

# Endogenously (Non-)Ricardian Beliefs\*

William A. Branch  
University of California, Irvine

Emanuel Gasteiger  
TU Wien

March 21, 2019

## Abstract

This paper develops a theory of endogenously (non-)Ricardian beliefs. That is, whether Ricardian Equivalence holds in an equilibrium depends on endogenous private sector beliefs. The novelty here is a restricted perceptions viewpoint: in complex forecasting environments, agents forecast aggregate variables with (potentially) misspecified 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 models is endogenous. Our formalization considers two predictors: in one rule Ricardian beliefs emerge 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 Ricardian or non-Ricardian equilibria. The theory suggests a novel interpretation of post-war U.S. inflation data as being generated by endogenous belief-driven regime change and a nuanced trade-off for monetary policy rules.

**JEL Classification:** D82; D83; E40; E50

**Keywords:** adaptive learning, misspecification, heterogeneous beliefs, fiscal theory of price level.

---

\*Gasteiger is also affiliated with ISCTE-IUL and gratefully acknowledges financial support from Fundação para a Ciência e a Tecnologia (UID/GES/00315/2019), the Berlin Economics Research Associates (BERA) program and the Fritz Thyssen Foundation. We thank seminar participants at U.C. Irvine, Universität Potsdam, Freie Universität Berlin, Humboldt-Universität zu Berlin, Tinbergen Institute, TU Wien, Universität Wien, the ‘Workshop on Expectations in Dynamic Macroeconomic Models’, the ‘CeNDEF@20 Workshop’ and the ‘Bamberg Behavioral Macroeconomics Workshop’ for many helpful comments.

# 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, taxes, and the other endogenous state variables. The departure point in this paper is to endow agents with a choice between two forecasting models: the first nests Ricardian beliefs – that is, where the private-sector holds beliefs that the path of future taxes will be sufficient to satisfy the government’s intertemporal budget constraint – 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 model. 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, rather than *policy-driven* regime change.

The design of monetary and fiscal policy often hinges on a question that has been of interest to economists for centuries, when does Ricardian equivalence hold or fail? The “fiscal theory of the price-level” literature identifies a set of conditions on policy rules under which Ricardian equivalence fails. For example, Leeper (1991) showed that, in a policy regime where fiscal policy does not guarantee fiscal solvency 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) incorporate uncertainty about the policy regimes to show that private sector (rational) beliefs play an important role in generating high inflation rates during the 1970’s.<sup>1</sup>

Breakthrough papers by Evans et al. (2009) and Eusepi and Preston (2018a) open 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 and/or the structural characteristics of the economy.<sup>2</sup> In these models, individuals have imperfect knowledge about whether the paths for primary surpluses will adjust to satisfy the govern-

---

<sup>1</sup>See Leeper and Leith (2016) for a recent overview of this extensive literature.

<sup>2</sup>See also Evans et al. (2012), Eusepi and Preston (2012), and Woodford (2013).

ment’s intertemporal budget constraint. Eusepi and Preston (2018a) develop their 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 the (non-)Ricardian property of beliefs is determined endogenously. The basic economic environment is New Keynesian where households and firms’ optimal decisions depend on their subjective beliefs about the paths of payoff-relevant aggregate variables. Policy is given by feedback rules for nominal interest rates and lump-sum taxes. The fiscal policy rule is Ricardian and guarantees that taxes are adjusted to satisfy the government’s long-run budget constraint. Monetary policy is described by a Taylor-rule that reflects a commitment to price stability. In a *temporary equilibrium*, without *a priori* imposing private-sector Ricardian beliefs, the aggregate state variables depend, in part, on the existing stock of debt and the contemporaneous primary surplus. A forecasting model linear in these variables, as well as the other state variables, nests the rational expectations equilibrium.

In this environment, we formalize our ideas by taking a step away from rational expectations, instead adopting 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 has the primary surplus as a predictor, does not.

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 be-

liefs about the possible paths 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 more 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, because of the self-referential features of the model, this predictor choice depends endogenously 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 interpretation of post-war U.S. inflation data. In particular, along a 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, estimated using the Extended Kalman Filter on U.S. data for the period 1960-2007:3, indicate that the data are consistent with a model that switches between mostly Ricardian and non-Ricardian agents. The estimates of the latent states, and evidence from the Survey of Professional Forecasters, suggest that the late 1980's-1990's was a period of non-Ricardian equilibria.<sup>3</sup>

This finding indicates that post-war U.S. inflation data, in addition to policy regime changes, may also be determined by endogenously (non-)Ricardian beliefs, which is of particular relevance for policy design. Ricardian policy has non-Ricardian wealth and income effects on aggregate demand due to endogenously (non-)Ricardian beliefs in addition to the disposable income and intertemporal substitution channel captured by existing standard frameworks for policy design. Therefore, our model can provide guidance for monetary and fiscal policy design. We explore this potential by turning to several counterfactual policy exercises, similar to Bianchi (2013). One set of counterfactuals looks at the consequences

---

<sup>3</sup>There is an extensive literature that models, for instance, the Great Inflation as arising from a policy regime change to a non-Ricardian set of policy rules and a fiscal theory of the price-level. In this paper, we focus on the model's one-step ahead predictions over the post-war period. A more complete examination would allow for endogenously (non-)Ricardian beliefs *and* regime-switching policy rules. The relative importance of these two channels, and how they might interact, is an open empirical question beyond the scope of the present study.

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 regime-switching in beliefs, which would have produced greater economic volatility, offsetting the usual stabilizing effect of a more hawkish policy.<sup>4</sup> 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. A fiscal policy that responds more strongly towards debt innovations would have produced counterfactually large output gaps, primary surpluses, and lower inflation throughout the 1980's and 1990's.

The paper proceeds as follows. The subsequent Section 2 presents the model and introduces the concept of endogenously non-Ricardian beliefs. Section 3 focuses on the primary theoretical results, while Section 4 presents the quantitative analysis. We discuss the connection of our paper to the existing literature in Section 5, while section 6 concludes.

## 2 Model

Woodford (2013) shows 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 Ricardian under rational expectations, i.e., active monetary/passive fiscal in the Leeper (1991) sense. We generalize the Woodford (2013) framework to see how, and whether, (non-)Ricardian beliefs arise in equilibrium.

### 2.1 Woodford's (2013) model

The setting is a New Keynesian model with subjective expectations, in particular, a simplified version of Preston (2005), based on Woodford (2003, ch.4).<sup>5</sup> Households and firms have subjective beliefs about payoff-relevant aggregate variables. Given these beliefs, households choose consumption, leisure, and one-period government debt, the only asset available to households, to solve their intertemporal optimization problem.<sup>6</sup> In Woodford's (2013)

---

<sup>4</sup>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.

<sup>5</sup>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.

<sup>6</sup>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

framework, households turn over wage-setting and labor supply decisions to a union and are obligated to supply labor to a firm on the union's terms.<sup>7</sup> Households also receive a lump-sum transfer of their share in firm profits.<sup>8</sup> This is a stylized assumption that renders the household's consumption rule analogous to the one in a model 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.

HOUSEHOLDS. Woodford (2013) derives an individual's consumption function, written recursively as

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

where  $v_t^i$  is a subjective composite variable that comprises all payoff-relevant aggregate variables over which a household formulates subjective beliefs:

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

The variables, written as log-deviations from steady-state,  $b_t^i, Y_t, \pi_t, i_t, \tau_t, \bar{c}_t$  are, respectively, the individual's holdings of real government debt, aggregate output, the inflation rate, the nominal interest rate, lump-sum taxes, and a preference shock. The government uses lump-sum taxes and debt to finance its consumption of an exogenous sequence  $G_t$ . The parameter  $0 < \beta < 1$  is the discount rate,  $\sigma$  is the elasticity of intertemporal substitution, and  $s_b \equiv \bar{b}/\bar{Y}$  is the steady-state debt-to-GDP ratio. The fiscal policy instrument is the real primary surplus  $s_t \equiv \tau_t - G_t$ , where  $G_t$  measures (exogenous) government purchases.

Following 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 or the other structural features of the economy, and so they may 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.

---

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.

<sup>7</sup>The union's negotiator seeks to maximize average expected lifetime utility.

<sup>8</sup>The shares in firms are illiquid, which makes government debt the only storable good. Eusepi and Preston (2018a) show that this assumption is consequential for non-Ricardian beliefs. Though, we abstract from these issues, it is worth bearing in mind that the issue is relevant within our non-Ricardian equilibrium.

Ricardian beliefs arise when the following condition on *beliefs* is satisfied

$$E_t^i \left\{ \sum_{T=t}^{\infty} \beta^{T-t} [s_T - s_b(\beta i_T - \pi_T)] \right\} = b_t. \quad (2)$$

Imposing Ricardian beliefs onto the consumption rule (1) implies that the potentially non-Ricardian effects, represented by current bond holdings (“wealth effect”) and the perceived present-value of the future real returns on bonds (“income effect”), do not impact consumption because (2) directly imposes that the household properly forecasts the path for future surpluses. Ricardian beliefs, therefore, lead to a consumption rule that depends only on the household’s subjective beliefs about future paths for disposable income and real interest rates. Conversely, by not *a priori* imposing Ricardian beliefs, households may perceive their current bond holdings as real wealth and a change in the expected path for future surpluses can have a real effect on consumption. See Appendix D or Woodford (2013) for details.<sup>9</sup>

**FIRMS.** Firms are monopolistically competitive and face a nominal pricing friction based on Calvo (1983). An individual firm  $j$  produces a differentiated good. With probability  $0 < \alpha < 1$  it will adjust its previous price by the long-run target rate of inflation, assumed to be zero and with probability  $1 - \alpha$  a firm receives an idiosyncratic signal to (optimally) reset the price. A firm  $j$  that can optimally reset price  $p_t^*(j)$ , relative to the previous aggregate price level  $p_{t-1}$ , will do so to satisfy the first-order condition, written recursively,

$$p_t^*(j) = (1 - \alpha\beta) (E_t^j p_t^{\text{opt}} - p_{t-1}) + (\alpha\beta) E_t^j p_{t+1}^*(j) + (\alpha\beta)\pi_t$$

where  $E_t^j p_T^{\text{opt}}$  is the perceived optimal price in  $T$ . The aggregate inflation dynamics are

$$\pi_t = (1 - \alpha)p_t^*, \quad \text{where} \quad p_t^* \equiv \int p_t^*(j) dj. \quad (3)$$

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

$$i_t = \phi_\pi \pi_t + \phi_y y_t + w_t, \quad (4)$$

where the monetary policy shock is  $w_t \sim \text{iid}(0, \sigma_w^2)$ .<sup>10</sup>

Fiscal policy is characterized by a Leeper (1991) rule for the real primary surplus:

$$s_t = \phi_b b_t + z_t, \quad (5)$$

---

<sup>9</sup>In all of the analysis below, the fiscal rule is *ex post* Ricardian, i.e., real primary surpluses will satisfy the government’s intertemporal constraint. However, out of equilibrium, non-Ricardian beliefs could be consistent with explosive debt. The consequences of this, and its implications for strategic behavior, is an old issue in the fiscal theory of the price level literature (cf., Bassetto, 2002).

<sup>10</sup> In the quantitative analysis we assume all exogenous shocks are stationary AR(1)’s.

where the surplus shock is  $z_t \sim \text{iid}(0, \sigma_z^2)$ . The government also faces a flow budget constraint

$$b_{t+1} = \beta^{-1}[b_t - s_b \pi_t - s_t] + s_b i_t. \quad (6)$$

The steady-state debt-to-GDP ratio  $s_b$  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.<sup>11</sup>

Throughout, the analysis focuses on the *active* monetary and *passive* fiscal policy regime:

$$\begin{aligned} 1 &< \phi_\pi + \frac{1 - \beta}{\kappa} \phi_y \\ (1 - \beta) &< \phi_b < 1. \end{aligned} \quad (7)$$

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). By abstracting from non-Ricardian policy or policy regime changes, provides us with a stark example of the potentially important role of endogenously time-varying Ricardian beliefs.

## 2.2 Temporary equilibrium with heterogeneous beliefs

The income-expenditure identity is given by

$$Y_t = \int c_t^i di + G_t. \quad (8)$$

Combining (1) and (8) with the bond-market clearing condition  $b_t \equiv \int b_t^i di$ , allows us to express aggregate demand as

$$Y_t = g_t + (1 - \beta)b_t + v_t - \sigma \pi_t, \quad (9)$$

where a composite exogenous disturbance  $g_t \equiv \bar{c}_t + G_t$ , such that  $g_t \sim \text{iid}(0, \sigma_g^2)$ .

To express the aggregate demand equation in explicit dependence of expectations, we aggregate,  $v_t = \int v_t^i di$ , and use (6) and (9) (see Appendix D for details), which yields

$$v_t^i = (1 - \beta)v_t + (1 - \beta)\beta(b_{t+1} - b_t) - \beta\sigma(i_t - \pi_t) + \beta E_t^i v_{t+1}. \quad (10)$$

Averaging over expectations in (10) and plugging into (9) yields the “IS equation” without *a priori* imposing Ricardian beliefs:

$$Y_t = g_t - \sigma i_t + (1 - \beta)b_{t+1} + \hat{E}_t v_{t+1}. \quad (11)$$

---

<sup>11</sup>This formulation arises in a cashless environment that allows us to abstract from the effect of monetary aggregates appearing in the consolidated budget constraint.

The aggregate expectations operator  $\hat{E}$  is defined as  $\hat{E}_t(x) = \int E_t^i(x) di$ , for any variable  $x$ .

Given that heterogeneous beliefs lead to a non-degenerate cross-sectional wealth and consumption distribution, some readers may be surprised that individual household bond holdings do not appear in the aggregate demand equation. However, this is a result of several simplifying assumptions in Woodford (2013). First, assumptions about the labor market and the distribution of firm profits imply that future non-financial income are a proportion of aggregate output, which is beyond the agent's control. This implies that household consumption decisions, in a temporary equilibrium, depend on expectations about variables that are also beyond their control. Second, a temporary equilibrium path, in this setting, consists of local perturbations around a non-stochastic steady-state in which all agents hold identical beliefs. It is in this sense that households' beliefs are not *too* heterogeneous. Finally, in the approximated economy, household debt holdings enter linearly and, as a result, individual bond holdings do not matter for the aggregate output path.<sup>12</sup>

On the firm side, after applying the law of iterated expectations, a firm  $j$  sets

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

and, an aggregate New Keynesian Phillips Curve results after averaging across all firms:

$$\pi_t = (1 - \alpha)\beta \hat{E}_t p_{t+1}^* + \kappa y_t + u_t,$$

where we define 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 shock as  $u_t \equiv \{[(1 - \alpha)(1 - \alpha\beta)]/\alpha\}\mu_t$ .

We can now define a *temporary equilibrium* for this economy.

**Definition 1** *Given a distribution of beliefs  $(E_t^i v_{t+1}, E_t^i p_{t+1}^*)_i$ , a temporary equilibrium is a triple  $(b_{t+1}, \pi_t, y_t)$  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*

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

where

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

---

<sup>12</sup>An extension to a setting where heterogeneous expectations give rise to a non-trivial aggregate role to the cross-sectional wealth distribution is interesting and potentially important. For instance, Giusto (2014) incorporates learning into a Krusell-Smith economy derives unique empirical implications. This is beyond the scope of the present study.

## 2.3 Model misspecification

A key insight from Eusepi and Preston (2018a) is that in environments where people have imperfect knowledge of the data generating process it is not reasonable to impose *a priori* that beliefs by individual  $i$  about future spending and taxes satisfies equation (2). Theoretically, we construct equilibria in which *equilibrium* beliefs may, or may not, satisfy this Ricardian belief condition. This section details the proposed theory of expectation formation.

### 2.3.1 A restricted perceptions approach

Under full-information rational expectations the equilibrium law of motion takes the form

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

where  $\eta_t$  is a vector of composite disturbances and  $A$  is conformable. It follows that in order to formulate rational expectations, the agents adopt linear forecast rules that depend on both the stock of beginning-of-period debt,  $b_t$ , and the primary surplus,  $s_t$ . Our theory contemplates environments where it is prohibitively costly, in computational/cognitive costs, to formulate forecasting models that incorporate both fiscal state variables,  $b_t, s_t$ .

Our proposed theory begins by assuming that the agents formulate expectations by *optimizing* their statistical forecasts given their information and abilities. Our perspective is informed by the econometric learning literature and the “cognitive consistency principle” of Evans and Honkapohja (2001): a consistent theory of expectation formation models economic agents like a good economist who forecasts from a well-specified econometric model. Like many professional forecasters who exist in complex forecasting environments and often face degrees-of-freedom limitations, our agents favor parsimonious models. Therefore, our key assumption is that agents will forecast from one of two parsimonious 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 it leads to a convenient formalization of endogenous (non-)Ricardian beliefs. When some fraction of agents forecast from a model that includes  $s_t$ , but not  $b_t$ , Ricardian equivalence fails in equilibrium. Conversely, when all agents include  $b_t$ , but not  $s_t$ , the (self-confirming) equilibrium features a weak form of Ricardian equivalence.

While restricting the set of regressors in agents’ econometric model is, admittedly, *ad hoc*, our equilibrium concept preserves many cross-equation restrictions that are a salient feature of rational expectations models. We do this as follows. All individuals and firms make a discrete 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 having been determined by the discrete choice between models. Thus, whether beliefs are misspecified or not is an equilibrium property and not imposed by the modeler.

Obviously, by relaxing the model-consistency of rational expectations, there are a number of plausible ways we could model misspecified beliefs. We briefly discuss reasons why our approach is natural for the issue at hand. First, the set of misspecified models nests the Ricardian equivalent outcome as a self-confirming equilibrium. It is, therefore, a convenient formalization of (non-)Ricardian beliefs since Ricardian equivalence is not ruled out *a priori*. Second, Section 4 presents evidence that the median forecast of the Survey of Professional Forecasters is consistent with a model that incorporates a single fiscal variable. Third, there is a long empirical tradition of using a limited number of fiscal indicators in empirical studies. For instance, many empirical studies include budget deficits, but not debt, in fiscal VAR's even though the data favor such specifications (Favero and Giavazzi, 2012). Kliesen and Thornton (2012) find that naively extrapolating the previous year's surplus can outperform the CBO's budget projections and Ericsson (2017) provides evidence of bias in government forecasts of the federal debt. Fourth, the two forecast models proposed here have an appealing interpretation in terms of endogenous paradigm shifts. Finally, for the agents to know that these models are misspecified requires them to step out of the context of their models, where forecast errors are orthogonal to their regressors, and to know the form of the model-consistent forecasting equation. However, model consistency in this context requires that agents hold a great deal of knowledge about the structural features of the economy such as beliefs, constraints, and decision rules of the other agents in the economy including the government and whether its surplus rule will adjust to satisfy the solvency constraint.

### 2.3.2 Misspecification equilibrium

Expectations are formed from one of the following forecasting models, or, 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  $\mathbf{Z}'_t = (v_t, p_t^*, b_{t+1})$ ,  $\eta_t$  is a perceived noise, and the coefficient matrix, for  $k = s, b$ ,

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

$\psi^k = (\psi_v^k, \psi_p^k)'$  and  $\Gamma^k$  is the coefficient for  $b_{t+1}$ .<sup>13</sup> In a *restricted perceptions equilibrium* (RPE) the coefficients will satisfy the least-squares orthogonality condition:

$$E x_{t-1}^k (\mathbf{Z}_t - \boldsymbol{\psi}^k x_{t-1}^k) = 0$$

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  $x_t^k$ . It follows that

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

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

We do not impose *a priori* which of the PLM's individuals and firms use to form expectations. Instead, we endow agents 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), extended to stochastic environments by Branch and Evans (2006b). 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.

Let  $n$  denote the fraction of agents who have selected model- $s$ , leaving  $1 - n$  of the population forecasting with model- $b$ .<sup>14</sup> They rank these choices by calculating the relative mean square error (MSE):

$$EU^k = -E [(\mathbf{Z}_t - E_t^k[\mathbf{Z}_t^k])]' \times \mathbf{W} \times E [(\mathbf{Z}_t - E_t^k[\mathbf{Z}_t^k])], \quad k = \{s, b\}, \quad (12)$$

where  $\mathbf{W}$  is a weighting matrix.<sup>15</sup> Consequently, we define relative predictor performance  $F(n) : [0, 1] \rightarrow \mathbb{R}$  as  $F(n) \equiv EU^s - EU^b$ .

<sup>13</sup>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.

<sup>14</sup>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. It would be straightforward to generalize this way, at the cost of an expanded state vector.

<sup>15</sup>The main results do not depend heavily on the weights, so for simplicity we set  $W = I$ .

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  $\omega$  denotes the “intensity of choice”. The MNL map states that 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  $\omega \rightarrow \infty$  warrants special attention. In this case, agents only select the best-performing statistical models. The following proposition, an immediate consequence of the continuity of  $T_\omega : [0, 1] \rightarrow [0, 1]$ , provides a set of conditions for the existence of (non-)Ricardian beliefs and multiple equilibria in the neoclassical case of a large  $\omega$ .<sup>16</sup>

**Proposition 1** *Let  $N_*(\omega) = \{n_* \mid n_* = T_\omega(n_*)\}$  denote the set of misspecification equilibria. As  $\omega \rightarrow \infty$ ,  $N_*$  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  $F(\hat{n}) = 0$ .*

**Remark 1** *Proposition 1 relies only on the continuity of  $F(n)$  and  $T(n)$ . If  $F(n)$  is monotonic then a stronger statement is possible: Proposition 1 then identifies the full set of misspecification equilibria. In the next section, we present a simple case that facilitates analytic results including conditions under which  $F(n)$  is monotonic. When  $F(n)$  is non-monotonic it is theoretically possible for there to exist multiple interior equilibria, though, in all of the numerical cases examined we found at most 3 misspecification equilibria.*

---

<sup>16</sup>All proofs are in Appendix B. The existence of a misspecification equilibrium depends, as well, on the existence of a unique RPE given  $n$ . Since this is a technical point, the Appendix provides a set of necessary and sufficient conditions for a unique RPE.

The first condition in Proposition 1 implies that the  $b$ -model forecasts best when all 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*.<sup>17</sup> 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 equilibria.<sup>18</sup>

The neoclassical limiting case,  $\omega \rightarrow \infty$ , is useful for identifying sufficient conditions for the existence of multiple equilibria. The multinomial approach, i.e., a finite  $\omega$ , has a venerable history in discrete decision making because it provides an elegant way of introducing randomness into discrete decision-making. Young (2004) shows that randomness in forecasting, much like mixed strategies in actions, provides robustness against model uncertainty and flexibility in self-referential economies. The intensity of choice parameter  $\omega$  is inversely related to the idiosyncratic random utility innovation and, thereby, parameterizes model uncertainty. In particular, larger values of  $\omega$  parameterize less model uncertainty with the neoclassical case  $\omega \rightarrow \infty$  representing no uncertainty at all. An extensive literature tests for dynamic predictor selection using empirical MNL models and typically finds finite values for  $\omega$ . In the theoretical analysis that follows we present results for both  $\omega \rightarrow \infty$  and for finite values. In the quantitative model, our baseline specification features a finite  $\omega$  and we compare the empirical properties across a range of large and small intensities of choice.

## 3 Theoretical results

### 3.1 A simple example

We begin with a special case, first exposted 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  $s_b = 0$ .

---

<sup>17</sup>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.

<sup>18</sup>Note that  $\hat{n}$  case is never gonna be stable under learning and therefore the discussion is centred around the other two equilibria.

The case where  $s_b > 0$  requires numerical analysis.

### 3.1.1 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

$$\begin{bmatrix} v_t \\ b_t \end{bmatrix} = \begin{bmatrix} \psi^s \\ \Gamma_b^s \end{bmatrix} s_{t-1} + \eta_t^s \Rightarrow \begin{bmatrix} E_t^1 v_{t+1} \\ E_t^1 b_{t+1} \end{bmatrix} = \begin{bmatrix} \psi^s \\ \Gamma_b^s \end{bmatrix} s_t \quad (13)$$

$$\begin{bmatrix} v_t \\ b_t \end{bmatrix} = \begin{bmatrix} \psi^b \\ \Gamma_b^b \end{bmatrix} b_{t-1} + \eta_t^b \Rightarrow \begin{bmatrix} E_t^2 v_{t+1} \\ E_t^2 b_{t+1} \end{bmatrix} = \begin{bmatrix} \psi^b \\ \Gamma_b^b \end{bmatrix} b_t. \quad (14)$$

In a restricted perceptions equilibrium, the coefficients in (13) and (14) are optimal, i.e., they satisfy the least-squares orthogonality conditions

$$\begin{aligned} E[s_t \eta_{t+1}^s] &= 0 \\ E[b_t \eta_{t+1}^b] &= 0. \end{aligned}$$

In this simple case, the key model equations are the surplus rule (5) and

$$\begin{aligned} b_{t+1} &= \beta^{-1} (b_t - s_t) \\ y_t &= v_t + (1 - \beta)b_t \\ v_t &= (1 - \beta) (b_{t+1} - b_t) + \hat{E}_t v_{t+1}. \end{aligned}$$

Next we introduce *extrinsic* heterogeneity in expectations. Assume that a fraction  $n \in [0, 1]$  of agents use (13) and  $1 - n$  use (14). We now show that, depending on the distribution  $n$ , (non-)Ricardian equilibria can emerge.

**Proposition 2 (Extrinsic Heterogeneity)** *In the special parametric case of the model, if the aggregate expectations operator is given by*

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

*then, for each  $n \in [0, 1]$ , there exists a unique restricted perceptions equilibrium with*

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

*where*

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

As a corollary, when  $n = 0$ , i.e., all agents forecast with the  $b$ -model, then beliefs are Ricardian *along* an equilibrium path, even though their beliefs are misspecified out of equilibrium. Thus, Ricardian equivalence is a self-confirming equilibrium in the sense of Sargent (1999), Cho et al. (2002), and Williams (2018).

**Corollary 1 (Weak Ricardian Equivalence)** *In the special parametric case, if all agents form expectations from the  $b$ -model (14), then there exists a unique restricted perceptions equilibrium with*

$$\begin{aligned} y_t &= -(\beta^{-1} - 1)z_t \\ \psi^b &= -(\beta^{-1} - 1)(1 - \phi_b). \end{aligned} \tag{15}$$

Woodford's (2013) result of the failure of Ricardian equivalence also arises as a special case of Proposition 2 when  $n = 1$ . Like the more general case of heterogeneous expectations, the equilibrium path for  $y_t$  depends directly on the transitory fiscal shock,  $z_t$ , as well as a persistent effect acting through  $b_t$ .

**Corollary 2 (Woodford (2013))** *In the special parametric case, if all agents form expectations from the  $s$ -model (13), then there exists a unique restricted perceptions equilibrium with*

$$\begin{aligned} y_t &= \left[ \frac{(1 - \beta)(1 + \beta - \phi_b)}{\beta(1 + \beta) + \phi_b} \right] b_t - \left[ \frac{\beta^{-1} - \beta}{\beta(1 + \beta) + \phi_b} \right] z_t \\ \psi^s &= -\frac{\beta^{-1}(1 - \beta)(1 - \beta^2 - \phi_b)}{(\beta + \beta^2 + \phi_b)} < \beta^{-1} - 1. \end{aligned}$$

**Remark 2** *Proposition 2 demonstrates the fragility of Ricardian equivalence, especially in a restricted perceptions environment. Even though all agents have misspecified forecasting models, when  $n = 0$  Ricardian equivalence arises as a self-confirming equilibrium. But, for any  $n > 0$  – including  $n \rightarrow 0$  – then neither type of agent holds Ricardian beliefs.*

### 3.1.2 Misspecification equilibrium

Proposition 2 shows that Ricardian equivalence depends fundamentally on the distribution of households across the two forecasting models. It is, therefore, important to pin down the value  $n$  endogenously within a *misspecification equilibrium*. The following result provides necessary and sufficient conditions for the existence of multiple misspecification equilibria in the limiting case  $\omega \rightarrow \infty$ .

**Theorem 1** Consider the special parametric case of the model. Let  $\omega \rightarrow \infty$  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

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

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

**Corollary 3** Let  $\omega \rightarrow \infty$ . The following results hold.

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

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

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

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

With a stricter set of conditions, we can guarantee that  $F(n)$  is monotonically increasing and there will exist multiple *interior*, i.e., non-Ricardian, misspecification equilibria for  $0 < \omega < \infty$ .

**Corollary 4** For finite  $\omega > 0$ , if  $\exists \underline{\phi}(\beta) > \phi_b$  and  $\underline{\phi}(\beta) < \phi_b < \bar{\phi}(\beta)$ , then there exists three misspecification equilibria  $n_l^*, n_h^*, \hat{n}$  where  $0 \leq n_l^* < \hat{n} < n_h^* \leq 1$ . For  $\omega$  sufficiently small, all equilibria are interior, i.e., non-Ricardian.

**Remark 3** 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, the special case of non-Ricardian equilibria is likely to arise in empirically plausible models in the form of multiple equilibria. The sufficient conditions for multiple interior equilibria in Corollary 4 guarantee that  $F(n)$  is monotonic and  $F(0) < 0, F(1) > 0$ . However, with finite  $\omega$  (possibly multiple) non-Ricardian equilibria can still exist even when  $F(1) < 0$  so long as  $F(n)$  is not too negative for some  $0 < n \leq 1$ .

Figure 1 illustrates the results in Theorem 1 and Corollary 3, i.e., the large  $\omega$  case. There are combinations of  $(\beta, \phi_b)$  consistent with multiple or unique equilibria. The large unshaded area in the lower half of the plot corresponds to active fiscal policy, i.e.,  $(1 - \beta) < \phi_b$ . 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  $(\beta, \phi_b)$  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.

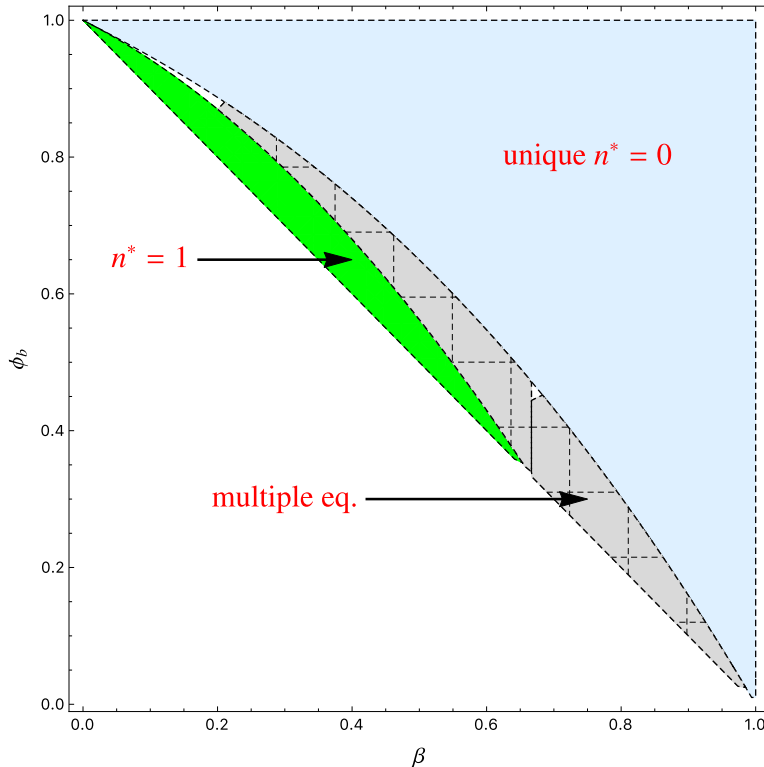


Figure 1: Equilibrium existence

Theorem 1, and its corollaries, is the main theoretical result of the paper: even though fiscal policy is passive, non-Ricardian beliefs can emerge endogenously. For  $\phi_b$  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 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  $\beta = 0.99$ ,  $\phi_b = 0.015$ , 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 a range of  $\omega$ . This figure clearly indicates the three misspecification equilibria. In the quantitative analysis, the interior misspecification equilibrium  $\hat{n}$  is unstable, and so the learning dynamics can feature recurrent switching between the basins of attraction of the Ricardian  $n^* = 0$  and non-Ricardian  $n^* = 1$  equilibrium.

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

$$\begin{aligned} -EU^s &= \mu_{v,1}^2 E[b_t^2] + (\mu_{v,2} - \psi^s(n))^2 E[s_t^2] + 2\mu_{v,1} (\mu_{v,2} - \psi^s(n)) E[b_t s_t] \\ -EU^b &= (\mu_{v,1} - \psi^b(n))^2 E[b_t^2] + \mu_{v,2}^2 E[s_t^2] + 2\mu_{v,2} (\mu_{v,1} - \psi^b(n)) E[b_t s_t]. \end{aligned}$$

Thus, a 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

$$\begin{aligned} F(n) &= [\psi^b(n) (\psi^b(n) - 2\mu_{v,1}) - \psi^s(n) (\psi^s(n) - 2\mu_{v,2}) \phi_b^2] E[b_t^2] \\ &\quad + 2 [\mu_{v,1} \psi^s(n) - \mu_{v,2} \psi^b(n)] E[b_t s_t] - \psi^s(n) [\psi^s(n) - 2\mu_{v,2}] \sigma_z^2. \end{aligned}$$

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  $\phi_b$ , 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, which is strengthened through the self-referential features of the model. For larger values of  $\phi_b$  the agents will always mass, for large  $\omega$ , onto the Ricardian predictor as the surplus and debt models are strongly correlated. Finally, when  $\phi_b$  takes middling values between these two extremes, then either (non-)Ricardian equilibrium can emerge.

### 3.1.3 Building intuition

We now develop intuition about the economic implications by comparing and contrasting the  $n = 0$  (Ricardian) and  $n = 1$  (non-Ricardian) restricted perceptions equilibria.

Revisit the (non-)Ricardian result in Proposition 2. 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

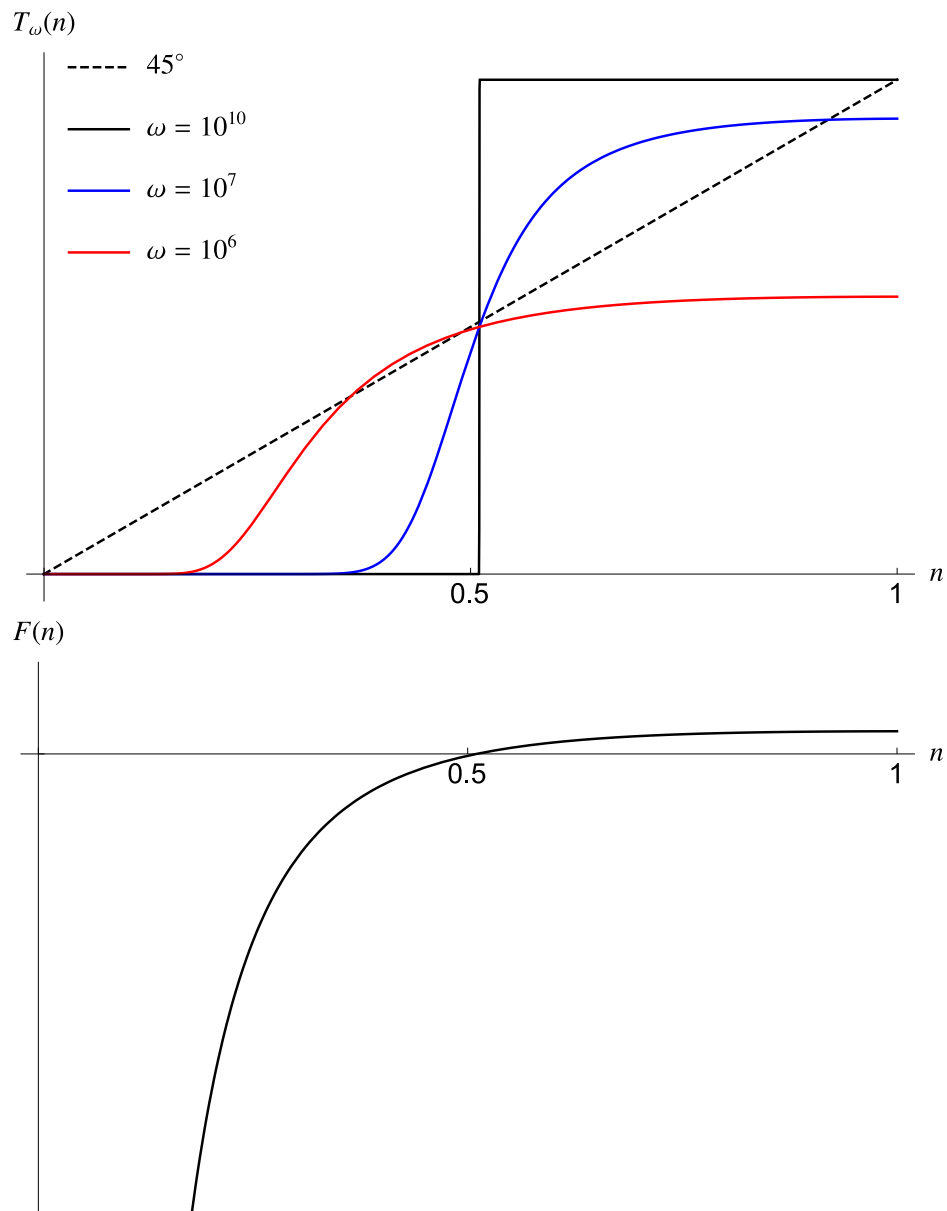
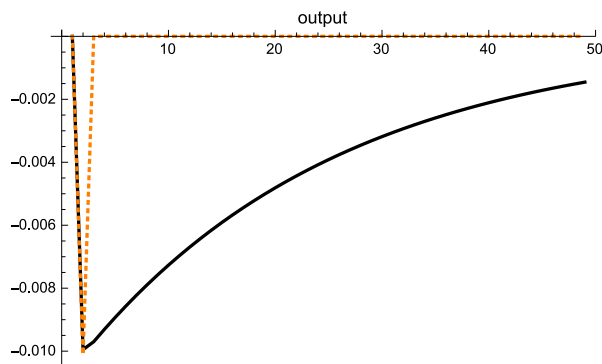
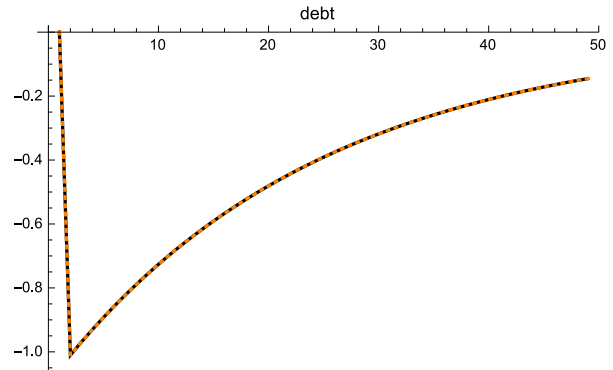


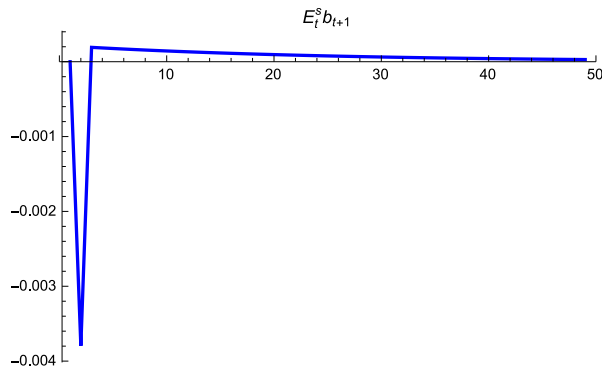
Figure 2: Multiple (non-)Ricardian equilibria in the special case. The top panel plots the T-map for various values of the intensity of choice  $\omega$ .



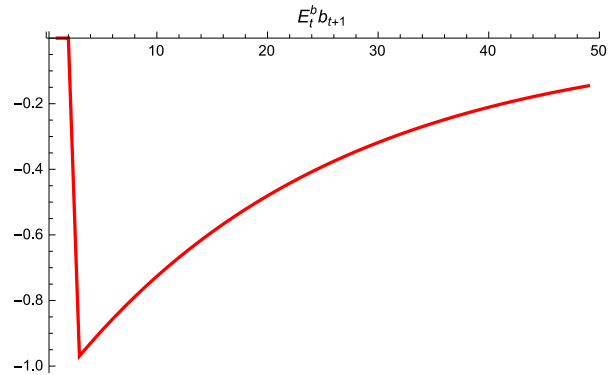
(a) Output response in  $n = 0$  (dashed) and  $n = 1$  RPE.



(b) Debt response in  $n = 0$  (dashed) and  $n = 1$  RPE.



(c) One-step ahead expected debt in  $n = 1$  RPE.



(d) One-step ahead expected debt in  $n = 0$  RPE.

Figure 3: Impulse responses in the special case.

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.

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  $n = 0$  (dashed line), this is the weak Ricardian equivalence result (NW-panel). The  $n = 1$  (solid line) initial impact is slightly smaller, however, it has a strong persistent component. Even though the paths for  $b_{t+1}$  are exogenous (NE-panel), in the Ricardian case the agents track the path of debt correctly (SE-panel), while non-Ricardian agents only correctly forecast debt on impact (SW-panel).

We gain further intuition by comparing consumption across the  $n = 0$  and  $n = 1$  restricted

perceptions equilibria. Let  $n = 0$ , then consumption by belief-type are given by

$$\begin{aligned} c_t^b(0) &= (1 - \beta)s_t - (1 - \beta)\tau_t - (\beta^{-1} - 1)z_t \\ c_t^s(0) &= (1 - \beta) \left[ b_t - \tau_t + \frac{(1 - \beta^2 - \phi_b)}{(1 - \beta^2 - 2\phi_b)}s_t - (\beta^{-1} - 1)z_t \right]. \end{aligned} \tag{16}$$

Equation (16) illustrates weak Ricardian equivalence. For the zero-mass of agents who forecast with the  $s_t$ -model and consume  $c_t^s(0)$ , the effect of surplus shocks have a predictable and persistent impact on consumption. Conversely, when  $n = 1$  Ricardian equivalence fails for both types of households:

$$\begin{aligned} c_t^b(1) &= \frac{(1 - \beta)}{(\beta + \beta^2 + \phi_b)} \{ [1 + \beta - \beta^2(1 - \phi_b)\phi_b - \phi_b^2(2\phi_b - 3)] b_t + (1 + \beta) s_t + (1 - \beta^{-1}) z_t \} \\ c_t^s(1) &= \left[ \frac{1 - \beta^2 - (1 - \beta)\phi_b}{\beta + \beta^2 + \phi_b} \right] b_t - (1 - \beta)\tau_t + (1 - \beta)s_t + \left[ \frac{-\beta^{-1} + \beta}{\beta + \beta^2 + \phi_b} \right] z_t. \end{aligned}$$

### 3.1.4 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?

We can address this question by studying the transitional learning dynamics of agents who adopt a forecast model that depends on both fiscal variables, but who use discounted least-squares to estimate the coefficients of their forecasting model. To conserve space we present the details in Appendix C. From this analysis, we see that the expected learning dynamics feature a path to the rational expectations equilibrium that crosses through the  $n = 1$  RPE, demonstrating that non-Ricardian beliefs can emerge from an escape dynamic of the type studied by Williams (2018). We conclude that even if agents did not face any computational/cognitive constraints, the RPE is a relevant concept as we can expect recurrent escapes near a non-Ricardian equilibrium even when all agents in the economy form forecasts from an equation that nests the correctly specified model. Moreover, for learning rules with sufficiently long memory the economy will persist near the RPE for long stretches of time. These dynamics are reminiscent of Cho and Kasa (2017).

## 3.2 Further results

The results in the special case are useful for clear intuition and analytic tractability. 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, we generalize the results to all small  $\sigma$  parameterizations when  $s_b = 0$ .

The key equations are now the surplus rule (5) and

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

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  $\sigma$ .

**Proposition 3** *For  $\sigma$  sufficiently small, there exists a  $\tilde{\phi}(\beta)$  such that multiple misspecification 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, though, that endogenously non-Ricardian beliefs do not depend directly on the monetary policy coefficient ( $\phi_\pi$ ). However, we show that  $\phi_\pi$  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.

### 3.3 Multiple equilibria in the full model

When  $s_b > 0$  then there exist parameterizations that cover all of the equilibrium possibilities in Proposition 1. In the quantitative analysis we explore the empirical implications of multiple equilibria with different degrees of non-Ricardian agents, i.e., cases where *both*  $n^* = n_i^*$  and  $n^* = n_h^*$ , as in Corollary 4. As seen in Figure 1, a calibrated model with  $\beta = .99$ , and large  $\omega$  will feature multiple equilibria  $n^* = 0$  and  $n^* = 1$  right at the border  $\phi_b > 1 - \beta$ . Thus, in the empirically plausible case of a finite intensity of choice  $\omega$ , non-Ricardian equilibria will emerge in the form of heterogeneous beliefs. For instance, Figure 4 plots the T-map for the numerical parameterization in Table 1 from the quantitative analysis in the subsequent section.

Figure 4 illustrates the possibility of 3 equilibrium values for  $n^*$ .<sup>19</sup> The two equilibria

---

<sup>19</sup>The calibrated model yields the case in Theorem 1 where  $F(0) < 0, F(1) < 0$ , and so Ricardian equivalence holds only in the limiting case  $\omega \rightarrow \infty$ . Because  $F(n) < 0$  for all  $n$  the equilibria feature  $n^* < 0.5$ . However, in the quantitative analysis we do not fix beliefs at their restricted perceptions equilibrium values and so a learning path will hover around, but depart from, the misspecification equilibrium values.

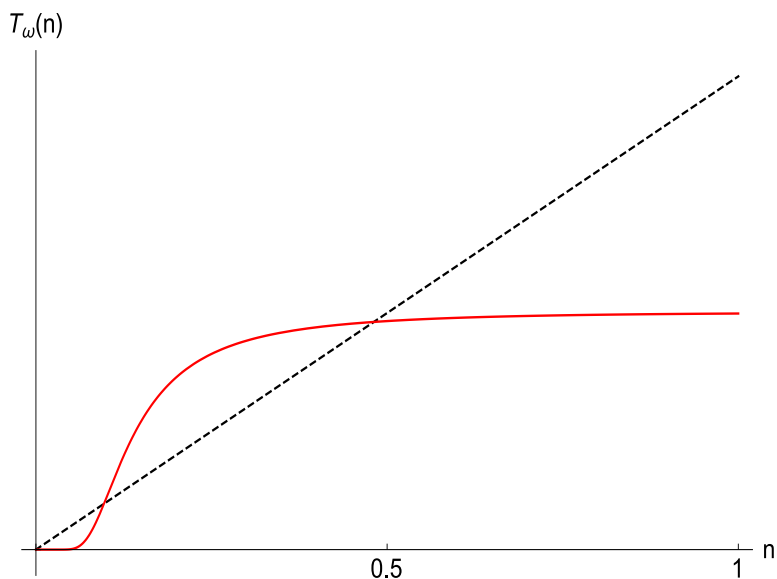


Figure 4: Illustration of multiple equilibria with calibration as in Table 1 and  $\omega \rightarrow \infty$ .

$0 = n_l^* < n_h^* < 1$  feature a slope of  $T_\omega$  at these points less than one, therefore, these two equilibria are the ones that will be stable under learning. We can anticipate real-time learning dynamics that endogenously switch between these regimes.

## 4 Quantitative results

Having proposed a theory of endogenously (non-)Ricardian beliefs, we now turn to a quantitative analysis to see whether U.S. post-war 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 describes well U.S. data on the output gap, inflation, and the primary real-surplus. Our estimates of the latent state dynamics suggest a sizable fraction of individuals and firms holding non-Ricardian beliefs, with the fraction increasing over time in response to low frequency drifts in the primary surplus. A counterfactual analysis explores the implications of alternative monetary policies.

### 4.1 Theory

This section generalizes the temporary equilibrium model to include a richer set of serially correlated disturbances and to follow Eusepi and Preston (2018a) in replacing the fixed RPE parameters with a real-time learning process.

The set of 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

$$E_t^k x_{j,t+1} = (\psi_{j,t-1}^k)' X_{k,t-1},$$

where, for  $j = v, p$  and  $k = s, b$ ,  $x_{j,t} \in \{v_t, p_t^*\}$ ,  $X_{s,t-1}' = (s_{t-1}, g_{t-1}, u_{t-1}, w_{t-1}, z_{t-1})$ , and  $X_{b,t-1}' = (b_{t-1}, g_{t-1}, u_{t-1}, w_{t-1}, z_{t-1})$ . The coefficients  $\psi_{j,t}^k$  are updated with constant gain least-squares:

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

where  $R_{k,t}$  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  $t - 1$ .

Similarly, we assume a recursive estimator for the distribution of agents across forecasting models:

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

where

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

Note that we allow for the possibility that gain parameters  $\gamma_1 \neq \gamma_2$ . Forecasters that are relatively more uncertain about the forecasting accuracies of the two models than they are about their model coefficient estimates would set  $\gamma_2 > \gamma_1$ .<sup>20</sup> And, the MNL law of motion delivers the real-time distribution of endogenously (non-)Ricardian beliefs:

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

---

<sup>20</sup>See Branch and Evans (2006a) for discussion and evidence from the Survey of Professional Forecasters.

## 4.2 Methodology

The Extended Kalman Filter (EKF) generates the data-implied one-step ahead predicted paths for the endogenous latent 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:

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

where the state vector is

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

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

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

and the parameter vector is

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

The measurement and state disturbances are  $\eta_t, \nu_t$  respectively. Our sample for the observed variables is 1955.1-2007.3.<sup>21</sup> We measure  $y_t$  as the log difference between output and the CBO's measure of potential output. We measure  $\pi_t$  from the PCE index. We compute  $b_t$  and  $s_t$  as the debt-GDP ratio and primary surplus-GDP ratio, respectively. To remain consistent with the model, all variables are measured as deviations from mean.

We estimate the (one step ahead) predicted state path  $E(X_{t+1}|Y_t, \Theta)$ . 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 covariances 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  $\Theta$  by using the Extended Kalman Filter to approximate the likelihood function. For example, Eusepi and Preston (2018a) use Bayesian techniques to estimate the posterior distribution of a richer model where adaptive learning generates temporary,

---

<sup>21</sup>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.

endogenous departures from Ricardian equivalence. Although the economic environment in this paper is more parsimonious, we follow Eusepi-Preston and parameterize the model according to their mean and 95% probability bounds on the posterior distribution: see Table 1. We leave a rigorous empirical test of the theory to future research and focus here on the quantitative implications of the theory. We view the analysis that follows as useful for the following reasons. In a more detailed model, Eusepi and Preston (2018a) find that learning in their benchmark estimated model does not produce sizeable non-Ricardian effects. Thus, we view calibrating parameters to their estimates as a conservative choice. As we will see, this benchmark parameterization yields state dynamics consistent with U.S. data. In the end, the approach pursued here facilitates a deeper understanding of the empirical implications of the model and leads us towards interesting counterfactual experiments. A

Table 1: Quantitative parameterization

Parameter	Posterior Mean
$\beta$	0.99
$\sigma$	1/7.7147
$s_b$	0.3
$\kappa$	0.003
$\alpha$	0.738
$\phi_\pi$	1.623
$\phi_y$	0.094
$\phi_b$	0.047
$\rho_g$	0.931
$\rho_u$	0.870
$\rho_w$	0.857
$\rho_z$	0.073
$\sigma_g$	0.526
$\sigma_u$	0.186
$\sigma_w$	0.197
$\sigma_z$	2.088
$\omega$	5.04
$\gamma_1$	0.005
$\gamma_2$	0.039

few comments are in order. We initially set the ‘intensity of choice’ parameter  $\omega = 5.04$ , a value that was estimated by Cornea-Madeira et al. (2017) from a benchmark New Keynesian model with heterogeneous expectations. We also examine robustness of our benchmark results to smaller and larger values of  $\omega$ . In the counterfactual experiments, we also consider

the case where  $\omega$  is large, in line with the analytic results presented earlier. The value of  $\gamma_2$  is in line with Eusepi and Preston (2018a), however  $\gamma_1$  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  $\psi$ ’s bounded in an appropriate neighborhood, and then considered values where  $\gamma_1 = \gamma_2$ . Larger values of  $\gamma_1$  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  $\phi_b$  and  $\rho_z$  also work against the theory of endogenously non-Ricardian beliefs: larger values of both  $\phi_b$  and  $\rho_z$  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  $\phi_\pi, \phi_y$  impact the results on (non-)Ricardian beliefs and macroeconomic outcomes.

### 4.3 Benchmark results

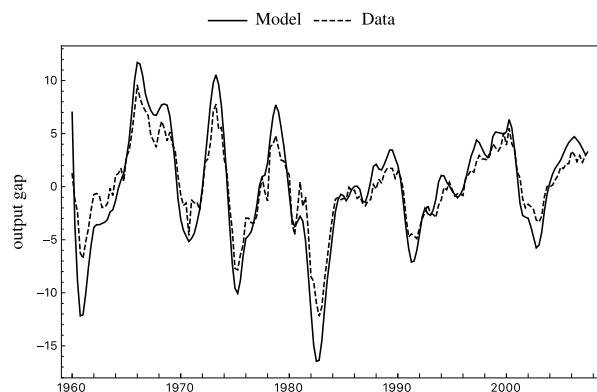
Figure 5 plots both the model-predicted state dynamics and actual U.S. data.<sup>22</sup> The solid lines are the model-predicted paths for the output gap, inflation, the primary government surplus, and the fraction of non-Ricardian agents. The fraction of non-Ricardian agents, in the benchmark parameterization, evolves over time and is below 0.50 for most of the sample, though never below 0.15. The model is closest to the Ricardian equilibrium during the 1960’s. In the latter half of the 1960’s, the percentage of non-Ricardian agents increases from about 15% to 30%. Subsequently, there is a steadily increasing fraction of individuals with non-Ricardian beliefs to over 50% in the second half of the 1990’s. The early 2000’s then feature another increase. The largest movements in  $n_t$  are precipitated by periods of a low frequency drift in the primary surplus. The top two panels show that the fluctuations in non-Ricardian beliefs produce predicted time-paths for the output gap and inflation that is in line with actual U.S. data over the period (dashed lines).<sup>23</sup>

Of course, the analytic results show that the number and nature of misspecification

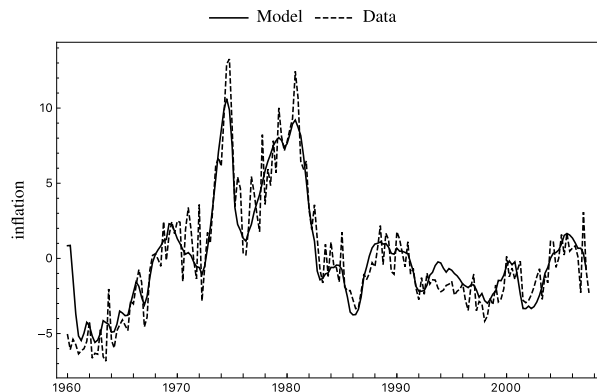
---

<sup>22</sup>We initialize the model over the period 1955-1960 by assuming that in period 1955.1 the economy is initially at the misspecification equilibrium  $n \approx 0.5$ . The latent states are simulated over the next 20 quarters. Using those initial values, we run the Extended Kalman Filter to estimate the predicted path.

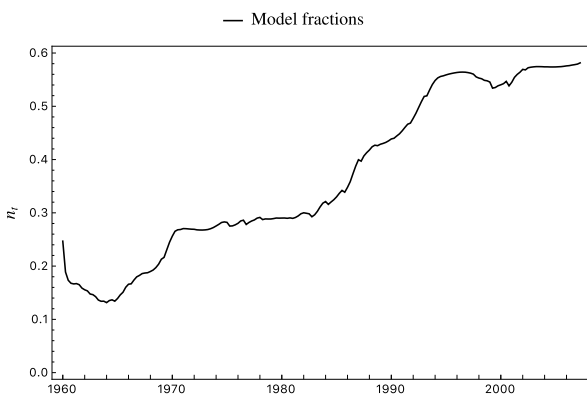
<sup>23</sup>The EKF produces the optimal one step ahead predictions of the state variables. An objective measure of fit is, of course, elusive. The visual fit of the data is partly the result of the predicted paths for the exogenous shocks. However, it should be noted that the fiscal shock process is very weakly autocorrelated. Therefore, the model does a plausible job with the surplus state variable, which 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. These figures also were computed after shutting down the observation noise and plotting the predicted paths as (non-linear) functions of the latent states.



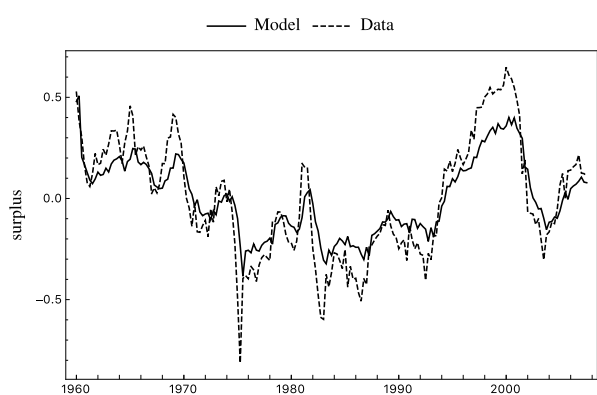
(a) Output gap.



(b) Inflation.



(c) Fraction of non-Ricardian.

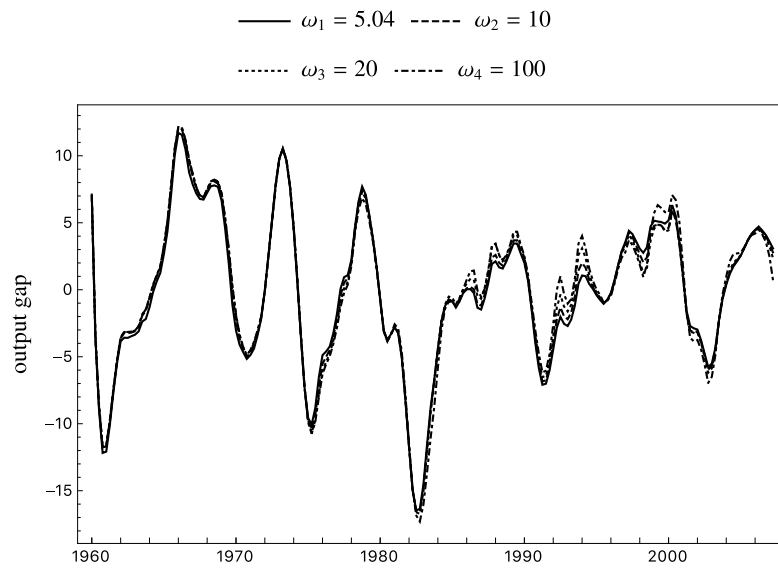


(d) Surplus.

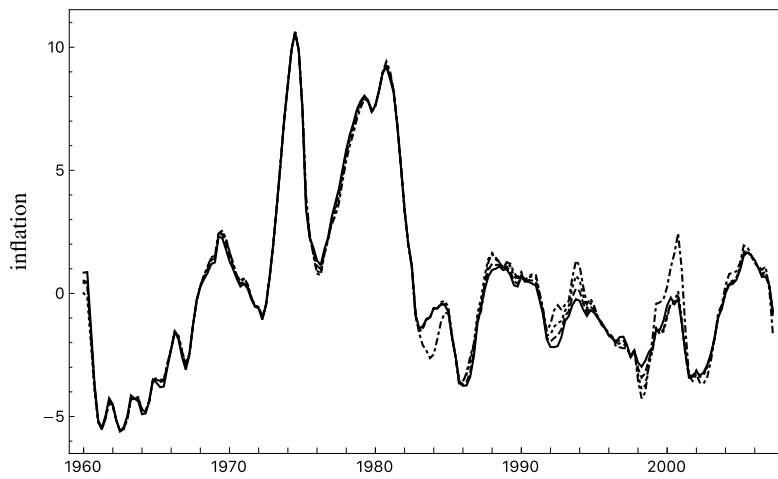
Figure 5: Predicted State Dynamics. Solid lines are model predicted state variables. Dashed lines are corresponding U.S. data.

equilibria depend in a critical way on the intensity of choice parameter  $\omega$ . Analogously, Figure 6 examines how the predicted state dynamics depend on different values for  $\omega$ . As  $\omega$  increases from the benchmark 5.04 (solid line) to 100, the volatility in predicted fractions of non-Ricardian agents increases (bottom panel), with a corresponding increase in economic volatility (top) and inflation (middle). The  $\omega = 100$  case predicts, at times, that nearly all agents use the surplus-model which results in counter-factually high rates of inflation.

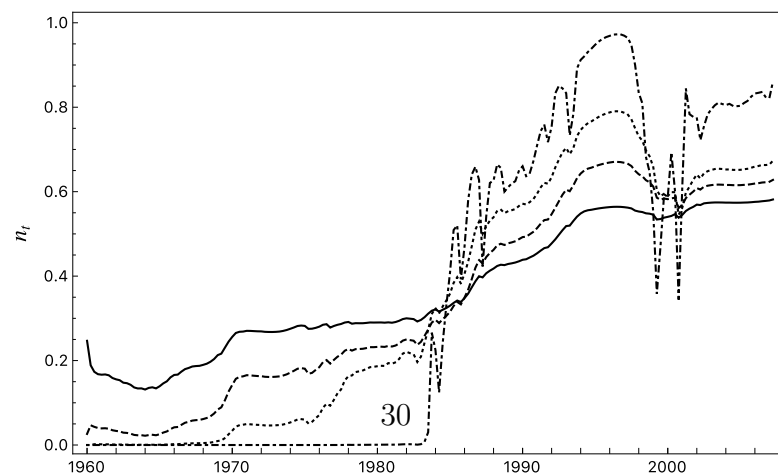
The results presented may be surprising in light of existing results that predict fiscally driven inflation in the 1970's and Ricardian equilibria following the beginning of the Great Moderation period. Our model predicts a non-trivial fraction of non-Ricardian beliefs



(a) Output gap.



(b) Inflation.



(c) Fraction of non-Ricardian.

Figure 6: Predicted state dynamics for different intensities of choice.

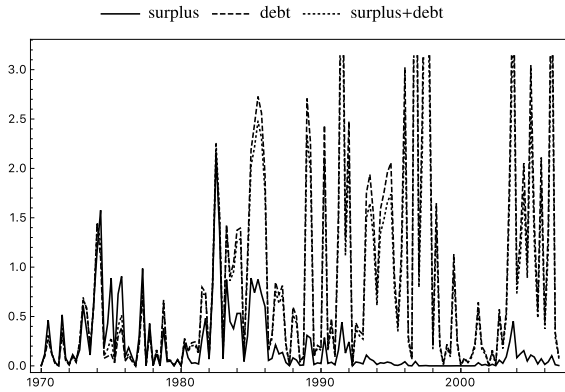
throughout the sample with the fraction adopting the surplus model increasing over time. In the context of the model, this is not entirely surprising as the largest movements in the estimated  $n_t$  correspond also with large drifts in the primary budget surplus. The self-referential dynamics of the economy reinforce the drift in surplus creating a feedback loop that leads a greater proportion of individuals to adopt the surplus model in light of its greater predictive power (measured in a geometrically weighted average mean-square forecast error).

To better assess the plausibility of these results, we present (informal) evidence from the Survey of Professional Forecasters (SPF). Although beyond the scope of this paper, a complete empirical analysis would use the empirical framework in Branch (2004) to analyze the probability that an individual-level survey forecast was made by a simple model with a restricted set of fiscal variables. As a first pass, we compute the statistical scores of the median SPF forecast across three different possible sets of forecasting model regressors: one that includes the primary surplus only, one that includes the debt, and one that includes both. Specifically, we compute moving averages of the statistical score  $Ex_{j,t-1}(\pi_t - \pi_{t-1,t}^e)$  where  $x_{j,t} \in \{s_t, b_t\}$ ,  $\pi_t$  is the PCE inflation rate, and  $\pi_{t-1,t}^e$  is the one-step ahead median SPF survey forecast. Notice that within a restricted perceptions equilibrium, the (time-)average score should be zero. Thus, if the surplus model leads to a lower, and near zero, score then this provides indirect evidence in favor of a restricted perceptions equilibrium with non-Ricardian beliefs. The results in Figure 7 shows that, beginning in the late 1980's, the median SPF is consistent with a greater share of forecasters using the primary surplus as the fiscal variable. In fact, in the late 1990's that score vector is near zero, as predicted by a non-Ricardian restricted perceptions equilibrium.

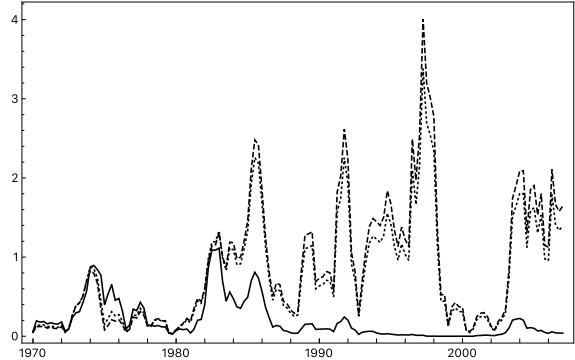
## 4.4 Counterfactual results

The model fits the data well in Figure 5. 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 focus on counterfactuals related to the monetary and fiscal policy rules.

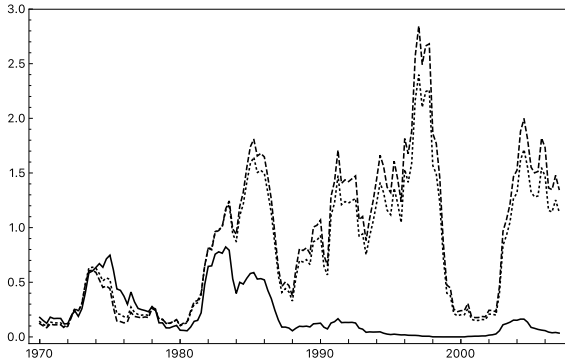
A change in the policy rule will have counterfactual implications for inflation, the output gap, and the primary surplus that operate, in part, through the endogenously (non-)Ricardian beliefs. We conduct three experiments: (1.) a more hawkish monetary policy that sets  $\alpha_\pi = 2.5$ ; (2.) a dovish monetary policy that sets  $\alpha_\pi = 1.005$ , near the determinacy boundary; and, (3.) a more passive fiscal policy rule, with  $\phi_b = 0.10$ , that adjusts the primary surplus more strongly in response to government debt. We can summarize



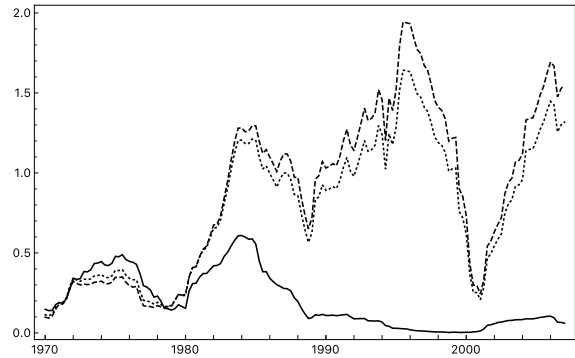
(a) Raw scores.



(b) 4-Qtr. Moving-average.



(c) 8-Qtr. Moving-average.



(d) 5-Yr. Moving-average.

Figure 7: Measured scores for different sets of predictors in the Survey of Professional Forecasters. Each panel computes the scores with different moving average lengths. A score close to zero is consistent with a restricted perceptions equilibrium.

our findings as follows First, there are subtle trade-offs in the design of monetary policy rules. Our counterfactuals demonstrate that a more active interest rate rule will produce a less volatile real debt process which, in turn, will have relatively lower predictive power with more agents adopting the surplus model, which produces a counterfactual with a more volatile fiscally-driven inflation. Conversely, a more dovish policy will make the real debt model relatively more attractive to agents but produce more volatility because the central bank is less aggressive in achieving price stability. We describe this nuanced trade-off as the “goldilocks” feature of monetary policy design: more activist monetary policy can produce more fiscally-driven inflation and less activist policy is more Ricardian but at the expense of higher volatility. Second, our counterfactual analysis shows that a more aggressive fiscal

policy stance will lead to a greater fraction of non-Ricardian agents and a more prominent role for fiscal factors in the inflation process. In these counterfactuals, a higher fraction of non-Ricardian agents and a higher  $\phi_b$  combine to increase the surplus during the 1970's and 1990's and lead to counterfactually low rates of inflation.

#### 4.4.1 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  $\phi_\pi$  for large  $\omega$ . However, even with large  $\omega$  the monetary policy rule coefficients impact the relative sizes of the basins of attraction.<sup>24</sup> To explore the implication, we present results from two 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. Panel (c.) demonstrates that the effect of such a policy would have led to a higher fraction of non-Ricardian agents, on average, throughout the sample period. During the 1960's and 1970's, panel (b.) shows that the counterfactual effect would have been a counterfactually lower, and less volatile, rate of inflation. However, during the 1990's the counterfactual predicts a large run up in the surplus-GDP ratio (see panel (d.)), coinciding with a large fraction of non-Ricardian agents who forecast future fiscal policy primarily using lagged surplus as a predictor, leads to a large 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. In a sense, the endogenous impact of the new policy rule on the extent of non-Ricardian beliefs generates less price stability and strengthens the fiscal impact on inflation.

Now consider the counterfactual with  $\phi_\pi = 1.005$ , which is right 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 panel (c.). In fact, for large  $\omega$  the more dovish monetary policy would lead to  $n = 0$  with all individuals and firms holding Ricardian beliefs. While this dovish policy leads to economic outcomes closer to the Ricardian equilibrium, the policy achieves more economic volatility. Since policymakers are less committed to price stability the result is

---

<sup>24</sup>This can be established formally as the  $\hat{n} = T_\omega(\hat{n})$  equilibrium, in the multiple misspecification equilibrium case, shifts with policy coefficients.

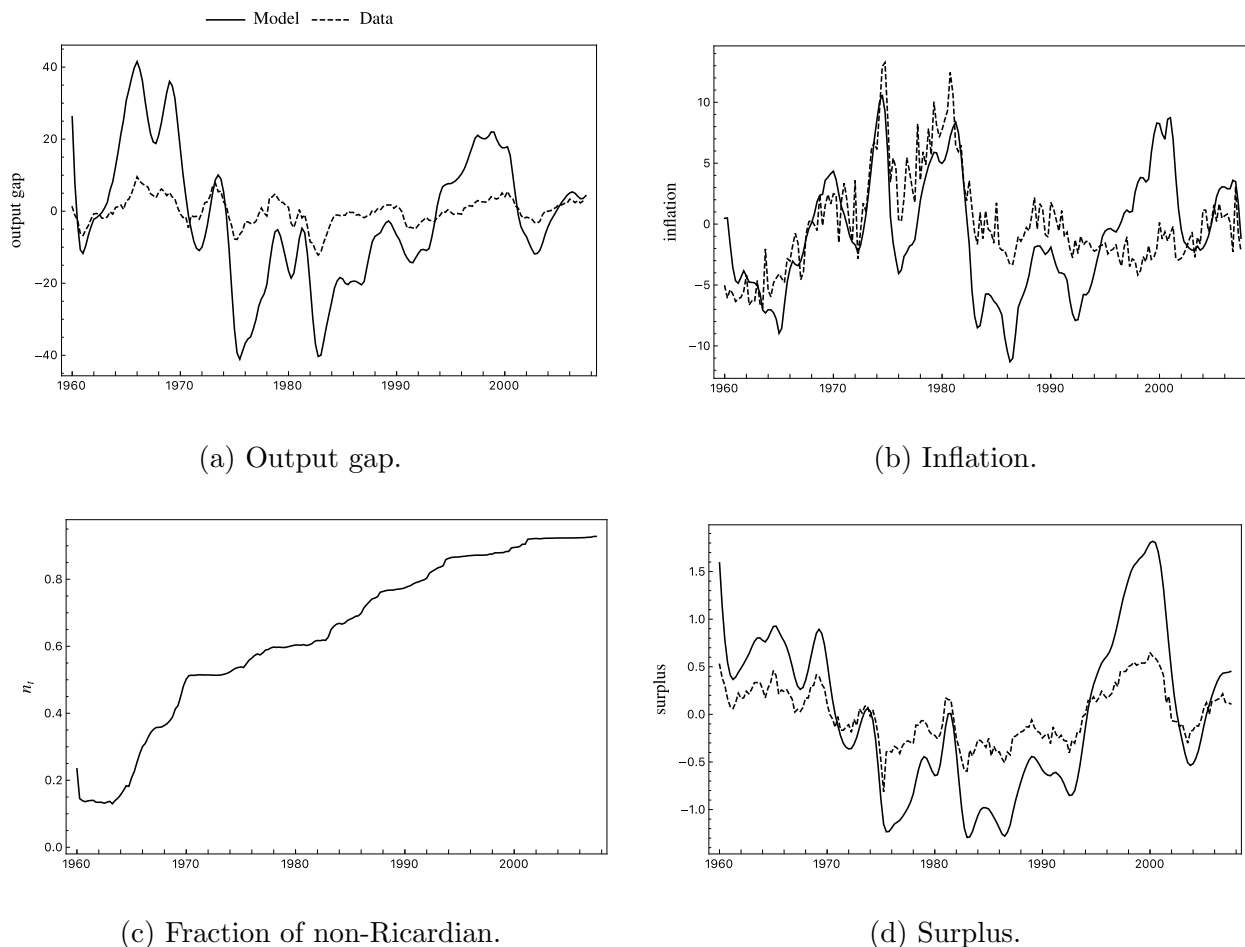


Figure 8: Counterfactual State Dynamics when  $\phi_\pi = 2.5$ .

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 then the economy could have coordinated on a Ricardian regime for inflation. But, a monetary policy rule that is less active against inflation would have led to greater economic volatility.<sup>25</sup> If, instead, policymakers had pursued a more hawkish policy rule, then inflation would have been non-Ricardian more 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-off faced by

<sup>25</sup>This is a standard result in learning models: see result 5a in Eusepi and Preston (2018b).



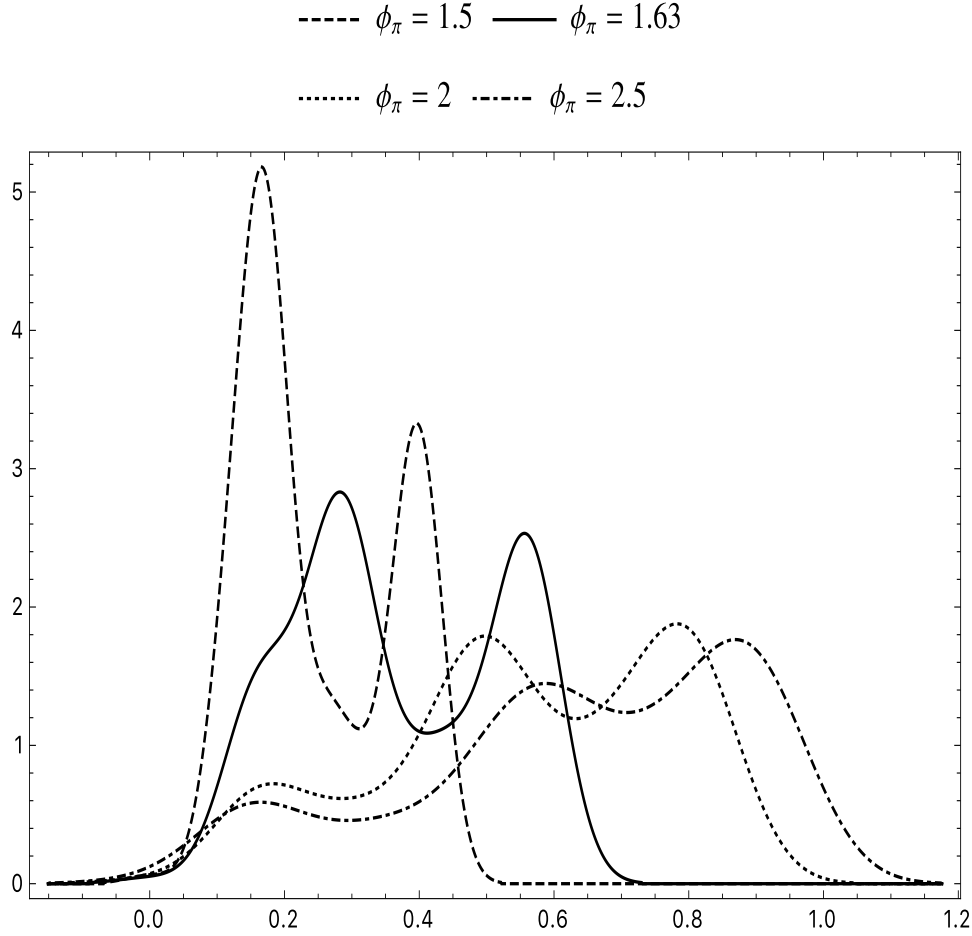


Figure 10: Counterfactual empirical distribution of Ricardian beliefs when  $\phi_\pi = 1.5, 1.63, 2.0, 2.5$

distribution towards the right, i.e. more agents hold non-Ricardian beliefs. Notice as well, the empirical distributions are bimodal and the spread between the peaks gets increases with  $\phi_\pi$ . This empirical feature captures the regime-switching between basins of attraction driven by the learning dynamics.

#### 4.4.2 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  $\phi_b$ , for moderate values of  $\phi_b$ . We now study the economic implications from a counterfactual

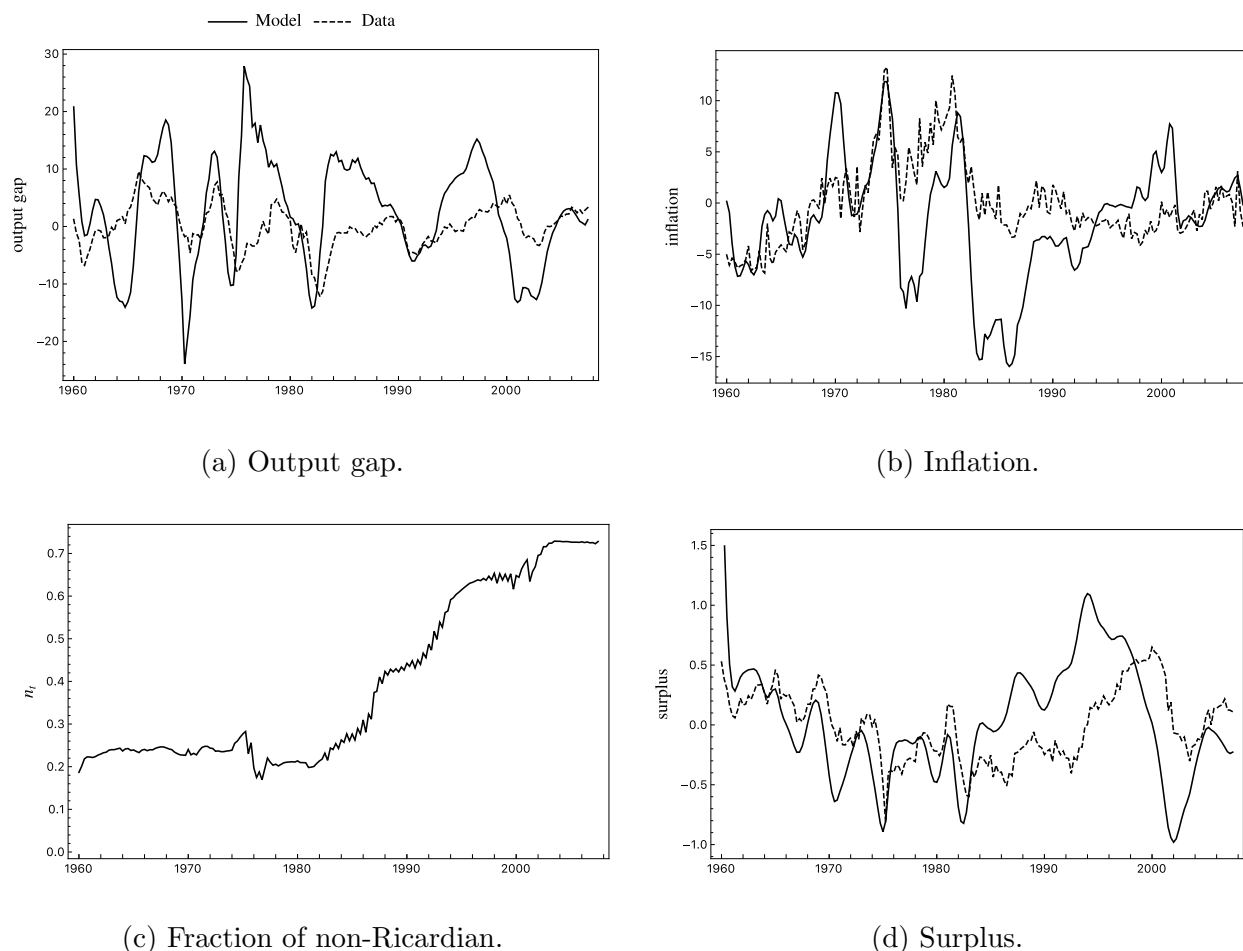


Figure 11: Counterfactual State Dynamics when  $\phi_b = 0.10$ .

analysis of larger values for  $\phi_b$  that lead to a greater extent of non-Ricardian beliefs.

Figure 11 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. By the end of the sample, the counterfactual predicts that a fiscal rule with  $\phi_b = 0.10$  would have converged near the non-Ricardian equilibrium, i.e.  $n^* = 1$ . To understand the counterfactual economic implications focus on the period during the late 1980's and throughout the 1990's. The  $\phi_b = 0.10$  fiscal rule would have produced a counterfactually large primary surplus. This strengthens the incentives of agents to forecast with the surplus model and, consequently, produced temporary equilibrium dynamics that are more non-Ricardian. This in turn produces counterfactually large output gaps in line with lower expected paths of primary surpluses. The smaller stock of debt would have produced much lower rates of inflation.

## 5 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 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 (cf., Evans and Honkapohja, 2007; 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 under adaptive learning.

The theory of 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), Adam (2005), 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. In short, this paper takes the theory of forecast misspecification in Branch and Evans (2006b) into the Eusepi and Preston (2018a) environment with fiscal and monetary policy interaction and generalizing the restricted perceptions beliefs in Woodford (2013). Our approach is also closely related to Adam (2005), who, in a business cycle model, imposes onto agents a choice between two forecasting models. One of these models is consistent with rational expectations, but only within a self-confirming equilibrium. Adam (2005) shows that misspecified models can be sustained in equilibrium and this has implications for the time-series properties of inflation and output. We also contribute to this literature by providing some support for our theory of expectation formation using survey data, following an extensive literature recently summarized in Coibion and Gorodnichenko (2018).

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 of 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.

Our paper is also related to a long-standing 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 research explain post-war U.S. inflation 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, which has two implications. First, when agents assign a positive probability to changes from the Ricardian policy regime to the non-Ricardian policy regime, then the beliefs imply failure of Ricardian equivalence and inflation is also a fiscal phenomenon. Second, 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. Lastly, our results do not suggest that policy regime change is an unimportant part of the inflation story. In fact more subtle changes, within the Ricardian policy regime, can generate belief-driven regime change.

Finally, our 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. In addition, their paper illustrates how the maturity structure of government debt has important implications for inflation in a non-Ricardian belief economy. 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.

## 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

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

- Adam, K. (2005). Learning To Forecast And Cyclical Behavior Of Output And Inflation. *Macroeconomic Dynamics*, 9(1):1–27.
- Bassetto, M. (2002). A Game-Theoretic View of the Fiscal Theory of the Price Level. *Econometrica*,

70(6):2167–2195.

- Bianchi, F. (2013). Regime Switches, Agents' Beliefs, and Post-World War II U.S. Macroeconomic Dynamics. *Review of Economic Studies*, 80(2):463–490.
- Bianchi, F. and Ilut, C. (2017). Monetary/Fiscal Policy Mix and Agents' Beliefs. *Review of Economic Dynamics*, 26:113–139.
- Branch, W. A. (2004). The Theory of Rationally Heterogeneous Expectations: Evidence from Survey Data on Inflation Expectations. *Economic Journal*, 114(497):592–621.
- 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. (2009). A New Keynesian Model with Heterogeneous Expectations. *Journal of Economic Dynamics and Control*, 33(5):1036–1051.
- 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(1):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.
- Coibion, O. and Gorodnichenko, Y. (2018). The Formation of Expectations, Inflation and the Phillips Curve. *Journal of Economic Literature*, 56(4):1447–1491.
- Cornea-Madeira, A., Hommes, C. H., and Massaro, D. (2017). Behavioral Heterogeneity in U.S. Inflation Dynamics. *Journal of Business and Economic Statistics*, forthcoming.
- Davig, T. and Leeper, E. M. (2006). Fluctuating Macro Policies and the Fiscal Theory. In Acemoglu, D., Rogoff, K. S., and Woodford, M., editors, *NBER Macroeconomics Annual*, volume 21, pages 247–316. MIT Press, Cambridge, MA.
- Ericsson, N. R. (2017). How Biased are U.S. Government Forecasts of the Federal Debt? *International Journal of Forecasting*, 33(2):543–559.
- 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*, 108(9):2551–2589.
- 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. (2009). Anticipated Fiscal Policy and Adaptive Learning. *Journal of Monetary Economics*, 56(7):930–953.

- 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.
- Favero, C. A. and Giavazzi, F. (2012). Measuring Tax Multipliers: the Narrative Method in Fiscal VARs. *American Economic Journal: Economic Policy*, 4(2):69–94.
- Gasteiger, E. (2018). Do Heterogeneous Expectations Constitute a Challenge for Policy Interaction? *Macroeconomic Dynamics*, 22(8):2107–2140.
- Giusto, A. (2014). Adaptive Learning and Distributional Dynamics in an Incomplete Markets Model. *Journal of Economic Dynamics and Control*, 40:317–333.
- Kliesen, K. L. and Thornton, D. L. (2012). How Good Are the Government’s Deficit and Debt Projections and Should We Care? *Federal Reserve Bank of St. Louis Review*, 94(1):21–39.
- 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(1):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. *Review of Economic Studies*, forthcoming.
- 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.
- Young, P. H. (2004). *Strategic Learning and Its Limits*. Oxford University Press, Oxford, UK.

# 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

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

$$b_{t+1} = \xi_1 b_t + \xi_2 s_t + \xi_3 u_t. \quad (\text{A.1.2})$$

Moreover, we can aggregate (10) and combine it with (4), (D.4), (A.1.1) and (A.1.2) to obtain

$$v_t = \mu_{v,1} b_t + \mu_{v,2} s_t + \mu_{v,3} u_t, \quad (\text{A.1.3})$$

and (D.4), (3), (A.1.1) and (A.1.2) imply that

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

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

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

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

Now, plug the PLM and ALM into (A.1.5), i.e.,

$$\begin{aligned} 0 &\stackrel{!}{=} E[s_t\eta_{t+1}] = E[s_t(\mathbf{z}_{t+1} - \psi^s s_t)] \\ \Leftrightarrow \psi^s E[s_t^2] &= E[s_t\mathbf{z}_{t+1}]. \end{aligned} \quad (\text{A.1.6})$$

Equation by equation, we obtain

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

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

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

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

Again, equation by equation, we obtain

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

$$\begin{aligned}
&\Leftrightarrow \psi_p^b E[b_t^2] = E[b_t (\mu_{p,1} b_{t+1} + \mu_{p,2} s_{t+1} + \mu_{p,3} u_{t+1})] \\
&\psi_p^b E[b_t^2] = \mu_{p,1} E[b_t b_{t+1}] + \mu_{p,2} E[b_t s_{t+1}] + \mu_{p,3} E[b_t u_{t+1}] \\
&\Leftrightarrow \psi_p^b = \mu_{p,1} \frac{E[b_t b_{t+1}]}{E[b_t^2]} + \mu_{p,2} \frac{E[b_t s_{t+1}]}{E[b_t^2]} + \mu_{p,3} \frac{E[b_t u_{t+1}]}{E[b_t^2]}. \tag{A.1.10}
\end{aligned}$$

The next step is to compute the moments. For this purpose, it is convenient to combine (A.1.2) and (5) in a VAR(1), i.e.,

$$\begin{bmatrix} 1 & -\xi_2 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} b_{t+1} \\ s_t \end{bmatrix} = \begin{bmatrix} \xi_1 & 0 \\ \phi_b & 0 \end{bmatrix} \begin{bmatrix} b_t \\ s_{t-1} \end{bmatrix} + \begin{bmatrix} \xi_3 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} u_t \\ z_t \end{bmatrix} \tag{A.1.11}$$

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

where  $\mathcal{Y}_t \equiv (b_{t+1}, s_t)'$  and  $\varepsilon_t \equiv (u_t, z_t)'$ .

Define the variance-covariance matrix  $\mathbf{\Omega} \equiv E[\mathcal{Y}_t \mathcal{Y}_t']$  and likewise  $\mathbf{\Sigma} \equiv E[\varepsilon_t \varepsilon_t']$ . Then we can compute

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

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

$$\mathbf{\Omega} = \begin{bmatrix} E[b_{t+1}^2] & E[b_{t+1}s_t] \\ E[s_t b_{t+1}] & E[s_t^2] \end{bmatrix}, \quad \mathbf{A}\mathbf{\Omega} = \begin{bmatrix} E[b_{t+1}b_t] & E[b_{t+1}s_{t-1}] \\ E[s_t b_t] & E[s_t s_{t-1}] \end{bmatrix}. \quad (\text{A.1.13})$$

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 (5) 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

$$\psi_v^s(n) = \mu_{v,1}\Gamma_b^s + \mu_{v,2}\phi_b\Gamma_b^s \quad (\text{A.1.14})$$

$$\psi_p^s(n) = \mu_{p,1}\Gamma_b^s + \mu_{p,2}\phi_b\Gamma_b^s \quad (\text{A.1.15})$$

$$\psi_v^b(n) = \mu_{v,1}\Gamma_b^b + \mu_{v,2}\phi_b\Gamma_b^b \quad (\text{A.1.16})$$

$$\psi_p^b(n) = \mu_{p,1}\Gamma_b^b + \mu_{p,2}\phi_b\Gamma_b^b. \quad (\text{A.1.17})$$

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  $\mathbf{A}$  and  $\mathbf{C}$  in (A.1.12) also depend on these coefficients.

## A.2 Computation of the misspecification equilibrium

Recall the objective (12). We combine (A.1.3) and (A.1.4) to

$$\mathbf{z}_t = \boldsymbol{\mu}_b b_t + \boldsymbol{\mu}_s s_t + \boldsymbol{\mu}_u u_t, \quad (\text{A.2.1})$$

where  $\mathbf{z}_t \equiv (v_t, p_t^*)'$ ,  $\boldsymbol{\mu}_b \equiv (\mu_{v,1}, \mu_{p,1})'$ ,  $\boldsymbol{\mu}_s \equiv (\mu_{v,2}, \mu_{p,2})'$ , and  $\boldsymbol{\mu}_u \equiv (\mu_{v,3}, \mu_{p,3})'$ . Moreover, we have

$$E^s[\mathbf{z}_t^s] = \psi^s(n)s_t, \quad \text{and} \quad (\text{A.2.2})$$

$$E^b[\mathbf{z}_t^b] = \psi^b(n)b_t. \quad (\text{A.2.3})$$

Thus, we can use (A.2.1) and (A.2.2) to compute

$$(\mathbf{z}_t - E^s[\mathbf{z}_t^s]) = \boldsymbol{\mu}_b b_t + \boldsymbol{\mu}_s s_t + \boldsymbol{\mu}_u u_t - \psi^s(n)s_t.$$

Under the assumption  $E[b_t u_t] = E[s_t u_t] = 0$ , it follows that

$$E[(\mathbf{z}_t - E^s[\mathbf{z}_t^s])'(\mathbf{z}_t - E^s[\mathbf{z}_t^s])] = E[[b_t' \boldsymbol{\mu}_b' + s_t' \boldsymbol{\mu}_s' + u_t' \boldsymbol{\mu}_u' - s_t' \psi^s(n)'] [\boldsymbol{\mu}_b b_t + \boldsymbol{\mu}_s s_t + \boldsymbol{\mu}_u u_t - \psi^s(n)s_t]]$$

$$\begin{aligned}
E[(\mathbf{z}_t - E^s[\mathbf{z}_t^s])'(\mathbf{z}_t - E^s[\mathbf{z}_t^s])] &= (\boldsymbol{\mu}'_b \boldsymbol{\mu}_b) E[b_t^2] \\
&\quad + [\boldsymbol{\mu}'_s \boldsymbol{\mu}_s + \psi^s(n)' \psi^s(n) - \boldsymbol{\mu}'_s \psi^s(n) - \psi^s(n)' \boldsymbol{\mu}_s] E[s_t^2] \\
&\quad + (\boldsymbol{\mu}'_u \boldsymbol{\mu}_u) E[u_t^2] + [\boldsymbol{\mu}'_b \boldsymbol{\mu}_s - \boldsymbol{\mu}'_b \psi^s(n) + \boldsymbol{\mu}'_s \boldsymbol{\mu}_b - \psi^s(n)' \boldsymbol{\mu}_b] E[b_t s_t].
\end{aligned}$$

In consequence, we obtain

$$\begin{aligned}
EU^s &= - [(\boldsymbol{\mu}'_b \boldsymbol{\mu}_b) E[b_t^2] + [\boldsymbol{\mu}'_s \boldsymbol{\mu}_s + \psi^s(n)' \psi^s(n) - \boldsymbol{\mu}'_s \psi^s(n) - \psi^s(n)' \boldsymbol{\mu}_s] E[s_t^2] \\
&\quad + (\boldsymbol{\mu}'_u \boldsymbol{\mu}_u) E[u_t^2] + [\boldsymbol{\mu}'_b \boldsymbol{\mu}_s - \boldsymbol{\mu}'_b \psi^s(n) + \boldsymbol{\mu}'_s \boldsymbol{\mu}_b - \psi^s(n)' \boldsymbol{\mu}_b] E[b_t s_t]].
\end{aligned}$$

Likewise, we can use (A.2.1) and (A.2.3) to compute

$$(\mathbf{z}_t - E^b[\mathbf{z}_t^b]) = \boldsymbol{\mu}_b b_t + \boldsymbol{\mu}_s s_t + \boldsymbol{\mu}_u u_t - \psi^b(n) b_t.$$

Therefore it follows that

$$E[(\mathbf{z}_t - E^b[\mathbf{z}_t^b])'(\mathbf{z}_t - E^b[\mathbf{z}_t^b])] = E[[b'_t \boldsymbol{\mu}'_b + s'_t \boldsymbol{\mu}'_s + u'_t \boldsymbol{\mu}'_u - b'_t \psi^b(n)'] [\boldsymbol{\mu}_b b_t + \boldsymbol{\mu}_s s_t + \boldsymbol{\mu}_u u_t - \psi^b(n) b_t]]$$

Again we use the assumption  $E[b_t u_t] = E[s_t u_t] = 0$  to obtain

$$\begin{aligned}
E[(\mathbf{z}_t - E^b[\mathbf{z}_t^b])'(\mathbf{z}_t - E^b[\mathbf{z}_t^b])] &= [\boldsymbol{\mu}'_b \boldsymbol{\mu}_b + \psi^b(n)' \psi^b(n) - \boldsymbol{\mu}'_b \psi^b(n) - \psi^b(n)' \boldsymbol{\mu}_b] E[b_t^2] + (\boldsymbol{\mu}'_s \boldsymbol{\mu}_s) E[s_t^2] \\
&\quad + (\boldsymbol{\mu}'_u \boldsymbol{\mu}_u) E[u_t^2] + [\boldsymbol{\mu}'_b \boldsymbol{\mu}_s + \boldsymbol{\mu}'_s \boldsymbol{\mu}_b - \boldsymbol{\mu}'_s \psi^b(n) - \psi^b(n)' \boldsymbol{\mu}_s] E[b_t s_t].
\end{aligned}$$

In consequence

$$\begin{aligned}
EU^b &= - [[\boldsymbol{\mu}'_b \boldsymbol{\mu}_b + \psi^b(n)' \psi^b(n) - \boldsymbol{\mu}'_b \psi^b(n) - \psi^b(n)' \boldsymbol{\mu}_b] E[b_t^2] + (\boldsymbol{\mu}'_s \boldsymbol{\mu}_s) E[s_t^2] \\
&\quad + (\boldsymbol{\mu}'_u \boldsymbol{\mu}_u) E[u_t^2] + [\boldsymbol{\mu}'_b \boldsymbol{\mu}_s + \boldsymbol{\mu}'_s \boldsymbol{\mu}_b - \boldsymbol{\mu}'_s \psi^b(n) - \psi^b(n)' \boldsymbol{\mu}_s] E[b_t s_t]].
\end{aligned}$$

Finally, one can define  $F(n) : [0, 1] \rightarrow \mathbb{R}$  as  $F(n) \equiv EU^s - EU^b$ , thus

$$\begin{aligned}
F(n) &= [\psi^b(n)' \psi^b(n) - \boldsymbol{\mu}'_b \psi^b(n) - \psi^b(n)' \boldsymbol{\mu}_b] E[b_t^2] \\
&\quad + [\boldsymbol{\mu}'_s \psi^s(n) + \psi^s(n)' \boldsymbol{\mu}_s - \psi^s(n)' \psi^s(n)] E[s_t^2] \\
&\quad + [\psi^s(n)' \boldsymbol{\mu}_b + \boldsymbol{\mu}'_b \psi^s(n) - \psi^b(n)' \boldsymbol{\mu}_s - \boldsymbol{\mu}'_s \psi^b(n)] E[b_t s_t].
\end{aligned}$$

## B Proofs

### B.1 Proof of Proposition 1

The proof to Proposition 1 is straightforward, but relies on the existence of a unique restricted perceptions equilibrium for an open set of  $n$ . The following Lemma provides the necessary and sufficient conditions for a unique RPE to exist.

Before stating the proposition, note first that the temporary equilibrium equations can be written in the form of an expectational difference equation:

$$X_t = A \begin{bmatrix} b_t \\ s_t \end{bmatrix} + B\hat{E}_t X_{t+1} + C\hat{\epsilon}_t$$

where  $\hat{\epsilon}_t$  is a vector of white noise shocks and  $A, B, C$  are conformable. Further, denote  $EX_t X_t' = \Omega$ ,  $\Gamma_1 = E \begin{bmatrix} b_t \\ s_t \end{bmatrix} \begin{bmatrix} b_{t-1} \\ s_{t-1} \end{bmatrix}'$ , and  $e_j$  is a  $(1 \times 2)$  unit vector with a 1 in the  $j$ th element.

**Lemma 1** *A unique restricted perceptions equilibrium exists for all  $n$  if and only if*

$$\Delta \equiv \det(I_4 - P' \otimes B) \neq 0$$

where

$$P = \Gamma_1' \left[ ne_1' (e_1 \Omega e_1')^{-1} e_1 + (1-n)e_2' (e_2 \Omega e_2')^{-1} e_2 \right]$$

**Proof.** In an RPE

$$E e_j \begin{bmatrix} b_{t-1} \\ s_{t-1} \end{bmatrix} \left( X_t - \psi^j e_j \begin{bmatrix} b_{t-1} \\ s_{t-1} \end{bmatrix} \right)' = 0$$

After plugging in for aggregate expectations into the expectational difference equation

$$X_t = \xi \begin{bmatrix} b_t \\ s_t \end{bmatrix} + C\hat{\epsilon}_t$$

where

$$\xi = A + nB\psi^s e_1 + (1-n)B\psi^b e_2$$

Using this notation,

$$\begin{aligned} \psi^{j'} &= (e_j \Omega e_j')^{-1} E e_j \begin{bmatrix} b_{t-1} \\ s_{t-1} \end{bmatrix} X_t' \\ &= (e_j \Omega e_j')^{-1} e_j \Gamma_1 \xi' \end{aligned}$$

After plugging in for  $\psi^s, \psi^b$  into  $\xi$ :

$$\begin{aligned}\xi &= A + B\xi\gamma'_1 \left[ ne'_1 (e_1\Omega e'_1)^{-1} e_1 + (1-n)e'_2 (e_2\Omega e'_2)^{-1} e_2 \right] \\ \Leftrightarrow \xi &= A + B\xi P\end{aligned}$$

It follows that

$$vec(\xi) = vec(A) + (P' \otimes B) vec(\xi)$$

Finally, the RPE coefficient is given by

$$vec(\xi) = (I_4 - (P' \otimes B))^{-1} vec(A)$$

and the stated conditions provides necessary and sufficient conditions for a unique  $\xi$ . ■

### Proof of Proposition 1.

The existence of a set of fixed points  $n_* = T_\omega(n_*)$  follow directly from applying Brouwer's theorem, since  $T_\omega : [0, 1] \rightarrow [0, 1]$  and  $F(n)$  is continuous provided there exists an RPE. Lemma 1 provides the requisite necessary and sufficient conditions. To complete the proof, we simply require establishing the existence of a unique RPE for an open set of  $n$ . This is straightforward as for  $n = 0$  or  $n = 1$  implies that  $\Delta = 0$  and  $\xi$  is continuous in  $n$ .

■

## B.2 Proof of Proposition 2 (Extrinsic Heterogeneity)

### Proof.

From the simplifications made in Section 3 it follows that (6) becomes

$$b_{t+1} = \beta^{-1}(b_t - s_t), \tag{B.2.1}$$

which can be written as (A.1.2) with  $\xi_1 \equiv \beta^{-1}$ ,  $\xi_2 \equiv -\beta^{-1}$ , and,  $\xi_3 = 0$ . Moreover, expectations are heterogeneous. Therefore (11) becomes

$$y_t = (1 - \beta)b_{t+1} + \int_0^1 \widehat{E}_t^i v_{t+1}^i di = (1 - \beta)b_{t+1} + n\psi_v^s s_t + (1 - n)\psi_v^b b_t, \tag{B.2.2}$$

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

$$\begin{aligned}v_t^i &= (1 - \beta)(b_{t+1} - b_t) + \widehat{E}_t^i v_{t+1}^i \\ \int_0^1 v_t^i di &= v_t = (1 - \beta)(b_{t+1} - b_t) + \int_0^1 \widehat{E}_t^i v_{t+1}^i di \\ &= (1 - \beta)(b_{t+1} - b_t) + n\psi_v^s s_t + (1 - n)\psi_v^b b_t\end{aligned} \tag{B.2.3}$$

$$\Leftrightarrow v_t = y_t - (1 - \beta)b_t \tag{B.2.4}$$

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

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.2.3) with (B.2.1) and (B.2.2) 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.2.1) and (B.2.2) to (B.2.4). Coefficients in (13) and (14) are required to satisfy the orthogonality conditions (A.1.7) and (A.1.9) respectively.

Under PF, i.e., assumption (7), 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) and we obtain

$$\Gamma_b^b = \frac{E[b_{t+1}b_t]}{E[b_t^2]} = \beta^{-1}(1 - \phi_b), \quad (\text{B.2.5})$$

where the linear projection  $E[b_{t+1}] = \Gamma_b^b b_t$  satisfies orthogonality condition (A.1.9). Likewise we can compute

$$\Gamma_b^s \equiv \frac{E[b_{t+1}s_t]}{E[s_t^2]} = \frac{-\beta^{-1}(1 - \beta^2 - \phi_b)}{1 - \beta^2 - 2\phi_b}, \quad (\text{B.2.6})$$

where the linear projection  $E[b_{t+1}] = \Gamma_b^s s_t$  satisfies orthogonality condition (A.1.7).

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

$$\begin{aligned} \psi_v^s &= [(\beta^1 - 1) - (1 - \beta) + (1 - n)\psi_v^b]\Gamma_b^s + [n\psi_v^s - (\beta^1 - 1)]\phi_b\Gamma_b^s \\ \psi_v^b &= [(\beta^1 - 1) - (1 - \beta) + (1 - n)\psi_v^b]\Gamma_b^b + [n\psi_v^s - (\beta^1 - 1)]\phi_b\Gamma_b^b. \end{aligned}$$

Rearranging terms yields

$$\psi_v^s = [(\beta^1 - 1)(1 - \beta - \phi_b)]\Gamma_b^s + [\phi_b n\psi_v^s + (1 - n)\psi_v^b]\Gamma_b^s \quad (\text{B.2.7})$$

$$\psi_v^b = [(\beta^1 - 1)(1 - \beta - \phi_b)]\Gamma_b^b + [\phi_b n\psi_v^s + (1 - n)\psi_v^b]\Gamma_b^b. \quad (\text{B.2.8})$$

Clearly,  $n = 0$  implies that (B.2.8) collapses to (B.3.3) below and  $n = 1$  implies that (B.2.7) collapses to (B.4.3) below.

Thus, we can solve for

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

This proves Proposition 2.

■

### B.3 Proof of Corollary 1 (Weak Ricardian Equivalence)

**Proof.**

Suppose that all agents have PLM (14). Then the TE dynamics are still governed by (B.2.1). Moreover, (11) is given by

$$y_t = (\beta^{-1} - 1)(b_t - s_t) + \psi_v^b b_t. \quad (\text{B.3.1})$$

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 (14), i.e.,  $v_t^i = v_t, \forall i$ , the implications for (10) are

$$\begin{aligned} v_t^i &= (1 - \beta)v_t^i + (1 - \beta)\beta[b_{t+1} - b_t] + \beta\psi_v^b b_t \\ \beta v_t^i &= (1 - \beta)\beta[\beta^{-1}(b_t - s_t) - b_t] + \beta\psi_v^b b_t \\ v_t^i &= (1 - \beta)[\beta^{-1}(b_t - s_t) - b_t] + \psi_v^b b_t \\ v_t^i &= (\beta^{-1} - 1)(b_t - s_t) + \psi_v^b b_t - (1 - \beta)b_t \\ v_t^i &= y_t - (1 - \beta)b_t. \end{aligned} \quad (\text{B.3.2})$$

The ALM is then given by (B.2.1), (B.3.1) and (B.3.2).

Now we can apply  $v_t \equiv \int v_t^i di$  to (B.3.2) and combine it with (B.3.1) 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,  $\mu_{v,3} = 0$ .

Due to (B.2.5), (A.1.16) is given by

$$\begin{aligned} \psi_v^b &= [(\beta^{-1} - 1)(1 - \beta - \phi_b) + \psi_v^b] \Gamma_b^b \\ 0 &= \psi_v^b - [(\beta^{-1} - 1)(1 - \beta - \phi_b) + \psi_v^b] \Gamma_b^b, \end{aligned} \quad (\text{B.3.3})$$

and  $\psi_v^b$  in (15) follows.

Notice that (B.3.1) together with (5) and (15) 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 Corollary 1.

■

### B.4 Proof of Corollary 2 (Woodford (2013))

**Proof.**

Suppose  $n = 1$ , i.e., all agents use PLM (13). In this case, (B.2.1) and (B.3.2) remain the same, however (11) becomes

$$y_t = -\sigma(\phi_\pi \pi_t) + (1 - \beta)\beta^{-1}(b_t - s_t) + \psi_v^s s_t \quad (\text{B.4.1})$$

$$y_t = (\beta^{-1} - 1)(b_t - s_t) + \psi_v^s s_t, \quad (\text{B.4.2})$$

where (B.4.1) 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.2.1), (B.4.2) and (B.3.2). Thus, we can apply  $v_t \equiv \int v_t^i di$  to (B.3.2) and combine it with (B.4.2) 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,  $\mu_{v,3} = 0$ .

Using (B.2.6), we obtain

$$\psi_v^s = [(\beta^{-1} - 1)(1 - \beta - \phi_b) + \psi_v^s \phi_b] \Gamma_b^s \quad (\text{B.4.3})$$

$$\Leftrightarrow \psi_v^s = \frac{(1 - \beta)(1 - \beta - \phi_b)\Gamma_b^s}{\beta(1 - \phi_b\Gamma_b^s)} = -\frac{\beta^{-1}(1 - \beta)(1 - \beta^2 - \phi_b)}{(\beta + \beta^2 + \phi_b)} \quad (\text{B.4.4})$$

$$\Leftrightarrow \psi_v^s < \beta^{-1} - 1. \quad (\text{B.4.5})$$

From (B.4.2), (5) and (B.4.4) to (B.4.5) follows that

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

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

Thus, as  $y_t$  depends on  $b_t$ , Ricardian equivalence fails. This proves Corollary 2.

■

## B.5 Proof of Theorem 1

**Proof.**

Recall that in this simplified version of the model (A.1.3) is given 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,  $\mu_{v,3} = 0$ . Moreover,  $\mu_{p,1} = \mu_{p,2} = \mu_{p,3} = 0$ .

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

$$F(n) = \left( \frac{(1 - \beta)^2 \sigma_z^2}{\beta^2 (1 - \beta^2 - n(1 + \beta - \phi_b) - 2\phi_b)^2} \right) \times \quad (\text{B.5.1})$$

$$(2n(1 - \beta^2 - \phi_b)(2(1 - \phi_b) + \beta(1 - \beta)) - (1 - \beta^2 - 2\phi_b)(4(1 - \phi_b) - \beta(2 + 3\beta)))$$

where  $\sigma_z^2 \equiv E[z_t z_t]$ . From (B.5.1) one can observe that the denominator of  $F(n)$  is always positive and whether  $F(n)$  is positive or negative depends on the numerator. Then it is straight-forward to verify Corollary 3 and therefore to prove Theorem 1.

■

## B.6 Proof of Corollary 4

**Proof.**

In the simple case of Corollary 2, we find that

$$F(n) = \sigma_z^2 (1-\beta)^2 \left[ \frac{2n(1-\beta^2-\phi_b)(2(1-\phi_b)+\beta(1-\beta)) - (1-\beta^2-2\phi_b)(4(1-\phi_b)-\beta(2-3\beta))}{\beta^2(\beta^2+n(1+\beta-\phi_b)+2\phi_b-1)^2} \right]$$

To prove the result, it suffices to provide conditions under which  $F(n)$  is monotonically increasing in  $n$ . In this case,  $T_\omega(n)$  is monotonically increasing on  $[0, 1]$  and, therefore according to the intermediate value theorem,  $T_\omega(n)$  has 3 fixed points, and for  $0 < \omega, \infty$  they satisfy  $0 < n_l^* < \hat{n} < n_h^* < 1$ , with  $F(\hat{n}) = 0$ .

To prove the sufficient conditions for monotonicity, first compute

$$\begin{aligned} F'(n) &= 2 \left[ \beta^2 (-1 + \beta^2 + n(1 + \beta - \phi_b) + 2\phi_b)^3 \right]^{-1} \\ &\times \left\{ (-1 + \beta^2 + 2\phi_b) [-n(1 + \beta - \phi_b) (-2 - \beta + \beta^2 + 2\phi_b)] \right. \\ &\quad \left. - 4\beta^2 + \beta^4 + \beta(\phi_b - 1) - 2(\phi_b - 1)^2 + \beta^2(6\phi_b - 4) \right\} \end{aligned}$$

Tedious algebra shows that  $F'(n)$  is monotonically increasing in  $n$  provided that  $\underline{\phi} < \phi_b \bar{\phi}$  where  $\underline{\phi}$  is the exact 2nd root of the polynomial

$$2z^3 - (13\beta^2 - \beta + 4)z^2 + (2 - 3\beta + 15\beta^2 + 11\beta^3 - 9\beta^4)z + 2\beta - 4\beta^2 - 7\beta^3 + 5\beta^4 + 5\beta^5 - \beta^6$$

and,  $\underline{\phi} > \bar{\phi}$  for all  $.0882 < \beta < 1$ .

■

## B.7 Proof of Proposition 3

**Proof.**

The TE dynamics in this case are (B.2.1),

$$y_t = -\sigma\phi_\pi\pi_t + (1 - \beta)b_{t+1} + n\psi_v^s s_t + (1 - n)\psi_v^b b_t \quad (\text{B.7.1})$$

$$\begin{aligned} \pi_t &= \kappa y_t + u_t + (1 - \alpha)\beta \int \widehat{E}_t^j p_{t+1}^*(j) dj \\ &= \kappa y_t + u_t + (1 - \alpha)\beta [n\psi_p^s s_t + (1 - n)\psi_p^b b_t] \end{aligned} \quad (\text{B.7.2})$$

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

$$v_t = (1 - \beta)(b_{t+1} - b_t) - \sigma(\phi_\pi - 1)\pi_t + n\psi_v^s s_t + (1 - n)\psi_v^b b_t \quad (\text{B.7.3})$$

$$p_t^*(j) = (1 - \alpha)p_t^* + (1 - \alpha\beta) [\xi y_t + \mu_t] + \alpha\beta \widehat{E}_t^j p_{t+1}^*(j) \quad (\text{B.7.4})$$

$$\begin{aligned} \int_0^1 p_t^*(j) dj &= p_t^* = \frac{(1 - \alpha\beta)}{\alpha} [\xi y_t + \mu_t] + \beta \int_0^1 \widehat{E}_t^j p_{t+1}^*(j) dj \\ p_t^* &= (1 - \alpha)^{-1} [\kappa y_t + u_t] + \beta [n\psi_p^s s_t + (1 - n)\psi_p^b b_t] \\ (1 - \alpha)^{-1} \pi_t &= (1 - \alpha)^{-1} [\kappa y_t + u_t] + \beta [n\psi_p^s s_t + (1 - n)\psi_p^b b_t] \\ \pi_t &= \kappa y_t + u_t + (1 - \alpha)\beta [n\psi_p^s s_t + (1 - n)\psi_p^b b_t]. \end{aligned} \quad (\text{B.7.5})$$

The ALM is then given by (13) to (14), (B.2.1), (B.7.1), (B.7.2), (B.7.3), and (B.7.4) to (B.7.5) for given policy (5) and (6).

Next, we can combine (B.7.1) and (B.7.2) to obtain (A.1.1) with coefficients

$$\begin{aligned} \delta_0 &\equiv \frac{(1 - \beta)}{(1 + \sigma\phi_\pi\kappa)}, & \delta_1 &\equiv \frac{(1 - n)(\psi_v^b - \sigma\phi_\pi(1 - \alpha)\beta\psi_p^b)}{(1 + \sigma\phi_\pi\kappa)}, \\ \delta_2 &\equiv \frac{n(\psi_v^s - \sigma\phi_\pi(1 - \alpha)\beta\psi_p^s)}{(1 + \sigma\phi_\pi\kappa)}, & \delta_3 &\equiv \frac{-\sigma\phi_\pi}{(1 + \sigma\phi_\pi\kappa)}. \end{aligned}$$

Moreover, we use (A.1.1) to eliminate  $y_t$  in (B.7.5), i.e.,

$$\begin{aligned} \pi_t &= \kappa [\delta_0 b_{t+1} + \delta_1 b_t + \delta_2 s_t + \delta_3 u_t] + u_t + (1 - \alpha)\beta [n\psi_p^s s_t + (1 - n)\psi_p^b b_t] \\ \pi_t &= [\kappa\delta_1 + (1 - \alpha)\beta(1 - n)\psi_p^b] b_t + [\kappa\delta_2 + (1 - \alpha)\beta n\psi_p^s] s_t \\ &\quad + [\kappa\delta_3 + 1] u_t + \kappa\delta_0 b_{t+1}. \end{aligned} \quad (\text{B.7.6})$$

Furthermore, we use (A.1.1), (A.1.2) and (B.7.6) to eliminate  $\pi_t$ ,  $y_t$  and  $b_{t+1}$  in (B.7.3), i.e.,

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

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

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

Moreover, we use (A.1.2) to eliminate  $b_{t+1}$  in (B.7.6), i.e.,

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

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

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

Thus, we can obtain (A.1.14), (A.1.15), (A.1.16) and (A.1.17) as

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

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

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

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

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

Based on the coefficients above, we can compute  $F(n)$  as outlined in Appendix A.2 above. Taking the limit as  $\sigma \rightarrow 0$ , and simplifying, produces the result in the text.

■

## C Connection to rational expectations/transitional learning dynamics: details

In this subsection, we address this concern 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  $\psi^s, \psi^b$  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

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

where

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

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  $S(\psi^s, \psi^b)$ . 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  $\Theta$  given by the following system of ordinary differential equations (O.D.E.’s)

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

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. 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. Over time, one can expect with high probability to see coefficient estimates  $\Theta$  that fluctuate around  $\Theta^*$ . The response of  $\Theta$  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.<sup>27</sup> The escape dynamics, therefore, can be thought of as re-initializing the mean

---

<sup>27</sup>See Williams (2018) for details and a comprehensive set of results and toolkit on escape dynamics in constant gain learning models.

dynamics. 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,  $\gamma$ , corresponds to a long stretch of real time.

Figure 12 plots the mean dynamics for a particular illustrative parameterization:  $\phi_y = 0.5$ ,  $\phi_b = 0.9$ ,  $\sigma = 2$ ,  $\sigma_z^2 = 1$ .<sup>28</sup> We then choose initial values for the  $\psi^s, \psi^b$  that are both above their rational expectations equilibrium values. The mean dynamics O.D.E. is then solved and Figure 12 plots the expected learning path ( $\psi^s$  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.

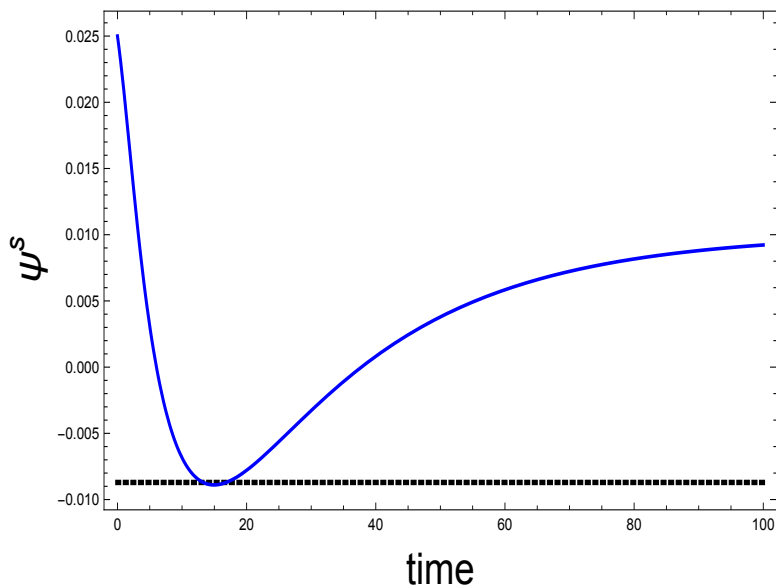


Figure 12: Expected learning dynamics for a correctly specified forecast model

The figure plots (solid line) the expected transition path for  $\psi^s$  while the dashed line is the value in an  $n = 1$  restricted perceptions equilibrium, that is the value for  $\psi^s$  and  $\psi^b$  that would arise in an  $n = 1$  RPE. The learning dynamics are expected to eventually

<sup>28</sup>For expositional ease, we present an example where the RPE and REE values are starkly far apart.

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  $\psi^s$ , the beliefs hover for a finite stretch of time at its  $n = 1$  RPE. This coincides with a path for  $\psi^b$  (not shown) that is also drifting down towards its RPE value. As the path for  $\psi^b$  continues to transition towards its REE value, this draws  $\psi^s$  away from its RPE value and back towards the rational expectations equilibrium.

We conclude from Figure 12 that even if agents did not face any computational/cognitive constraints the RPE is a relevant concept as we can expect recurrent escapes near a non-Ricardian equilibrium even when all agents in the economy form forecasts from a correctly specified model. Moreover, for small gains  $\gamma$ , the economy will persist near the RPE for long stretches of time. These dynamics are reminiscent of Cho and Kasa (2017).

The mean dynamics in Figure 12 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. Therefore, the theory in this paper is complementary to theirs, while providing an equilibrium explanation for the phenomenon of non-Ricardian beliefs.

## D Additional Model Details

This Appendix provides additional details on the model and its derivations. The reader is referred to Woodford (2013) for more complete details. All parameters and variables are explained in the paper above.

HOUSEHOLDS. Woodford (2013) derives an individual’s consumption function,

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

The first two terms in (D.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  $\sigma$ , captures an intertemporal substitution 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.

Then imposing Ricardian beliefs (2) onto the consumption rule (D.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:

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

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 (D.1) so that

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

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

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

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.<sup>29</sup>

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

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

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

$$\begin{aligned} p_t^*(j) &= (1 - \alpha\beta) (E_t^j p_t^{\text{opt}} - p_{t-1}) + (\alpha\beta) E_t^j p_{t+1}^*(j) + (\alpha\beta) \pi_t, \quad \text{where} \quad (\text{D.2}) \\ E_t^j p_{t+1}^*(j) &\equiv (1 - \alpha\beta) \sum_{T=t}^{\infty} (\alpha\beta)^{T-t} (E_t^j p_{T+1}^{\text{opt}} - p_t). \end{aligned}$$

**TEMPORARY EQUILIBRIUM WITH HETEROGENEOUS BELIEFS.** Recall the consumption function recursion:

$$\begin{aligned} c_t^i &= (1 - \beta) b_t^i + (1 - \beta)(Y_t - \tau_t) - \beta[\sigma - (1 - \beta)s_b] i_t - (1 - \beta)s_b \pi_t + \beta \bar{c}_t + \beta E_t^i v_{t+1}^i \\ v_t^i &= (1 - \beta)(Y_t - \tau_t) - [\sigma - (1 - \beta)s_b](\beta i_t - \pi_t) - (1 - \beta)\bar{c}_t + \beta E_t^i v_{t+1}^i. \end{aligned}$$

Because the continuation variable  $v_t^i$  consists of aggregate variables that are beyond the household's control, and the agents understand their optimal consumption plan and perceived budget constraints, we can instead write the consumption function as follows

$$\begin{aligned} c_t^i &= (1 - \beta) [b_t^i + Y_t - \tau_t - s_b \pi_t] - [\sigma - (1 - \beta)s_b] i_t + \beta E_t^i \hat{v}_{t+1} \\ \hat{v}_t &= (1 - \beta) [Y_t - \tau_t - \bar{c}_t] - [\sigma - (1 - \beta)s_b] (\beta i_t - \pi_t) + \beta \hat{v}_{t+1} \end{aligned}$$

and  $v_t^i = E_t^i \hat{v}$  by construction. As in the text, combining the  $\hat{v}_t$  recursion with the goods market clearing condition and the government's flow budget constraint confirms that  $\hat{v}_t = v_t$ , and the aggregation result in the main text follows immediately. The ease with which the heterogeneous beliefs aggregate follows from the infinite-horizon learning consumption which depends on household  $i$ 's subjective *forecasts* of aggregate variables beyond their control. An example of where aggregation of heterogeneous beliefs is made more difficult by higher-order beliefs is provided by Branch and McGough (2009).

---

<sup>29</sup>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 beliefs. We show below how these consumption rules can be aggregated with heterogeneous agents.

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

$$p_t^{\text{opt}} = p_t + \xi (Y_t - Y_t^n) + \mu_t, \quad (\text{D.3})$$

where  $\xi > 0$  is a composite term of structural parameters measuring the output elasticity of a firm's optimal price.<sup>30</sup> The exogenous random variable  $Y_t^n$  is the natural level of output that captures exogenous demand shocks and  $\mu_t$  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 (D.3) and (3) into (D.2) that

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

Again averaging across firms, using (3), 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

$$\pi_t = (1 - \alpha)\beta \int E_t^j p_{t+1}^*(j) dj + \kappa y_t + u_t. \quad (\text{D.4})$$

As for the households, after applying the law of iterated expectations a firm  $j$  sets

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

and, an aggregate New Keynesian Phillips Curve results after averaging across all firms:

$$\pi_t = (1 - \alpha)\beta \hat{E}_t p_{t+1}^* + \kappa y_t + u_t.$$

---

<sup>30</sup>The term is defined in Woodford (2003, ch.3-4).