrix} b_{t-1} + \eta_t^b \Rightarrow \begin{bmatrix} E_t^2 v_{t+1} \\ E_t^2 b_{t+1} \end{bmatrix} = \begin{bmatrix} \psi^b \\ \Gamma_b^b \end{bmatrix} b_t \tag{19}$$

In a restricted perceptions equilibrium, the coefficients in (18) and (19) are optimal, i.e., they satisfy the least-squares orthogonality conditions

$$E[s_t \eta_{t+1}^s] = 0 \tag{20}$$

$$E[b_t \eta_{t+1}^b] = 0 \tag{21}$$

A brief remark about a timing assumption. Here, we follow Woodford (2013), in assuming that agents project the state variables onto the *lagged* regressors. We could alternatively assume that they regress the state onto *contemporaneous* regressors and it would not greatly impact the equilibrium results. However, the timing convention followed here has two benefits. First, it simplifies many of the analytic expressions. Second, in the quantitative analysis below, we implement a real-time learning version of the model and the timing avoids a potential multicollinearity problem.

Weak Ricardian equivalence.

We first show that when $n = 0$, i.e., all agents forecast with the b-model, then Ricardian Equivalence holds. This is even though we do not impose *a priori* that beliefs are Ricardian, but with this PLM agents will hold Ricardian beliefs *along* an equilibrium path, even though their beliefs are misspecified out of equilibrium. Thus, Ricardian equivalence here is a self-confirming equilibrium in the sense of Sargent (1999), Cho et al. (2002), and Williams (2018).

With homogeneous expectations, in this simple case, the key model equations are:

$$\begin{aligned} b_{t+1} &= \beta^{-1}(b_t - s_t) \\ y_t &= v_t + (1-\beta) b_t \\ v_t &= (1-\beta)(b_{t+1} - b_t) + E_t^j v_{t+1} \\ s_t &= \phi_b b_t + w_t \end{aligned}$$

With $n = 0$, $E_t^2 v_{t+1} = \psi^b b_t$, and the actual law of motion is

$$\begin{aligned} v_t &= (1-\beta)(b_{t+1} - b_t) + \psi^b b_t \\ &= (1-\beta)\beta^{-1}(b_t - s_t) + \psi^b b_t \end{aligned}$$

One can find ψ^b, Γ^b by solving the pair of orthogonality conditions

$$\begin{aligned} E\left(b_{t+1} - \Gamma^b b_t\right) b_t &= 0 \\ E\left(v_{t+1} - \psi^b b_t\right) b_t &= 0 \end{aligned}$$

Straightforward algebra leads to

$$\Gamma^b = \frac{E[b_{t+1}b_t]}{E[b_t^2]} = \beta^{-1}(1-\phi_b)$$

It follows that

$$\psi^b = \frac{E v_{t+1} s_t}{E b_t^2} = \left[(\beta^{-1}-1)(1-\beta-\phi_b) + \psi^b\right]\Gamma_b^b$$

$$\Leftrightarrow \psi^b = \frac{(1-\beta)(1-\beta-\phi_b)\Gamma^b}{\beta(1-\Gamma^b)} = -\beta^{-1}(1-\beta)(1-\phi_b) \tag{22}$$

The following result is a direct consequence.

PROPOSITION 2. *In the special parametric case, if all agents form expectations from the b-model* (19)*, then there exists a unique restricted perceptions equilibrium with*

$$y_t = -(\beta^{-1}-1)z_t \tag{23}$$

$$\psi^b = -\beta^{-1}(1-\beta)(1-\phi_b). \tag{24}$$

All proofs are in the Appendix. Proposition 2 provides a *weak Ricardian equivalence* result: by forecasting with the most recently observed b_t instead of b_{t+1}, then innovations to the surplus only have a contemporaneous impact on output. Thus, only unexpected fiscal innovations matter for the real economy. This will differ from the $0 < n \leq 1$ cases where the beliefs of the agents do not just so happen to coincide approximately to the rational expectations equilibrium. In these cases, Ricardian beliefs do not arise endogenously. This first polar case of $n = 0$ shows that Ricardian beliefs can arise endogenously and then Ricardian equivalence **holds**, as with rational expectations, under a passive fiscal/active monetary policy regime. It is important to note here that Ricardian equivalence is stronger than an RPE, it arises as part of a self-confirming equilibrium where beliefs are consistent with the government's intertemporal budget constraint on, but not off, the equilibrium path.

Endogenously non-Ricardian beliefs.

Now consider the case $n = 1$, where all households forecast with the s-model. This is exactly the example presented in Woodford (2013, Section 4.1.2.). Using the same set of equilibrium equations as the previous example, it follows that expectations are $E_t^1 v_{t+1} = \psi^s s_t$, and the actual law of motion is

$$\begin{aligned} v_t &= (1-\beta)\left(b_{t+1} - b_t\right) + \psi^s s_t \\ &= (1-\beta)\beta^{-1}\left(b_t - s_t\right) + \psi^s s_t \end{aligned}$$

Solving the corresponding least-squares orthogonality conditions pin down the RPE values for the belief coefficients:

$$\Gamma^s \equiv \frac{E[b_{t+1}s_t]}{E[s_t^2]} = \frac{-\beta^{-1}(1-\beta^2-\phi_b)}{(1-\beta^2-2\phi_b)}, \tag{25}$$

and,

$$\begin{aligned} \psi^s &= \left[(\beta^{-1}-1)(1-\beta-\phi_b) + \psi^s\phi_b\right]\Gamma^s \\ &= -\frac{\beta^{-1}(1-\beta)(1-\beta^2-\phi_b)}{(\beta+\beta^2+\phi_b)} \end{aligned} \tag{26}$$

Notice, in particular, that the expression for ψ^s implies that $\psi^s < \beta^{-1} - 1$.

PROPOSITION 3. *In the special parametric case of the model, if agents form expectations from the s-model* (18)*, then there exists a unique restricted perceptions equilibrium with*

$$y_t = \left[\frac{(1-\beta)(1+\beta-\phi_b)}{\beta(1+\beta)+\phi_b}\right] b_t - \left[\frac{\beta^{-1}-\beta}{\beta\left(1+\beta\right)+\phi_b}\right] z_t \tag{27}$$

$$\psi^s = -\frac{\beta^{-1}(1-\beta)(1-\beta^2-\phi_b)}{(\beta+\beta^2+\phi_b)} < \beta^{-1} - 1. \tag{28}$$

Proposition 3 replicates the result in Woodford (2013, Section 4.1.2.). The solution for y_t, (27), shows that a fiscal policy shock, z_t, has both transitory and persistent real effects, as y_t depends on both b_t and z_t. Below, we dive deeper into comparing and contrasting the effect of fiscal shocks across equilibria.

The failure of Ricardian equivalence with the s-model is due to the forecast errors by the agents using the s-forecasting model. It is important to stress, however, that within the context of their restricted model the agents cannot detect their misspecification while the economy is in a restricted perceptions equilibrium. This finding is related to Eusepi and Preston (2018a), who find the possible failure of Ricardian Equivalence in an economy that is comparable but with a different theory of expectation formation. In particular, the agents in Eusepi-Preston are learning about the long-run aspects of monetary and fiscal policy and the failure of Ricardian equivalence arises because of temporary variations in agents' econometric learning rule. In this paper, the failure of Ricardian equivalence can arise as an equilibrium outcome. This point is more salient when we consider the misspecification equilibrium refinement and show that, under certain conditions, agents will prefer the non-Ricardian to the Ricardian model.

Heterogeneous beliefs.

The above results are based on homogeneous expectations with (potentially) misspecified PLMs (18) or (19). Next we introduce *extrinsic* heterogeneity in expectations with aggregate expectations derived from a weighted average of the two forecast models (18) and (19). Assume now that a fraction $n \in [0,1]$ of agents use (18) and $1-n$ use (19). Our objective is to show that, depending on the distribution n, (non-)Ricardian equilibria can emerge.

We have

$$\hat{E}_t[v_{t+1}] = nE_t^1 v_{t+1} + (1-n)E_t^2 v_{t+1} = n\boldsymbol{\psi}^s s_t + (1-n)\boldsymbol{\psi}^b b_t, \tag{29}$$

where the aggregate expectations operator $\hat{E} = nE^1 + (1-n)E^2$. We have the following result.

PROPOSITION 4. *In the special parametric case of the model, if a fraction n of the agents forms expectations from the s- model* (18) *and $1-n$ agents form expectations from the b-model* (19)*, then for each $n \in [0,1]$ there exists a unique restricted perceptions equilibrium with*

$$y_t = \left[\phi_b n\psi^s(n) + (1-n)\psi^b(n) + \left(\beta^{-1}-1\right)(1-\phi_b)\right] b_t - \left[(\beta^{-1}-1) - n\psi^s(n)\right] z_t$$

where

$$\psi^s(n) = \frac{\beta^{-1}(1-\beta)(1-\beta^2-\phi_b)}{[1-\beta^2-n(1+\beta-\phi_b)-2\phi_b]}$$
$$\psi^b(n) = \frac{-\beta^{-1}(1-\beta)(1-\beta^2-2\phi_b)(1-\phi_b)}{[1-\beta^2-n(1+\beta-\phi_b)-2\phi_b]}$$

This proposition demonstrates how fragile Ricardian equivalence can be, especially in a restricted perceptions environment. Even though all agents have misspecified forecasting models, when $n = 0$ then Ricardian equivalence arises as a self-confirming equilibrium. But, for any $n > 0$, even n arbitrarily close to zero, then neither type of agent will hold Ricardian beliefs. Below, we show that under fairly general conditions it may be that the $n = 1$ is the more likely equilibrium outcome.

4.1.2 Misspecification equilibrium

The restricted perceptions equilibrium analysis above shows that Ricardian equivalence depends fundamentally on n the distribution of households across the two forecasting models. It is important to pin down the value n endogenously as an equilibrium object. We accomplish this by characterizing the set of misspecification equilibria. The following result provides necessary and sufficient conditions for the existence of multiple misspecification equilibria.

THEOREM 1. *Consider the special parametric case of the model. Let $\omega \to \infty$ and $\beta > 2/3$. There exists multiple misspecification equilibria, $n^* \in \{0, \hat{n}, 1\}$, if and only if*

$$\underline{\phi}(\beta) < \phi_b < \bar{\phi}(\beta)$$

where

$$\begin{aligned}\underline{\phi}(\beta) &= \max\left\{1-\beta, \frac{1}{4}\left(4-2\beta-3\beta^2\right)\right\} \\ \bar{\phi}(\beta) &= \frac{1}{4}\left[\left(2-3\beta-2\beta^2\right)+\sqrt{4+4\beta+5\beta^2-4\beta^3}\right]\end{aligned}$$

We can similarly characterize the necessary and sufficient conditions for unique (non-)Ricardian equilibria.

COROLLARY 1. *Let $\omega \to \infty$. The following results hold.*

i. A unique misspecification equilibrium $n^* = 1$ exists if and only if

$$1-\beta < \phi_b \leq \frac{1}{4}\left(4-2\beta-3\beta^2\right)$$

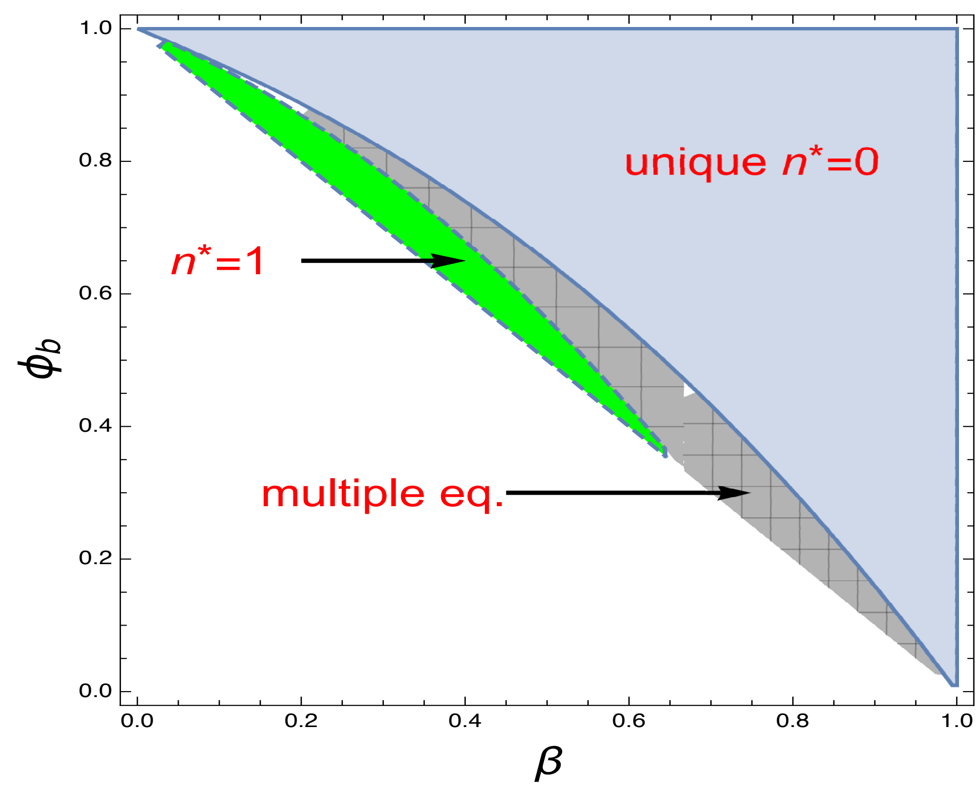
ii. A unique misspecification equilibrium $n^* = 0$ exists if and only if

$$\bar{\phi}(\beta) < \phi_b < 1$$

REMARK. *The existence of the unique non-Ricardian equilibrium,* $n^* = 1$*, requires that* $\beta < 2/3 \Leftrightarrow 1 - \beta < \frac{1}{4}(4 - 2\beta - 3\beta^2)$*. Thus, in the special case non-Ricardian equilibria is likely to arise in empirically plausible models in the form of multiple equilibria.*

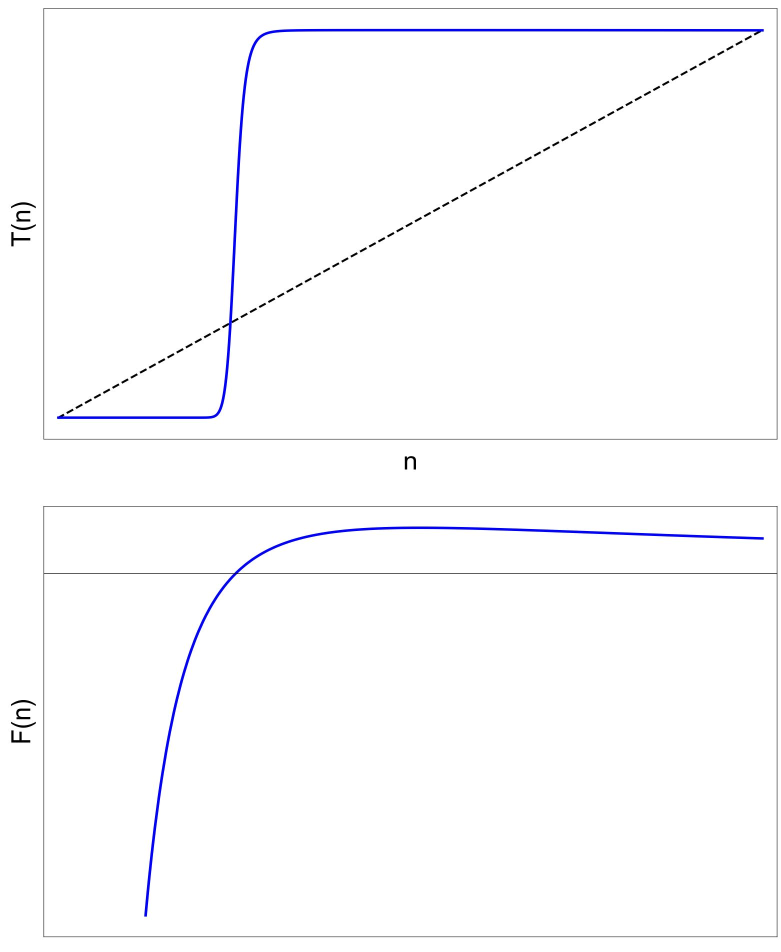
Figure 1 illustrates the results in Theorem 1 and Corollary 1. In the graph, are the combinations of (β, ϕ_b) consistent with multiple or unique equilibria. The large unshaded area in the lower half of the plot corresponds to active fiscal policy, i.e. $\phi_b > 1 - \beta$. The restriction to Ricardian policy rules out equilibria in this region. Then, moving outward from the origin, the shaded area with a dashed-boundary consists of the pairs of (β, ϕ_b) consistent with a unique non-Ricardian equilibrium. The next shaded area, with grid lines, corresponds to the existence of multiple equilibria. Finally, the outermost shaded area is where a self-confirming Ricardian equilibrium, $n^* = 0$, is the unique misspecification equilibrium

Figure 1: Equilibrium existence



Theorem 1 is the main theoretical result of the paper: even though fiscal policy is passive, non-Ricardian beliefs can emerge endogenously. For ϕ_b within a certain range $[\underline{\phi}, \bar{\phi}]$ then the non-Ricardian outcome can be sustained in a misspecification equilibrium. Most interestingly, for these fiscal policy rules there exists multiple misspecification equilibria with existence also of a Ricardian equilibrium $n^* = 0$. As we discuss below the case of multiple equilibria leads to interesting model dynamics that offer an alternative interpretation to regime-switching non-Ricardian policy effects. As an example, Figure 2 plots the T-map $T_\omega(n)$ and the relative predictor fitness function $F(n)$ when $\beta = 0.99, \phi_b = 0.02$, and $\sigma_z = 1$. In the bottom plot, it is evident that $F(0) < 0$ and $F(1) > 0$, which implies the existence of both Ricardian and non-Ricardian equilibria, respectively. The top panel plots the T-map for various values of n when $\omega \to \infty$. This figure clearly indicates the three misspecification equilibria. In the quantitative analysis, the interior misspecification equilibrium is unstable, and so the learning dynamics can feature recurrent switching between the basins of attraction of the two (non-)Ricardian equilibria.

Figure 2: Multiple (non-)Ricardian equilibria in the special case.



Why would individuals ever prefer the non-Ricardian forecasting model? An closer examination of $F(n)$ provides the intuition. The predictor fitness measures, in the simple case, are

$$
\begin{aligned}
-EU^s &= \mu_{v1}^2 Eb_t^2 + \left(\mu_{v2} - \psi^2\right)^2 Es_t^2 + 2\mu_{v1}\left(\mu_{v2} - \psi_s\right) Eb_t s_t \\
-EU^b &= \left(\mu_{v1} - \psi^b\right)^2 Eb_t^2 + \mu_{v2} Es_t^2 + 2\mu_{v2}\left(\mu_{v1} - \psi^b\right) Eb_t s_t
\end{aligned}
$$

Thus, a given model's predictor fitness depends, essentially, on three components. First, the distance between the belief parameter and the corresponding coefficient in the actual law of motion. Second, how volatile the missing component is from their forecasting model. Third, a term that is best interpreted as the omitted variable bias component of the prediction error. These distances are all weighted by the corresponding equilibrium covariances of the state variables.

After calculating the differences between these predictor fitness functions leads to

$$
F(n) = \left[\psi^b\left(\psi^b - 2\mu_{v1}\right) - \psi^s\left(\psi^s - 2\mu_{v2}\right)\phi_b^2\right] Eb_t^2 + 2\left[\mu_{v1}\psi^s - \mu_{v2}\psi^b\right] Eb_t s_t - \psi^s\left[\psi^s - 2\mu_{v2}\right]\sigma_z^2
$$

The fraction of agents who use the surplus-model then depends a balancing of how well the surplus model captures the serial correlation of the debt process and the additional predictive power from the surplus model conditioning directly on the z_t innovation. For small values of ϕ_b, the surplus- and debt-models are weakly correlated and so $n^* = 1$ can emerge as the unique equilibrium as it best captures the contemporaneous z_t innovations. Such an equilibrium is self-fulfilling in the sense that with $n = 1$ the indirect effect of z_t is strengthened through the self-referential features of the model. For larger values of ϕ_b the agents will always mass onto the Ricardian predictor as the surplus and debt models become more correlated the potential forecasting advantage of the surplus model is reduced. Finally, when ϕ_b takes middling values between these two extremes, then either (non-)Ricardian equilibrium can emerge. Now the debt-model is sufficiently correlated with the surplus-model that an agent will prefer it if $n = 0$. Similarly, the debt and surplus-models are not too closely correlated so that when $n = 1$, the agents will prefer the surplus model that captures the contemporaneous innovation in its forecast.

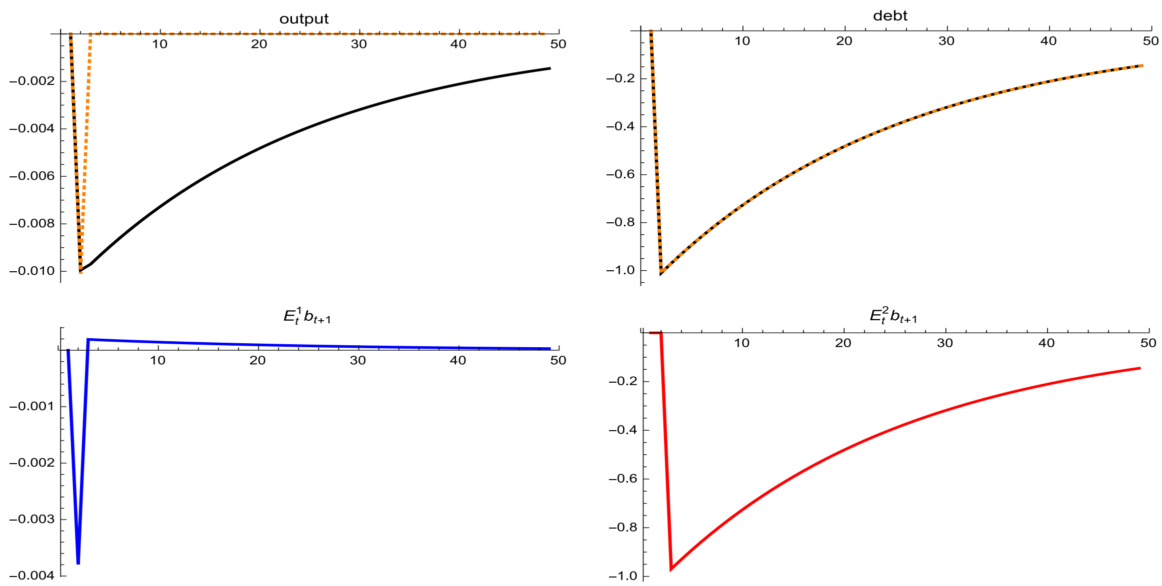
To sum up: for empirically plausible parameterizations of the special case, a Ricardian self-confirming equilibrium is always supported as a misspecification equilibrium. For a range of fiscal policy parameters there can also exist a non-Ricardian equilibrium. The non-Ricardian equilibrium is the unique outcome for ϕ_b sufficiently small and close to the border between active and passive fiscal policy.

4.1.3 Building intuition

We now develop intuition about the economic implications by comparing and contrasting the $n = 0$ and $n = 1$ RPE. These polar cases feature, of course, one case ($n = 0$) where weak Ricardian equivalence holds and the other ($n = 1$) where Ricardian equivalence does not hold.

Revisit the (non-)Ricardian results in Propositions 2 and 3. We can compare the restricted perceptions equilibrium paths for the state variables y_t, b_{t+1} as well as expectations about future debt $E_t^j b_{t+1}$. Figure 3 plots the impulse response functions to a one-percent innovation in z_t at $t = 1$ in both the $n = 0$ and $n = 1$ RPE's. For illustrative purposes, the figure sets $\beta = 0.99, \phi_b = 0.05$. Although, in this simple case there is a unique misspecification equilibrium at $n = 0$, this comparison is nevertheless informative.

Figure 3: Impulse responses in the special case.



The impact of an innovation $z_1 = 1$ produces a (slightly) larger, but purely transitory, contractionary effect on y_t in the Ricardian belief case $n = 0$, this is the weak Ricardian equivalence result. The $n = 1$ initial impact is slightly smaller, however, it has a strong persistent component. This stark example is informative because, as the NE-panel illustrates, the paths for b_{t+1} are the same across the two equilibria: the very restrictive case under consideration has government debt following a stationary exogenous process. The bottom two panels plot the expected paths for bonds in each of the two equilibria, i.e., $E_t^j b_{t+1}, j = 1, 2$. The SE-panel is for the Ricardian case and, as the figure shows, the Ricardian beliefs track the path of debt correctly. The SW-panel, though, shows remarkably different forecasts for debt. The non-Ricardian agents correctly forecast that debt decreases on impact, however, since the impact is small on the surplus (with $\phi_b = 0.05$), and their forecasting coefficient $\Gamma^s < 0$, a lower debt on impact implies a decrease in the surplus in the next period ($s_2 = \phi_b b_1$) and the expected path of debt increases and then slowly transitions back to its steady-state value.

The distinction between Ricardian and non-Ricardian equivalence in this restricted perceptions environment comes down to the difference in expectations across the two models. We gain further intuition by computing consumption along the impulse path. Denote the consumption function for this simplified case by $c_t^i(n)$. Quick computations lead to

$$c_t^i(n) = (1-\beta)b_t^i + (1-\beta)(y_t - \tau_t) + \beta \widehat{E}_t^i v_{t+1}^i.$$

In the case $n=0$, one can use PLM (19), (24), (23) and (7) to obtain

$$c_t^b(0) = (1-\beta)s_t - (1-\beta)\tau_t - (\beta^{-1}-1)z_t. \tag{30}$$

Equation (30) illustrates both the crowding out of private consumption by government spending and weak Ricardian equivalence.[13] The effect of tax shocks on consumption for these agents is the purely transitory unobserved component $(\beta^{-1}-1)z_t$. In contrast, the consumption function for a zero-mass agent who uses the s-model (18) while the aggregate variables are generated by the $n=0$ RPE is

$$c_t^s(0) = (1-\beta)\left[b_t - \tau_t + \frac{(1-\beta^2-\phi_b)}{(1-\beta^2-2\phi_b)}s_t - (\beta^{-1}-1)z_t\right].$$

This (zero-mass) agent features a weaker crowding out effect, there is a positive wealth effect of government debt, and predictable tax movements have impact on consumption.

We elaborate on the intuition via the diametrically opposite RPE with $n=1$. By combining the s-model (18), (28), (27) and (7) we obtain

$$c_t^s(1) = \left[\frac{1-\beta^2-(1-\beta)\phi_b}{\beta+\beta^2+\phi_b}\right]b_t - (1-\beta)\tau_t + (1-\beta)s_t + \left[\frac{-\beta^{-1}+\beta}{\beta+\beta^2+\phi_b}\right]z_t$$

Similarly, the consumption function of a zero-mass agent who uses the b-model (19) within the $n=1$ RPE is

$$c_t^b(1) = \frac{(1-\beta)}{(\beta+\beta^2+\phi_b)}\left[1+\beta-\beta^2(1-\phi_b)\phi_b - \phi_b^2(2\phi_b-3)\right]b_t + \left[\frac{1-\beta^2}{\beta+\beta^2+\phi_b}\right]s_t + \left[\frac{-\beta^{-1}+\beta}{\beta+\beta^2+\phi_b}\right]z_t$$

It follows that Ricardian equivalence fails in an $n=1$ RPE both for households with the s-model as well as the b-model.

[13] Given that $s_t \equiv \tau_t - G_t$, one can rewrite (30) as $c_t^b = -(1-\beta)G_t - (\beta^{-1}-1)z_t$.

4.1.4 Connection to rational expectations

An obvious objection is that the results presented hinge on the restricted perceptions restriction to forecasting models with only a single fiscal variable: what happens if the agents have a forecasting model with both b_t and s_t which nests the rational expectations equilibrium? In this subsection, we address this through the lens of econometric learning, e.g., Evans and Honkapohja (2001), and relax the parsimony assumption by assuming that agents form their expectations via a correctly specified model

$$v_t = \psi^s s_t + \psi^b b_t$$

We continue to maintain the imperfect knowledge assumptions, including not *a priori* imposing Ricardian beliefs, and further assume that the belief coefficients ψ^s, ψ^b are real-time estimates from a constant gain learning model, a form of discounted least-squares. With this perceived law of motion, the actual law of motion implied by these beliefs can be written as

$$v_t = S\left(\psi^s, \psi^b\right)' \begin{bmatrix} s_t \\ b_t \end{bmatrix} - \left(1/\left(1 + \sigma^{-1}\phi_y^{-1}\right)\right) g_t$$

where

$$S\left(\psi^s, \psi^b\right) = \frac{1}{1+\sigma\phi_y} \begin{bmatrix} -\beta^{-1}\left(\psi^s\phi_b + \psi^b + 1 - \beta\right) \\ \beta^{-1}\left(\psi^s\phi_b + \psi^b\right) + (1-\beta)\left(\beta^{-1} - 1\right) - \sigma\phi_y\left(1-\beta\right) \end{bmatrix}$$

The S-map has the usual interpretation: given a perceived law of motion with coefficients $(\psi^s \psi^b)'$ the corresponding coefficients in the actual law of motion implied by these beliefs are $S\left(\psi^s, \psi^b\right)$. A rational expectations equilibrium is a fixed point of the "S-map", i.e., $\Theta^* = S(\Theta^*), \Theta' = (\psi^s, \psi^b)$.

We can solve for the "mean dynamics" associated to the constant gain learning dynamics as a (small gain) approximation to the expected transitional learning dynamics. Adapting the stochastic recursive approximation results in Evans and Honkapohja (2001) it is possible to show that, across sequences of increasingly smaller gain parameters, the learning dynamics weakly converge to the expected path for Θ given by the following system of ordinary differential equations (o.d.e's.)

$$\begin{aligned} \dot{\Theta} &= R^{-1}M\left(S\left(\Theta\right) - \Theta\right) \\ \dot{R} &= M - R \end{aligned}$$

where

$$M = E\begin{bmatrix} s_t \\ b_t \end{bmatrix}\begin{bmatrix} s_t & b_t \end{bmatrix}$$

The mean dynamics are the solution path, for a given initial condition $\Theta(0)$, to this system of o.d.e.'s.

The mean dynamics are useful for understanding the qualitative nature of learning dynamics. In particular, standard results in the literature show that constant gain learning dynamics are distributed asymptotically normal with a mean equal to the rational expectations equilibrium values and a variance that is proportional to the size of the gain parameter. Thus, over time one can expect with high probability to see coefficient estimates Θ that fluctuate around Θ^*. The response of Θ to a particular sequence of unlikely shocks is described by the "escape dynamics", which provide the "most likely unlikely" path away from the rational expectations equilibrium, and then the mean dynamics describe the transition path back to the equilibrium.[14] The escape dynamics, therefore, can be thought of as re-initializing the mean dynamics. Thus, we can use different starting values for the mean-dynamics to characterize the type of learning paths that we might actually observe.

We use these insights to show that the learning dynamics in the case of fully specified perceived laws of motion will be drawn, for a finite stretch of time, towards the $n = 1$ restricted perceptions equilibrium. The mean dynamics are derived from a continuous time approximation of the real-time learning dynamics and the application of a law of large numbers, however, it is straightforward to convert the notional time in the O.D.E. to actual discrete time according to $t = \gamma^{-1}\tau^{\gamma}$, so that a small constant gain γ corresponds to a long stretch of real time.

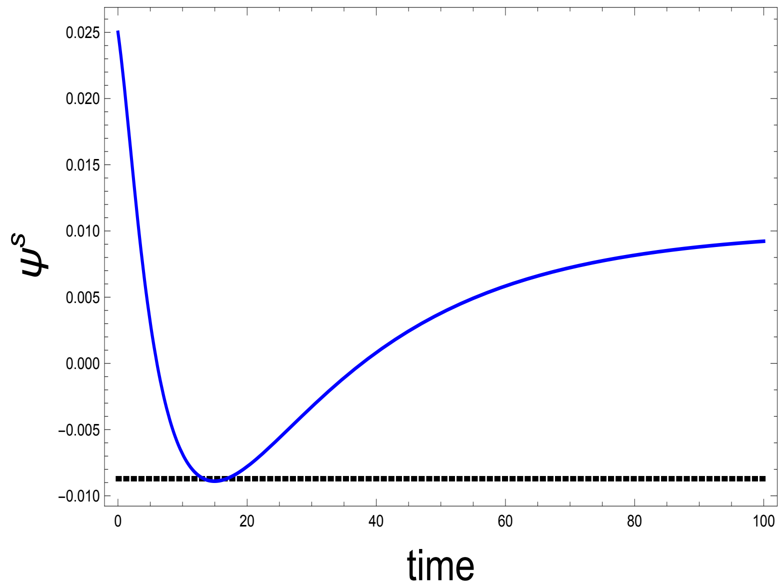
Figure 4 plots the mean dynamics for a particular illustrative parameterization: $\phi_y = 0.5, \phi_b = 0.9, \sigma = 2, \sigma_z^2 = 1$.[15] We then choose initial values for the ψ^s, ψ^b that are both above their rational expectations equilibrium values. The mean dynamics O.D.E. is then solved and Figure 4 plots the expected learning path (ψ^s shown). The experiment is to imagine an "escape" that has driven beliefs above their rational expectations equilibrium values and use the solution to the mean dynamics O.D.E. to trace out how the economy is most likely to respond.

The figure plots (solid line) the expected transition path for ψ^s while the the dashed line is the value in an $n = 1$ restricted perceptions equilibrium, that is the value for ψ^s and ψ^b that would arise in an $n = 1$ RPE. The learning dynamics are expected to eventually converge to the rational expectations equilibrium. However, with a small constant gain, that speed of convergence can be quite slow. Interestingly, along the transition path for ψ^s, the beliefs hover for a finite stretch of time at its $n = 1$ RPE. This coincides with a path for ψ^b (not shown) that is also drifting down towards its RPE value. As the path for ψ^b continues to transition towards its REE value, this helps draws ψ^s away from its RPE value and back

[14]See Williams (2018) for details and a comprehensive set of results and toolkit on escape dynamics in constant gain learning models.

[15]For expositional ease, we present an example where the RPE and REE values are starkly far apart.

Figure 4: Expected learning dynamics for a correctly specified forecast model



towards the rational expectations equilibrium.

Thus, we conclude from Figure 4 that the RPE is both theoretically appealing and realistic as we can expect recurrent escapes near the $n = 1$ RPE even when all agents in the economy form forecasts from a bivariate model that nests the rational expectations equilibrium. Moreover, for small gains γ, the economy will persist near the RPE for long stretches of time. Even without the restricted perceptions restriction, since the fully specified model nests the RPE which acts as an escape point and impacts the learning dynamics even in the standard set up. The restricted perceptions equilibrium, then, is a reasonable approximation to an even more sophisticated forecasting model for the agents.

The mean dynamics in Figure 4 also help to better understand the connection between this paper and Eusepi and Preston (2018a). In their model, beliefs nest the rational expectations equilibrium however the agents attempt to learn about the long-run stances of fiscal and monetary policy. They show how learning dynamics can generate fluctuations with non-Ricardian effects. These non-Ricardian effects are strengthened in economies with a high steady state debt/output ratio. The theory of expectation formation here emphasizes restricted perceptions which require the agents to estimate the relevant auto- and cross-covariances which in combination gives scope for escape dynamics. The mean dynamics results suggest loosely that the convincing non-Ricardian effects in Eusepi and Preston (2018a) can also be impacted by the existence of a non-Ricardian restricted perceptions equilibrium. Therefore, our theoretical and quantitative analysis is complementary to their paper, while providing also an equilibrium explanation for the phenomenon of non-Ricardian beliefs.

4.2 Further results

The results in the special case are useful for clear intuition and analytic tractability. However, the results are robust to all cases where $s_b = 0$. While analytic results are not possible when $s_b > 0$, we use numerical analysis to show that the insights from the special case carry over but also with more equilibrium possibilities. In this subsection generalize the results to all parameterizations where $s_b = 0$.

The key equations are now:

$$\begin{aligned}
b_{t+1} &= \beta^{-1}(b_t - s_t) \\
y_t &= v_t + (1-\beta)b_t + g_t \\
v_t &= (1-\beta)(b_{t+1} - b_t) + n\psi_v^s(n)s_t + (1-n)\psi_v^b(n)b_t \\
\pi_t &= \kappa y_t + (1-\alpha)\beta n\psi_{p^*}^s(n)s_t + (1-n)\psi_{p^*}^b b_t + u_t \\
i_t &= \phi_\pi \pi_t + w_t \\
s_t &= \phi_b b_t + z_t
\end{aligned}$$

where $\phi_\pi > 1$ and $1-\beta < \phi_b < 1$, i.e., active monetary/passive fiscal policy. We are able to prove the following result in the case of small σ.

PROPOSITION 5. *For σ sufficiently small, there exists a $\tilde{\phi}(\beta)$ such that multiple mis-specification equilibria exist provided that*

$$1-\beta < \phi_b < \tilde{\phi}(\beta)$$

.

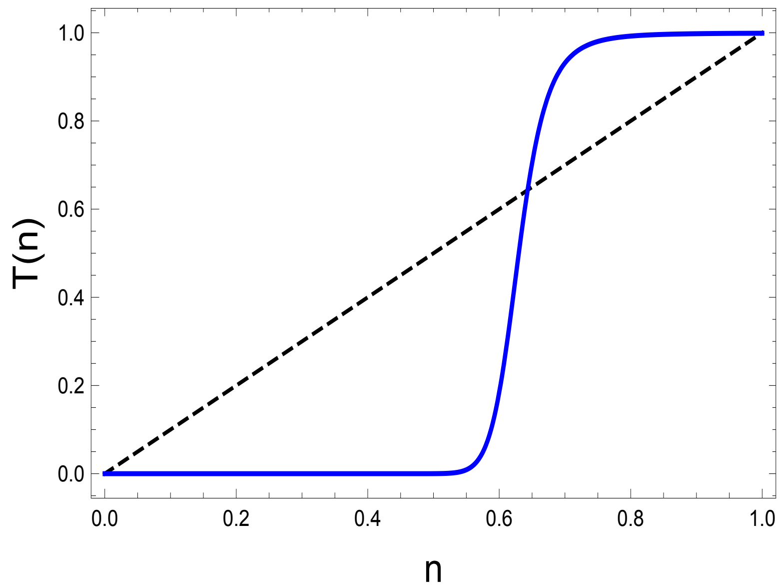
This generalizes the previous results to a New Keynesian model and monetary policy that adheres to a Taylor-type rule. Notice from the proposition that endogenously non-Ricardian beliefs and, conversely, self-confirming Ricardian beliefs, depend on the extent to which the fiscal policy rule is passive and does not depend directly on the monetary policy coefficient (ϕ_π). However, we show that ϕ_π can have a qualitative and quantitative impact on the dynamics when $s_b > 0$, a case that we consider numerically in the remainder of the paper.

4.3 Multiple equilibria

When $s_b > 0$ then there exist parameterizations that cover all of the equilibrium possibilities in Proposition 1. In the quantitative analysis we are particularly interested in the possibility of multiple equilibria, i.e., cases where *both* $n^* = 0$ and $n^* = 1$ are theoretical possibilities.

To illustrate, Figure 5 plots the T-map for the numerical parameterization in Table 1 from the quantitative analysis in the subsequent section.[16]

Figure 5: Illustration of multiple equilibria with calibration as in Table 1 and $\omega \rightarrow$.



Recall that a misspecification equilibrium is a fixed point of the T-map: $n^* = T_\omega(n^*)$. Figure 5 illustrates the possibility of 3 equilibrium values for n^*. The polar cases $n^* = 0, n^* = 1$ and an interior equilibrium $\hat{n}$ which features $F(\hat{n}) = 0$. Since the slope of the interior equilibrium is greater than one, the (non-)Ricardian equilibria are the ones that will be stable under learning. We can, therefore, anticipate real-time learning dynamics that endogenously switch regimes by hopping between the basins of attractions for the (non-)Ricardian equilibria. These sorts of dynamics play a key part of the quantitative analysis.

5 Quantitative results

Having proposed a theory of endogenously (non-)Ricardian beliefs, we now turn to a quantitative analysis to see whether macroeconomic data are consistent with non-Ricardian beliefs. We first demonstrate that a New Keynesian model with active monetary policy, passive fiscal policy, and endogenously (non-)Ricardian beliefs describe well U.S. data on output gap, inflation, and the primary real-surplus. Our estimates of the underlying state dynamics suggest a sizable fraction of individuals and firms hold non-Ricardian expectations, with the

[16]The T-map here continues to assume iid innovations, while the quantitative analysis includes serially correlated shocks. As a result, the multiple equilibria here require a larger value of $\kappa = 0.03$.

fraction increasing over time. Counterfactual analysis explores the implications of alternative monetary policies.

5.1 Theory

This section generalizes the theoretical model while maintaining the previous assumptions that monetary policy is active and fiscal policy is passive. To explain the data we also follow Eusepi and Preston (2018a) in replacing the fixed RPE parameters with a constant gain learning process that assumes agents estimate model coefficients and the relative forecast fitness, in real-time: after new data are realized, agents estimate the parameters of their model using discounted least-squares, update their estimate of the relative forecast accuracy of the two models, and then select a forecasting model. We also expand to a richer set of exogenous, serially correlated disturbances.

The actual laws of motion are given by,

$$
\begin{aligned}
b_{t+1} &= \beta^{-1}\left(b_t - s_b\pi_t - s_t\right) + s_b i_t \\
y_t &= v_t - \sigma\pi_t + g_t + (1-\beta)b_t \\
v_t &= (1-\beta)\left(b_{t+1} - b_t\right) - \sigma\left(i_t - \pi_t\right) + \hat{E}_t v_{t+1} \\
\pi_t &= \kappa y_t + (1-\alpha)\beta\hat{E}_t p^*_{t+1} + u_t \\
i_t &= \phi_\pi \pi_t + \phi_y y_t + w_t \\
s_t &= \phi_b b_t + z_t
\end{aligned}
$$

where, for a variable x, $\hat{E}_t x = n_{t-1}E^1_t x + (1-n_{t-1})E^2_t x$. The exogenous shocks are uncorrelated, stationary AR(1) processes:

$$
\begin{aligned}
g_t &= \rho_g g_{t-1} + \varepsilon_{gt} \\
u_t &= \rho_u u_{t-1} + \varepsilon_{ut} \\
w_t &= \rho_w w_{t-1} + \varepsilon_{wt} \\
z_t &= \rho_z z_{t-1} + \varepsilon_{zt}
\end{aligned}
$$

with $\varepsilon_{jt} \sim \text{iid}(0, \sigma_j^2)$ and $E\varepsilon_j\varepsilon_j' = 0, j' \neq j$. Extending the two restricted forecasting models to this more general environment, we can write

$$E_t^1 v_{t+1} = \left(\psi_{v,t-1}^s\right)' \begin{bmatrix} s_t \\ g_t \\ u_t \\ w_t \\ z_t \end{bmatrix}, \; E_t^2 v_{t+1} = \left(\psi_{v,t-1}^b\right)' \begin{bmatrix} b_t \\ g_t \\ u_t \\ w_t \\ z_t \end{bmatrix}$$

$$E_t^1 p_{t+1}^* = \left(\psi_{p^*,t-1}^s\right)' \begin{bmatrix} s_t \\ g_t \\ u_t \\ w_t \\ z_t \end{bmatrix}, \; E_t^2 p_{t+1}^* = \left(\psi_{p^*,t-1}^s\right)' \begin{bmatrix} b_t \\ g_t \\ u_t \\ w_t \\ z_t \end{bmatrix}$$

The coefficients $\psi_{j,t}^k, j = v, p^*, k = s, b$ are updated with constant gain least-squares:

$$\begin{aligned} \psi_{j,t}^k &= \psi_{j,t-1}^k + \gamma_1 R_{k,t}^{-1} X_{k,t-1} \left(x_{k,t} - \left(\psi_{j,t-1}^k\right)' X_{k,t-1} \right) \\ R_{k,t} &= R_{k,t-1} + \gamma_1 \left(X_{k,t-1} X_{k,t-1}' - R_{k,t-1} \right) \end{aligned}$$

where $x_{k,t} \in \{v_t, p_t^*\}$, $X_{s,t-1}' = (s_{t-1}, g_{t-1}, u_{t-1}, w_{t-1}, z_{t-1})$, $X_{b,t-1}' = (b_{t-1}, g_{t-1}, u_{t-1}, w_{t-1}, z_{t-1})$, and $R_{k,t}$ is the sample estimate of the regressor covariance matrix $EX_{k,t}X_{k,t}'$. The parameter $0 < \gamma_1 < 1$ is the "constant gain" as it governs the responsiveness of parameter updating to recent forecast errors. The discounted least-squares places a geometrically declining weight, $(1-\gamma_1)^t$ on recent data observations. The timing implicit in these learning rules is consistent with the previous analysis: expectations are formed at the beginning of t using coefficient estimates based on all observable information through $t-1$.

A similar recursive estimator for the distribution of agents across forecasting models, n, can be derived. Re-writing (17) as

$$EU_t^j = -MSE_{v,t}^j - MSE_{p^*,t}^j$$

where

$$MSE_{k,t}^j = MSE_{k,t-1}^j + \gamma_2 \left[\left(x_{k,t} - \left(\psi_{j,t-1}^k\right)' x_{k,t-1} \right)^2 - MSE_{k,t-1}^j \right]$$

Note we allow for the possibility that gain parameters $\gamma_1 \neq \gamma_2$. A forecaster that is relatively more uncertain about the forecasting accuracies of the two models than they are about their model coefficient estimates would set $\gamma_2 > \gamma_1$.[17] And, the MNL law of motion delivers the real-time distribution of endogenously (non-)Ricardian beliefs:

$$n_t = \frac{1}{2} \left\{ \tanh \left[\frac{\omega}{2} \left(EU_t^s - EU_t^b \right) \right] + 1 \right\}$$

[17] See Branch and Evans (2006a) for discussion and evidence from the Survey of Professional Forecasters.

5.2 Methodology

The Extended Kalman Filter (EKF) generates the data-implied one-step ahead predicted paths for the endogenous state variables. After plugging in the policy rules, expectations, and recursive updating equations for the learning rules, the model can be written in non-linear state space form:

$$\begin{aligned} X_t &= g\left(X_{t-1}, \Theta\right) + Q(X_{t-1}, \Theta)v_t \\ Y_t &= f(X_t, \eta_t) \end{aligned}$$

where the state vector

$$X_t' = \left(b_{t+1}, g_t, u_t, w_t, z_t, n_t, MSE_{st}, MSE_{bt}, \text{vec}\left(\psi_t^s\right), \text{vec}\left(\psi_t^b\right), \text{vec}\left(R_{st}\right), \text{vec}\left(R_{bt}\right), b_t\right),$$

$\text{vec}\left(\cdot\right)$ is the vectorization operator, the observation variables are

$$Y_t' = \left(y_t, \pi_t, s_t, b_{t+1}\right),$$

and the parameter vector

$$\Theta' = \left(\kappa, \alpha, \phi_\pi, \phi_y, \phi_b, \rho_g, \rho_u, \rho_w, \rho_z, \sigma_g, \sigma_u, \sigma_w, \sigma_z, \omega, \gamma_1, \gamma_2\right).$$

The measurement and state disturbances are η_t, v_t respectively. Our sample for the observed variables is 1955.1-2007.3.[18] We measure y_t as the log difference between output and the CBO's measure of potential output. We measure π_t from the PCE index. We compute b_t and s_t as the debt-GDP ratio and primary surplus-GDP ratio, respectively. All variables are HP-filtered and detrended.

We estimate the (one step ahead) predicted state path $E\left(X_{t+1}|Y_t, \Theta\right)$. Since the state transition and measurement equations are highly non-linear in the belief state variables, an approximation of the non-linear state-space model is necessary. The Extended Kalman Filter (EKF) mimics the linear Kalman Filter by naturally extending the prediction steps to the non-linear state space. The non-linearity creates a difficulty for calculating the co-variances of the state and measurement variables that the EKF overcomes with a first-order approximation to these moments.

One could use Bayesian methods to uncover posterior estimates of the structural parameters of interest Θ by using the Extended Kalman Filter to approximate the likelihood function. However, since our modeling environment is closely related to Eusepi and Preston (2018a), who use Bayesian techniques to estimate the posterior distribution of a model where adaptive learning generates temporary, endogenous departures from Ricardian equivalence, we follow this seminal work and parameterize the model according to the mean and 95% probability bounds on the posterior distribution: see Table 1. As we will see, this benchmark parameterization yields state dynamics consistent with U.S. data.

[18]We end the sample before the ZLB episode as incorporating an effective lower bound on interest rates is beyond the scope of the present paper, but is the topic of future research.

Table 1: Quantitative parameterization

Parameter	Posterior Mean
β	0.99
σ	1/7.7147
s_b	0.3
κ	0.003
α	0.738
ϕ_π	1.623
ϕ_y	0.094
ϕ_b	0.047
ρ_g	0.931
ρ_u	0.870
ρ_w	0.857
ρ_z	0.073
σ_g	0.526
σ_u	0.186
σ_w	0.197
σ_z	2.088
ω	1
γ_1	0.005
γ_2	0.039

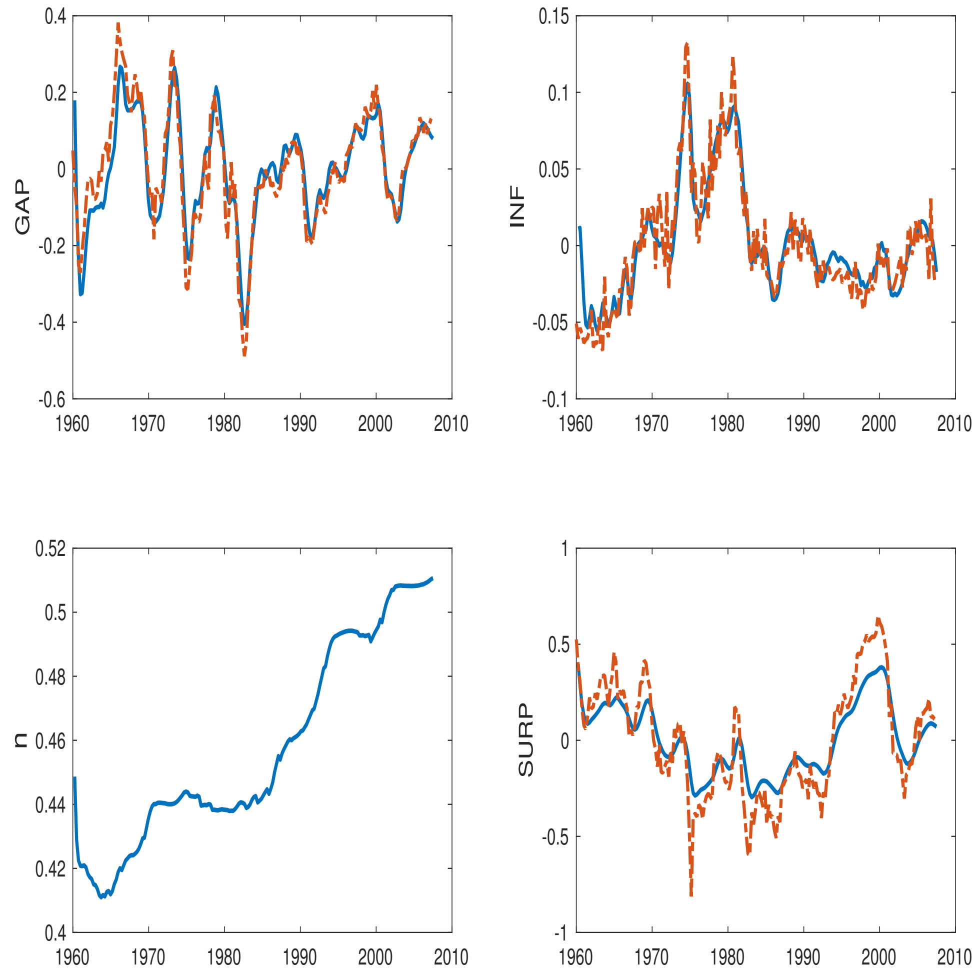
A few comments are in order. We initially normalize the 'intensity of choice' parameter $\omega = 1$. In the counterfactual experiments, we also consider the case where ω is large, in line with the misspecification equilibrium results presented earlier. A value of $\omega = 1$ in this model can be interpreted as a low "intensity of choice" as the T-map $T_{\omega=1}(n)$ is close to linear. The value of γ_2 is in line with Eusepi and Preston (2018a), however γ_1 is on the smaller side of what is often estimated in the literature. We fixed the value $\gamma_1 = 0.005$ to be sure that the learning dynamics remain bounded. Alternatively, we could have imposed a "projection facility" that keeps the values of the ψ's bounded in an appropriate neighborhood, and then considered values where $\gamma_1 = \gamma_2$. Larger values of γ_1 lead to more volatile belief parameter updating and more frequent switching between misspecified equilibria, when they exist. Thus, the relatively small value of $\gamma_1 = 0.005$ is a conservative choice. The small estimated values for the parameters ϕ_b and ρ_z also work against the theory of endogenously non-Ricardian beliefs: larger values of both ϕ_b and ρ_z increase the set of structural parameters consistent with multiple misspecification equilibria. While the estimated parameters for the Taylor rule are in line with estimates in the literature, our main counterfactuals involve how different values for ϕ_π, ϕ_y impact the results on (non-)Ricardian beliefs and macroeconomic outcomes.

5.3 Benchmark results

To begin, Figure 6 plots both the model-predicted state dynamics and actual U.S. data. We use the Extended Kalman Filter to compute (smoothed) $E\left(X_{t+1}|Y_t, \Theta\right)$ for the period 1960-2007.3, and initialized the model over the period 1955-1960 as follows. For period 1955.1 we assume that the economy is initially at the misspecification equilibrium $n \approx 0.5$ and compute the corresponding RPE. The variables are all initialized at this RPE and then we simulate the state dynamics, including the endogenous beliefs and n_t, over the next 20 quarters. Using those initial values, we run the Extended Kalman Filter to estimate the predicted path, which is plotted in the panels of Figure 6.

Moving clockwise from the NW panel, the solid lines are the model-predicted paths for the output gap, inflation, the primary government surplus, and the fraction of non-Ricardian agents. Figure 6 demonstrates that the fraction of non-Ricardian agents, in the benchmark parameterization, is below 0.50 for most of the sample, though throughout the fraction of non-Ricardian agents is never below 0.41. There are several notable periods of an evolving extent of non-Ricardian beliefs. In the first half of the 1960's, the fraction of non-Ricardian agents is decreasing, but in the latter half of the decade, there is an approximately 3% increase in the percentage of non-Ricardian agents. Throughout the 1970's and 1980's, the model predicts a steadily increasing fraction of individuals with non-Ricardian beliefs to nearly 50% in the second half of the 1990's. Then in the early 2000's there is another increase to where

Figure 6: Predicted State Dynamics. Solid lines are model predicted state variables. Dashed lines are corresponding U.S. data.



$n > 0.5$ and then holding roughly that value for the rest of the sample. These switches are bounded, in part, because of the intensity of choice normaliztion. The switches are also in line with stretches of time where the surplus process encounters a sequence of shocks in the same direction, i.e., a "most likely unlikely sequence."[19]

The top two panels show that these fluctuations in Ricardian and non-Ricardian beliefs produce predicted time-paths for the output gap and inflation that is roughly in line with actual U.S. data over the period (dashed lines).[20] In particular, the model does a very good

[19]See Cho et al. (2002) for formal details behind escape dynamics following, in their terms, a "most likely unlikely sequence."

[20]The EKF produces the optimal one step ahead predictions of the state variables. An objective measure of

job in matching the magnitudes and timing of fluctuations in inflation and the output gap. The SE panel plots the surplus-GDP ratio. The model-implied surplus (solid line) captures the broad trends seen in the data, but does not quite match the observed volatility

5.4 Counterfactual results

The model fits the data well in Figure 6. To dive deeper into these results, and better understand how the learning dynamics play an important role, we turn to several counterfactual exercises. Throughout, a counterfactual is constructed by extracting the predicted shocks from the benchmark model simulations presented above. Then, assuming the same realization of shocks, we can alter one or two structural parameters, calculate the model predicted state path again with the new value for the parameters $\hat{\Theta}$. We first show counterfactuals for a larger intensity of choice ω, and then we focus on the monetary and fiscal policy implications.

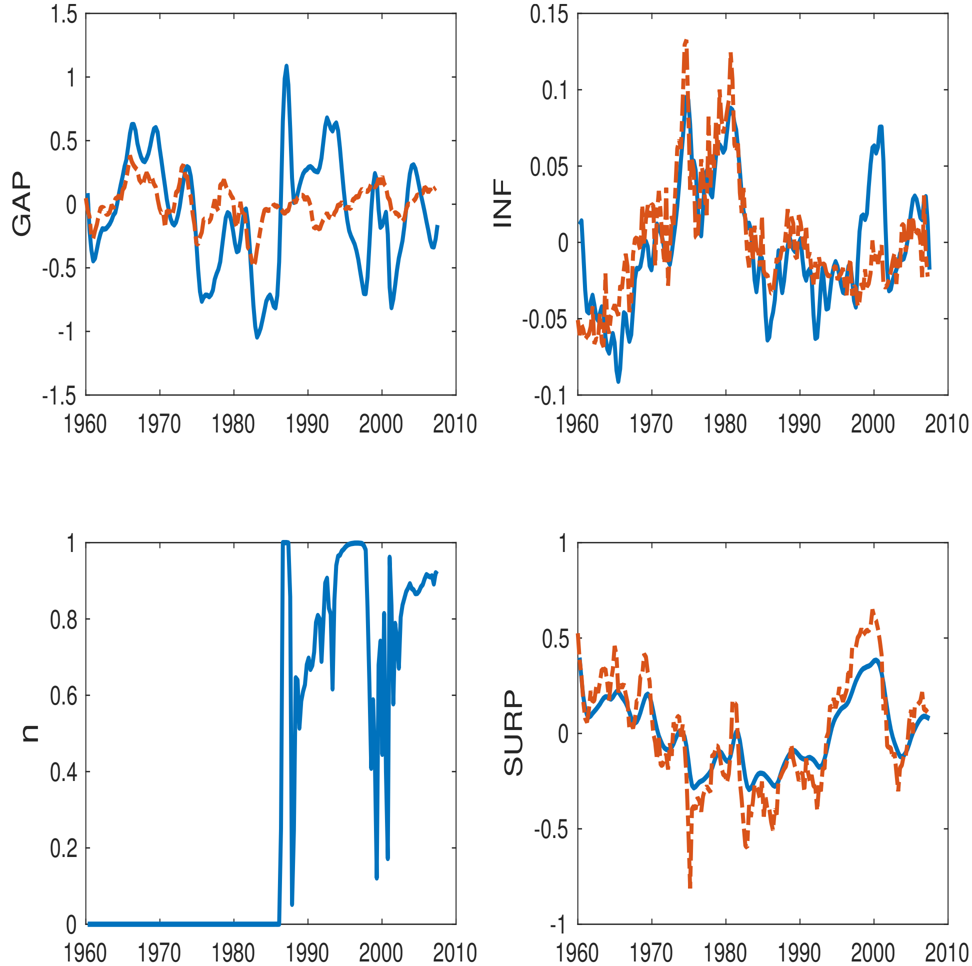
5.4.1 The 'intensity of choice'

Recall from the theoretical analysis that as $\omega \to \infty$ then one can expect to observe either $n_t = 0, 1$ or switching between the two if there are multiple misspecification equilibria as forecasters will only select the best performing statistical models. To see how a stronger 'intensity of choice' in forecast model selection would impact the economy, we present a counterfactual where $\omega = 300$. Figure 7 plots the model implied state dynamics.

With the same sequence of shocks, the most likely path for n_t, given the observed data, has the predicted pattern. Until the late 1980's, it is estimated that all agents have Ricardian beliefs. This is not surprising from Figure 6 where, throughout this period, the fraction $n < 0.5$, which implies that $EU^1 < EU^2$ in all periods. Thus, larger values of ω will decrease n and, when the intensity of choice is large enough, feature $n = 0$. However, beginning in the late 1980's there are regime-changes between periods where all agents have non-Ricardian beliefs, most agents have non-Ricardian beliefs, or very few have non-Ricardian beliefs. While the timing is in line with Figure 6 the alignment is not perfect. That is because changing ω leads to a counterfactual path for the entire state with the exception of the exogenous shocks.

fit is, of course, elusive. The visual fit of the data is partly the result of the predicted paths for the exogenous shocks. However, it should be noted that the fiscal shock process is very weakly autocorrelated and so that the model does a plausible job with the surplus state variable suggests that the internal propagation of shocks is an important component. Ideally we would use a formal model comparison to test our model against others in the literature including regime-switching policy models and belief-driven models. However, such an exercise is beyond the scope of this paper.

Figure 7: Counterfactual State Dynamics when $\omega = 300$. Solid lines are model predicted state variables. Dashed lines are corresponding U.S. data.



The remaining panels demonstrate how the counterfactual path with a large intensity of choice exhibits substantially greater output gap volatility and a relatively poor fit for inflation post 1980. The counterfactual path for the surplus is similar to the benchmark case. A natural question is what accounts for the regime-change in the late 1980's in both Figure 6 and the counterfactual in Figure 7. An intuitive interpretation is that the shift in decreasing primary surpluses throughout the sample until the mid 1980's, and then an increasing surplus throughout the 1990's, created the kind of sequence of shocks– the "most likely unlikely" sequence – that led individuals to discover that non-Ricardian beliefs provide a better forecast and then this is reinforced through the self-referential features of the economy. The switching between $n = 0$ and $n = 1$ equilibrium neighborhoods lead to substantial volatility. This is because after each switch the agents' learning process, essentially, starts anew.

5.5 Monetary policy

Taylor-rule coefficients that react relatively more strongly to inflation innovations lead to longer and more frequent spells with (non-)Ricardian beliefs. Counterfactual analysis establishes this result.

Recall from the theoretical analysis that the existence of RPE with non-Ricardian beliefs is independent of the monetary policy coefficient ϕ_π for large ω. However, even with large ω the monetary policy rule coefficients impact the relative sizes of the basins of attraction.[21] To explore the implication, we present results from several counterfactual experiments. As above, we take as given the exogenous shocks from the benchmark path and then estimate the predicted paths under two scenarios: a small policy coefficient $\phi_\pi = 1.005$, and, a large value $\phi_\pi = 2.5$. See Figure 8 where $\phi_\pi = 2.5$.

This counterfactual asks the question of what would have happened to the economy had policymakers placed a substantially higher weight on reacting to inflation innovations. The SW-panel demonstrates that the effect of such a policy would have been to lead to a slightly higher fraction of non-Ricardian agents, on average, throughout the sample period. During the 1960's and 1970's, the NE-panel shows that the counterfactual effect would have been less volatile inflation. However, during the 1990's when there was a large run up in the surplus-GDP ratio, this coincides with a larger fraction of non-Ricardian agents, by forecasting future fiscal policy primarily using the surplus, leads to a substantial increase in both the inflation rate and a positive output gap. In fact, the more aggressive monetary policy would lead to a counterfactually large economic expansion.

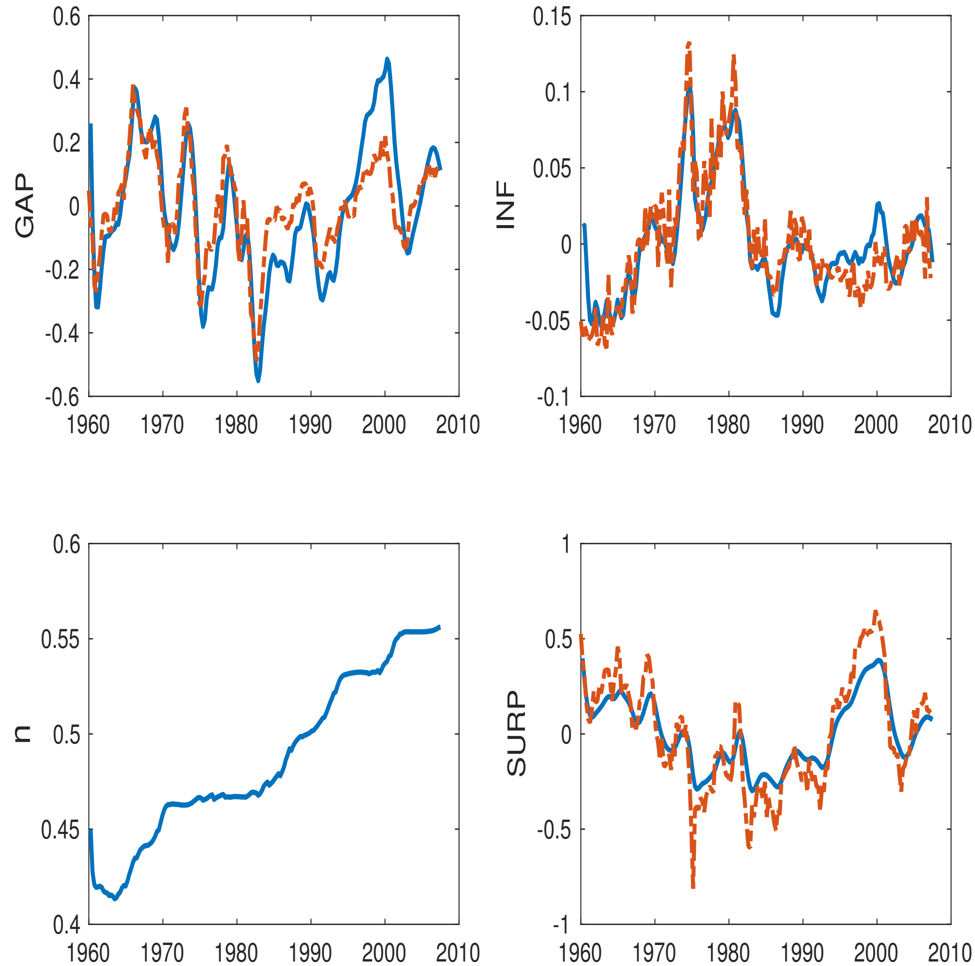
Now consider the counterfactual with a value of $\phi_\pi = 1.005$ at the edge of the active monetary/passive fiscal determinacy region (Figure 9). In this counterfactual exercise, there is a significant decrease in the fraction of agents with non-Ricardian beliefs throughout the sample: see the SW panel. In fact, for large ω the more dovish monetary policy would lead to $n = 0$ with all individuals and firms holding Ricardian beliefs. The combination of a less aggressive monetary policy response to inflation and an endogenously higher fraction of Ricardian agents, leads to substantially greater volatility and a large deflation/negative output gap during the 1990's as the surplus is increasing substantially.

Notice the nuanced trade-off faced by policymakers here. A monetary policy rule could be tuned to be more, or less, hawkish. If policymakers had adopted a less hawkish policy rule that would have coordinated on a Ricardian regime for inflation. But, by being less active against inflation the economic volatility would have been higher.[22] If instead policymakers had pursued a more hawkish policy rule, then inflation would have been non-Ricardian more

[21]This can be established formally as the $\hat{n} = T_\omega(\hat{n})$ equilibrium, in the multiple misspecification equilibrium case, shifts with policy coefficients.

[22]This is a standard result in learning models: see result 5a in Eusepi and Preston (2018b).

Figure 8: Counterfactual State Dynamics when $\phi_\pi = 2.5$. Solid lines are model predicted state variables. Dashed lines are corresponding U.S. data.

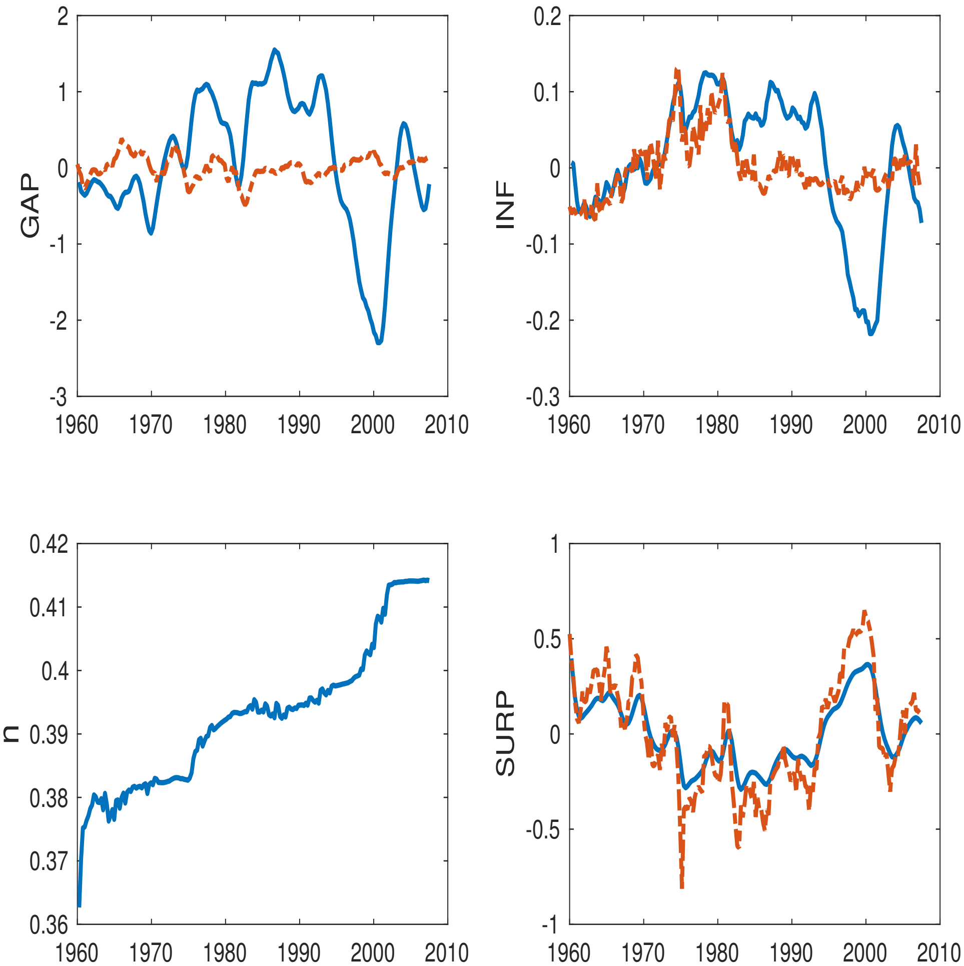


often, again with possibly higher volatility. This "goldilocks" prescription for monetary policy is a novel finding in the learning literature and suggests that learning models can illuminate a complex, nuanced trade-offs faced by policymakers.

5.6 Monetary policy and endogenously Ricardian beliefs

The claim at the beginning of the previous section is that dovish monetary policy would have led to more agents with non-Ricardian beliefs and hawkish monetary policy would have led to fewer agents with Ricardian beliefs. To assess this claim, we again hold the exogenous shocks fixed to their benchmark path, consider a variety of alternative policy rule coefficients, and

Figure 9: Counterfactual State Dynamics when $\phi_\pi = 1.005$. Solid lines are model predicted state variables. Dashed lines are corresponding U.S. data.

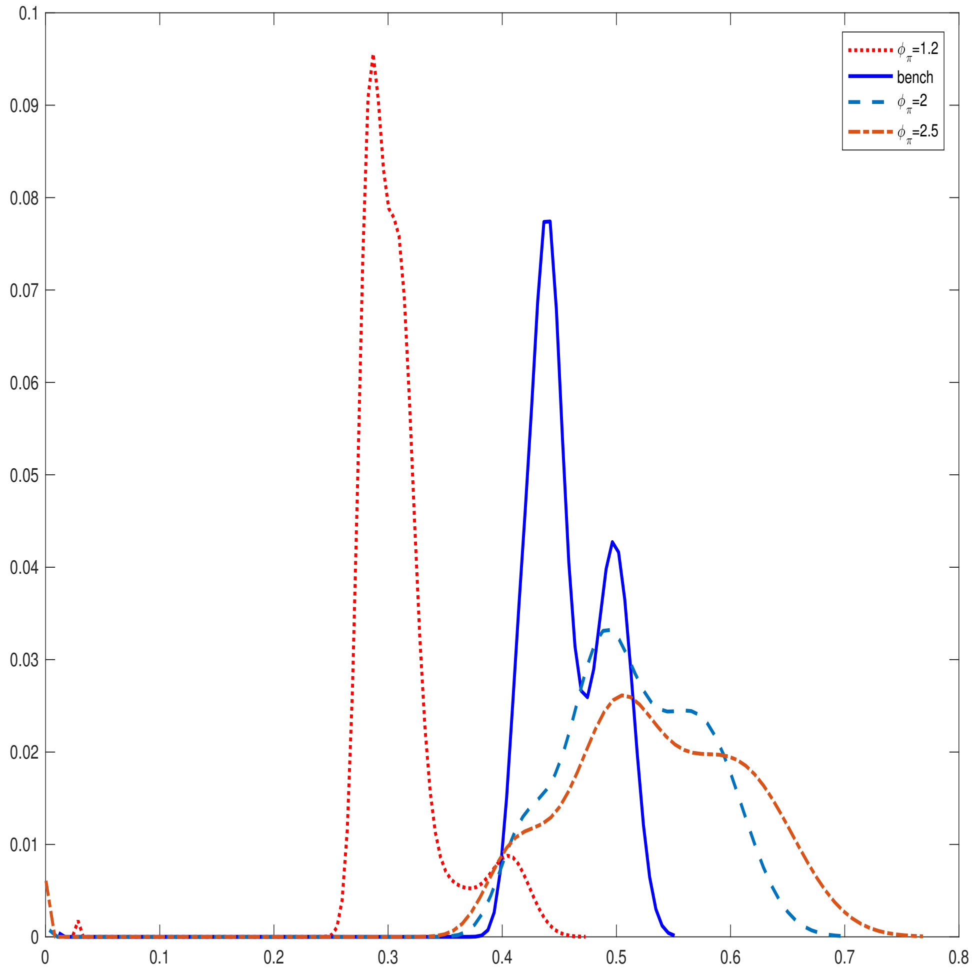


then plot the empirical distribution of n_t, the fraction of agents with non-Ricardian beliefs.

Figure 10 plots the empirical distributions from the counterfactual exercises of setting the inflation reaction coefficient to a range of plausible values, in particular $\phi_\pi = 1.2, 1.63, 2.0, 2.5$. Evidently, a more hawkish monetary policy rule shifts the empirical distribution towards more agents holding non-Ricardian beliefs. In fact, if the value for the intensity of choice ω was also large, then a more hawkish policy can feature regime-switching between the $n = 0$ and $n = 1$ misspecification equilibria: see Figure 11. That is, rather than a drifting share of Ricardian agents as in Figure 6, Figure 11 suggests that there would be abrupt, recurrent endogenous switching between the Ricardian and non-Ricardian regimes.[23]

[23]Similar findings arise if the central bank were to have altered the coefficient on the output gap ϕ_y.

Figure 10: Counterfactual empirical distribution of Ricardian beliefs when $\phi_\pi = 1.2, 2.0, 2.5$

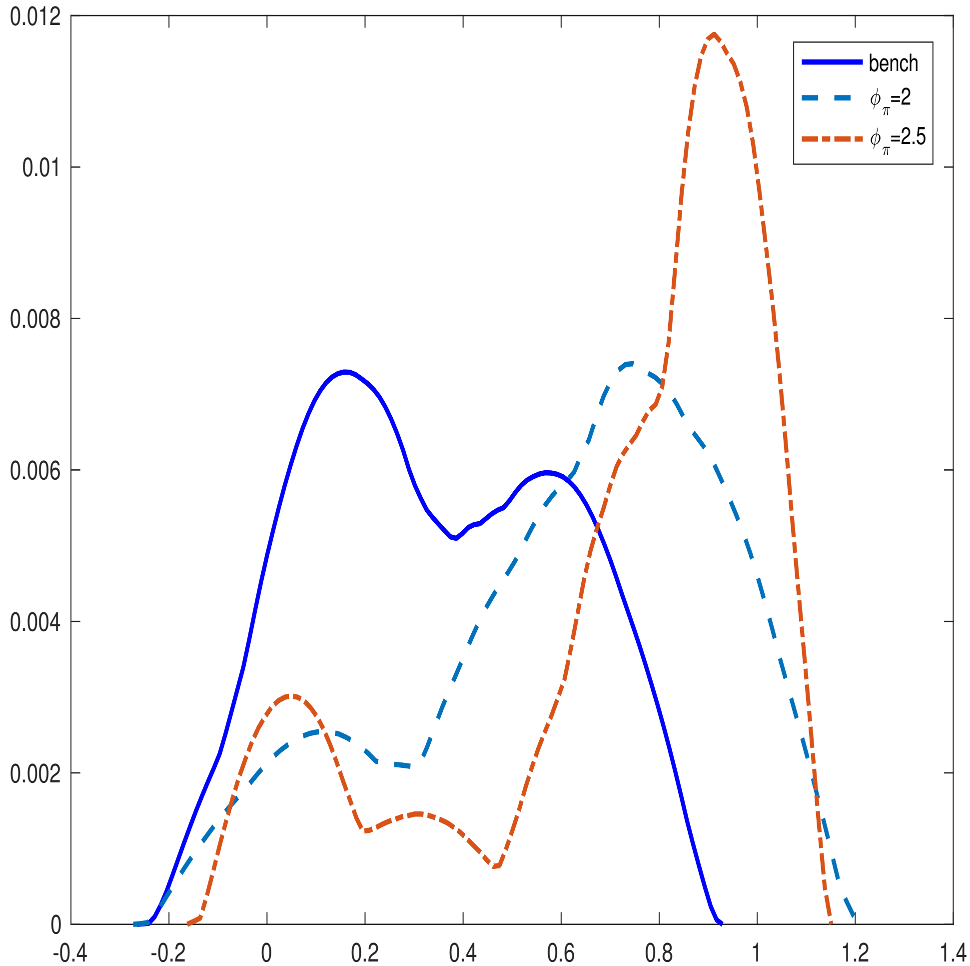


5.7 Fiscal policy

The theoretical results show that the region for which an $n = 1$ misspecification equilibrium (i.e., all agents are non-Ricardian) exists is increasing in the fiscal policy coefficient ϕ_b. As predicted, a counterfactual analysis of larger values for ϕ_b lead to more non-Ricardian beliefs and a greater frequency spent at the $n = 1$ non-Ricardian equilibrium. Recall, as well, that larger values of ϕ_b produce less serial correlation in the primary surplus and government debt processes.

Figure 12 confirms this prediction over the sample period. There is a higher average fraction of n and the time period when $n > 0.5$ arrives earlier. However, the panels in the figure also show that such outcomes are not in line with actual U.S. economic experience. The

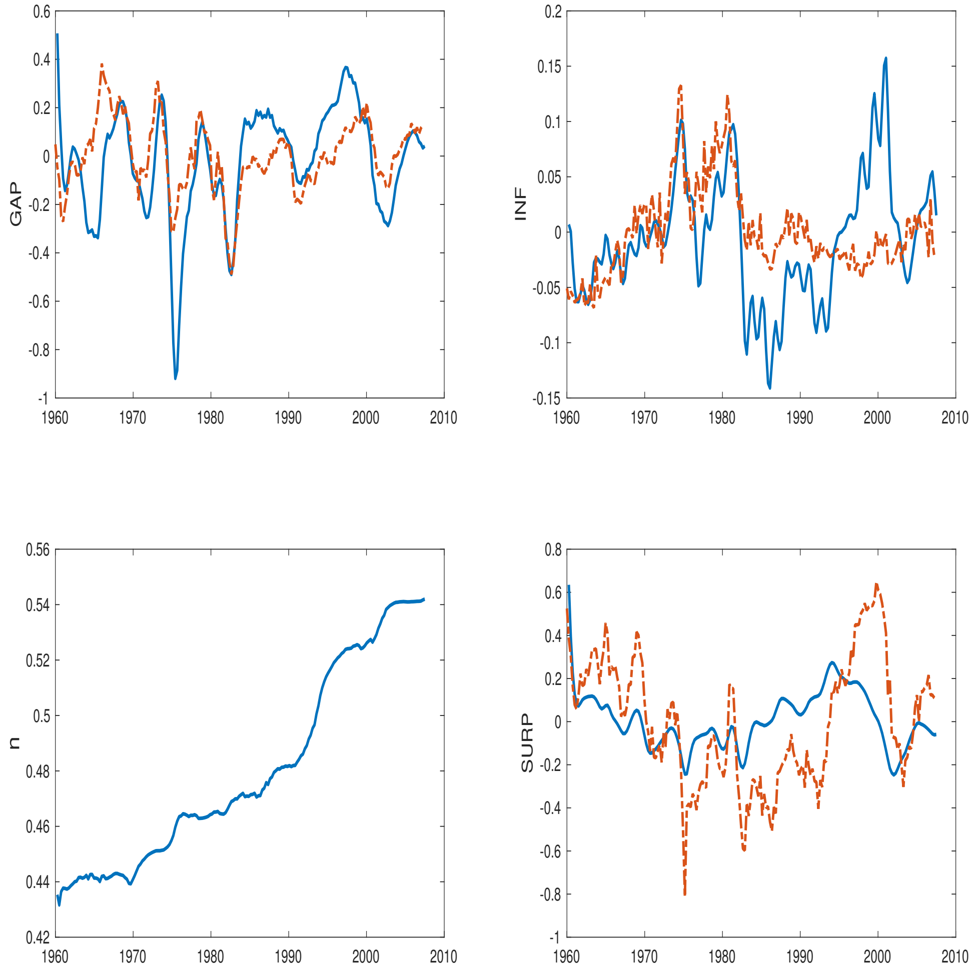
Figure 11: Counterfactual empirical distribution of Ricardian beliefs when $\phi_\pi = 1.2, 2.0, 2.5$ and $\omega = 300$.



larger value for ϕ_b actually produces too little volatility – it decreases the serial correlation in surplus and debt – in the surplus-GDP ratio and does not capture the overall movements either. During the 1970's this leads to counterfactually small output gaps, and post 1980 inflation dynamics that do not approximate actual inflation.

Plotting the empirical distributions for n, though, do show how a more aggressive fiscal policy would lead to more non-Ricardian beliefs. Figure 13 plots the empirical distributions if $\phi_b = 0.02, 0.10, 0.15$. It is clear from Figure 13 that for passive fiscal policy rules with $\phi_b = 0.02 - 0.10$ that the distribution of Ricardian beliefs are fairly tightly distributed in the interval $[0.4, 0.6]$. However, if $\phi_b = 0.15$ instead, then the economy would have spent a substantial fraction of time at the non-Ricardian equilibrium $n = 1$, as well as exhibit

Figure 12: Counterfactual State Dynamics when $\phi_b = 0.10$. Solid lines are model predicted state variables. Dashed lines are corresponding U.S. data.

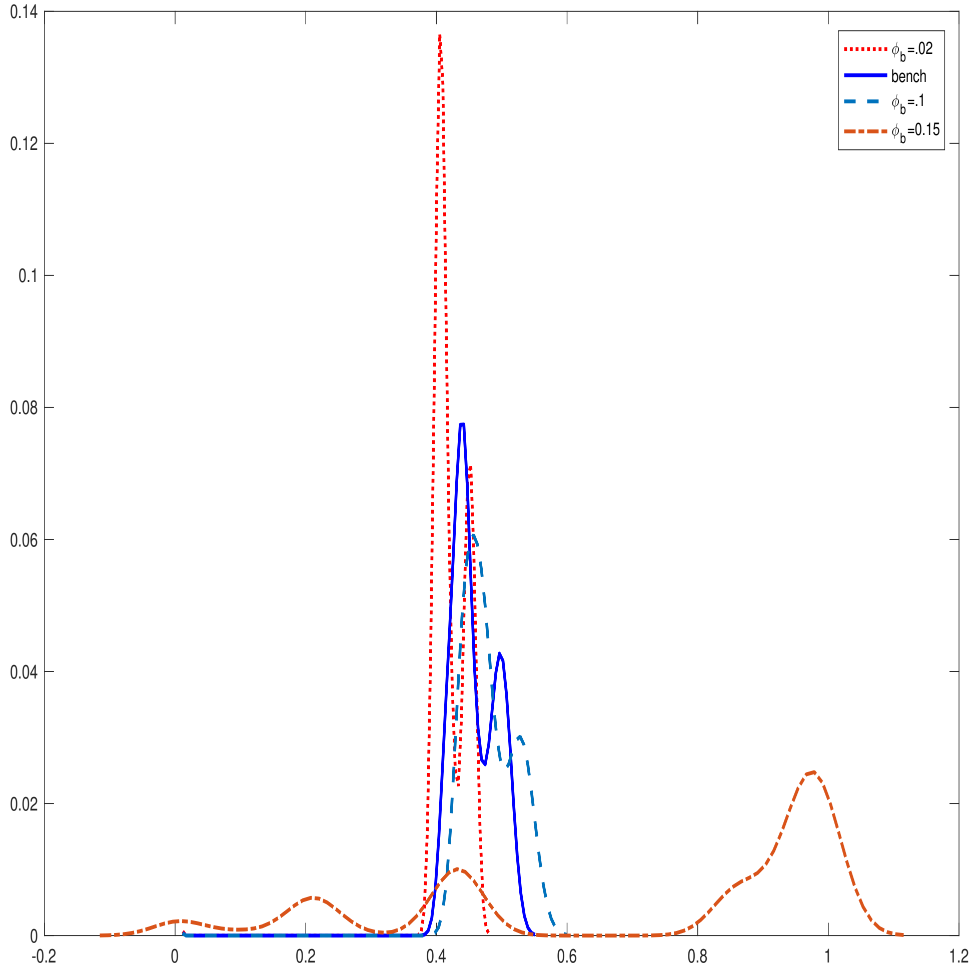


regime-switches with the economy frequently being near the Ricardian equilibrium $n = 0$.

6 Conclusion

This paper proposes a theory of expectation formation, based on restricted perceptions, that produces endogenously (non-)Ricardian beliefs. The building blocks of our paper come from the theory of non-Ricardian beliefs when individuals have imperfect knowledge about the long-run consequences of fiscal and monetary policy, first proposed by Eusepi and Preston (2018a). We follow Woodford (2013) and give the households and firms restricted perceptions

Figure 13: Counterfactual empirical distribution of Ricardian beliefs when $\phi_b = 0.02, 0.043, 0.10, 0.15$.



by allowing them to form expectations from models that include only a single fiscal variable – either the existing stock of government bonds or the primary surplus – while model-consistent rational expectations would condition on all relevant state variables. Despite the forecast model misspecification, in a restricted perceptions equilibrium agents' beliefs are optimal within the restricted class. The set of forecast models entertained by agents are natural. First, in complex forecasting environments with many state variables and potential degrees of freedom limitations, forecasters typically embrace parsimonious models. Second, as we show, the forecast models presented to agents are a natural formalization of endogenous (non-)Ricardian beliefs. When all agents forecast with the "debt-model" then Ricardian equivalence emerges as a self-confirming equilibrium. On the other hand, with some positive fraction of "surplus-model" forecasters then Ricardian equivalence fails.

These results highlight the fragile nature of Ricardian equivalence and motivate our central interest in focusing on misspecification equilibria as a refinement that endogenizes the distribution of agents across these two forecasting models. We provide necessary and sufficient conditions for (non-)Ricardian beliefs to emerge endogenously in a misspecification equilibrium. Throughout, the government is committed to a policy regime where taxes are adjusted to meet the government's intertemporal obligations and monetary policy is conducted via a Taylor rule. Our main theoretical results are as follows. If fiscal policy adjusts the primary surplus sufficiently strongly to the existing stock of government debt (while still remaining passive) then the non-Ricardian equilibrium can emerge as the unique misspecification equilibrium. Conversely, a weaker adjustment of the surplus leads to a unique Ricardian equilibrium. For some parameterizations of the model it is also possible for there to exist multiple misspecification equilibria, with the simultaneous existence of Ricardian and non-Ricardian equilibria.

This latter result motivates the quantitative exercise presented in the paper. Using the estimates in Eusepi and Preston (2018a), we show that multiple equilibria may exist in the U.S. economy and a real-time learning formulation where beliefs endogenously switch between the Ricardian and non-Ricardian belief regimes provide an alternative interpretation to the findings of regime-switching monetary/fiscal policy explanation of inflation in the U.S. We estimate the extent of (non-)Ricardian beliefs using the data-implied predicted paths of the endogenous state variables. Our estimates lead us to conclude that time-varying non-Ricardian beliefs is a potentially important component of U.S. inflation dynamics.

References

Berardi, M. and Galimberti, J. (2017). On the Initialization of Adaptive Learning in Macroeconomic Models. *Journal of Economic Dynamics and Control*, 78:26–53.

Bianchi, F. and Ilut, C. (2017). Monetary/Fiscal Policy Mix and Agents' Beliefs. *Review of Economic Dynamics*, 26:113–139.

Branch, W. A. (2006). Restricted Perceptions Equilibria and Learning in Macroeconomics. In Colander, D., editor, *Post-Walrasian Macroeconomics: Beyond the Dynamic, Stochastic General Equilibrium Model*. Cambridge University Press.

Branch, W. A., Davig, T., and McGough, B. (2008). Monetary-Fiscal Policy Interactions under Implementable Monetary Policy Rules. *Journal of Money, Credit and Banking*, 40(5):1095–1102.

Branch, W. A. and Evans, G. W. (2006a). A Simple Recursive Forecasting Model. *Economic Letters*, 91:158–166.

Branch, W. A. and Evans, G. W. (2006b). Intrinsic Heterogeneity in Expectation Formation. *Journal of Economic Theory*, 127(1):264–295.

Branch, W. A. and Evans, G. W. (2011). Monetary Policy and Heterogeneous Expectations. *Economic Theory*, 47(2-3):365–393.

Branch, W. A. and McGough, B. (2018). Heterogeneous Expectations and Micro-Foundations in Macroeconomics. In Hommes, C. H. and LeBaron, B., editors, *Handbook of Computational Economics*, volume 4. Elsevier, Amsterdam.

Brock, W. A. and Hommes, C. H. (1997). A Rational Route to Randomness. *Econometrica*, 65(5):1059–1095.

Bullard, J. B. and Mitra, K. (2002). Learning about Monetary Policy Rules. *Journal of Monetary Economics*, 49(6):1105–1129.

Calvo, G. A. (1983). Staggered Prices in a Utility-Maximizing Framework. *Journal of Monetary Economics*, 12(3):383–398.

Cho, I.-K. and Kasa, K. (2015). Learning and model validation. *Review of Economic Studies*, 82:45–82.

Cho, I.-K. and Kasa, K. (2017). Gresham's law of model averaging. *American Economic Review, forthcoming.*

Cho, I.-K., Williams, N., and Sargent, T. J. (2002). Escaping Nash Inflation. *Review of Economic Studies*, 69:1–40.

Cochrane, J. H. (2001). Long-Term Debt and Optimal Policy in the Fiscal Theory of the Price Level. *Econometrica*, 69(1):69–116.

Davig, T. and Leeper, E. M. (2006). Fluctuating Macro Policies and the Fiscal Theory. In Acemoğlu, D., Rogoff, K. S., and Woodford, M., editors, *NBER Macroeconomics Annual*, volume 21, pages 247–316. MIT Press, Cambridge, MA.

Eusepi, S. and Preston, B. (2011). Learning the Fiscal Theory of the Price Level: Some Consequences of Debt Management Policy. *Journal of the Japanese and International Economies*, 25(4):358–379.

Eusepi, S. and Preston, B. (2012). Debt, Policy Uncertainty and Expectations Stabilization. *Journal of the European Economic Association*, 10(4):860–886.

Eusepi, S. and Preston, B. (2018a). Fiscal Foundations of Inflation: Imperfect Knowledge. *American Economic Review*, forthcoming.

Eusepi, S. and Preston, B. (2018b). The Science of Monetary Policy: An Imperfect Knowledge Perspective. *Journal of Economic Literature*, 56(1):3–59.

Evans, G. W. and Honkapohja, S. (2001). *Learning and Expectations in Macroeconomics*. Frontiers of Economic Research. Princeton University Press, Princeton, NJ.

Evans, G. W. and Honkapohja, S. (2003). Expectations and the Stability Problem for Optimal Monetary Policies. *Review of Economic Studies*, 70(4):807–824.

Evans, G. W. and Honkapohja, S. (2007). Policy Interaction, Learning, and the Fiscal Theory of Prices. *Macroeconomic Dynamics*, 11(5):665–690.

Evans, G. W., Honkapohja, S., and Mitra, K. (2012). Does Ricardian Equivalence Hold When Expectations are not Rational? *Journal of Money, Credit and Banking*, 44(7):1259–1283.

Gasteiger, E. (2018). Do Heterogeneous Expectations Constitute a Challenge for Policy Interaction? *Macroeconomic Dynamics*, 22(8).

Kreps, D. M. (1998). Anticipated Utility and Dynamic Choice. In Jacobs, D. P., Kalai, E., and Kamien, M. I., editors, *Frontiers of Research in Economic Theory: The Nancy L. Schwartz Memorial Lectures, 1983-1997*. Cambridge University Press, Cambridge, UK.

Leeper, E. M. (1991). Equilibria under Active and Passive Monetary and Fiscal Policies. *Journal of Monetary Economics*, 27(1):129–147.

Leeper, E. M. and Leith, C. (2016). Understanding Inflation as a Joint Monetary-Fiscal Phenomenon. In Taylor, J. B. and Uhlig, H., editors, *Handbook of Macroeconomics*, volume 2, chapter 30, pages 2305–2415. Elsevier, Amsterdam, North-Holland.

Orphanides, A. and Williams, J. C. (2005). Imperfect Knowledge, Inflation Expectations, and Monetary Policy. In Bernanke, B. S. and Woodford, M., editors, *The Inflation-Targeting Debate*, volume 32 of *Studies in Business Cycles*, pages 201–234. The University of Chicago Press, Chicago and London.

Preston, B. (2005). Learning about Monetary Policy Rules when Long-Horizon Expectations Matter. *International Journal of Central Banking*, 1(2):81–126.

Sargent, T. J. (1999). *The Conquest of American Inflation*. Princeton University Press, Princeton, NJ.

Sargent, T. J. (2008). Evolution and intelligent design. *American Economic Review*, 98:5–37.

Sims, C. A. (1994). A Simple Model for Study of the Determination of the Price Level and the Interaction of Monetary and Fiscal Policy. *Economic Theory*, 4(3):381–399.

Sims, C. A. (2011). Stepping on a Rake: the Role of Fiscal Policy in the Inflation of the 1970s. *European Economic Review*, 55(1):48–56.

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

Williams, N. (2018). Escape Dynamics in Learning Models. Working Paper.

Woodford, M. (2001). Fiscal Requirements for Price Stability. *Journal of Money, Credit and Banking*, 33(3):669–728.

Woodford, M. (2003). *Interest and Prices: Foundations of a Theory of Monetary Policy*. Princeton University Press, Princeton, NJ.

Woodford, M. (2013). Macroeconomic Analysis Without the Rational Expectations Hypothesis. *Annual Review of Economics*, 5:303–346.

A Methodological framework

A.1 Computation of the restricted perceptions equilibrium

For a given distribution of PLMs, n, for all versions of the model the RPE can be computed in a similar way. First, we can re-organize the ALM to obtain

$$y_t = \delta_0 b_{t+1} + \delta_1 b_t + \delta_2 s_t + \delta_3 u_t \tag{A.1.1}$$

$$b_{t+1} = \xi_1 b_t + \xi_2 s_t + \xi_3 u_t. \tag{A.1.2}$$

Moreover, we can aggregate (12) and combine it with (6), (16), (A.1.1) and (A.1.2) to obtain

$$v_t = \mu_{v,1} b_t + \mu_{v,2} s_t + \mu_{v,3} u_t. \tag{A.1.3}$$

Likewise we can use (16), (4), (A.1.1) and (A.1.2) to obtain

$$p_t^* = \mu_{p,1} b_t + \mu_{p,2} s_t + \mu_{p,3} u_t. \tag{A.1.4}$$

Next, recall that PLMs are given by

$$\mathbf{z}_t = \psi^s s_{t-1} + \eta_t$$

$$\mathbf{z}_t = \psi^b b_{t-1} + \eta_t,$$

where $\mathbf{z}_t \equiv (v_t, p_t^*)'$, $\psi^s \equiv (\psi_v^s, \psi_p^s)'$, $\psi^b \equiv (\psi_v^b, \psi_p^b)'$ and $\eta_t \equiv (\eta_{v,t}, \eta_{p,t})'$. This implies four orthogonality conditions that can be written as

$$0 \stackrel{!}{=} E[s_{t-1}\eta_t] = E[s_t \eta_{t+1}] \tag{A.1.5}$$

$$0 \stackrel{!}{=} E[b_{t-1}\eta_t] = E[b_t \eta_{t+1}].$$

Next, as before, plug the PLM and ALM into (A.1.5), i.e.,

$$0 \stackrel{!}{=} E[s_t \eta_{t+1}] = E[s_t(\mathbf{z}_{t+1} - \psi^s s_t)]$$

$$\Leftrightarrow \psi^s E[s_t^2] = E[s_t \mathbf{z}_{t+1}]. \tag{A.1.6}$$

Equation by equation, we obtain

$$\Leftrightarrow \psi_v^s E[s_t^2] = E\left[s_t\left(\mu_{v,1} b_{t+1} + \mu_{v,2} s_{t+1} + \mu_{v,3} u_{t+1}\right)\right]$$

$$\psi_v^s E[s_t^2] = \mu_{v,1} E\left[s_t b_{t+1}\right] + \mu_{v,2} E\left[s_t s_{t+1}\right] + \mu_{v,3} E\left[s_t u_{t+1}\right]$$

$$\Leftrightarrow \psi_v^s = \mu_{v,1} \frac{E[s_t b_{t+1}]}{E[s_t^2]} + \mu_{v,2} \frac{E[s_t s_{t+1}]}{E[s_t^2]} + \mu_{v,3} \frac{E[s_t u_{t+1}]}{E[s_t^2]} \quad \text{and} \tag{A.1.7}$$

$$
\begin{aligned}
\Leftrightarrow \psi_p^s E[s_t^2] &= E\left[s_t\left(\mu_{p,1}b_{t+1} + \mu_{p,2}s_{t+1} + \mu_{p,3}u_{t+1}\right)\right] \\
\psi_p^s E[s_t^2] &= \mu_{p,1}E\left[s_t b_{t+1}\right] + \mu_{p,2}E\left[s_t s_{t+1}\right] + \mu_{p,3}E\left[s_t u_{t+1}\right] \\
\Leftrightarrow \psi_p^s &= \mu_{p,1}\frac{E[s_t b_{t+1}]}{E[s_t^2]} + \mu_{p,2}\frac{E[s_t s_{t+1}]}{E[s_t^2]} + \mu_{p,3}\frac{E[s_t u_{t+1}]}{E[s_t^2]}. \qquad \text{(A.1.8)}
\end{aligned}
$$

Likewise plug the PLM and ALM into (A.1.6), i.e.,

$$
\begin{aligned}
0 &\overset{!}{=} E[b_t \eta_{t+1}] = E[b_t(\mathbf{z}_{t+1} - \psi^b b_t)] \\
\Leftrightarrow \psi^b E[b_t^2] &= E[b_t \mathbf{z}_{t+1}].
\end{aligned}
$$

Again, equation by equation, we obtain

$$
\begin{aligned}
\Leftrightarrow \psi_v^b E[b_t^2] &= E\left[b_t\left(\mu_{v,1}b_{t+1} + \mu_{v,2}s_{t+1} + \mu_{v,3}u_{t+1}\right)\right] \\
\psi_v^b E[b_t^2] &= \mu_{v,1}E\left[b_t b_{t+1}\right] + \mu_{v,2}E\left[b_t s_{t+1}\right] + \mu_{v,3}E\left[b_t u_{t+1}\right] \\
\Leftrightarrow \psi_v^b &= \mu_{v,1}\frac{E[b_t b_{t+1}]}{E[b_t^2]} + \mu_{v,2}\frac{E[b_t s_{t+1}]}{E[b_t^2]} + \mu_{v,3}\frac{E[b_t u_{t+1}]}{E[b_t^2]} \qquad \text{and} \qquad \text{(A.1.9)} \\
\Leftrightarrow \psi_p^b E[b_t^2] &= E\left[b_t\left(\mu_{p,1}b_{t+1} + \mu_{p,2}s_{t+1} + \mu_{p,3}u_{t+1}\right)\right] \\
\psi_p^b E[b_t^2] &= \mu_{p,1}E\left[b_t b_{t+1}\right] + \mu_{p,2}E\left[b_t s_{t+1}\right] + \mu_{p,3}E\left[b_t u_{t+1}\right] \\
\Leftrightarrow \psi_p^b &= \mu_{p,1}\frac{E[b_t b_{t+1}]}{E[b_t^2]} + \mu_{p,2}\frac{E[b_t s_{t+1}]}{E[b_t^2]} + \mu_{p,3}\frac{E[b_t u_{t+1}]}{E[b_t^2]}. \qquad \text{(A.1.10)}
\end{aligned}
$$

The next step is to compute the moments. For this purpose, it is convenient to combine (A.1.2) and (7) in a VAR(1), i.e.,

$$
\begin{bmatrix} 1 & 0 \\ -\phi_b & 1 \end{bmatrix}\begin{bmatrix} b_{t+1} \\ s_{t+1} \end{bmatrix} = \begin{bmatrix} \xi_1 & \xi_1 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} b_t \\ s_t \end{bmatrix} + \begin{bmatrix} \xi_3 & 0 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} u_t \\ z_{t+1} \end{bmatrix} \qquad \text{(A.1.11)}
$$

$$
\Leftrightarrow \mathcal{Y}_t = \mathbf{A}\mathcal{Y}_{t-1} + \mathbf{C}\varepsilon_t, \qquad \text{(A.1.12)}
$$

where $\mathcal{Y}_t \equiv (b_t, s_t)'$ and $\varepsilon_t \equiv (u_{t-1}, z_t)'$.

Define the variance-covariance matrix $\mathbf{\Omega} \equiv E[\mathcal{Y}_t\mathcal{Y}_t']$ and likewise $\mathbf{\Sigma} \equiv E[\varepsilon_t\varepsilon_t']$, both constant $\forall t$. Then we can compute

$$
\begin{aligned}
\mathbf{\Omega} &= E[(\mathbf{A}\mathcal{Y}_{t-1} + \mathbf{C}\varepsilon_t)(\mathbf{A}\mathcal{Y}_{t-1} + \mathbf{C}\varepsilon_t)'] = \mathbf{A}E[\mathcal{Y}_{t-1}\mathcal{Y}_{t-1}']\mathbf{A}' + \mathbf{C}E[\varepsilon_t\varepsilon_t']\mathbf{C}' \\
\mathbf{\Omega} &= \mathbf{A}\mathbf{\Omega}\mathbf{A}' + \mathbf{C}\mathbf{\Sigma}\mathbf{C}' \\
\Leftrightarrow \operatorname{vec}(\mathbf{\Omega}) &= \left[\mathbf{I} - \mathbf{A}\otimes\mathbf{A}\right]^{-1}\left(\mathbf{C}\otimes\mathbf{C}\right)\operatorname{vec}(\mathbf{\Sigma})
\end{aligned}
$$

Moreover, the auto-covariance matrix is defined as $E[\mathcal{Y}_t\mathcal{Y}_{t-1}']$, thus

$$
E[\mathcal{Y}_t\mathcal{Y}_{t-1}'] = E[(\mathbf{A}\mathcal{Y}_{t-1}\mathcal{Y}_{t-1}' + \mathbf{C}\varepsilon_t\mathcal{Y}_{t-1}')] = \mathbf{A}E[\mathcal{Y}_{t-1}\mathcal{Y}_{t-1}'] = \mathbf{A}\mathbf{\Omega}.
$$

Notice that

$$\mathbf{\Omega} = \begin{bmatrix} E[b_t^2] & E[b_t s_t] \\ E[s_t b_t] & E[s_t^2] \end{bmatrix}, \qquad \mathbf{A\Omega} = \begin{bmatrix} E[b_{t+1} b_t] & E[b_{t+1} s_t] \\ E[s_{t+1} b_t] & E[s_{t+1} s_t] \end{bmatrix}. \tag{A.1.13}$$

Recall definitions $\Gamma_b^s \equiv E[b_{t+1}s_t]/E[s_t^2]$ and $\Gamma_b^b \equiv E[b_{t+1}b_t]/E[b_t^2]$ as well as $E[b_{t+1}s_t] = E[s_t b_{t+1}]$, $E[b_{t+1}b_t] = E[b_t b_{t+1}]$, $E[s_{t+1}s_t] = E[s_t s_{t+1}]$, and that $E[s_t u_{t+1}] = E[b_t u_{t+1}] = 0$. Moreover, recall that (7) implies that $E[s_t s_{t+1}] = \phi_b E[s_t b_{t+1}]$ and that $E[b_t s_{t+1}] = \phi_b E[b_t b_{t+1}]$. Thus, we can rewrite (A.1.7), (A.1.8), (A.1.9) and (A.1.10) as

$$\psi_v^s(n) = \mu_{v,1}\Gamma_b^s + \mu_{v,2}\phi_b\Gamma_b^s \tag{A.1.14}$$
$$\psi_p^s(n) = \mu_{p,1}\Gamma_b^s + \mu_{p,2}\phi_b\Gamma_b^s$$
$$\psi_v^b(n) = \mu_{v,1}\Gamma_b^b + \mu_{v,2}\phi_b\Gamma_b^b \tag{A.1.15}$$
$$\psi_p^b(n) = \mu_{p,1}\Gamma_b^b + \mu_{p,2}\phi_b\Gamma_b^b. \tag{A.1.16}$$

These conditions can be solved for $\psi_v^s(n)$, $\psi_p^s(n)$, $\psi_v^b(n)$, and $\psi_p^b(n)$. In case for $s_b > 0$, this can only be achieved numerically as matrices **A** and **C** in (A.1.12) also depend on these coefficients.

A.2 Computation of the misspecification equilibrium

Given the objective (17), we use (A.1.3) and (7) to compute

$$v_{t+1} = \mu_{v,1}b_{t+1} + \mu_{v,2}s_{t+1} + \mu_{v,3}u_{t+1}$$
$$v_{t+1} = [\mu_{v,1} + \mu_{v,2}\phi_b]b_{t+1} + \mu_{v,2}z_{t+1} + \mu_{v,3}u_{t+1}.$$

More concise, this is

$$v_{t+1} = \Upsilon_{v,b}b_{t+1} + \Upsilon_{v,z}z_{t+1} + \Upsilon_{v,u}u_{t+1}, \qquad \text{where} \tag{A.2.1}$$
$$\Upsilon_{v,b}(\psi_v^s(n), \psi_v^b(n), n) = \big[\mu_1\left(\psi_v^b(n), \xi_1(\delta_1(\psi_v^b(n), n)), \delta_1(\psi_v^b(n), n), n\right)$$
$$+\mu_2\left(\psi_v^s(n), \xi_2(\delta_2(\psi_v^s(n), n)), \delta_2(\psi_v^s(n), n), n\right)\phi_b\big] \equiv \Upsilon_{v,b}$$
$$\Upsilon_{v,z}(\psi_v^s(n), n) = \mu_2\left(\psi_v^s(n), \xi_2(\delta_2(\psi_v^s(n), n)), \delta_2(\psi_v^s(n), n), n\right) \equiv \Upsilon_{v,z}$$
$$\Upsilon_{v,u} \equiv \mu_{v,3}.$$

Next, we use (A.1.4) and (7) to compute

$$p_{t+1}^* = \mu_{p,1}b_{t+1} + \mu_{p,2}s_{t+1} + \mu_{p,3}u_{t+1}$$
$$p_{t+1}^* = [\mu_{p,1} + \mu_{p,2}\phi_b]b_{t+1} + \mu_{p,2}z_{t+1} + \mu_{p,3}u_{t+1}.$$

More concise, this is

$$
\begin{aligned}
p^*_{t+1} &= \Upsilon_{p,b} b_{t+1} + \Upsilon_{p,z} z_{t+1} + \Upsilon_{p,u} u_{t+1}, \qquad \text{where} \qquad \text{(A.2.2)}\\
\Upsilon_{p,b}(\psi^s_p(n), \psi^b_p(n), n) &= \left[\mu_1\left(\psi^b_v(n), \xi_1(\delta_1(\psi^b_v(n), n)), \delta_1(\psi^b_v(n), n), n\right)\right.\\
&\quad \left. +\mu_2\left(\psi^s_v(n), \xi_2(\delta_2(\psi^s_v(n), n)), \delta_2(\psi^s_v(n), n), n\right)\phi_b\right] \equiv \Upsilon_{p,b}\\
\Upsilon_{p,z}(\psi^s_p(n), n) &= \mu_2\left(\psi^s_v(n), \xi_2(\delta_2(\psi^s_v(n), n)), \delta_2(\psi^s_v(n), n), n\right) \equiv \Upsilon_{p,z}\\
\Upsilon_{p,u} &\equiv \mu_{p,3}.
\end{aligned}
$$

We can combine (A.2.1) and (A.2.2) to obtain

$$\mathbf{z}_{t+1} = \Upsilon_b b_{t+1} + \Upsilon_z z_{t+1} + \Upsilon_u u_{t+1}, \qquad \text{(A.2.3)}$$

where $\Upsilon_b \equiv (\Upsilon_{v,b}, \Upsilon_{p,b})'$, $\Upsilon_z \equiv (\Upsilon_{v,z}, \Upsilon_{p,z})'$ and $\Upsilon_u \equiv (\Upsilon_{v,u}, \Upsilon_{p,u})'$. Moreover, recall

$$
\begin{aligned}
E[z^1_{t+1}] &= \psi^s(n) s_t, \qquad \text{and} \qquad \text{(A.2.4)}\\
E[z^2_{t+1}] &= \psi^b(n) b_t. \qquad \text{(A.2.5)}
\end{aligned}
$$

Thus, we can use (A.2.3) and (A.2.4) to compute

$$
\begin{aligned}
(\mathbf{z}_{t+1} - E[z^1_{t+1}]) =& \Upsilon_b b_{t+1} + \Upsilon_z z_{t+1} + \Upsilon_u u_{t+1} - \psi^s(n) s_t\\
(\mathbf{z}_{t+1} - E[z^1_{t+1}])'(\mathbf{z}_{t+1} - E[z^1_{t+1}]) =& (\Upsilon_b'\Upsilon_b) b^2_{t+1} + (\Upsilon_z'\Upsilon_z) z^2_{t+1} + (\Upsilon_u'\Upsilon_u) u^2_{t+1}\\
&+ (\psi^s(n)'\psi^s(n)) s^2_t\\
&+ 2(\Upsilon_b'\Upsilon_z) b_{t+1} z_{t+1} + 2(\Upsilon_b'\Upsilon_u) b_{t+1} u_{t+1}\\
&+ 2(\Upsilon_z'\Upsilon_u) z_{t+1} u_{t+1} - 2(\Upsilon_b'\psi^s(n)) b_{t+1} s_t\\
&- 2(\Upsilon_z'\psi^s_v(n)) z_{t+1} s_t - 2(\Upsilon_u'\psi^s_v(n)) u_{t+1} s_t \qquad \Big| E[\cdot]\\
E[(\mathbf{z}_{t+1} - E[z^1_{t+1}])'(\mathbf{z}_{t+1} - E[z^1_{t+1}])] =& (\Upsilon_b'\Upsilon_b) E[b^2_{t+1}] + (\Upsilon_z'\Upsilon_z) E[z^2_{t+1}] + (\Upsilon_u'\Upsilon_u) E[u^2_{t+1}]\\
&+ (\psi^s(n)'\psi^s(n)) E[s^2_t] - 2(\Upsilon_b'\psi^s(n)) E[b_{t+1} s_t]\\
\Leftrightarrow EU^1 =& - \left[(\Upsilon_b'\Upsilon_b) E[b^2_{t+1}] + (\Upsilon_z'\Upsilon_z)\sigma^2_z + (\Upsilon_u'\Upsilon_u)\sigma^2_u\right.\\
&\left. + (\psi^s(n)'\psi^s(n)) E[s^2_t] - 2(\Upsilon_b'\psi^s(n)) E[b_{t+1} s_t]\right].
\end{aligned}
$$

Likewise, we can use (A.2.3) and (A.2.5) to compute

$$
\begin{aligned}
(\mathbf{z}_{t+1} - E[z^2_{t+1}]) =& \Upsilon_b b_{t+1} + \Upsilon_z z_{t+1} + \Upsilon_u u_{t+1} - \psi^b(n) b_t\\
(\mathbf{z}_{t+1} - E[z^2_{t+1}])'(\mathbf{z}_{t+1} - E[z^2_{t+1}]) =& (\Upsilon_b'\Upsilon_b) b^2_{t+1} + (\Upsilon_z'\Upsilon_z) z^2_{t+1} + (\Upsilon_u'\Upsilon_u) u^2_{t+1}\\
&+ (\psi^b(n)'\psi^b(n)) b^2_t\\
&+ 2(\Upsilon_b'\Upsilon_z) b_{t+1} z_{t+1} + 2(\Upsilon_b'\Upsilon_u) b_{t+1} u_{t+1}\\
&+ 2(\Upsilon_z'\Upsilon_u) z_{t+1} u_{t+1} - 2(\Upsilon_b'\psi^b(n)) b_{t+1} b_t\\
&- 2(\Upsilon_z'\psi^b_v(n)) z_{t+1} b_t - 2(\Upsilon_u'\psi^b_v(n)) u_{t+1} b_t \qquad \Big| E[\cdot]
\end{aligned}
$$

$$
\begin{aligned}
E[(\mathbf{z}_{t+1}-E[z_{t+1}^2])'(\mathbf{z}_{t+1}-E[z_{t+1}^2])] =& (\Upsilon_b'\Upsilon_b)E[b_{t+1}^2]+(\Upsilon_z'\Upsilon_z)E[z_{t+1}^2]+(\Upsilon_u'\Upsilon_u)E[u_{t+1}^2]\\
&+(\psi^b(n)'\psi^b(n))E[b_t^2]-2(\Upsilon_b'\psi^b(n))E[b_{t+1}b_t]\\
\Leftrightarrow EU^2 =& -\left[(\Upsilon_b'\Upsilon_b)E[b_{t+1}^2]+(\Upsilon_z'\Upsilon_z)\sigma_z^2+(\Upsilon_u'\Upsilon_u)\sigma_u^2\right.\\
&\left.+(\psi^b(n)'\psi^b(n))E[b_t^2]-2(\Upsilon_b'\psi^b(n))E[b_{t+1}b_t]\right].
\end{aligned}
$$

Next, one can define relative forecast performance as $F(n):[0,1]\to\mathbb{R}$ as $F(n)\equiv EU^1-EU^2$, thus

$$
\begin{aligned}
F(n) =& -(\Upsilon_b'\Upsilon_b)E[b_{t+1}^2]-(\Upsilon_z'\Upsilon_z)\sigma_z^2-(\Upsilon_u'\Upsilon_u)\sigma_u^2-(\psi^s(n)'\psi^s(n))E[s_t^2]\\
&+2(\Upsilon_b'\psi^s(n))E[b_{t+1}s_t]+(\Upsilon_b'\Upsilon_b)E[b_{t+1}^2]+(\Upsilon_z'\Upsilon_z)\sigma_z^2+(\Upsilon_u'\Upsilon_u)\sigma_u^2\\
&+(\psi^b(n)'\psi^b(n))E[b_t^2]-2(\Upsilon_b'\psi^b(n))E[b_{t+1}b_t]\\
F(n) =& -(\psi^s(n)'\psi^s(n))E[s_t^2]+2(\Upsilon_b'\psi^s(n))E[b_{t+1}s_t]-2(\Upsilon_b'\psi^b(n))E[b_{t+1}b_t]\\
&+(\psi^b(n)'\psi^b(n))E[b_t^2].
\end{aligned}
$$

We can rearrange this in the following way

$$
F(n) = -(\psi^s(n)'\psi^s(n))E[s_t^2]+2\Upsilon_b'\left(\psi^s(n)E[b_{t+1}s_t]-\psi^b(n)E[b_{t+1}b_t]\right)+(\psi^b(n)'\psi^b(n))E[b_t^2]
$$

$$
F(n) = E[s_t^2]\left[-(\psi^s(n)'\psi^s(n))+2\Upsilon_b'\left(\psi^s(n)\Gamma_b^s-\psi^b(n)\Gamma_b^bQ\right)+(\psi^b(n)'\psi^b(n))Q\right],
$$

where we make use of the definitions of Γ_b^s and Γ_b^b from above and define $Q\equiv E[b_t^2]/E[s_t^2]$.

B Proofs

B.1 Proof of Proposition 2

Proof. Suppose that all agents have PLM (19). From the simplifications made in Section 4 it follows that (8) becomes

$$
b_{t+1} = \beta^{-1}(b_t - s_t), \tag{B.1.1}
$$

which can be written as (A.1.2) with $\xi_1\equiv\beta^{-1}$, $\xi_2\equiv-\beta^{-1}$, and, $\xi_3=0$. Moreover, (13) is given by

$$
y_t = (\beta^{-1}-1)(b_t-s_t)+\psi_v^b b_t. \tag{B.1.2}
$$

Thus, coefficients in (A.1.1) are given by $\delta_0\equiv(1-\beta)$, $\delta_1\equiv\psi_v^b$, $\delta_2=0$, and, $\delta_3=0$.

Given homogeneous beliefs based on (19), i.e., $v_t^i = v_t, \forall i$, the implications for (12) are

$$\begin{aligned} v_t^i &= (1-\beta)v_t^i + (1-\beta)\beta[b_{t+1} - b_t] + \beta\psi_v^b b_t \\ \beta v_t^i &= (1-\beta)\beta[\beta^{-1}(b_t - s_t) - b_t] + \beta\psi_v^b b_t \\ v_t^i &= (1-\beta)[\beta^{-1}(b_t - s_t) - b_t] + \psi_v^b b_t \\ v_t^i &= (\beta^{-1} - 1)(b_t - s_t) + \psi_v^b b_t - (1-\beta)b_t \\ v_t^i &= y_t - (1-\beta)b_t. \end{aligned} \tag{B.1.3}$$

The ALM is then given by (B.1.1), (B.1.2) and (B.1.3).

Now we can apply $v_t \equiv \int v_t^i di$ to (B.1.3) and combine it with (B.1.2) to obtain (A.1.3) with coefficients $\mu_{v,1} \equiv [(\beta^{-1} - 1) - (1-\beta) + \psi_v^b]$, $\mu_{v,2} \equiv -(\beta^{-1} - 1)$, and, $\mu_{v,3} = 0$.

Under PF, i.e., assumption (9), we have $0 < \beta^{-1}(1-\phi_b) < 1$ and $\{b_t\}$ follows a stationary AR(1) process. Thus, we can compute the unconditional moments following the steps outlined in (A.1.11) to (A.1.13). Thus, we obtain

$$\Gamma_b^b = \frac{E[b_{t+1}b_t]}{E[b_t^2]} = \beta^{-1}(1-\phi_b). \tag{B.1.4}$$

where the linear projection $E[b_{t+1}] = \Gamma_b^b b_t$ satisfies an orthogonality condition. In consequence, (A.1.15) is given by

$$\begin{aligned} \psi_v^b &= \left[(\beta^{-1} - 1)(1-\beta-\phi_b) + \psi_v^b\right]\Gamma_b^b \\ 0 &= \psi_v^b - \left[(\beta^{-1} - 1)(1-\beta-\phi_b) + \psi_v^b\right]\Gamma_b^b, \end{aligned} \tag{B.1.5}$$

and (22) follows.

Notice that (B.1.2) together with (7) and (22) imply that

$$y_t = -(\beta^{-1} - 1)z_t,$$

thus, Ricardian equivalence holds in the sense that y_t depends not on b_t, but only z_t. Despite transitory effects of the surplus shock on aggregate output, there are no real effects of public debt. This proves Proposition 2.

□

B.2 Proof of Proposition 3

Proof. Suppose $n = 1$, i.e., all agents use PLM (18). In this case, (B.1.1) and (B.1.3) remain the same, however (13) becomes

$$y_t = -\sigma(\phi_\pi \pi_t) + (1-\beta)\beta^{-1}(b_t - s_t) + \psi_v^s s_t \tag{B.2.1}$$

$$y_t = (\beta^{-1} - 1)(b_t - s_t) + \psi_v^s s_t, \tag{B.2.2}$$

where (B.2.1) can be written as (A.1.1) with $\delta_0 \equiv (1-\beta)$, $\delta_1 = 0$, $\delta_2 \equiv \psi_v^s$, and, $\delta_3 = 0$.

The ALM is then given by (B.1.1), (B.2.2) and (B.1.3). Thus, we can apply $v_t \equiv \int v_t^i di$ to (B.1.3) and combine it with (B.2.2) to obtain (A.1.3) with coefficients $\mu_{v,1} \equiv [(\beta^{-1}-1) - (1-\beta)]$, $\mu_{v,2} \equiv \psi_v^s - (\beta^{-1}-1)$, and, $\mu_{v,3} = 0$.

Moments are computed as for (B.1.4) and we obtain (25), where the linear projection $E[b_{t+1}] = \Gamma_b^s s_t$ satisfies an orthogonality condition.

In consequence, (A.1.14) is given by (26) and therefore

$$\Leftrightarrow \psi_v^s = \frac{(1-\beta)(1-\beta-\phi_b)\Gamma_b^s}{\beta(1-\phi_b\Gamma_b^s)} = -\frac{\beta^{-1}(1-\beta)(1-\beta^2-\phi_b)}{(\beta+\beta^2+\phi_b)} \tag{B.2.3}$$

$$\Leftrightarrow \psi_v^s < \beta^{-1} - 1. \tag{B.2.4}$$

From (B.2.2), (7) and (B.2.3) to (B.2.4) follows that

$$y_t = \left[(\beta^{-1}-1)(1-\phi_b) + \phi_b\psi_v^s\right] b_t - \left[(\beta^{-1}-1) - \psi_v^s\right] z_t$$

$$y_t = \left[\frac{(1-\beta)(1+\beta-\phi_b)}{\beta(1+\beta)+\phi_b}\right] b_t - \left[(\beta^{-1}-1) - \psi_v^s\right] z_t.$$

Thus, as y_t depends on b_t, Ricardian equivalence fails. This proves Proposition 3.

□

B.3 Proof of Proposition 4

Proof. Consider the simplifications made in Section 4. Thus the TE dynamics are still governed by (B.1.1). Moreover, as expectations are heterogeneous. Therefore (13) becomes

$$y_t = (1-\beta)b_{t+1} + \int_0^1 \widehat{E}_t^i v_{t+1}^i di = (1-\beta)b_{t+1} + n\psi_v^s s_t + (1-n)\psi_v^b b_t, \tag{B.3.1}$$

for given expectations on $\{p_t^*(j), v_t^i\}$, i.e.,

$$v_t^i = (1-\beta)(b_{t+1} - b_t) + \widehat{E}_t^i v_{t+1}^i$$

$$\int_0^1 v_t^i di = v_t = (1-\beta)(b_{t+1} - b_t) + \int_0^1 \widehat{E}_t^i v_{t+1}^i di$$

$$= (1-\beta)(b_{t+1} - b_t) + n\psi_v^s s_t + (1-n)\psi_v^b b_t \tag{B.3.2}$$

$$\Leftrightarrow v_t = y_t - (1-\beta)b_t \tag{B.3.3}$$

$$p_t^*(j) = 0.$$

Thus, coefficients in (A.1.1) are given by $\delta_0 \equiv (1-\beta)$, $\delta_1 \equiv (1-n)\psi_v^b$, $\delta_2 \equiv n\psi_v^s$, and, $\delta_3 = 0$. Moreover, we can combine (B.3.2) with (B.1.1) and (B.3.1) to obtain (A.1.3) with coefficients $\mu_{v,1} \equiv [(\beta^{-1}-1)-(1-\beta)+(1-n)\psi_v^b]$, $\mu_{v,2} \equiv n\psi_v^s - (\beta^{-1}-1)$, and, $\mu_{v,3} = 0$.

The ALM is then given by (B.1.1) and (B.3.1) to (B.3.3). Coefficients in (18) and (19) are required to satisfy the orthogonality conditions (20) and (21) respectively.

Notice that unconditional moments are computed as explained in Appendix A.1 above and therefore the coefficients (25) and (B.1.4) are still true. Thus, we can obtain (A.1.14) and (A.1.14) as

$$\psi_v^s = [(\beta^1-1)-(1-\beta)+(1-n)\psi_v^b]\Gamma_b^s + [n\psi_v^s - (\beta^1-1)]\phi_b\Gamma_b^s$$
$$\psi_v^b = [(\beta^1-1)-(1-\beta)+(1-n)\psi_v^b]\Gamma_b^b + [n\psi_v^s - (\beta^1-1)]\phi_b\Gamma_b^b.$$

Rearranging terms yields

$$\psi_v^s = [(\beta^1-1)(1-\beta-\phi_b)]\Gamma_b^s + [\phi_b n\psi_v^s + (1-n)\psi_v^b]\Gamma_b^s \tag{B.3.4}$$
$$\psi_v^b = [(\beta^1-1)(1-\beta-\phi_b)]\Gamma_b^b + [\phi_b n\psi_v^s + (1-n)\psi_v^b]\Gamma_b^b. \tag{B.3.5}$$

Clearly, $n=0$ implies that (B.3.5) collapses to (B.1.5) and $n=1$ implies that (B.3.4) collapses to (26).

Thus, we can solve for

$$\Leftrightarrow \psi_v^s(n) = \frac{(1-\beta)\Gamma_b^s(1-\beta-\phi_b)}{\beta\left[1-\left(\phi_b n\Gamma_b^s+(1-n)\Gamma_b^b\right)\right]} = \frac{(1-\beta^2-\phi_b)\beta^{-1}(1-\beta)}{[1-\beta^2-n(1+\beta-\phi_b)-2\phi_b]}$$
$$\Leftrightarrow \psi_v^b(n) = \frac{(1-\beta)\Gamma_b^b(1-\beta-\phi_b)}{\beta\left[1-\left(\phi_b n\Gamma_b^s+(1-n)\Gamma_b^b\right)\right]} = \frac{-(1-\beta^2-2\phi_b)(1-\phi_b)\beta^{-1}(1-\beta)}{[1-\beta^2-n(1+\beta-\phi_b)-2\phi_b]}.$$

This proves Proposition 4.

□

B.4 Proof of Proposition 5

Proof. v_{t+1} is given by combining (29), (B.1.1) and (7), i.e.,

$$\begin{aligned}
v_{t+1} &= (1-\beta)(b_{t+2}-b_{t+1}) + n\psi_v^s(n)s_{t+1} + (1-n)\psi_v^b(n)b_{t+1} \\
v_{t+1} &= (1-\beta)(\beta^{-1}(b_{t+1}-s_{t+1})-b_{t+1}) + n\psi_v^s(n)s_{t+1} + (1-n)\psi_v^b(n)b_{t+1} \\
v_{t+1} &= \left[(1-\beta)(\beta^{-1}-1)+(1-n)\psi_v^b(n)\right]b_{t+1} + \left[n\psi_v^s(n)-(\beta^{-1}-1)\right]s_{t+1} \\
v_{t+1} &= \left[\left[(1-\beta)(\beta^{-1}-1)+(1-n)\psi_v^b(n)\right] + \phi_b\left[n\psi_v^s(n)-(\beta^{-1}-1)\right]\right]b_{t+1} \\
&\quad + \left[n\psi_v^s(n)-(\beta^{-1}-1)\right]z_{t+1}.
\end{aligned}$$

More concise, this is (A.2.1) with coefficients

$$\Upsilon_b(\psi_v^s(n), \psi_v^b(n), n) = \left[\left[(1-\beta)(\beta^{-1}-1) + (1-n)\psi_v^b(n)\right] + \phi_b\left[n\psi_v^s(n) - (\beta^{-1}-1)\right]\right] \equiv \Upsilon_b$$

$$\Upsilon_z(\psi_v^s(n), n) = \left[n\psi_v^s(n) - (\beta^{-1}-1)\right] \equiv \Upsilon_z$$
$$\Upsilon_u = 0.$$

Thus, we can compute $F(n)$ as outlined in Appendix A.2 above. We can also express $F(n)$ explicitly by plugging in, i.e.,

$$F(n) = \frac{(1-\beta)^2\sigma_z^2(2-\beta^2-2\phi_b)(1-\beta^2-2\phi_b)}{\beta^2(1-\beta^2-n(1+\beta-\phi_b)-2\phi_b)^2}, \tag{B.4.1}$$

where $\sigma_z^2 \equiv E[z_t z_t]$. From (B.4.1) one can observe that the denominator is always positive and whether $F(n)$ is positive or negative depends on the term $(2-\beta^2-2\phi_b)(1-\beta^2-2\phi_b)$.

Next, based on (B.4.1), we can characterize the behavior of $F(0)$ and $F(1)$ depending on ϕ_b for passive fiscal policy, i.e., condition (8) and $\beta \in (0,1)$, as follows

$$\begin{aligned}
F(1) &> 0, && \text{if} && \left(1-\frac{\beta^2}{2}\right) < \phi_b < 1 \\
F(0) &> 0, && \text{if} && \left(1-\frac{\beta^2}{2}\right) < \phi_b < 1 \\
F(1) &< 0, && \text{if} && (1-\beta) < \phi_b < \left(1-\frac{\beta^2}{2}\right) \\
F(0) &< 0, && \text{if} && (1-\beta) < \phi_b < \left(1-\frac{\beta^2}{2}\right).
\end{aligned}$$

This proves Proposition 1.

□

B.5 Proof of Proposition 6

Proof. The TE dynamics in this case are (B.1.1),

$$y_t = -\sigma\phi_\pi\pi_t + (1-\beta)b_{t+1} + n\psi_v^s s_t + (1-n)\psi_v^b b_t \tag{B.5.1}$$

$$\pi_t = \kappa y_t + u_t + (1-\alpha)\beta \int \widehat{E}_t^j p_{t+1}^*(j)dj$$
$$= \kappa y_t + u_t + (1-\alpha)\beta\left[n\psi_p^s s_t + (1-n)\psi_p^b b_t\right] \tag{B.5.2}$$

for given expectations on $\{p_t^*(j), v_t^i\}$, i.e.,

$$v_t = (1-\beta)(b_{t+1} - b_t) - \sigma(\phi_\pi - 1)\pi_t + n\psi_v^s s_t + (1-n)\psi_v^b b_t \quad \text{(B.5.3)}$$

$$p_t^*(j) = (1-\alpha)p_t^* + (1-\alpha\beta)\left[\xi y_t + \mu_t\right] + \alpha\beta \widehat{E}_t^j p_{t+1}^*(j) \quad \text{(B.5.4)}$$

$$\int_0^1 p_t^*(j)dj = p_t^* = \frac{(1-\alpha\beta)}{\alpha}\left[\xi y_t + \mu_t\right] + \beta \int_0^1 \widehat{E}_t^j p_{t+1}^*(j)dj$$

$$p_t^* = (1-\alpha)^{-1}\left[\kappa y_t + u_t\right] + \beta\left[n\psi_p^s s_t + (1-n)\psi_p^b b_t\right]$$

$$(1-\alpha)^{-1}\pi_t = (1-\alpha)^{-1}\left[\kappa y_t + u_t\right] + \beta\left[n\psi_p^s s_t + (1-n)\psi_p^b b_t\right]$$

$$\pi_t = \kappa y_t + u_t + (1-\alpha)\beta\left[n\psi_p^s s_t + (1-n)\psi_p^b b_t\right]. \quad \text{(B.5.5)}$$

The ALM is then given by (18) to (19), (B.1.1), (B.5.1), (B.5.2), (B.5.3), and (B.5.4) to (B.5.5) for given policy (7) and (8).

Next, we can combine (B.5.1) and (B.5.2) to obtain (A.1.3) with coefficients

$$\delta_0 \equiv \frac{(1-\beta)}{(1+\sigma\phi_\pi\kappa)}, \quad \delta_1 \equiv \frac{(1-n)\left(\psi_v^b - \sigma\phi_\pi(1-\alpha)\beta\psi_p^b\right)}{(1+\sigma\phi_\pi\kappa)},$$

$$\delta_2 \equiv \frac{n\left(\psi_v^s - \sigma\phi_\pi(1-\alpha)\beta\psi_p^s\right)}{(1+\sigma\phi_\pi\kappa)}, \quad \delta_3 \equiv \frac{-\sigma\phi_\pi}{(1+\sigma\phi_\pi\kappa)}.$$

Moreover, we use (A.1.3) to eliminate y_t in (B.5.5), i.e.,

$$\begin{aligned}
\pi_t =& \kappa\left[\delta_0 b_{t+1} + \delta_1 b_t + \delta_2 s_t + \delta_3 u_t\right] + u_t + (1-\alpha)\beta\left[n\psi_p^s s_t + (1-n)\psi_p^b b_t\right] \\
\pi_t =& \left[\kappa\delta_1 + (1-\alpha)\beta(1-n)\psi_p^b\right] b_t + \left[\kappa\delta_2 + (1-\alpha)\beta n\psi_p^s\right] s_t \\
&+ \left[\kappa\delta_3 + 1\right] u_t + \kappa\delta_0 b_{t+1}. \quad \text{(B.5.6)}
\end{aligned}$$

Furthermore, we use (A.1.3), (A.1.2) and (B.5.6) to eliminate π_t, y_t and b_{t+1} in (B.5.3), i.e.,

$$\begin{aligned}
v_t =& \left[(1-n)\psi_v^b - (1-\beta)\right] b_t + n\psi_v^s s_t - \sigma(\phi_\pi - 1)\pi_t + (1-\beta)b_{t+1} \\
v_t =& \left[(1-n)\psi_v^b - (1-\beta) - \sigma(\phi_\pi - 1)\left[\kappa\delta_1 + (1-\alpha)\beta(1-n)\psi_p^b\right]\right] b_t \\
&+ \left[n\psi_v^s - \sigma(\phi_\pi - 1)\left[\kappa\delta_2 + (1-\alpha)\beta n\psi_p^s\right]\right] s_t \\
&- \sigma(\phi_\pi - 1)\left[\kappa\delta_3 + 1\right] u_t \\
&+ \left[(1-\beta) - \sigma(\phi_\pi - 1)\kappa\delta_0\right] b_{t+1} \\
v_t =& \left[(1-n)\psi_v^b - (1-\beta) - \sigma(\phi_\pi - 1)\left[\kappa\delta_1 + (1-\alpha)\beta(1-n)\psi_p^b\right]\right] b_t \\
&+ \left[n\psi_v^s - \sigma(\phi_\pi - 1)\left[\kappa\delta_2 + (1-\alpha)\beta n\psi_p^s\right]\right] s_t \\
&- \sigma(\phi_\pi - 1)\left[\kappa\delta_3 + 1\right] u_t \\
&+ \Xi\left[\xi_1 b_t + \xi_2 s_t + \xi_3 u_t\right], \qquad \text{where} \qquad \Xi \equiv \left[(1-\beta) - \sigma(\phi_\pi - 1)\kappa\delta_0\right].
\end{aligned}$$

More concise, this is (A.1.3) with coefficients

$$\begin{aligned}
\mu_{v,1} &\equiv \left[(1-n)\psi_v^b - (1-\beta) - \sigma(\phi_\pi - 1)\left[\kappa\delta_1 + (1-\alpha)\beta(1-n)\psi_p^b\right] + \xi_1\Xi\right] \\
\mu_{v,2} &\equiv \left[n\psi_v^s - \sigma(\phi_\pi - 1)\left[\kappa\delta_2 + (1-\alpha)\beta n\psi_p^s\right] + \xi_2\Xi\right] \\
\mu_{v,3} &\equiv \left[\xi_3\Xi - \sigma(\phi_\pi - 1)\left[\kappa\delta_3 + 1\right]\right].
\end{aligned}$$

Moreover, we use (A.1.2) to eliminate b_{t+1} in (B.5.6), i.e.,

$$\begin{aligned}
\pi_t = &\left[\kappa\delta_1 + (1-\alpha)\beta(1-n)\psi_p^b\right] b_t + \left[\kappa\delta_2 + (1-\alpha)\beta n\psi_p^s\right] s_t + \left[\kappa\delta_3 + 1\right] u_t \\
&+ \kappa\delta_0\left[\xi_1 b_t + \xi_2 s_t + \xi_3 u_t\right],
\end{aligned}$$

which, can be used to obtain (A.1.4) with coefficients

$$\begin{aligned}
\mu_{p,1} &\equiv \left[\kappa\left(\delta_1 + \delta_0\xi_1\right) + (1-\alpha)\beta(1-n)\psi_p^b\right]/(1-\alpha) \\
\mu_{p,2} &\equiv \left[\kappa\left(\delta_2 + \delta_0\xi_2\right) + (1-\alpha)\beta n\psi_p^s\right]/(1-\alpha) \\
\mu_{p,3} &\equiv \left[\kappa\left(\delta_3 + \delta_0\xi_3\right) + 1\right]/(1-\alpha).
\end{aligned}$$

Thus, we can obtain (A.1.14) to (A.1.16) as

$$\begin{aligned}
\psi_v^s(n) =& \frac{(1-\beta)\Gamma_b^s\left[(1+\kappa\sigma)(1-\phi_b) + \beta^2\left(\phi_b n\Gamma_b^s + (1-n)\Gamma_b^b\right) - \beta\left(1 + (1-\phi_b)\left(\phi_b n\Gamma_b^s + (1-n)\Gamma_b^b\right) + \kappa\sigma\phi_\pi\right)\right]}{\beta\left[1 - \left(\phi_b n\Gamma_b^s + (1-n)\Gamma_b^b\right)\left(1 + \kappa\sigma + \beta\left(1 - \left(\phi_b n\Gamma_b^s + (1-n)\Gamma_b^b\right)\right)\right) + \kappa\sigma\phi_\pi\right]} \\
=& \left[(1-\beta^2-\phi_b)(1-\beta)\left(n(1+\beta-\phi_b)(1-\beta-\phi_b)^2 + (1-\beta^2-2\phi_b)\left(\phi_b(1-\beta-\phi_b) - \kappa\sigma(\phi_b + (\beta\phi_\pi - 1))\right)\right)\right]/\mathcal{D}
\end{aligned}$$

$$\begin{aligned}
\mathcal{D} \equiv& \beta\left[-n^2(\beta^2-(1-\phi_b)^2)^2 + n(1+\beta-\phi_b)(1-\beta-\phi_b)(1-\beta^2-2\phi_b)(1-\beta-\kappa\sigma-2\phi_b)\right. \\
&\left.+(1-\beta^2-2\phi_b)^2(\phi_b(1-\beta-\phi_b) - \kappa\sigma(\phi_b + (\beta\phi_\pi - 1)))\right]
\end{aligned}$$

$$\begin{aligned}
\psi_p^s(n) =& \frac{(1-\beta)\Gamma_b^s\kappa\left[1 - \phi_b - \beta\left(\phi_b n\Gamma_b^s + (1-n)\Gamma_b^b\right)\right]}{(1-\alpha)\beta\left[1 - \left(\phi_b n\Gamma_b^s + (1-n)\Gamma_b^b\right)\left(1 + \kappa\sigma + \beta\left(1 - \left(\phi_b n\Gamma_b^s + (1-n)\Gamma_b^b\right)\right)\right) + \kappa\sigma\phi_\pi\right]} \\
=& \left[(1-\beta^2-\phi_b)n(1-\beta)\kappa(1+\beta-\phi_b)(1-\beta-\phi_b)\right]/\left[(1-\alpha)\mathcal{D}\right]
\end{aligned}$$

$$\begin{aligned}
\psi_v^b(n) =& \frac{(1-\beta)\Gamma_b^b\left[(1+\kappa\sigma)(1-\phi_b) + \beta^2\left(\phi_b n\Gamma_b^s + (1-n)\Gamma_b^b\right) - \beta\left(1 + (1-\phi_b)\left(\phi_b n\Gamma_b^s + (1-n)\Gamma_b^b\right) + \kappa\sigma\phi_\pi\right)\right]}{\beta\left[1 - \left(\phi_b n\Gamma_b^s + (1-n)\Gamma_b^b\right)\left(1 + \kappa\sigma + \beta\left(1 - \left(\phi_b n\Gamma_b^s + (1-n)\Gamma_b^b\right)\right)\right) + \kappa\sigma\phi_\pi\right]} \\
=& \left[-(1-\phi_b)(1-\beta^2-\phi_b)(1-\beta)\left(n(1+\beta-\phi_b)(1-\beta-\phi_b)^2\right.\right. \\
&\left.\left.+(1-\beta^2-2\phi_b)\left(\phi_b(1-\beta-\phi_b) - \kappa\sigma(\phi_b + (\beta\phi_\pi - 1))\right)\right)\right]/\mathcal{D}
\end{aligned}$$

$$\begin{aligned}
\psi_p^b(n) =& \frac{(1-\beta)\Gamma_b^b\kappa\left[1 - \phi_b - \beta\left(\phi_b n\Gamma_b^s + (1-n)\Gamma_b^b\right)\right]}{(1-\alpha)\beta\left[1 - \left(\phi_b n\Gamma_b^s + (1-n)\Gamma_b^b\right)\left(1 + \kappa\sigma + \beta\left(1 - \left(\phi_b n\Gamma_b^s + (1-n)\Gamma_b^b\right)\right)\right) + \kappa\sigma\phi_\pi\right]} \\
=& \left[-(1-\phi_b)(1-\beta^2-\phi_b)n(1-\beta)\kappa(1+\beta-\phi_b)(1-\beta-\phi_b)\right]/\left[(1-\alpha)\mathcal{D}\right].
\end{aligned}$$

Next, we use (A.1.3) and (8) to compute

$$v_{t+1} = \mu_{v,1} b_{t+1} + \mu_{v,2} s_{t+1} + \mu_{v,3} u_{t+1}$$
$$v_{t+1} = [\mu_{v,1} + \mu_{v,2}\phi_b] b_{t+1} + \mu_{v,2} z_{t+1} + \mu_{v,3} u_{t+1}.$$

More concise, this is

$$\begin{aligned} v_{t+1} =& \Upsilon_{v,b} b_{t+1} + \Upsilon_{v,z} z_{t+1} + \Upsilon_{v,u} u_{t+1}, \qquad \text{where} \qquad \text{(B.5.7)} \\ \Upsilon_{v,b}\left(\psi_v^s(n), \psi_p^s(n), \psi_v^b(n), \psi_p^b(n), n\right) =& \left[\mu_{v,1}\left(\psi_v^b(n), \psi_p^b(n), \delta_1\left(\psi_v^b(n), \psi_p^b(n), n\right), n\right)\right. \\ & \left. +\mu_{v,2}\left(\psi_v^s(n), \psi_p^s(n), \delta_2\left(\psi_v^s(n), \psi_p^s(n), n\right), n\right)\phi_b\right] \equiv \Upsilon_{v,b} \end{aligned}$$

$$\begin{aligned} \Upsilon_{v,z}\left(\psi_v^s(n), \psi_p^s(n), n\right) =& \mu_{v,2}\left(\psi_v^s(n), \psi_p^s(n), \delta_2\left(\psi_v^s(n), \psi_p^s(n), n\right), n\right) \equiv \Upsilon_{v,z} \\ \Upsilon_{v,u} \equiv& \mu_{v,3}. \end{aligned}$$

Moreover, we use (A.1.4) and (8) to compute

$$p^*_{t+1} = \mu_{p,1} b_{t+1} + \mu_{p,2} s_{t+1} + \mu_{p,3} u_{t+1}$$
$$p^*_{t+1} = [\mu_{p,1} + \mu_{p,2}\phi_b] b_{t+1} + \mu_{p,2} z_{t+1} + \mu_{p,3} u_{t+1}.$$

More concise, this is

$$\begin{aligned} p^*_{t+1} =& \Upsilon_{p,b} b_{t+1} + \Upsilon_{p,z} z_{t+1} + \Upsilon_{p,u} u_{t+1}, \qquad \text{where} \qquad \text{(B.5.8)} \\ \Upsilon_{p,b}\left(\psi_v^s(n), \psi_p^s(n), \psi_v^b(n), \psi_p^b(n), n\right) =& \left[\mu_{p,1}\left(\psi_p^b(n), \delta_1\left(\psi_v^b(n), \psi_p^b(n), n\right), n\right)\right. \\ & \left. +\mu_{p,2}\left(\psi_p^s(n), \delta_2\left(\psi_v^s(n), \psi_p^s(n), n\right), n\right)\phi_b\right] \equiv \Upsilon_{p,b} \\ \Upsilon_{p,z}\left(\psi_v^s(n), \psi_p^s(n), n\right) =& \mu_{p,2}\left(\psi_p^s(n), \delta_2\left(\psi_v^s(n), \psi_p^s(n), n\right), n\right) \equiv \Upsilon_{p,z} \\ \Upsilon_{p,u} \equiv& \mu_{p,3}. \end{aligned}$$

Thus, we can combine (B.5.7) and (B.5.8) to obtain

$$\mathbf{z}_{t+1} = \Upsilon_b b_{t+1} + \Upsilon_z z_{t+1} + \Upsilon_u u_{t+1},$$

where $\Upsilon_b \equiv (\Upsilon_{v,b}, \Upsilon_{p,b})'$, $\Upsilon_z \equiv (\Upsilon_{v,z}, \Upsilon_{p,z})'$ and $\Upsilon_u \equiv (\Upsilon_{v,u}, \Upsilon_{p,u})'$. In consequence, we can compute $F(n)$ as outlined in Appendix A.2 above. We can also express $F(n)$ explicitly by plugging in, i.e.,

$$\begin{aligned} F(n) =& (1-\beta)^2\sigma_z^2\left(2-\beta^2-2\phi_b\right)\left(1-\beta^2-2\phi_b\right)\times \\ & \Big[(1-\alpha)^2\left(\beta^2+2\phi_b-1\right)^2\left[\kappa\sigma(\beta\phi_\pi+\phi_b-1)+\phi_b(\beta+\phi_b-1)\right]^2 + n^2(\beta-\phi_b+1)^2(\beta+\phi_b-1)^2 \\ & \left[\alpha^2(\beta+\phi_b-1)^2 - 2\alpha(\beta+\phi_b-1)^2 + (\beta+\phi_b)^2 - 2\beta + \kappa^2 - 2\phi_b + 1\right] \\ & +2(1-\alpha)^2 n(\beta-\phi_b+1)\left(\beta^2+2\phi_b-1\right)(\beta+\phi_b-1)^2\left[\kappa\sigma(\beta\phi_\pi+\phi_b-1)+\phi_b(\beta+\phi_b-1)\right]\Big]/\mathcal{D} \\ \mathcal{D} =& (1-\alpha)^2\beta^2\Big[\left(\beta^2+2\phi_b-1\right)^2\left[\kappa\sigma(\beta\phi_\pi+\phi_b-1)+\phi_b(\beta+\phi_b-1)\right] + n^2\left(\beta^2-(\phi_b-1)^2\right)^2 \\ & +n(\beta-\phi_b+1)(\beta+\phi_b-1)\left(\beta^2+2\phi_b-1\right)(\beta+\kappa\sigma+2\phi_b-1)\Big]^2. \end{aligned}$$

Notice that the denominator $\mathcal{D}$ is always positive. So sign changes can only be expected from the numerator. With the numerator one can show that the same conditions on passive fiscal policy are necessary for $F(1) > 0$ and $F(0) > 0$ and for $F(1) < 0$ and $F(0) < 0$. This proves Proposition 5.

$\square$