Endogenously (Non-)Ricardian Beliefs*

William A. Branch
University of California, Irvine

Emanuel Gasteiger
TU Wien

August 11, 2020

Abstract

We propose an interpretation of U.S. inflation as arising from *belief-driven* regime shifts. The findings follow from a restricted perceptions theory where Ricardian equivalence, or its failure, arises endogenously. Agents face a choice between two forecasting models: in one, Ricardian beliefs emerge as a self-confirming equilibrium; the other features non-Ricardian beliefs. In a New Keynesian model, there may be multiple equilibria, where there is simultaneously (non-)Ricardian beliefs. Estimates from the model provide evidence for endogenously (non-)Ricardian beliefs even when fiscal policy is Ricardian. The 1970's were a period of rising non-Ricardian beliefs, the 1980's were broadly Ricardian, and the 1990's were non-Ricardian, though, without the manifestation into high inflation.

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 (UIDB/00315/2020), the Berlin Economics Research Associates (BERA) program and the Fritz Thyssen Foundation. We are grateful for many constructive comments from an Associate Editor and a Referee. 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 novel interpretation of U.S. inflation data as arising from *belief-driven* regime shifts, rather than *policy-driven* regime change. Our empirical findings follow from 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 (2018) 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 point of departure 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 with failur of Ricardian equivalence. Model estimates identify a self-reinforcing feedback between fiscal imbalances and endogenously (non-)Ricardian beliefs that helps explain the U.S. inflation experience.

The design of monetary and fiscal policy often hinges on whether Ricardian equivalence holds or fails (c.f. Leeper, 1991). Davig and Leeper (2006) provide evidence in favor of a model of inflation driven by policy regime switches that evolve between (non-)Ricardian 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. Breakthrough papers by Evans et al. (2009) and Eusepi and Preston (2018) open a new avenue for research into the implications for inflation in an environment where the private sector has imperfect knowledge about whether the paths for primary surpluses will adjust to satisfy the government's intertemporal budget constraint. 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 even though the policy regime itself is Ricardian.

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 primary surpluses. 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.

Our ideas are formalized 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. 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. In the former, equilibrium beliefs about the possible paths for future debt are correct and real variables display a (weak) Ricardian equivalence. Conversely, the surplus-based model leads to a restricted perceptions equilibrium where Ricardian equivalence fails. After estimating the model, the main results show that beliefs are endogenously heterogeneous, time-varying, and non-Ricardian.

We estimate the model using Bayesian techniques. To make the model empirically realistic, we generalize the model by allowing for the agents to estimate their forecasting equations in real-time, following a line of research beginning with Milani (2007). The endogenous beliefs, though, render the model highly non-linear and present substantial computational complexity during estimation. We develop a MCMC algorithm and approximate the likelihood function with a particle filter. Empirical estimates identify that, on average, there are non-Ricardian beliefs in the U.S. economy, though the extent varies over time periods. For instance the late 1960's and 1970's feature increasing degrees of non-Ricardian beliefs and the late 1970's and early 1980's are largely consistent with a Ricardian regime. One novel finding is that the 1990's are consistent with an equilibrium with endogenously non-Ricardian beliefs, though without manifesting higher inflation. The estimated Bayes factors prefer the model with endogenously non-Ricardian beliefs when compared to the version of the model with fully rational expectations and with fully non-Ricardian expectations.

The estimated state dynamics yield the following interpretation of U.S. inflation. Going from the 1960's into the 1970's, U.S. fiscal policy was expansionary and ran substantial fiscal imbalances. Those trends, we estimate, led to an increasing fraction of households and firms with non-Ricardian beliefs, reaching a peak in the early 1970's. In turn, inflation trended up and stabilized the real debt-to-GDP ratio. As the inflation rate peaked and real debt-to-GDP stabilized, we estimate that the economy transitioned to a Ricardian regime in the early 1980's. However, the large fiscal balances of the 1980's again destabilized real debt ratios and led to another period of endogenously non-Ricardian beliefs that peaked in the early 1990's. The transition to non-Ricardian beliefs in the 1990's, though, did not translate into substantially higher inflation as fiscal authorities ran primary surpluses, which had the effect of helping to anchor expectations by mitigating demand effects.

The final part of our analysis explores the policy implications. The message from this paper is that endogenously (non-)Ricardian beliefs can account for U.S. inflation data even without policy regime-change. However, rolling window estimates provide a nuanced view on evolving beliefs versus policy rules. Allowing for a less active monetary policy leads our estimates of the fiscal rule to be less passive and a lower degree of non-Ricardian beliefs. Counterfactual experiments show that, while still within an active monetary/passive fiscal regime, there is scope for policy coordination via the endogeneity of non-Ricardian beliefs.

2. Model

We begin by generalizing 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, based on Preston (2005), where households and firms have subjective beliefs about payoff-relevant aggregate variables.[1] Given these beliefs,

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

households choose consumption, leisure, and one-period government debt, the only asset available to households, to solve their intertemporal optimization problem. In Woodford (2013) 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. Households also receive a lump-sum transfer of their share in firm profits.[2] 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

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

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, $\sigma > 0$ is the elasticity of intertemporal substitution, and $s_b \equiv \bar{b}/\bar{Y} \geq 0$ is the steady-state debt-to-GDP ratio. The fiscal policy instrument is the real primary surplus $s_t \equiv \tau_t - G_t$.

Following Eusepi and Preston (2018), 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.

[2] The shares in firms are illiquid, which makes government debt the only storable good. Eusepi and Preston (2018) 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

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

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.[3]

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,

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

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

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

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

$$i_t = \phi_\pi \pi_t + \phi_y y_t + w_t, \tag{4}$$

where the monetary policy shock is $w_t \sim \text{iid}(0, \sigma_w^2)$.[4]

[3] 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).

[4] In the quantitative analysis we assume all exogenous shocks are stationary AR(1)'s.

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

$$s_t = \phi_b b_t + z_t, \tag{5}$$

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

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

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.[5]

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

$$1 < \phi_\pi + \frac{1-\beta}{\kappa}\phi_y \tag{7}$$

$$(1-\beta) < \phi_b < 1. \tag{8}$$

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).

2.2 Temporary equilibrium with heterogeneous beliefs

The income-expenditure identity is given by

$$Y_t = \int c_t^i di + G_t. \tag{9}$$

Combining (1) and (9) with the bond-market clearing condition $b_t \equiv \int b_t^i di$, computing $v_t \equiv \int v_t^i di$, and averaging over expectations, allows us to express aggregate demand as the "IS equation" without *a priori* imposing Ricardian beliefs:

$$Y_t = g_t - \sigma i_t + (1-\beta) b_{t+1} + \hat{E}_t v_{t+1}, \tag{10}$$

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

where $g_t \equiv \bar{c}_t + G_t$ is a composite exogenous disturbance, such that $g_t \sim \text{iid}(0, \sigma_g^2)$. 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.[6]

On the firm side, applying the law of iterated expectations and aggregating across all firms, results in an aggregate New Keynesian Phillips Curve:

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

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 $u_t \equiv \{[(1-\alpha)(1-\alpha\beta)]/\alpha\}\mu_t$. Y_t^n and μ_t denote the natural rate of output and the markup deviations respectively.

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

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

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

[6] 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 but beyond the scope of the present study.

2.3 Model misspecification

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 η_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 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 endogenoulsy (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 (SCE) 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 is a convenient formalization of (non-)Ricardian beliefs since Ricardian equivalence is not ruled out *a priori*. Second, there is a long empirical tradition of using a limited number of fiscal indicators in empirical studies (e.g. Favero and Giavazzi, 2012). Third, 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):

$$PLM_s: \ \mathbf{Z}_t = \boldsymbol{\psi}^s s_{t-1} + \eta_t \Rightarrow E_t^s \mathbf{Z}_{t+1} = \boldsymbol{\psi}^s s_t \tag{11}$$

$$PLM_b: \ \mathbf{Z}_t = \boldsymbol{\psi}^b b_{t-1} + \eta_t \Rightarrow E_t^b \mathbf{Z}_{t+1} = \boldsymbol{\psi}^b b_t, \tag{12}$$

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

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

$\psi^k = \left(\psi_v^k, \psi_p^k\right)'$ and Γ^k is the coefficient for b_{t+1}.[7] In a RPE the coefficients will satisfy the least-squares orthogonality condition:

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

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

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

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

We do not impose *a priori* which of the PLM's individuals and firms use to form expectations. Instead, we confront them with a discrete choice: they can forecast with the *s-model* or the *b-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 (2006). 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 the s-model, leaving $1-n$ of the population forecasting with the b-model.[8] They rank these choices by calculating the relative mean square error (MSE):

$$EU^k = -E\left[(\mathbf{Z}_t - E_t^k[\mathbf{Z}_t^k])\right]' \times \mathbf{W} \times E\left[(\mathbf{Z}_t - E_t^k[\mathbf{Z}_t^k])\right], \qquad k = \{s, b\}, \tag{13}$$

where $\mathbf{W}$ is a weighting matrix.[9] Consequently, we define relative predictor performance $F(n) : [0,1] \to \mathbb{R}$ as $F(n) \equiv EU^s - EU^b$.

Building on Brock and Hommes (1997), we assume that the distribution of agents across

[7]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.

[8]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.

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

the two forecasting models, n, is pinned down according to the multinomial logit (MNL) map (see, e.g., Branch and Evans, 2006)

$$n = \frac{1}{2}\left\{\tanh\left[\frac{\omega}{2}F(n)\right] + 1\right\} \equiv T_\omega(n), \tag{14}$$

where ω denotes the "intensity of choice". The MNL map – also, the "T-map" – states that the fraction of agents adopting the s-model, 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* $n_* = T_\omega(n_*)$.

An immediate consequence of the continuity of $T_\omega : [0,1] \to [0,1]$ is that there exists a ME: see Appendix B for analytic details on existence. The neoclassical case $\omega \to \infty$ warrants special attention. In this case, agents only select the best-performing statistical models. It turns out that, in this case, one can learn quite a bit about the set of MEs by studying the end points to $F(n)$. For instance, when $F(0) < 0, F(1) < 0$ – that is, the b-model forecasts best when all agents use the b-model or if they all use the s-model – then $n_* = 0$ is a ME. Conversely, when $F(0) > 0$ and $F(1) > 0$ then $n_* = 1$ is a ME. Outside of these polar cases, there is also the possibility of multiple MEs, $n = \{0, \hat{n}, 1\}$, for some $0 < \hat{n} < 1$, that arises when $F(0) < 0, F(1) > 0$. This case will make repeated appearances in various points of the remainder of this paper. As we will see, the $n = 0$ ME can be thought of as a SCE with *weakly Ricardian beliefs* and the $n = 1$ will correspond to *non-Ricardian beliefs*.[10] 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.[11]

The neoclassical limiting case, $\omega \to \infty$, is useful for identifying sufficient conditions for the existence of multiple equilibria. The multinomial approach, i.e., a finite ω, 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 ω is inversely related to the idiosyncratic random utility innovation and, thereby, parameterizes model uncertainty. In particular, larger values of ω parameterize less model uncertainty with the neoclassical

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

[11] Note that $T'_\omega(\hat{n}) > 1$ so that the $\hat{n}$ case is never going to be stable under learning and therefore the discussion is centered around the other two equilibria.

case $\omega \to \infty$ representing no uncertainty at all. An extensive literature tests for dynamic predictor selection using empirical MNL models and typically finds finite values for ω. The parameter ω is an object in our estimation below.

3. Theoretical results: a simple example

This section presents analytic results for a special case, first exposited by Woodford (2013), which reduces the fiscal variables, b_t, s_t, to follow exogenous processes, $\pi_t = 0$ for all t, and households have a simple permanent income problem to solve; that is, $\phi_y = s_b = \kappa = 0$ and $\alpha = 1$. We further shut down all of the exogenous disturbances except for the fiscal policy shock z_t. While not directly empirically relevant, the simple example facilitates a deeper understanding of the mechanism identified in the empirical analysis. Appendix B generalizes the results.

3.1 Restricted perceptions equilibria

We begin by characterizing the RPEs with *extrinsic heterogeneity*, i.e., with an exogenous distribution n. In this special case households need only forecast the continuation variable v_{t+1} using a single regressor, either s_t or b_t. The key model equations are the surplus rule (5) and

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

Assume that a fraction $n \in [0,1]$ of agents use the b-model and $1-n$ use the s-model. Depending on the distribution n, (non-)Ricardian equilibria can emerge.

Proposition 1 (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

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

where

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

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 SCE in the sense of Sargent (1999).

Corollary 4 (Weak Ricardian Equivalence) *In the special parametric case, if all agents form expectations from the b-model* (12)*, then there exists a unique restricted perceptions equilibrium with*

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

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

Corollary 5 (Woodford (2013)) *In the special parametric case, if all agents form expectations from the s-model* (11)*, then there exists a unique restricted perceptions equilibrium with*

$$\begin{aligned} y_t &= \left[\frac{(1-\beta)(1+\beta-\phi_b)}{\beta(1+\beta)+\phi_b}\right] b_t - \left[\frac{\beta^{-1}-\beta}{\beta(1+\beta)+\phi_b}\right] z_t \\ \psi^s &= -\frac{\beta^{-1}(1-\beta)(1-\beta^2-\phi_b)}{(\beta+\beta^2+\phi_b)} < \beta^{-1}-1. \end{aligned} \tag{16}$$

Remark 6 *Proposition 1 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 $n\to 0$ – then neither type of agent holds Ricardian beliefs.*

3.2 Misspecification equilibrium

Proposition 1 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 ME. The following result provides necessary and sufficient conditions for the existence of multiple MEs in the limiting case $\omega \to \infty$.

Theorem 7 (Multiple Misspecification Equilibria) *Consider the special parametric case of the model. Let $\omega \to \infty$ and $\beta > 2/3$. There exist multiple misspecification equilibria, $n^* \in \{0, \hat{n}, 1\}$, if and only if*

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

where

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

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

Corollary 8 (Unique Misspecification Equilibria) *Let $\omega \to \infty$. The following results hold.*

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

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

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

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

Remark 9 *The main drawback to the example here is that public debt is real and exogenous, thereby lacking important feedback/valuation effects, through beliefs, onto the inflation process. While the theoretical results hold in a simple case of the model, it is illustrative for several reasons. First, it gives insights into conditions under which endogenous non-Ricardian beliefs can arise in equilibrium. Second, the special case under consideration here is presented for tractability, but the results generalize in more empirically realistic versions of the model. For instance, Appendix B presents a generalization of Theorem 7 to a New*

Keynesian model and monetary policy that adheres to a Taylor-type rule. This generalization continues for economies with small steady-state debt-to-GDP ratios. Multiple equilibria in the case of large ω also arises, as we demonstrate below, in the estimated version of the full model. Moreover, valuation effcts, operating through bliefs, provide feedback effects between public debt, output and prices that are important empirically.

Figure 1 illustrates the results in Theorem 7 and Corollary 8, i.e., the large ω case. There are combinations of (β, ϕ_b) consistent with multiple or unique equilibria. The large unshaded area in the lower half of the plot corresponds to active fiscal policy, i.e., $\phi_b > (1-\beta)$. The restriction to Ricardian policy rules out equilibria in this region. Then, moving outward from the origin, the shaded area with a dashed-boundary consists of the pairs of (β, ϕ_b) consistent with a unique non-Ricardian equilibrium. The next shaded area, with grid lines, corresponds to the existence of multiple equilibria. Finally, the outermost shaded area is where a Ricardian SCE, $n^* = 0$, is the unique ME.

Theorem 7, and its corollaries, sets up the main result of the paper: even though fiscal policy is passive, non-Ricardian beliefs can emerge endogenously. For ϕ_b within a certain range $[\underline{\phi}, \bar{\phi}]$ the non-Ricardian outcome can be sustained in a ME. Most interestingly, for these fiscal policy rules there exist multiple MEs 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 ω. This figure clearly indicates the three MEs.

Why would individuals ever prefer the non-Ricardian forecasting model? There is a direct effect on expectations that depends on 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. Additionally, there is an indirect effect that stems from the self-referential property of New Keynesian models, i.e. the equilibrium weight on the primary surplus and the debt state variables are determined jointly along with the objects n and $\psi^k, k = \{s, b\}$. The result in Theorem 7 says that, even within a "passive" fiscal regime, how responsive primary surpluses are to past debt levels is important for the existence of non-Ricardian beliefs. Non-Ricardian beliefs will exist if the policy coefficient ϕ_b is moderately passive. This is a point we will return to when discussing the empirical results.

3.3 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 address this possibility in Appendix C and find 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 a correctly specified model for a finite stretch of time.

4. Empirical results

We now turn to our main interest: an empirial assessment of the role played by endogenously (non-)Ricardian beliefs. The approach pursued is to use Bayesian methods to estimate the model's key parameters, to use the estimated Bayes factors to assess model fit *vis a vis* a restriction to fully rational expectations or fully Ricardian beliefs and, finally, to estimate the evolution of belief regimes over the sample.

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 (2018) in replacing the fixed RPE parameters with a real-time learning process. Extending the two restricted forecasting models to this more general environment, we can write

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

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 shocks follow uncorrelated stationary AR(1) processes with parameters ρ_l, σ_l and $\sigma_{ll'} = 0$ for all $l, l' \in \{g, u, w, z\}, l' \neq l$.

We follow the econometric learning literature (Marcet and Sargent, 1989; Evans and Honkapohja, 2001) by including beliefs in the set of state variables in order to have a model that can plausibly account for the persistence in U.S. macroeconomic data. Many modelers instead specify a medium-scale New Keynesian model that incorporates mechanical forms of persistence through, for instance, habits and price indexation. However, it is well known (see Milani, 2007) that estimated New Keynesian models prefer specifications where inertia is generated endogenously through beliefs. Here we allow, but do not impose, that the coefficients in the forecasting equations and the forecasting performance are adjusted in real-time from a recursive estimator that features misspecification equilibria as stable limit points. Besides providing a better empirical fit, the real-time learning approach also has the added advantage of capturing endogenous belief-driven regime shifts in our empirical model, while also remaining consistent with our restricted perceptions viewpoint.

The particular econometric learning process adopted here is a recursive Bayesian model based on Evans and Honkapohja (2001):

$$\psi^k_{j,t} = \psi^k_{j,t-1} + \gamma_1 \Gamma X_{k,t-1} \left(x_{j,t} - \left(\psi^k_{j,t-1}\right)' X_{k,t-1} \right) \tag{17}$$

$$MSE^k_{j,t} = MSE^k_{j,t-1} + \gamma_2 \left[\left(x_{j,t} - \left(\psi^k_{j,t-1}\right)' X_{k,t} \right)^2 - MSE^k_{j,t-1} \right] \tag{18}$$

$$EU^k_t = -MSE^k_{v,t} - MSE^k_{p,t} \tag{19}$$

$$n_t = \frac{1}{2} \left\{ \tanh \left[\frac{\omega}{2} \left(EU^s_t - EU^b_t \right) \right] + 1 \right\}. \tag{20}$$

Equation (17) is called a generalized stochastic gradient algorithm that emerges from a Bayesian time-varying parameter model where the coefficients follow a random-walk. With appropriate priors on parameter drift, the stochastic gradient emerges from the Kalman Filter. The stochastic gradient has several desirable properties including being robust to uncertainty about the data generating process and being optimal for risk-sensitive loss functions (see Evans et al., 2010). While for our purposes, the stochastic gradient algorithm significantly eases the computational cost of the estimation procedure.[12] The parameter $0 < \gamma_1 < 1$ is the "constant gain" as it governs the responsiveness of parameter updating to recent forecast errors. While the parameter Γ controls the direction of drift in the parameters. In our empirical application we adopt the standard stochastic gradient with $\Gamma = I$.[13] It is evident from the stochastic gradient algorithm that it nests the RPE. Equation (18) is a simple recursive estimator of the mean-squared forecast errors where past errors are geometrically discounted at rate $(1-\gamma_2)$. The learning gains, γ_1, γ_2, are key objects in the estimation as they control the relative speed of coefficient updating and model selection. If these estimated gain coefficients

[12] Another commonly employed version of econometric learning is based on a constant gain least-squares algorithm. This algorithm looks similar but replaces the parameter Γ with a recursively estimated covariance matrix for the regressors. With the particle filter, this increases the computational cost by several orders of magnitude.

[13] As in Sargent and Williams (2005), it is an interesting topic for future research to consider the role different priors about the direction of drift impacts time-varying Ricardian beliefs.

are significantly different, the dynamical system can be thought of as a "fast-slow" system, a feature that is studied extensively in the literature on large deviations.[14] The idea is that slow moving variables can generate large fluctuations that occur infrequently but with high probability. The estimates presented below find that model selection occurs relatively fast and coefficients evolve relatively slow. 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 $t-1$.

4.2 Empirical methodology

After plugging in the policy rules, expectations, and recursive updating equations for the learning rules, the model can be written in non-linear state space form:

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

where the state vector is

$$X_t' = \left(b_{t+1}, \pi_t, y_t, v_t, s_t, g_t, u_t, w_t, z_t, n_t, MSE_{st}, MSE_{bt}, \mathrm{vec}\left(\psi_t^s\right), \mathrm{vec}\left(\psi_t^b\right)\right),$$

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

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

and the parameter vector is

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

The measurement and state disturbances are η_t, ν_t respectively. Our sample for the observed variables is 1960.1-2007.3.[15] We measure y_t as the log difference between chain-weighted real GDP and the CBO's measure of potential output. We measure π_t from the PCE index. We compute b_t and s_t as the debt-to-GDP ratio and primary surplus-to-GDP ratio, respectively. To remain consistent with the model, all variables are measured as deviations from mean.

The goal of the empirical exercise is to identify reasonable values for the parameters in

[14] See, for instance, Bouchet et al. (2016).

[15] 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.

Θ in order to explain the U.S. economy over the sample period. We are especially interested in inferences on the latent state variable, n_t, measuring the extent of (non-)Ricardian beliefs over the period. To this end, we adopt Bayesian methods and use a simulation-based technique called the particle filter to approximate the likelihood function $p(Y_t|\Theta)$. The endogenous predictor selection and learning renders the state-space non-linear making analytic calculation of the likelihood intractable. In place of a Kalman Filter, we adopt the Bootstrap particle filter as described in Herbst and Schorfheide (2015). The particle filter is a way to produce recursive approximations to the distribution of the latent state variables X_t. Our algorithm samples from the posterior distribution through an adapted Random-Walk Metropolis Hastings (RWMH) MCMC technique, with the particle filter based estimate of the likelihood function. A detailed discussion of the MCMC-particle filter algorithm and its implementation is in Appendix E.

4.3 Parameter estimates

Table 1 reports the means, as well as the fifth and ninety-fifth percentiles, of the marginal posterior distributions of the parameters. In the estimation, we fixed the parameters $\beta = 0.99, s_b = 0.30$, and $\sigma = 2.00$. The parameter estimates are mostly in line with previous estimates in the literature, with a few notable exceptions described below.

We begin by briefly discussing the key parameters associated with endogenously (non-)Ricardian beliefs. The "intensity of choice" parameter, ω, has a mean estimate of 6.301. This is close to the value of 5.04 that was estimated by Cornea-Madeira et al. (2019) from a benchmark New Keynesian model with heterogeneous expectations. The speed of adjustment in estimating the relative forecasting accuracies of the two models, γ_2, has a mean estimate of 0.091. This implies that model selection is based on an average of past forecast errors with a geometric decay of 0.909. The interpretation of this value is the effective memory size for model selection is approximately 11 quarters. The 5/95% intervals imply a range of memory from 9.7-12.7 quarters. The mean estimated gain for coefficient updating, γ_1 is 0.002. This value is on the lower end of previously reported estimates, though the same value is used in Eusepi and Preston (2011). For instance, Eusepi and Preston (2018) report a value of 0.035, closer to the high end, but for a different learning model and without the forecast model selection. The relative values for γ_1, γ_2 imply that there is a fast-slow learning dynamic implied by these estimates. Model selection takes place at a higher speed, in response to past data, than coefficient estimation. This implies that after a model switch, coefficient estimates will adjust slowly as agents acquire more data in the regime and actual forecasts will lag model selection.

The estimated Taylor-rule policy parameters are in line with previous estimates (e.g. Eusepi and Preston, 2018; Del Negro et al., 2015; Justiniano et al., 2011) and suggest a strong active monetary policy. While the estimated reaction coefficients are a bit higher than these earlier papers, the coefficients are not as large as what is reported in the active monetary/passive fiscal regime in Bianchi and Ilut (2017). The fiscal policy reaction coefficient is close to the mean value reported in Eusepi and Preston (2018). The estimated shock processes are also similar to existing estimates. One difference is that the estimated process for the aggregate demand shock is less volatile and persistent than typical estimates. However, that is because we include serially correlated monetary policy shocks, whose innovations have the highest variance among the set of shocks considered. The estimated fiscal policy shocks are also the most persistent.

Two parameters whose estimates differ somewhat from many estimates in the literature involve the slope of the Phillips curve and the degree of price rigidity in the economy. For the latter, our estimate is that prices are updated, on average, every 2.4 quarters, more frequently than the typical range of 4-7 quarters. This is not an unexpected finding because the departure from full rational expectations often implies a lower estimated degree of nominal rigidities. Finally, the slope of the Phillips curve is on the high side, and considerably higher than in Eusepi and Preston (2018).

We can assess the empirical relevance of the non-Ricardian belief mechanism via the Bayes factor comparing our model to a version that fixes the fraction n at the temporary Ricardian equilibrium, i.e. $n_t = 0, 1$ in all periods. When we make this comparison we find that the difference in log marginal likelihoods is in favor of the model with endogenously (non-)Ricardian beliefs and whose difference totals 5.7180 when $n = 0$ and 7.026 when $n = 1$. The Bayes factor takes the product of the ratios of the marginal likelihoods and the prior probabilities of each of the two models. The empirical evidence in favor of the model with non-Ricardian beliefs is substantial in the sense that the data will prefer that model over the Ricardian or fully non-Ricardian versions for any prior ratio below 304.29 and 1,125.40, respectively. Thus, a researcher would have to be *a priori* more confident in a rational expectations, i.e. self-confirming expectations, model by a factor of over 300 in order to reject the model with non-Ricardian beliefs.[16]

[16] Of course, this model comparison is just on beliefs within the environment under consideration here. One might have a subjective prior for a different environment and need to compute the associated marginal likelihood. Recall from earlier, also, that the $n = 0$ RPE is observationally equivalent to rational expectations.

4.4 Ricardian beliefs and the U.S. economy

This section compares model-computed state dynamics to U.S. data and interprets the results via impulse response functions to a contractionary fiscal policy shock. We begin by estimating the (one step ahead) predicted state path $E\left(X_{t+1}|Y_t, \Theta\right)$ via the particle filter. Figure 3 plots the results. In panels (a)-(e) the solid lines are the median model-implied paths and the dashed lines are the corresponding data series. The shaded areas represent the 60% credible interval. Finally, panel (f) overlays the estimated fraction of agents estimated to hold non-Ricardian beliefs (right axis) with the model-implied inflation path (left axis).

The first row of panels in Figure 3 demonstrate that the model captures well the sample paths of the output gap and inflation. The model implied output gap is slightly more volatile than the data. This is particularly evident with the model predicting a, roughly, 10% deeper recession in 1982. In panel (b) it is evident that the model-implied path for inflation is very close to the data, though over-predicts peak inflation slightly. Panels (c) and (d) similarly show a close alignment between model-implied paths and the data for the real government debt and the real primary surplus. In particular, the model captures that the stance of fiscal policy turns expansionary during the decades of the 1970's and 1980's and turns contractionary during the mid-1990's.

The bottom row of panels, (e) and (f), illustrate our estimates for non-Ricardian beliefs and how they correlate with U.S. inflation. Over the sample period, we estimate a fraction of individuals and firms forming forecasts from the s-model to be 0.33. There is substantial volatility, though, with a standard deviation in the fraction n_t of 0.14. Qualitatively, the period between the late 1970's and early 1980's can broadly be described as Ricardian, with a fraction $n \approx 0$. From the mid-1960's and mid 1980's there are prolonged periods where the fraction with non-Ricardian beliefs increases over time, peaking at nearly 60% around 1993.

However, as explained in the previous section, the fast-slow learning dynamic implies that actual forecasts will adjust with a lag as model coefficients are updated slowly following a significant regime-switch like during the late 1970's-early 1980's and the late 1980's. Therefore, in panel (f) we can see that the rise in inflation during the 1970's arises, in part, as fiscal policy turns expansionary and there is an increasing fraction of agents with non-Ricardian beliefs. These beliefs have inertia that helps contribute to the further rise in inflation throughout the 1970's. There is an abrupt endogenous increase in the fraction of agents forecasting with the Ricardian model, and as actual forecasts evolve with new data, agents are not incorporating the expansionary fiscal policies into their beliefs which helps contribute to the collapse of inflation. Conversely, the mid 1980's through 1990's are characterized by a greater fraction adopting the non-Ricardian model, while fiscal policy also shifts its stance towards more contractionary policy. Here the non-Ricardian beliefs do not manifest into significantly higher inflation. The non-Ricardian beliefs help reinforce the stabilizing effects on demand from the policies and anchor inflation during the 1990's. The inertia in beliefs, and how the fiscal policy shocks manifest differently during the 1990's, is reminiscent of findings in Bianchi and Ilut (2017).

We can also gain greater insights into the estimated mechanism by examining impulse responses to a contractionary fiscal policy shock. The impulse response function is non-linear, depending on the fraction of agents forecasting with the s-model n, as well as the recursively updated model coefficients, and sequence of shocks. To present these impulse responses we proceed by fixing n at a subset of the values plotted in Figure 3 and fixing the resulting coefficients at their RPE values. These results are plotted in Figure 4. Each (color-coded) line represents a different value for n, with the blue line representing $n = 0$, or the fully Ricardian SCE.

Evidently from the top two panels, (a) and (b), the estimated effect of an unanticipated contractionary fiscal policy shock depends on the degree of non-Ricardian beliefs at the time. When $n = 0$, the fiscal policy policy shocks have essentially no impact on the output gap or inflation.[17] Similarly, when more than half of the agents have non-Ricardian beliefs ($n \geq 0.5$) then there is a stronger economic reaction to the fiscal policy shock with both inflation and the output gap dropping. One can understand how expectations impact this finding by looking at the bottom two rows of Figure 4. The middle rows, (c) and (d), plot the expectations of inflation by agents with the non-Ricardian and Ricardian forecast models, respectively. The bottom row plots these same cross-type expectations for the continuation value of financial and non-financial wealth v_t. When the model is in a Ricardian equilibrium ($n = 0$) there is no impact on either agent-type's expectations of future inflation, and as seen in panel (b), these beliefs are self-confirming. However, the bottom panels show that the fiscal policy shock does impact expectations about future wealth as the change in the real primary surplus impacts expectations about future after-tax income. For agents with fully Ricardian beliefs (panel (f)), the unanticipated increase in the primary surplus leads them to expect higher future disposable income. However, in economies that are non-Ricardian the presence of the non-Ricardian agents lead to non-Ricardian effects, through the lower output gap, that mediates those expectations for the agents in panel (f). For the non-Ricardian expectations-type, the positive surplus shock leads them to always expect lower continuation wealth in the near

[17] Recall the earlier weak Ricardian equivalence result. In these figures there are small, transitory impacts on output and inflation when $n = 0$.

term. For all agents, as the economy recovers and the agents understand that the increase in surplus leads to an increase in disposable income, their expected continuation values v increase before returning to steady-state.

These impulse responses also help partially explain how the model generates the rise of inflation during the 1970's and the anchoring of expectations during the 1990's. During the 1970's, the primary surplus shifts towards a sustained period of deficits. During the 1970's the increasing fraction of agents forecasting based on those surpluses helped generate inflation, along with other, but especially mark-up, shocks. During the 1980's there were larger primary deficits but the switch toward Ricardian beliefs, and the inertia in the forecasting rules, mitigated those effects from impacting inflation. Conversely, during the 1990's fiscal policy turned toward sustained primary surpluses and an increasing fraction forecasting with the surplus model helped to reduce and anchor inflation expectations.[18]

4.5 What drives non-Ricardian beliefs?

This section uses counterfactual analysis to provide a deeper understanding into which shocks are important drivers of evolving (non-)Ricardian beliefs. Of course, beliefs are endogenous and depend on the particular history of exogenous shocks as well as the cross-equation restrictions with the endogenous variables. To disentangle these effects we consider alternate economies with different histories of shocks. The counterfactuals constructed hold the parameters fixed at their mean values and re-estimate the mean model-implied paths shutting down one or more of the exogenous shocks while holding the other shocks to their originally estimated paths. Table 2a presents the counterfactual results. Each row identifies a particular shock combination and lists the mean and standard deviation for inflation and the predictor proportions.

Fiscal policy and mark-up shocks are the two most important drivers of volatility in beliefs. Fiscal policy shocks, in particular, are part of 7 of the 10 most volatile n counterfactual shock combinations. While the mark-up shocks are in the top 4 scenarios and 6 of the top 10. In those counterfactuals where there is low volatility in beliefs, then the fraction n tends to fluctuate around 0.5.

Figure 5 presents the first 3 rows of Table 2a graphically as a time-series comparison. The solid line is the estimated path for the fraction using the surplus model, the dashed line is the counterfactual with only mark-up and fiscal policy shocks, and the dotted line

[18] Appendix F presents further evidence, external to the model, from the Survey of Professional Forecasters, that forecasts are consistent with a non-Ricardian regime in the 1990's.

consists of the mark-up, monetary, and fiscal policy shocks. The counterfactuals suggest that the initial run up in non-Ricardian beliefs, and the resulting inflation, were primarily from mark-up and fiscal policy shocks. By the early 1980's the economy switched to a Ricardian equilibrium, the counterfactual suggests without the monetary shocks the economy would have remained longer in a non-Ricardian regime. The period from the mid-1980's through mid-1990's, conversely, were partly driven by a period of increasingly expansionary fiscal policy shocks.

We use two other counterfactuals to disentangle the role played by real-time learning. The first counterfactual, plotted in panels (a)-(b) of Figure 6 fixes the shocks and the parameters at their mean values except setting $\gamma_1 = 0.03$. This counterfactual undoes the estimated fast-slow learning dynamic and considers the implications to the output gap and inflation if learning is fast. The right panel shows that inflation volatility is counterfactually high. Inflation is predicted in this alternate economy to be greater during the 1960's, with vast over-shooting during the 1970's: the counterfactual implication is a dramatic collapse from the 1970's inflation. The counterfactual scenario also predicts that faster learning would have led to a second high inflation episode during the late 1990's.

Figure 6, panels (c)-(d), constructs a counterfactual scenario that shuts down learning entirely. This figure was constructed by holding the parameters and shocks fixed at their estimated mean values, allowing for n to evolve endogenously, but for each value of n_t setting the model coefficients equal to their RPE values. The left panel plots the counterfactual fraction with non-Ricardian beliefs while the right panel is the corresponding inflation path. Clearly, in this scenario, beliefs tend to be more Ricardian but with more volatility in the later sample periods. The counterfactual scenario substantially under-predicts inflation during the 1970's and early 1980's and over-predicts during the 1990's. The model with no learning does well to explain inflation during the late 1980's and 2000's.

4.6 Policy change and Ricardian beliefs

So far, the results presented impose policy rules with time and state-invariant coefficients. There is, however, evidence that the stance of monetary and fiscal policy evolves over time. Davig and Leeper (2006), Bianchi (2013), and Bianchi and Ilut (2017) identify regime-switching policy rules that alternate between active/passive and passive/active monetary/fiscal policies. One finding is that the high inflation of the 1970's is characterized by passive monetary and active fiscal policy. The volatile latent monetary policy shocks we estimate indicate that there may be omitted factors from the policy rule specification and time-varying reaction coefficients is one possibility.

A full estimation with regime-switching policy and beliefs is beyond the scope of this paper. However, we now present rolling-window estimates of time-varying non-Ricardian beliefs even with time-varying monetary and fiscal policies. To make the analysis computationally feasible we shut down the endogenous selection of beliefs and estimate the fraction n as a structural parameter.[19] We consider rolling 20-year windows and use Bayesian techniques to estimate the structural parameters, including n, over each rolling sample, advancing the window by 8 quarters.[20] We then can get a rough estimate of how policy and beliefs evolved over time.

The top panels in Figure 7 plot the rolling window estimates for the monetary policy rule reaction coefficients that align with previous findings. The period ending in 1981 features a significantly less active monetary policy estimated to be at the lower bound of what would satisfy the "Taylor principle." In subsequent windows it is estimated to be near 2 but declines during the later 1990's. The reaction coefficient to the output gap is estimated to be below 0.10 for most windows. Panel (d) plots the fiscal policy reaction coefficient ϕ_b and estimates increases through the mid-1980's before declining during later windows. This decline corresponds to a relatively more active fiscal policy.

We now take a brief detour to better understand the estimated beliefs in panel (c). It is instructive to understand how a change in the monetary policy rule via ϕ_π or the fiscal policy rule ϕ_b would impact the endogenous model selection. Panels (e) and (f) plot the T-maps computed with the estimates in Table 1, but for a range of ϕ_π. The left panel (e) is for a large intensity of choice $\omega = 100$ and the right panel is for the estimated $\omega = 6.301$. The T-map exhibits multiple equilibria in panel (e) and a unique equilibrium in (f). In the estimated version of the model, changes in the monetary policy rule have a very modest impact on the equilibrium fraction of non-Ricardian beliefs. In panel (e), since the $n = 0$ and $n = 1$ equilibria are stable under learning, the interior equilibrium marks the boundary to the basins of attraction for the $n = 0$ and $n = 1$ equilibria. A more active monetary policy rule shrinks the basin of attraction for the non-Ricardian equilibria. A similar comparative static arises for ϕ_y. Recall from Theorem 1, however, that the comparative effect from ϕ_b is non-monotonic. A more aggressive monetary policy response tends to make the economy more Ricardian and a moderately more aggressive fiscal policy, strengthening the correlation between the primary surplus and endogenous state variables, tends to make the economy more non-Ricardian.

[19] By shutting down the real-time predictor selection and focusing on the RPE in each window we can analytically evaluate the likelihood function with the Kalman Filter.

[20] As in, for example, Del Negro and Schorfheide (2006).

However, a much more aggressive fiscal policy makes beliefs Ricardian.

Now let's return to panel (c) of Figure 7. First, notice that the average degree of non-Ricardian beliefs is significantly higher than in the benchmark estimation. Of course, the model is different and without learning there are fewer natural state variables, so this specification leads the data to prefer a higher fraction using the s-model. Second, over the windows estimated with declining ϕ_π and declining ϕ_b there is a downward trend in n, as anticipated by panel (f) of Figure 7. During the 1970's the policy coefficients are moving in ways that have an ambiguous theoretical impact on n. The estimates, though, are consistent with a story where monetary policy is less active for the beginning of the great inflation period and then increasingly non-Ricardian beliefs play a subsequent role in maintaining the high inflation. These results are complementary to existing estimates, e.g. Bianchi (2013), who finds that the high inflation of the 1970's was the result of less active monetary policy and non-Ricardian beliefs/policy.

Finally, we conclude this section with policy implications of non-Ricardian beliefs. In particular, is it the case that non-Ricardian beliefs are associated with economic instability? The answer to this question depends on the policy stance that led to more non-Ricardian beliefs and how that impacts the endogenous response of inflation and the output gap. To illustrate, Table 2b reports results from policy counterfactuals. We fix the exogenous shocks at their estimated paths, consider the alternative monetary/fiscal policy stances in the table, and then calculate the counterfactual paths for the economy. The table reports on the average fraction n and the standard deviations of inflation and the fraction of non-Ricardian beliefs. The first row is for a counterfactual where both monetary and fiscal policy are near the boundary of the active/passive monetary/fiscal region. Here the average value for n is low and near a Ricardian equilibrium, but with high volatility in n and inflation. The smaller value for ϕ_b dominates the impact on n making the average fraction of non-Ricardian beliefs small. But, because monetary policy is not very aggressive against inflation there is high inflation volatility. In the second row, fiscal policy's ϕ_b is fixed at its estimated mean value while $\phi_\pi = 1.05$. In this case, there is a slight increase in the extent of Ricardian beliefs and a higher inflation volatility, but not to the same extent as the first counterfactual. Moving the other way, with more aggressive monetary and fiscal policies there is a higher average value for n, but with lower volatility in beliefs and inflation. Since beliefs are endogenous to policy, one cannot simply say that non-Ricardian beliefs are destabilizing. We saw this as well in Figure 3 where the economy was becoming non-Ricardian and inflation was becoming anchored. One implication is that even within an active/passive policy regime, there is scope for policy coordination.

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). The first to characterize fiscal and monetary policy interaction under adaptive learning are Evans and Honkapohja (2007) and Eusepi and Preston (2012). Gasteiger (2018) directly extends these frameworks to include heterogeneous expectations. Evans et al. (2012) examine the conditions under which Ricardian equivalence holds or fails under adaptive learning. The theory of restricted perceptions is related to a wide variety of applications of misspecified models, e.g. Sargent (1999), Adam (2005), Branch and Evans (2006), Sargent (2008), Branch and McGough (2018), and Cho and Kasa (2015). This paper builds on an insight from Woodford (2013) where an example of a RPE is considered that leads to a failure of Ricardian equivalence. In short, our paper extends the theory of forecast misspecification in Branch and Evans (2006) into the Eusepi and Preston (2018) environment with fiscal and monetary policy interaction and generalizing the beliefs in Woodford (2013).

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 explains 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. 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. 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 shifts. We explored this point by taking rolling window estimates of the policy coefficients and non-Ricardian beliefs. We found that the estimated fraction of non-Ricardian beliefs evolves along with policy rules in a way predicted by theory. While beyond the scope of the present study, we leave open quantifying policy-regime change versus belief-regime change.

Our theory builds on Eusepi and Preston (2018) 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.

Finally, we briefly discuss what is important in the mechanism here that is not in the Fiscal Theory of the Price Level (FTPL). In both the FTPL and imperfect knowledge models, the debt valuation channel matters for public debt through the government's budget constraint. The FTPL is built on that constraint serving as an equilibrium solvency conition. Here the government's fiscal policy guarantees that its budget constraint is satisfied ex-post, and so the constraint acts like a state transition equation. What is key here is less about the policy and more about how agents forecast future prices and wealth.

6. Conclusion

This paper provides evidence that endogenously (non-)Ricardian beliefs play a role in U.S. inflation dynamics. 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 (2018). We follow Woodford (2013) and give agents restricted perceptions by restricting forecast models to 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 both state variables. Despite the misspecification, in equilibrium agents' beliefs are optimal within the restricted class. The set of forecast models are a natural formalization of endogenous (non-)Ricardian beliefs. When the distribution of beliefs is determined endogenously, there can simultaneously exist (non-)Ricardian equilibria even though the monetary-fiscal regime is Ricardian in the Leeper-sense.

We estimate the model, with endogenously (non-)Ricardian forecasting models chosen in real-time by individuals and firms. The model is non-linear and we formulate a Bayesian MCMC estimator where the likelihood function is approximated by a bootstrap particle filter. We estimate a time-varying fraction of agents that use a non-Ricardian model and compare the model fit to the fully Ricardian and non-Ricardian versions of the model. We find that the data substantially prefer time-varying non-Ricardian beliefs. Our estimates imply that the late 1960's-1970's are characterized by a regime with non-Ricardian beliefs, corresponding to the period of high inflation in the U.S. The early 1980's, though, are consistent with a

Ricardian regime and the stabilization of inflation. Surprisingly, we find that the late 1980's and 1990's also exhibit a high degree of non-Ricardian beliefs. This period, though, did not manifest in higher inflation as fiscal policy moved to running primary surpluses which helped to anchor private-sector expectations.

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.

Bouchet, F., Grafke, T., Tangarife, T., and Vanden-Ejinden, E. (2016). Large Deviations in Fast-Slow Systems. *Journal of Statistical Physics*, 1062:793–812.

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

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*, 107(11):3589–3616.

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

Cornea-Madeira, A., Hommes, C. H., and Massaro, D. (2019). Behavioral Heterogeneity in U.S. Inflation Dynamics. *Journal of Business and Economic Statistics*, 37(2):288–300.

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

Del Negro, M., Giannoni, M. P., and Schorfheide, F. (2015). Inflation in the Great Recession and New Keynesian Models. *American Economic Journal: Macroeconomics*, 7(1):168–196.

Del Negro, M. and Schorfheide, F. (2006). How Good is What You've Got? DSGE-VAR as a Toolkit for Evaluating DSGE Models. *Federal Reserve Bank of Atlanta Economic Review*, 91(2):21–37.

Eusepi, S. and Preston, B. (2011). Expectations, Learning, and Business Cycle Fluctuations. *American Economic Review*, 101(6):2844–2872.

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. (2018). Fiscal Foundations of Inflation: Imperfect Knowledge. *American Economic Review*, 108(9):2551–2589.

Evans, G. W. and Honkapohja, S. (2001). *Learning and Expectations in Macroeconomics*. Princeton University Press, Princeton, New Jersey.

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.

Evans, G. W., Honkapohja, S., and Williams, N. (2010). Generalized Stochastic Gradient Learning. *International Economic Review*, 51(1):237–262.

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.

Herbst, E. P. and Schorfheide, F. (2015). *Bayesian Estimation of DSGE Models*. Princeton University Press.

Justiniano, A., Primiceri, G. E., and Tambalotti, A. (2011). Investment Shocks and the Relative Price of Investment. *Review of Economic Dynamics*, 14(1):102–121.

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

Marcet, A. and Sargent, T. J. (1989). Convergence of Least-Squares Learning Mechanisms in Self-Referential Linear Stochastic Models. *Journal of Economic Theory*, 48(2):337–368.

Milani, F. (2007). Expectations, Learning and Macroeconomic Persistence. *Journal of Monetary Economics*, 54(7):2065–2082.

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.

Sargent, T. J. and Williams, N. (2005). Impacts of Priors on Convergence and Escapes from Nash Inflation. *Review of Economic Dynamics*, 8(2):360–391.

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. (2019). Escape Dynamics in Learning Models. *Review of Economic Studies*, 86(2):882–912.

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.

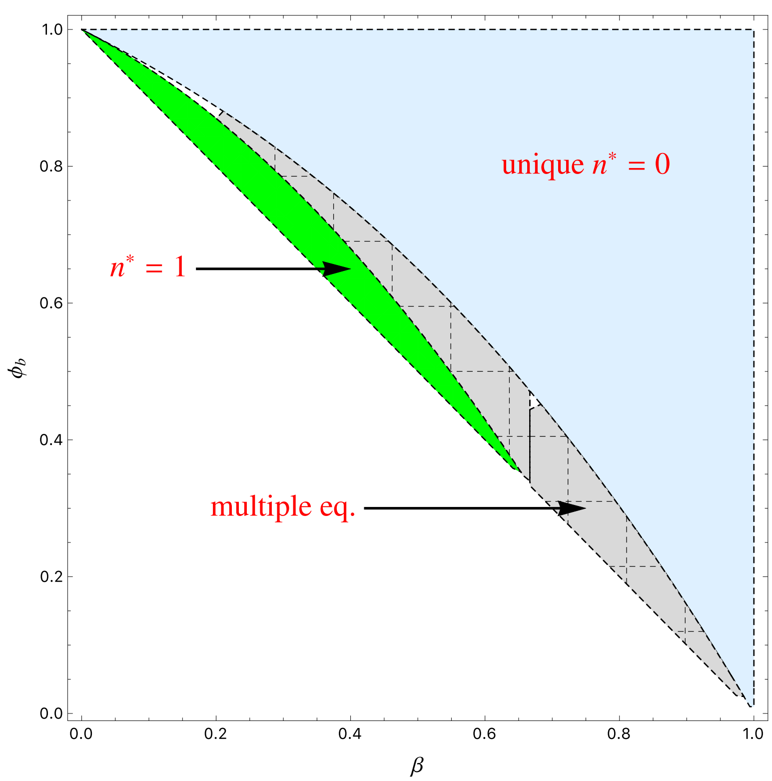


Figure 1: Equilibrium existence

Table 1: Posterior distribution of parameters

		90% credible region	
Parameter	Mean	5%	95%
Structural parameters			
κ	0.392	0.351	0.433
α	0.584	0.538	0.630
Policy parameters			
ϕ_π	1.854	1.665	2.042
ϕ_y	0.119	0.105	0.132
ϕ_b	0.062	0.034	0.091
Exogenous shocks			
ρ_g	0.050	0.007	0.093
ρ_u	0.343	0.296	0.389
ρ_w	0.305	0.262	0.347
ρ_z	0.860	0.814	0.907
$100\sigma_g$	0.631	0.429	0.833
$100\sigma_u$	0.141	0.002	0.281
$100\sigma_w$	1.440	1.359	1.526
$100\sigma_z$	0.235	0.117	0.353
Learning parameters			
ω	6.301	4.428	8.174
γ_1	0.002	0.000	0.003
γ_2	0.091	0.079	0.103

Table 2: Counterfactuals.

Shocks	$E\pi$	σ_π	En	σ_n
g-u-w-z	1.95	3.40	0.33	0.14
u-w-z	1.97	2.14	0.42	0.12
u-z	1.85	2.43	0.44	0.10
g-u-z	1.70	2.11	0.43	0.07
g-w-z	1.93	3.56	0.49	0.07
g-u-w	1.60	2.88	0.42	0.06
g-w	1.87	3.59	0.46	0.04
w-z	2.05	1.65	0.48	0.04
g-z	1.90	1.84	0.49	0.033
u-w	1.68	1.77	0.46	0.032
w	1.97	1.55	0.49	0.014
z	2.04	0.29	0.50	0.003
g-u	1.64	1.88	0.50	0.0008
u	1.74	2.19	0.50	0.0005
g	1.91	1.79	0.50	0.0003

(a) Counterfactual moments under different shock combination scenarios.

(ϕ_π, ϕ_b)	σ_π	En	σ_n
(1.05,0.015)	20.3957	0.10	0.157
(1.05, 0.062)	5.8097	0.34	0.155
(1.854,0.062)	3.40	0.33	0.141
(2.5,0.062)	2.94	0.31	0.125
(2.5,0.20)	2.62	0.45	0.051

(b) Policy counterfactuals.

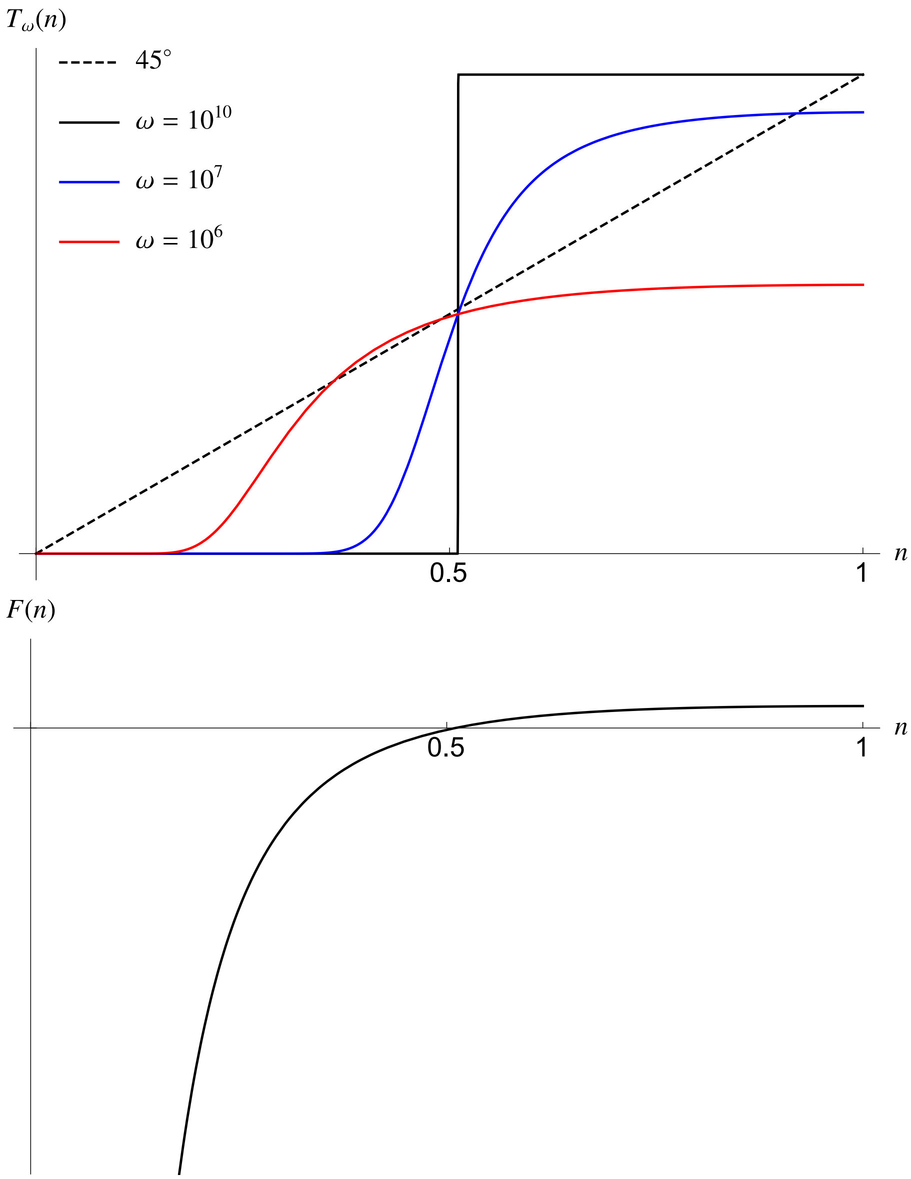
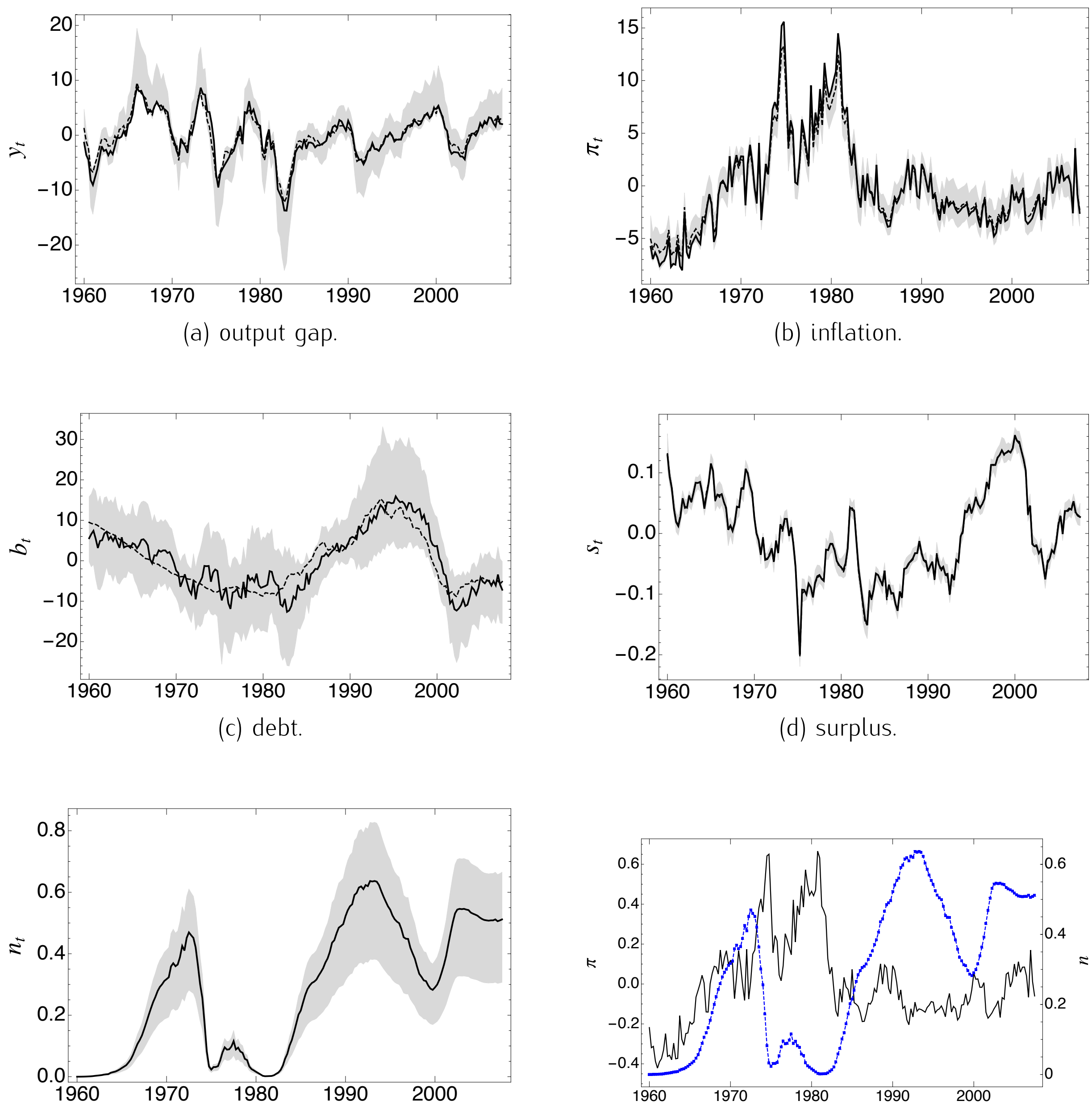


Figure 2: Multiple (non-)Ricardian equilibria in the special case. The top panel plots the T-map for various values of the intensity of choice ω.



(a) output gap.

(b) inflation.

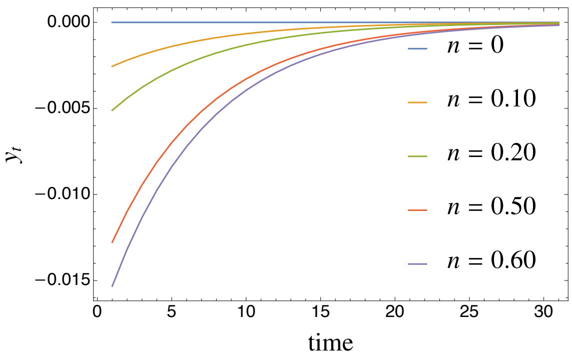
(c) debt.

(d) surplus.

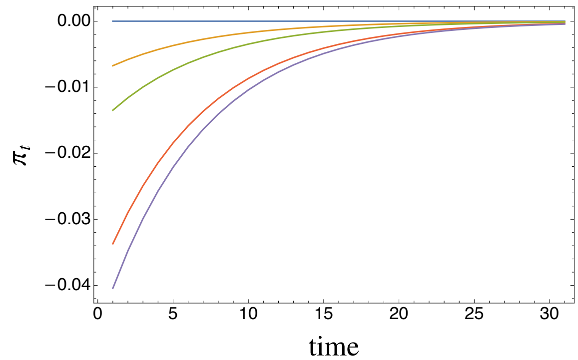
(e) fraction w/ non-Ricardian beliefs.

(f) inflation and non-Ricardian beliefs.

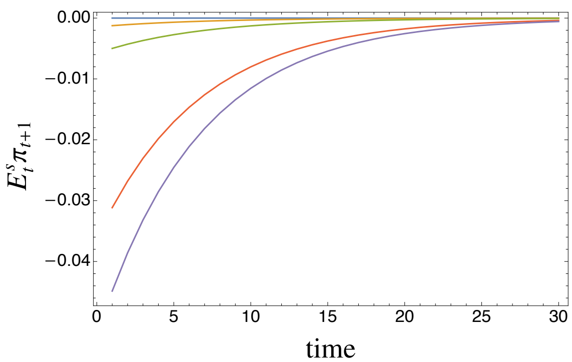
Figure 3: State dynamics. Solid line is the one step-ahead particle filtered state estimates, dashed line is the data.



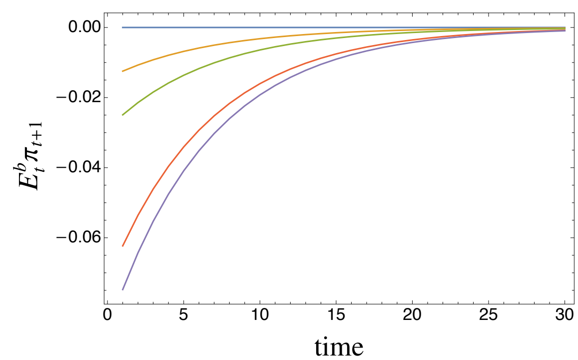
(a) output gap.



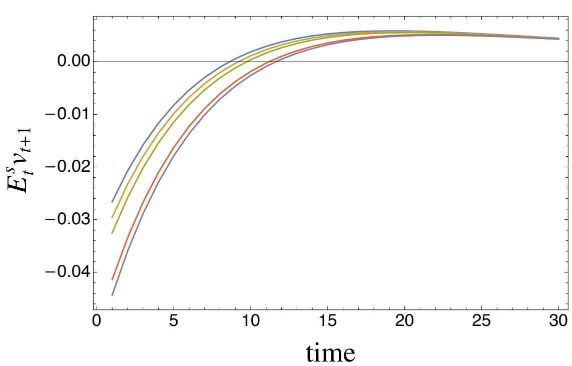
(b) inflation.



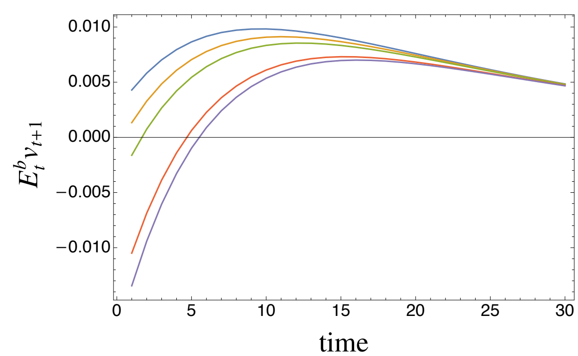
(c) expected inflation $E_t^s \pi_{t+1}$.



(d) expected inflation $E_t^b \pi_{t+1}$.



(e) expected continuation value $E_t^s v_{t+1}$.



(f) expected continuation value $E_t^b v_{t+1}$.

Figure 4: Impulse responses to a 1-unit contractionary fiscal policy shock. Each impulse response is computed for a different fraction of non-Ricardian beliefs, n, holding beliefs fixed at their RPE values.

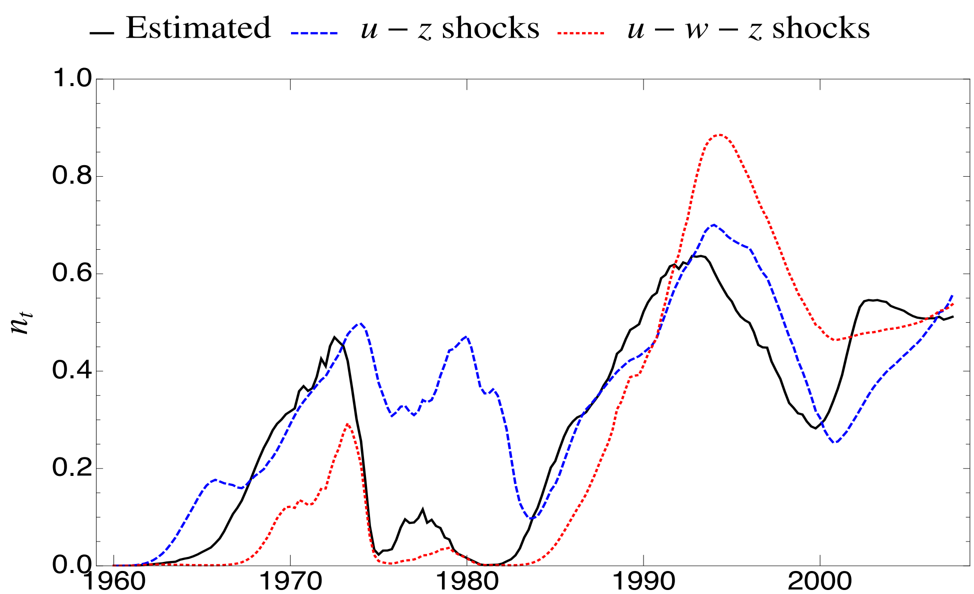
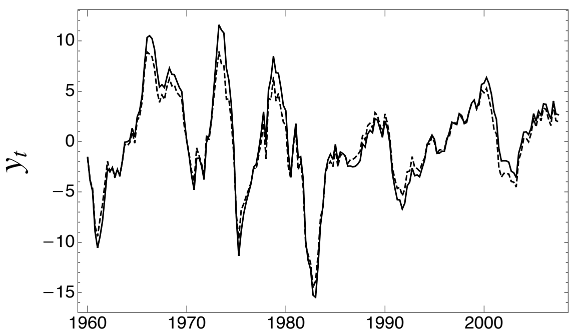
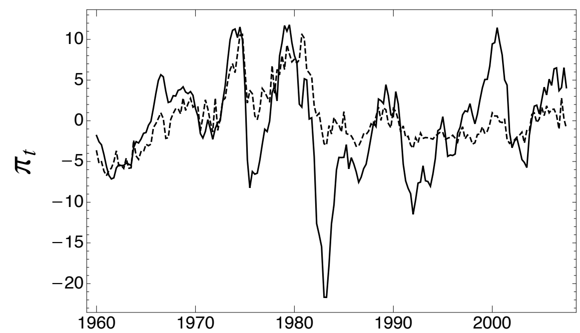


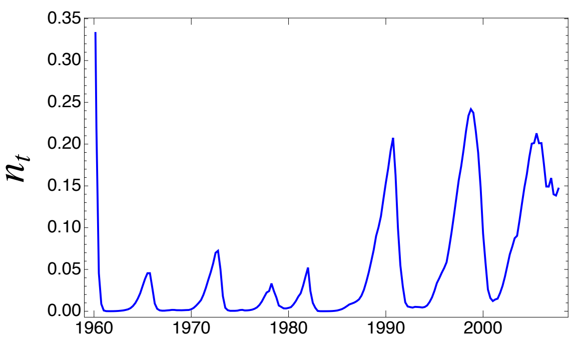
Figure 5: Comparison of estimated non-Ricardian beliefs to counterfactual shock scenarios.



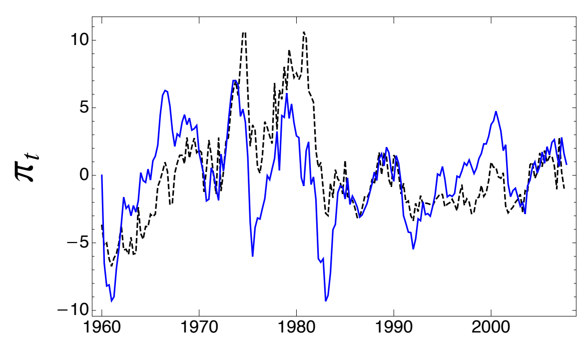
(a) fast learning ($\gamma = 0.03$): output gap.



(b) fast learning ($\gamma = 0.03$): inflation.

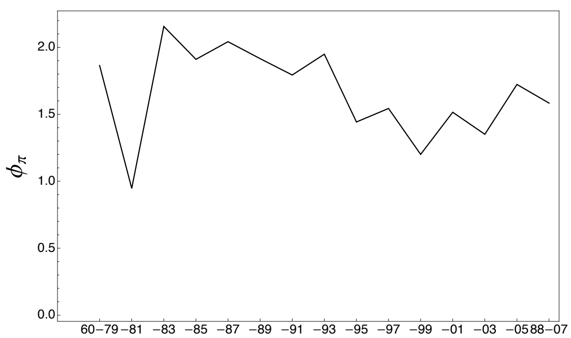


(c) no-learning: fraction non-Ricardian.

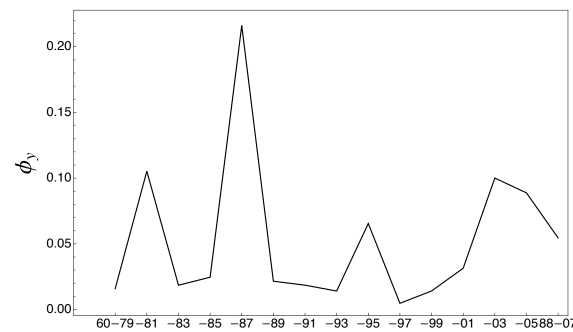


(d) no-learning: inflation.

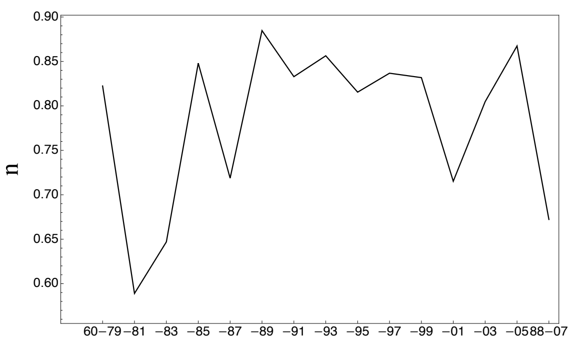
Figure 6: Counterfactuals. Panels (a)-(b) from counterfactual with higher learning gain $\gamma = 0.03$. Panels (c) and (d) from counterfactual with no learning. Solid lines are the counterfactual paths, dashed lines are the estimated paths.



(a) policy coefficient ϕ_π.

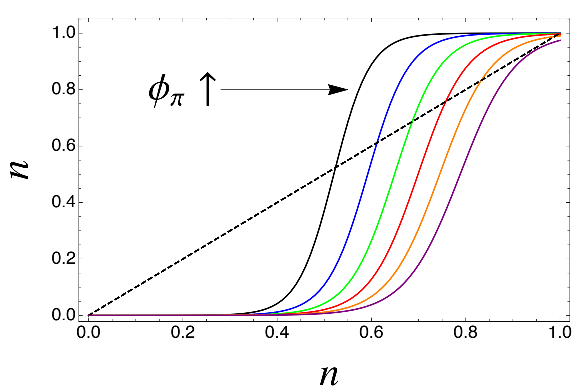


(b) policy coefficient ϕ_y.

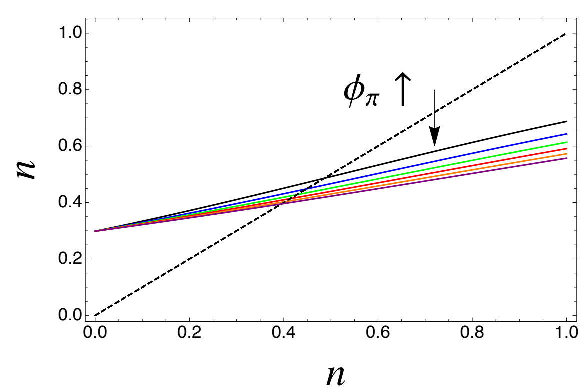


(c) fraction w/ non-Ricardian beliefs.

(d) policy coefficient ϕ_b.



(e) $\omega = 100$.



(f) $\omega = 6.31$.

Figure 7: Rolling window estimation results. Panels (a)–(d) coefficient estimates. Panels (e)–(f) corresponding T-maps with $\omega = 100$ and $\omega = 6.03$, resp.

A. Methodological framework

A.1 Computation of the restricted perceptions equilibrium

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

$$y_t = \delta_0 b_{t+1} + \delta_1 b_t + \delta_2 s_t + \delta_3 u_t \tag{A.1.1}$$
$$b_{t+1} = \xi_1 b_t + \xi_2 s_t + \xi_3 u_t. \tag{A.1.2}$$

Moreover, we can aggregate (D.12) and combine it with (4), (D.15), (A.1.1) and (A.1.2) to obtain

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

and (D.15), (3), (A.1.1) and (A.1.2) imply that

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

We combine (A.1.3) and (A.1.4) to

$$\mathbf{z}_t = \boldsymbol{\mu}_b b_t + \boldsymbol{\mu}_s s_t + \boldsymbol{\mu}_u u_t, \tag{A.1.5}$$

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 $\boldsymbol{\mu}_u \equiv (\mu_{v,3}, \mu_{p,3})'$.

Next, recall that PLMs are given by

$$\mathbf{z}_t = \psi^s s_{t-1} + \eta_t \tag{A.1.6}$$
$$\mathbf{z}_t = \psi^b b_{t-1} + \eta_t, \tag{A.1.7}$$

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

$$0 \stackrel{!}{=} E[s_{t-1}\eta_t] = E[s_t \eta_{t+1}] \tag{A.1.8}$$
$$0 \stackrel{!}{=} E[b_{t-1}\eta_t] = E[b_t \eta_{t+1}]. \tag{A.1.9}$$

Now, plug the PLM (A.1.6) and ALM (A.1.5) into (A.1.8), i.e.,

$$0 \stackrel{!}{=} E[s_t \eta_{t+1}] = E[s_t(\mathbf{z}_{t+1} - \psi^s s_t)] \tag{A.1.10}$$

$$\Leftrightarrow \psi^s E[s_t^2] = E[s_t \mathbf{z}_{t+1}]. \tag{A.1.11}$$

Equation by equation, we obtain

$$\Leftrightarrow \psi_v^s E[s_t^2] = E\left[s_t\left(\mu_{v,1} b_{t+1} + \mu_{v,2} s_{t+1} + \mu_{v,3} u_{t+1}\right)\right] \tag{A.1.12}$$

$$\psi_v^s E[s_t^2] = \mu_{v,1} E\left[s_t b_{t+1}\right] + \mu_{v,2} E\left[s_t s_{t+1}\right] + \mu_{v,3} E\left[s_t u_{t+1}\right] \tag{A.1.13}$$

$$\Leftrightarrow \psi_v^s = \mu_{v,1} \frac{E[s_t b_{t+1}]}{E[s_t^2]} + \mu_{v,2} \frac{E[s_t s_{t+1}]}{E[s_t^2]} + \mu_{v,3} \frac{E[s_t u_{t+1}]}{E[s_t^2]} \qquad \text{and} \tag{A.1.14}$$

$$\Leftrightarrow \psi_p^s E[s_t^2] = E\left[s_t\left(\mu_{p,1} b_{t+1} + \mu_{p,2} s_{t+1} + \mu_{p,3} u_{t+1}\right)\right] \tag{A.1.15}$$

$$\psi_p^s E[s_t^2] = \mu_{p,1} E\left[s_t b_{t+1}\right] + \mu_{p,2} E\left[s_t s_{t+1}\right] + \mu_{p,3} E\left[s_t u_{t+1}\right] \tag{A.1.16}$$

$$\Leftrightarrow \psi_p^s = \mu_{p,1} \frac{E[s_t b_{t+1}]}{E[s_t^2]} + \mu_{p,2} \frac{E[s_t s_{t+1}]}{E[s_t^2]} + \mu_{p,3} \frac{E[s_t u_{t+1}]}{E[s_t^2]}. \tag{A.1.17}$$

Likewise plug the PLM (A.1.7) and ALM (A.1.5) into (A.1.9), i.e.,

$$0 \stackrel{!}{=} E[b_t \eta_{t+1}] = E[b_t(\mathbf{z}_{t+1} - \psi^b b_t)] \tag{A.1.18}$$

$$\Leftrightarrow \psi^b E[b_t^2] = E[b_t \mathbf{z}_{t+1}]. \tag{A.1.19}$$

Again, equation by equation, we obtain

$$\Leftrightarrow \psi_v^b E[b_t^2] = E\left[b_t\left(\mu_{v,1} b_{t+1} + \mu_{v,2} s_{t+1} + \mu_{v,3} u_{t+1}\right)\right] \tag{A.1.20}$$

$$\psi_v^b E[b_t^2] = \mu_{v,1} E\left[b_t b_{t+1}\right] + \mu_{v,2} E\left[b_t s_{t+1}\right] + \mu_{v,3} E\left[b_t u_{t+1}\right] \tag{A.1.21}$$

$$\Leftrightarrow \psi_v^b = \mu_{v,1} \frac{E[b_t b_{t+1}]}{E[b_t^2]} + \mu_{v,2} \frac{E[b_t s_{t+1}]}{E[b_t^2]} + \mu_{v,3} \frac{E[b_t u_{t+1}]}{E[b_t^2]} \qquad \text{and} \tag{A.1.22}$$

$$\Leftrightarrow \psi_p^b E[b_t^2] = E\left[b_t\left(\mu_{p,1} b_{t+1} + \mu_{p,2} s_{t+1} + \mu_{p,3} u_{t+1}\right)\right] \tag{A.1.23}$$

$$\psi_p^b E[b_t^2] = \mu_{p,1} E\left[b_t b_{t+1}\right] + \mu_{p,2} E\left[b_t s_{t+1}\right] + \mu_{p,3} E\left[b_t u_{t+1}\right] \tag{A.1.24}$$

$$\Leftrightarrow \psi_p^b = \mu_{p,1} \frac{E[b_t b_{t+1}]}{E[b_t^2]} + \mu_{p,2} \frac{E[b_t s_{t+1}]}{E[b_t^2]} + \mu_{p,3} \frac{E[b_t u_{t+1}]}{E[b_t^2]}. \tag{A.1.25}$$

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.,

$$\begin{bmatrix} 1 & -\xi_2 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} b_{t+1} \\ s_t \end{bmatrix} = \begin{bmatrix} \xi_1 & 0 \\ \phi_b & 0 \end{bmatrix} \begin{bmatrix} b_t \\ s_{t-1} \end{bmatrix} + \begin{bmatrix} \xi_3 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} u_t \\ z_t \end{bmatrix} \tag{A.1.26}$$

$$\Leftrightarrow \mathcal{Y}_t = \mathbf{A}\mathcal{Y}_{t-1} + \mathbf{C}\varepsilon_t, \tag{A.1.27}$$

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

$$\mathbf{\Omega} = E[(\mathbf{A}\mathcal{Y}_{t-1} + \mathbf{C}\varepsilon_t)(\mathbf{A}\mathcal{Y}_{t-1} + \mathbf{C}\varepsilon_t)'] = \mathbf{A}E[\mathcal{Y}_{t-1}\mathcal{Y}_{t-1}']\mathbf{A}' + \mathbf{C}E[\varepsilon_t\varepsilon_t']\mathbf{C}' \tag{A.1.28}$$

$$\mathbf{\Omega} = \mathbf{A}\mathbf{\Omega}\mathbf{A}' + \mathbf{C}\mathbf{\Sigma}\mathbf{C}' \tag{A.1.29}$$

$$\Leftrightarrow \mathrm{vec}(\mathbf{\Omega}) = [\mathbf{I} - \mathbf{A}\otimes\mathbf{A}]^{-1}(\mathbf{C}\otimes\mathbf{C})\,\mathrm{vec}(\mathbf{\Sigma}) \tag{A.1.30}$$

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

$$E[\mathcal{Y}_t\mathcal{Y}_{t-1}'] = E[(\mathbf{A}\mathcal{Y}_{t-1}\mathcal{Y}_{t-1}' + \mathbf{C}\varepsilon_t\mathcal{Y}_{t-1}')] = \mathbf{A}E[\mathcal{Y}_{t-1}\mathcal{Y}_{t-1}'] = \mathbf{A}\mathbf{\Omega}. \tag{A.1.31}$$

Notice that

$$\mathbf{\Omega} = \begin{bmatrix} E[b_{t+1}^2] & E[b_{t+1}s_t] \\ E[s_tb_{t+1}] & E[s_t^2] \end{bmatrix}, \qquad \mathbf{A}\mathbf{\Omega} = \begin{bmatrix} E[b_{t+1}b_t] & E[b_{t+1}s_{t-1}] \\ E[s_tb_t] & E[s_ts_{t-1}] \end{bmatrix}. \tag{A.1.32}$$

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 $E[b_{t+1}s_t] = E[s_tb_{t+1}]$, $E[b_{t+1}b_t] = E[b_tb_{t+1}]$, $E[s_{t+1}s_t] = E[s_ts_{t+1}]$, and that $E[s_tu_{t+1}] = E[b_tu_{t+1}] = 0$. Moreover, recall that (5) implies that $E[s_ts_{t+1}] = \phi_bE[s_tb_{t+1}]$ and that $E[b_ts_{t+1}] = \phi_bE[b_tb_{t+1}]$. Thus, we can rewrite (A.1.14), (A.1.17), (A.1.22) and (A.1.25) as

$$\psi_v^s(n) = \mu_{v,1}\Gamma_b^s + \mu_{v,2}\phi_b\Gamma_b^s \tag{A.1.33}$$

$$\psi_p^s(n) = \mu_{p,1}\Gamma_b^s + \mu_{p,2}\phi_b\Gamma_b^s \tag{A.1.34}$$

$$\psi_v^b(n) = \mu_{v,1}\Gamma_b^b + \mu_{v,2}\phi_b\Gamma_b^b \tag{A.1.35}$$

$$\psi_p^b(n) = \mu_{p,1}\Gamma_b^b + \mu_{p,2}\phi_b\Gamma_b^b. \tag{A.1.36}$$

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.1.27) also depend on these coefficients.

A.2 Computation of the misspecification equilibrium

Recall the objective (13). Moreover, we have

$$E^s[\mathbf{z}_t^s] = \psi^s(n)s_t, \qquad \text{and} \tag{A.2.1}$$

$$E^b[\mathbf{z}_t^b] = \psi^b(n)b_t. \tag{A.2.2}$$

Thus, we can use (A.1.5) and (A.2.1) to compute

$$(\mathbf{z}_t - E^s[\mathbf{z}_t^s]) = \boldsymbol{\mu}_b b_t + \boldsymbol{\mu}_s s_t + \boldsymbol{\mu}_u u_t - \psi^s(n)s_t. \tag{A.2.3}$$

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

$$E[(\mathbf{z}_t - E^s[\mathbf{z}_t^s])'(\mathbf{z}_t - E^s[\mathbf{z}_t^s])] = E[[b_t'\boldsymbol{\mu}_b' + s_t'\boldsymbol{\mu}_s' + u_t'\boldsymbol{\mu}_u' - s_t'\psi^s(n)']\,[\boldsymbol{\mu}_b b_t + \boldsymbol{\mu}_s s_t + \boldsymbol{\mu}_u u_t - \psi^s(n)s_t]] \tag{A.2.4}$$

$$E[(\mathbf{z}_t - E^s[\mathbf{z}_t^s])'(\mathbf{z}_t - E^s[\mathbf{z}_t^s])] = (\boldsymbol{\mu}_b'\boldsymbol{\mu}_b)E[b_t^2] \tag{A.2.5}$$

$$+ [\boldsymbol{\mu}_s'\boldsymbol{\mu}_s + \psi^s(n)'\psi^s(n) - \boldsymbol{\mu}_s'\psi^s(n) - \psi^s(n)'\boldsymbol{\mu}_s]E[s_t^2] \tag{A.2.6}$$

$$+ (\boldsymbol{\mu}_u'\boldsymbol{\mu}_u)E[u_t^2] + [\boldsymbol{\mu}_b'\boldsymbol{\mu}_s - \boldsymbol{\mu}_b'\psi^s(n) + \boldsymbol{\mu}_s'\boldsymbol{\mu}_b - \psi^s(n)'\boldsymbol{\mu}_b]E[b_t s_t]. \tag{A.2.7}$$

In consequence, we obtain

$$EU^s = -\left[(\boldsymbol{\mu}_b'\boldsymbol{\mu}_b)E[b_t^2] + [\boldsymbol{\mu}_s'\boldsymbol{\mu}_s + \psi^s(n)'\psi^s(n) - \boldsymbol{\mu}_s'\psi^s(n) - \psi^s(n)'\boldsymbol{\mu}_s]E[s_t^2]\right. \tag{A.2.8}$$

$$\left. + (\boldsymbol{\mu}_u'\boldsymbol{\mu}_u)E[u_t^2] + [\boldsymbol{\mu}_b'\boldsymbol{\mu}_s - \boldsymbol{\mu}_b'\psi^s(n) + \boldsymbol{\mu}_s'\boldsymbol{\mu}_b - \psi^s(n)'\boldsymbol{\mu}_b]E[b_t s_t]\right]. \tag{A.2.9}$$

Likewise, we can use (A.1.5) and (A.2.2) to compute

$$(\mathbf{z}_t - E^b[\mathbf{z}_t^b]) = \boldsymbol{\mu}_b b_t + \boldsymbol{\mu}_s s_t + \boldsymbol{\mu}_u u_t - \psi^b(n)b_t. \tag{A.2.10}$$

Therefore it follows that

$$E[(\mathbf{z}_t - E^b[\mathbf{z}_t^b])'(\mathbf{z}_t - E^b[\mathbf{z}_t^b])] = E\big[\big[b_t'\boldsymbol{\mu}_b' + s_t'\boldsymbol{\mu}_s' + u_t'\boldsymbol{\mu}_u' - b_t'\psi^b(n)'\big]\,\big[\boldsymbol{\mu}_b b_t + \boldsymbol{\mu}_s s_t + \boldsymbol{\mu}_u u_t - \psi^b(n) b_t\big]\big]. \tag{A.2.11}$$

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

$$E[(\mathbf{z}_t - E^b[\mathbf{z}_t^b])'(\mathbf{z}_t - E^b[\mathbf{z}_t^b])] = [\boldsymbol{\mu}_b'\boldsymbol{\mu}_b + \psi^b(n)'\psi^b(n) - \boldsymbol{\mu}_b'\psi^b(n) - \psi^b(n)'\boldsymbol{\mu}_b]E[b_t^2] + (\boldsymbol{\mu}_s'\boldsymbol{\mu}_s)E[s_t^2] \tag{A.2.12}$$

$$+ (\boldsymbol{\mu}_u'\boldsymbol{\mu}_u)E[u_t^2] + [\boldsymbol{\mu}_b'\boldsymbol{\mu}_s + \boldsymbol{\mu}_s'\boldsymbol{\mu}_b - \boldsymbol{\mu}_s'\psi^b(n) - \psi^b(n)'\boldsymbol{\mu}_s]E[b_t s_t]. \tag{A.2.13}$$

In consequence

$$EU^b = -\big[[\boldsymbol{\mu}_b'\boldsymbol{\mu}_b + \psi^b(n)'\psi^b(n) - \boldsymbol{\mu}_b'\psi^b(n) - \psi^b(n)'\boldsymbol{\mu}_b]E[b_t^2] + (\boldsymbol{\mu}_s'\boldsymbol{\mu}_s)E[s_t^2] \tag{A.2.14}$$

$$+(\boldsymbol{\mu}_u'\boldsymbol{\mu}_u)E[u_t^2] + [\boldsymbol{\mu}_b'\boldsymbol{\mu}_s + \boldsymbol{\mu}_s'\boldsymbol{\mu}_b - \boldsymbol{\mu}_s'\psi^b(n) - \psi^b(n)'\boldsymbol{\mu}_s]E[b_t s_t]\big]\,. \tag{A.2.15}$$

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

$$F(n) = \big[\psi^b(n)'\psi^b(n) - \boldsymbol{\mu}_b'\psi^b(n) - \psi^b(n)'\boldsymbol{\mu}_b\big]\,E[b_t^2] \tag{A.2.16}$$

$$+ [\boldsymbol{\mu}_s'\psi^s(n) + \psi^s(n)'\boldsymbol{\mu}_s - \psi^s(n)'\psi^s(n)]\,E[s_t^2] \tag{A.2.17}$$

$$+ \big[\psi^s(n)'\boldsymbol{\mu}_b + \boldsymbol{\mu}_b'\psi^s(n) - \psi^b(n)'\boldsymbol{\mu}_s - \boldsymbol{\mu}_s'\psi^b(n)\big]\,E[b_t s_t]. \tag{A.2.18}$$

B. Proofs

B.1 Proof of Proposition 2 (Existence)

Proposition 2 *Let $N_*(\omega) = \{n_* \mid n_* = T_\omega(n_*)\}$ denote the set of misspecification equilibria. As $\omega \to \infty$, N_* has one of the following properties:*

[noitemsep]

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 10 *Proposition 2 relies only on the continuity of $F(n)$ and $T(n)$. If $F(n)$ is monotonic then a stronger statement is possible: Proposition 2 then identifies the full set of misspecification equilibria. In Section 3, 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.*

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

Lemma 11 *A unique restricted perceptions equilibrium exists for all n if and only if*

$$\triangle \equiv \det\left(I_4 - P' \otimes B\right) \neq 0$$

where

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

Proof. In an RPE

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

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,

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

After plugging in for ψ^s, ψ^b into ξ:

$$\begin{aligned} \xi &= A + B\xi\Gamma_1' \left[n e_1' \left(e_1 \Omega e_1'\right)^{-1} e_1 + (1-n) e_2' \left(e_2 \Omega e_2'\right)^{-1} e_2 \right] \\ \Leftrightarrow \xi &= A + B\xi P \end{aligned}$$

It follows that

$$\mathrm{vec}(\xi) = \mathrm{vec}(A) + (P' \otimes B)\,\mathrm{vec}(\xi)$$

Finally, the RPE coefficient is given by

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

and the stated conditions provides necessary and sufficient conditions for a unique ξ. ■

Proof of Proposition 2.

The existence of a set of fixed points $n_* = T_\omega(n_*)$ follow directly from applying Brouwer's theorem, since $T_\omega : [0,1] \to [0,1]$ and $F(n)$ is continuous provided there exists an RPE. Lemma 11 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 $\triangle = 0$ and ξ is continuous in n.

■

B.2 Proof of Proposition 1 (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 (10) becomes

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

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

$$v_t^i = (1-\beta)(b_{t+1} - b_t) + E_t^i v_{t+1}^i \tag{B.2.3}$$

$$\int_0^1 v_t^i di = v_t = (1-\beta)(b_{t+1} - b_t) + \int_0^1 E_t^i v_{t+1}^i di \tag{B.2.4}$$

$$= (1-\beta)(b_{t+1} - b_t) + n\psi_v^s s_t + (1-n)\psi_v^b b_t \tag{B.2.5}$$

$$\Leftrightarrow v_t = y_t - (1-\beta)b_t \tag{B.2.6}$$

$$p_t^*(j) = 0. \tag{B.2.7}$$

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.2.5) 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.2.6). Coefficients in (11) and (12) are required to satisfy the orthogonality conditions (A.1.14) and (A.1.22) respectively.

Under PF, i.e., assumption (8), 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.1.26) to (A.1.32) and we obtain

$$\Gamma_b^b = \frac{E[b_{t+1}b_t]}{E[b_t^2]} = \beta^{-1}(1-\phi_b), \tag{B.2.8}$$

where the linear projection $E[b_{t+1}] = \Gamma_b^b b_t$ satisfies orthogonality condition (A.1.22). Likewise we can compute

$$\Gamma_b^s \equiv \frac{E[b_{t+1}s_t]}{E[s_t^2]} = \frac{-\beta^{-1}\left(1-\beta^2-\phi_b\right)}{1-\beta^2-2\phi_b}, \tag{B.2.9}$$

where the linear projection $E[b_{t+1}] = \Gamma_b^s s_t$ satisfies orthogonality condition (A.1.14).

Thus, we can obtain (A.1.33) and (A.1.35) as

$$\psi_v^s = [(\beta^1-1)-(1-\beta)+(1-n)\psi_v^b]\Gamma_b^s + [n\psi_v^s-(\beta^1-1)]\phi_b\Gamma_b^s \tag{B.2.10}$$
$$\psi_v^b = [(\beta^1-1)-(1-\beta)+(1-n)\psi_v^b]\Gamma_b^b + [n\psi_v^s-(\beta^1-1)]\phi_b\Gamma_b^b. \tag{B.2.11}$$

Rearranging terms yields

$$\psi_v^s = [(\beta^1-1)(1-\beta-\phi_b)]\Gamma_b^s + [\phi_b n\psi_v^s+(1-n)\psi_v^b]\Gamma_b^s \tag{B.2.12}$$
$$\psi_v^b = [(\beta^1-1)(1-\beta-\phi_b)]\Gamma_b^b + [\phi_b n\psi_v^s+(1-n)\psi_v^b]\Gamma_b^b. \tag{B.2.13}$$

Clearly, $n=0$ implies that (B.2.13) collapses to (B.3.7) below and $n=1$ implies that (B.2.12) collapses to (B.4.3) below.

Thus, we can solve for

$$\Leftrightarrow \psi_v^s(n) = \frac{(1-\beta)\Gamma_b^s(1-\beta-\phi_b)}{\beta\left[1-\left(\phi_b n\Gamma_b^s+(1-n)\Gamma_b^b\right)\right]} = \frac{(1-\beta^2-\phi_b)\beta^{-1}(1-\beta)}{[1-\beta^2-n(1+\beta-\phi_b)-2\phi_b]} \tag{B.2.14}$$
$$\Leftrightarrow \psi_v^b(n) = \frac{(1-\beta)\Gamma_b^b(1-\beta-\phi_b)}{\beta\left[1-\left(\phi_b n\Gamma_b^s+(1-n)\Gamma_b^b\right)\right]} = \frac{-(1-\beta^2-2\phi_b)(1-\phi_b)\beta^{-1}(1-\beta)}{[1-\beta^2-n(1+\beta-\phi_b)-2\phi_b]}. \tag{B.2.15}$$

This proves Proposition 1.

■

B.3 Proof of Corollary 4 (Weak Ricardian Equivalence)

Proof.

Suppose that all agents have PLM (12). Then the TE dynamics are still governed by (B.2.1). Moreover, (10) is given by

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

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 (12), i.e., $v_t^i = v_t, \forall i$, the implications for (D.12) are

$$v_t^i = (1-\beta)v_t^i + (1-\beta)\beta[b_{t+1} - b_t] + \beta\psi_v^b b_t \tag{B.3.2}$$

$$\beta v_t^i = (1-\beta)\beta[\beta^{-1}(b_t - s_t) - b_t] + \beta\psi_v^b b_t \tag{B.3.3}$$

$$v_t^i = (1-\beta)[\beta^{-1}(b_t - s_t) - b_t] + \psi_v^b b_t \tag{B.3.4}$$

$$v_t^i = (\beta^{-1} - 1)(b_t - s_t) + \psi_v^b b_t - (1-\beta)b_t \tag{B.3.5}$$

$$v_t^i = y_t - (1-\beta)b_t. \tag{B.3.6}$$

The ALM is then given by (B.2.1), (B.3.1) and (B.3.6).

Now we can apply $v_t \equiv \int v_t^i di$ to (B.3.6) 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.2.8), (A.1.35) is given by

$$\psi_v^b = \left[(\beta^{-1} - 1)(1 - \beta - \phi_b) + \psi_v^b\right]\Gamma_b^b \tag{B.3.7}$$

$$0 = \psi_v^b - \left[(\beta^{-1} - 1)(1 - \beta - \phi_b) + \psi_v^b\right]\Gamma_b^b, \tag{B.3.8}$$

and ψ_v^b in (15) follows.

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

$$y_t = -(\beta^{-1} - 1)z_t, \tag{B.3.9}$$

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 4.

■

B.4 Proof of Corollary 5 (Woodford (2013))

Proof.

Suppose $n = 1$, i.e., all agents use PLM (11). In this case, (B.2.1) and (B.3.6) remain the same, however (10) becomes

$$y_t = -\sigma(\phi_\pi \pi_t) + (1-\beta)\beta^{-1}(b_t - s_t) + \psi_v^s s_t \tag{B.4.1}$$
$$y_t = (\beta^{-1} - 1)(b_t - s_t) + \psi_v^s s_t, \tag{B.4.2}$$

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.3.6). Thus, we can apply $v_t \equiv \int v_t^i di$ to (B.3.6) 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.2.9), we obtain

$$\psi_v^s = \left[(\beta^{-1} - 1)(1 - \beta - \phi_b) + \psi_v^s \phi_b\right] \Gamma_b^s \tag{B.4.3}$$
$$\Leftrightarrow \psi_v^s = \frac{(1-\beta)(1-\beta-\phi_b)\Gamma_b^s}{\beta(1-\phi_b\Gamma_b^s)} = -\frac{\beta^{-1}(1-\beta)(1-\beta^2-\phi_b)}{(\beta+\beta^2+\phi_b)} \tag{B.4.4}$$
$$\Leftrightarrow \psi_v^s < \beta^{-1} - 1. \tag{B.4.5}$$

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

$$y_t = \left[(\beta^{-1} - 1)(1-\phi_b) + \phi_b \psi_v^s\right] b_t - \left[(\beta^{-1} - 1) - \psi_v^s\right] z_t$$
$$y_t = \left[\frac{(1-\beta)(1+\beta-\phi_b)}{\beta(1+\beta)+\phi_b}\right] b_t - \left[(\beta^{-1} - 1) - \psi_v^s\right] z_t. \tag{B.4.6}$$

Thus, as y_t depends on b_t, Ricardian equivalence fails. This proves Corollary 5.

■

B.5 Proof of Theorem 7 (Multiple Equilibria)

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.,

$$F(n) = \left(\frac{(1-\beta)^2 \sigma_z^2}{\beta^2 \left(1-\beta^2 - n(1+\beta-\phi_b) - 2\phi_b\right)^2} \right) \times \tag{B.5.1}$$
$$\left(2n(1-\beta^2-\phi_b)(2(1-\phi_b)+\beta(1-\beta)) - (1-\beta^2-2\phi_b)(4(1-\phi_b)-\beta(2+3\beta))\right) \tag{B.5.2}$$

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 8 and therefore to prove Theorem 7.

■

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

B.6 Proof of Corollary 12

Proof.

In the simple case of Corollary 2, we find that

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

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

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

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

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

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

■

B.7 Proof of Proposition 3

Proposition 3 *For σ and $s_b \geq 0$ sufficiently small, there exists a $\tilde{\phi}(\beta)$ such that multiple misspecification equilibria exist provided that*

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

Proof.

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

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

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

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

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

$$p_t^*(j) = (1-\alpha)p_t^* + (1-\alpha\beta)\left[\xi y_t + \mu_t\right] + \alpha\beta E_t^j p_{t+1}^*(j) \tag{B.7.4}$$

$$\int_0^1 p_t^*(j)dj = p_t^* = \frac{(1-\alpha\beta)}{\alpha}\left[\xi y_t + \mu_t\right] + \beta \int_0^1 E_t^j p_{t+1}^*(j)dj \tag{B.7.5}$$

$$p_t^* = (1-\alpha)^{-1}\left[\kappa y_t + u_t\right] + \beta\left[n\psi_p^s s_t + (1-n)\psi_p^b b_t\right] \tag{B.7.6}$$

$$(1-\alpha)^{-1}\pi_t = (1-\alpha)^{-1}\left[\kappa y_t + u_t\right] + \beta\left[n\psi_p^s s_t + (1-n)\psi_p^b b_t\right] \tag{B.7.7}$$

$$\pi_t = \kappa y_t + u_t + (1-\alpha)\beta\left[n\psi_p^s s_t + (1-n)\psi_p^b b_t\right]. \tag{B.7.8}$$

The ALM is then given by (11) to (12), (B.2.1), (B.7.1), (B.7.2), (B.7.3), and (B.7.4) to (B.7.8) for given policy (5) and (6).

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

$$\delta_0 \equiv \frac{(1-\beta)}{(1+\sigma\phi_\pi\kappa)}, \quad \delta_1 \equiv \frac{(1-n)\left(\psi_v^b - \sigma\phi_\pi(1-\alpha)\beta\psi_p^b\right)}{(1+\sigma\phi_\pi\kappa)}, \tag{B.7.9}$$

$$\delta_2 \equiv \frac{n\left(\psi_v^s - \sigma\phi_\pi(1-\alpha)\beta\psi_p^s\right)}{(1+\sigma\phi_\pi\kappa)}, \quad \delta_3 \equiv \frac{-\sigma\phi_\pi}{(1+\sigma\phi_\pi\kappa)}. \tag{B.7.10}$$

Moreover, we use (A.1.1) to eliminate y_t in (B.7.8), i.e.,

$$\pi_t = \kappa\left[\delta_0 b_{t+1} + \delta_1 b_t + \delta_2 s_t + \delta_3 u_t\right] + u_t + (1-\alpha)\beta\left[n\psi_p^s s_t + (1-n)\psi_p^b b_t\right] \tag{B.7.11}$$

$$\begin{aligned}\pi_t = &\left[\kappa\delta_1 + (1-\alpha)\beta(1-n)\psi_p^b\right] b_t + \left[\kappa\delta_2 + (1-\alpha)\beta n\psi_p^s\right] s_t \\ &+ \left[\kappa\delta_3 + 1\right] u_t + \kappa\delta_0 b_{t+1}.\end{aligned} \tag{B.7.12}$$

Furthermore, we use (A.1.1), (A.1.2) and (B.7.12) to eliminate π_t, y_t and b_{t+1} in (B.7.3), i.e.,

$$v_t = \left[(1-n)\psi_v^b - (1-\beta)\right] b_t + n\psi_v^s s_t - \sigma(\phi_\pi - 1)\pi_t + (1-\beta)b_{t+1} \tag{B.7.13}$$

$$\begin{aligned}v_t = &\left[(1-n)\psi_v^b - (1-\beta) - \sigma(\phi_\pi - 1)\left[\kappa\delta_1 + (1-\alpha)\beta(1-n)\psi_p^b\right]\right] b_t \\ &+ \left[n\psi_v^s - \sigma(\phi_\pi - 1)\left[\kappa\delta_2 + (1-\alpha)\beta n\psi_p^s\right]\right] s_t \\ &- \sigma(\phi_\pi - 1)\left[\kappa\delta_3 + 1\right] u_t \\ &+ \left[(1-\beta) - \sigma(\phi_\pi - 1)\kappa\delta_0\right] b_{t+1}\end{aligned} \tag{B.7.14}$$

$$\begin{aligned}v_t = &\left[(1-n)\psi_v^b - (1-\beta) - \sigma(\phi_\pi - 1)\left[\kappa\delta_1 + (1-\alpha)\beta(1-n)\psi_p^b\right]\right] b_t \\ &+ \left[n\psi_v^s - \sigma(\phi_\pi - 1)\left[\kappa\delta_2 + (1-\alpha)\beta n\psi_p^s\right]\right] s_t \\ &- \sigma(\phi_\pi - 1)\left[\kappa\delta_3 + 1\right] u_t \\ &+ \Xi\left[\xi_1 b_t + \xi_2 s_t + \xi_3 u_t\right], \qquad \text{where} \qquad \Xi \equiv \left[(1-\beta) - \sigma(\phi_\pi - 1)\kappa\delta_0\right].\end{aligned} \tag{B.7.15}$$

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

$$\mu_{v,1} \equiv \left[(1-n)\psi_v^b - (1-\beta) - \sigma(\phi_\pi - 1)\left[\kappa\delta_1 + (1-\alpha)\beta(1-n)\psi_p^b\right] + \xi_1\Xi\right] \tag{B.7.16}$$

$$\mu_{v,2} \equiv \left[n\psi_v^s - \sigma(\phi_\pi - 1)\left[\kappa\delta_2 + (1-\alpha)\beta n\psi_p^s\right] + \xi_2\Xi\right] \tag{B.7.17}$$

$$\mu_{v,3} \equiv \left[\xi_3\Xi - \sigma(\phi_\pi - 1)\left[\kappa\delta_3 + 1\right]\right]. \tag{B.7.18}$$

Moreover, we use (A.1.2) to eliminate b_{t+1} in (B.7.12), i.e.,

$$\begin{aligned}\pi_t = &\left[\kappa\delta_1 + (1-\alpha)\beta(1-n)\psi_p^b\right] b_t + \left[\kappa\delta_2 + (1-\alpha)\beta n\psi_p^s\right] s_t + \left[\kappa\delta_3 + 1\right] u_t \\ &+ \kappa\delta_0\left[\xi_1 b_t + \xi_2 s_t + \xi_3 u_t\right],\end{aligned} \tag{B.7.19}$$

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

$$\mu_{p,1} \equiv \left[\kappa\left(\delta_1 + \delta_0\xi_1\right) + (1-\alpha)\beta(1-n)\psi_p^b\right]/(1-\alpha) \tag{B.7.20}$$

$$\mu_{p,2} \equiv \left[\kappa\left(\delta_2 + \delta_0\xi_2\right) + (1-\alpha)\beta n\psi_p^s\right]/(1-\alpha) \tag{B.7.21}$$

$$\mu_{p,3} \equiv \left[\kappa\left(\delta_3 + \delta_0\xi_3\right) + 1\right]/(1-\alpha). \tag{B.7.22}$$

Thus, we can obtain (A.1.33), (A.1.34), (A.1.35) and (A.1.36) as

$$\begin{aligned}\psi_v^s(n) =& \frac{(1-\beta)\Gamma_b^s\left[(1+\kappa\sigma)(1-\phi_b) + \beta^2\left(\phi_b n\Gamma_b^s + (1-n)\Gamma_b^b\right) - \beta\left(1 + (1-\phi_b)\left(\phi_b n\Gamma_b^s + (1-n)\Gamma_b^b\right) + \kappa\sigma\phi_\pi\right)\right]}{\beta\left[1 - \left(\phi_b n\Gamma_b^s + (1-n)\Gamma_b^b\right)\left(1 + \kappa\sigma + \beta\left(1 - \left(\phi_b n\Gamma_b^s + (1-n)\Gamma_b^b\right)\right)\right) + \kappa\sigma\phi_\pi\right]}\\ =& \left[(1-\beta^2-\phi_b)(1-\beta)\left(n(1+\beta-\phi_b)(1-\beta-\phi_b)^2 + (1-\beta^2-2\phi_b)\left(\phi_b(1-\beta-\phi_b) - \kappa\sigma(\phi_b + (\beta\phi_\pi - 1))\right)\right)\right]/\mathcal{D}\end{aligned} \tag{B.7.23}$$

$$\begin{aligned}\mathcal{D} \equiv& \beta\left[-n^2(\beta^2 - (1-\phi_b)^2)^2 + n(1+\beta-\phi_b)(1-\beta-\phi_b)(1-\beta^2-2\phi_b)(1-\beta-\kappa\sigma-2\phi_b)\right.\\ &\left.+(1-\beta^2-2\phi_b)^2(\phi_b(1-\beta-\phi_b) - \kappa\sigma(\phi_b + (\beta\phi_\pi - 1)))\right]\end{aligned} \tag{B.7.24}$$

$$\begin{aligned}\psi_p^s(n) =& \frac{(1-\beta)\Gamma_b^s\kappa\left[1-\phi_b - \beta\left(\phi_b n\Gamma_b^s + (1-n)\Gamma_b^b\right)\right]}{(1-\alpha)\beta\left[1 - \left(\phi_b n\Gamma_b^s + (1-n)\Gamma_b^b\right)\left(1 + \kappa\sigma + \beta\left(1 - \left(\phi_b n\Gamma_b^s + (1-n)\Gamma_b^b\right)\right)\right) + \kappa\sigma\phi_\pi\right]}\\ =& \left[(1-\beta^2-\phi_b)n(1-\beta)\kappa(1+\beta-\phi_b)(1-\beta-\phi_b)\right]/\left[(1-\alpha)\mathcal{D}\right]\end{aligned} \tag{B.7.25}$$

$$\begin{aligned}\psi_v^b(n) =& \frac{(1-\beta)\Gamma_b^b\left[(1+\kappa\sigma)(1-\phi_b) + \beta^2\left(\phi_b n\Gamma_b^s + (1-n)\Gamma_b^b\right) - \beta\left(1 + (1-\phi_b)\left(\phi_b n\Gamma_b^s + (1-n)\Gamma_b^b\right) + \kappa\sigma\phi_\pi\right)\right]}{\beta\left[1 - \left(\phi_b n\Gamma_b^s + (1-n)\Gamma_b^b\right)\left(1 + \kappa\sigma + \beta\left(1 - \left(\phi_b n\Gamma_b^s + (1-n)\Gamma_b^b\right)\right)\right) + \kappa\sigma\phi_\pi\right]}\\ =& \left[-(1-\phi_b)(1-\beta^2-\phi_b)(1-\beta)\left(n(1+\beta-\phi_b)(1-\beta-\phi_b)^2\right.\right.\\ &\left.\left.+(1-\beta^2-2\phi_b)\left(\phi_b(1-\beta-\phi_b) - \kappa\sigma(\phi_b + (\beta\phi_\pi - 1))\right)\right)\right]/\mathcal{D}\end{aligned} \tag{B.7.26}$$

$$\begin{aligned}\psi_p^b(n) =& \frac{(1-\beta)\Gamma_b^b\kappa\left[1-\phi_b - \beta\left(\phi_b n\Gamma_b^s + (1-n)\Gamma_b^b\right)\right]}{(1-\alpha)\beta\left[1 - \left(\phi_b n\Gamma_b^s + (1-n)\Gamma_b^b\right)\left(1 + \kappa\sigma + \beta\left(1 - \left(\phi_b n\Gamma_b^s + (1-n)\Gamma_b^b\right)\right)\right) + \kappa\sigma\phi_\pi\right]}\\ =& \left[-(1-\phi_b)(1-\beta^2-\phi_b)n(1-\beta)\kappa(1+\beta-\phi_b)(1-\beta-\phi_b)\right]/\left[(1-\alpha)\mathcal{D}\right].\end{aligned} \tag{B.7.27}$$

Based on the coefficients above, we can compute $F(n)$ as outlined in Appendix A.2 above. Taking the limit as $\sigma \to 0$, 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 ψ^s, ψ^b are real-time estimates from a constant gain learning model, a form of discounted least-squares. With this perceived law of motion, the actual law of motion implied by these beliefs can be written as

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

where

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

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 $S\left(\psi^s, \psi^b\right)$. A rational expectations equilibrium is a fixed point of the "S-map", i.e., $\Theta^* = S(\Theta^*), \Theta' = (\psi^s, \psi^b)$.

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

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

where

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

and R is an estimate of the variance-covarience matrix used as a weight when updating Θ. 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 Θ that fluctuate around Θ^*. The response of Θ to a particular sequence of unlikely shocks is described by the "escape dynamics", which provide the "most likely unlikely" path away from the rational expectations equilibrium, and then the mean dynamics describe the transition path back to the equilibrium.[21] The escape dynamics, therefore, can be thought of as re-initializing the mean 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, γ, corresponds to a long stretch of real time.

Figure A1 plots the mean dynamics for a particular illustrative parameterization: $\phi_y = 0.5$, $\phi_b = 0.9$, $\sigma = 2$, $\sigma_z^2 = 1$.[22] We then choose initial values for the ψ^s, ψ^b that are both above their rational expectations equilibrium values. The mean dynamics O.D.E. is then solved and the Figure plots the expected learning path (ψ^s shown). The experiment is to imagine an "escape" that has driven beliefs above their rational expectations equilibrium values and use the solution to the mean dynamics O.D.E. to trace out how the economy is most likely to respond.

The figure plots (solid line) the expected transition path for ψ^s while the the dashed line

[21]See Williams (2019) for details and a comprehensive set of results and toolkit on escape dynamics in constant gain learning models.

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

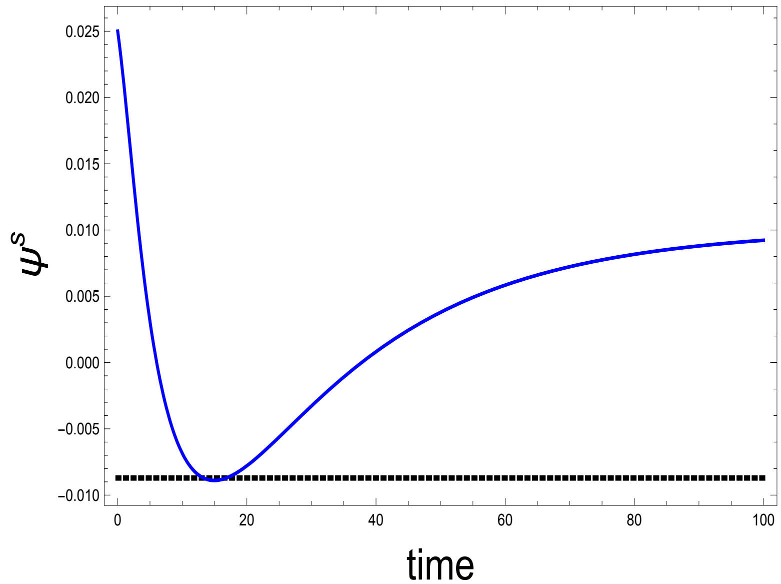


Figure A1: Expected learning dynamics for a correctly specified forecast model

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

We conclude from the mean dynamics Figure 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 γ, 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 the Figure also help to better understand the connection between this paper and Eusepi and Preston (2018). 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,

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

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 σ, captures an intertemporal substitttution 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:

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

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

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

where the subjective composite variable v_t^i is defined as

$$v_t^i \equiv \sum_{T=t}^{\infty} \beta^{T-t} E_t^i \{(1-\beta)(Y_T - \tau_T) - [\sigma - (1-\beta)s_b](\beta i_T - \pi_T) - (1-\beta)\bar{c}_T\}. \tag{D.3}$$

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

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

$$v_t^i = (1-\beta)(Y_t - \tau_t) - [\sigma - (1-\beta)s_b](\beta i_t - \pi_t) - (1-\beta)\bar{c}_t + \beta E_t^i v_{t+1}^i. \tag{D.4}$$

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.[23]

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

[23] 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.

condition

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

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

$$p_t^*(j) = (1-\alpha\beta)\left(E_t^j p_t^{\text{opt}} - p_{t-1}\right) + (\alpha\beta)E_t^j p_{t+1}^*(j) + (\alpha\beta)\pi_t, \qquad \text{where} \tag{D.6}$$

$$E_t^j p_{t+1}^*(j) \equiv (1-\alpha\beta)\sum_{T=t}^{\infty}(\alpha\beta)^{T-t}\left(E_t^j p_{T+1}^{\text{opt}} - p_t\right). \tag{D.7}$$

Temporary equilibrium with heterogeneous beliefs. Recall the consumption function recursion:

$$c_t^i = (1-\beta)b_t^i + (1-\beta)(Y_t - \tau_t) - \beta[\sigma - (1-\beta)s_b]i_t - (1-\beta)s_b\pi_t + \beta\bar{c}_t + \beta E_t^i v_{t+1}^i \tag{D.8}$$

$$v_t^i = (1-\beta)(Y_t - \tau_t) - [\sigma - (1-\beta)s_b](\beta i_t - \pi_t) - (1-\beta)\bar{c}_t + \beta E_t^i v_{t+1}^i. \tag{D.9}$$

The latter two can be written as

$$c_t^i = (1-\beta)b_t^i + \bar{c}_t - \sigma\pi_t + v_t^i, \tag{D.10}$$

which, together with $b_t \equiv \int b_t^i di$, $v_t \equiv \int v_t^i di$ and (9) yields aggregate demand

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

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

To express the aggregate demand equation in explicit dependence of expectations, we apply (6) and (D.11) to the v_t^i recursion and obtain

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

Averaging over expectations in (D.12) and plugging into (D.11) yields (10). Thus, 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, 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).

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

$$p_t^{\text{opt}} = p_t + \xi \left(Y_t - Y_t^n \right) + \mu_t, \tag{D.13}$$

where $\xi > 0$ is a composite term of structural parameters measuring the output elasticity of a firm's optimal price.[24] The exogenous random variable Y_t^n is the natural level of output that captures exogenous demand shocks and μ_t represents disturbances to the desired markup over marginal cost.

As the firm's price is a decision variable, it is natural to impose that $E_t^j p_t^{\text{opt}} = p_t^{\text{opt}}$. It follows, then, from plugging (D.13) and (3) into (D.6) that

$$p_t^*(j) = (1-\alpha) p_t^* + (1-\alpha\beta) \left[\xi y_t + \mu_t \right] + \alpha\beta E_t^j p_{t+1}^*(j). \tag{D.14}$$

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

$$\pi_t = (1-\alpha)\beta \int E_t^j p_{t+1}^*(j) dj + \kappa y_t + u_t. \tag{D.15}$$

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

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

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

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

[24] The term is defined in Woodford (2003, ch.3-4).

E. Bayesian estimation details

Recall the nonlinear state-space model:

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

where the state vector is

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

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

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

and the parameter vector is

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

The measurement and state disturbances are η_t, ν_t respectively. The 4 exogenous shocks follow a linear transition equation with a diagonal matrix whose diagonals are the respective AR(1) coefficients.

In brief, the particle filter, like the Kalman filter, operates in both a prediction and update steps. The first step, is given the previous period's particle approximation, is to draw a large number of innovations and then iterate on the state transition equation to yield a predicted next-period state. The predicted particles are then re-weighted based on the most recent data observation, this is the updating step. The updated weights are used to directly compute the likelihood value.

The Bootstrap particle filter, while conceptually straightforward, introduces several computational challenges. First, a stable approximation of the likelihood function requires a large number of particles. Even with efficient parallelization and vectorization each computation of the likelihood approximation is time-consuming. In our application, this is especially true since the state-vector consists of 34 variables. Adopting stochastic gradient learning versus constant gain learning lowers the computational cost considerably. With constant gain least-squares, the particle filter algorithm would require inverting the regressor covariance matrix 10^{12} times. Second, measurement noise in the observation equation is necessary for an accurate particle filter approximation of the likelihood function. The consequence is that the parameter estimates are estimated with greater uncertainty. We follow Herbst and Schorfheide (2015) in defining measurement error so that $\Sigma_\eta = 0.25 \times diag\left[Var\left(Y_T\right)\right]$.

To approximate the likelihood function we use the Bootstrap Particle Filter, as developed in Herbst and Schorfheide (2015):

Algorithm 13 *Bootstrap Particle Filter.*

1. Initialization *Draw the initial particles* $e_0^j \overset{iid}{\sim} p(s_0)$ *and let* $W_0^j, j = 1, ..., M$.
2. Recursion *For* $t = 1, ..., T$*:*
   (a) Forecasting s_t. *Iterate on the state-transition equation:*
   $$\hat{s}_t^j = g(s_{t-1}^j, \Theta) + Q(s_{t-1}^j, \Theta)\varepsilon_t^j,\ \ \varepsilon_t^j \sim F_\varepsilon(\cdot, \Theta)$$
   (b) Forecasting y_t. *Define the weights*
   $$\hat{w}_t^j = p(y_t|\hat{s}_t^j, \Theta)$$
   where
   $$p(\cdot|\cdot) \approx \frac{1}{M}\sum_{j=1}^{M} \hat{w}_t^j W_{t-1}^j$$
   and
   $$\hat{w}_t^j = (2\pi)^{-n/2}\left|\Sigma_u\right|^{-1/2} \times \exp\left\{-\frac{1}{2}\left[y_t - f(\hat{s}_t^j)\right]' \Sigma_u^{-1}\left[y_t - f(\hat{s}_t^j)\right]'\right\}$$
   where Σ_u *is the covariance matrix for the measurement errors.*
   (c) Updating. *Normalize weights:*
   $$\hat{W}_t^j = \frac{\hat{w}_t^j W_{t-1}^j}{\frac{1}{M}\sum_{j=1}^{M} \hat{w}_t^j W_{t-1}^j}$$
   (d) Selection. *Define* $\hat{n} = M/\left(M^{-1}\sum_{j=1}^{M}(\hat{W}_t^j)^2\right)$ *and let* r_t *be an indicator variable whenever* $\hat{n}_t < M/2$.

Case 1: $r_t = 1$ *Let* $s_t^j,\ j = 1, ..., M$ *denote* M *iid draws from a multinomial distribution with support points/weights* $\left\{\hat{s}_t^j, \hat{W}_t^j\right\}$ *and set* $W_t^j = 1, \forall M$.

Case 2: $r_t = 0$ *Then set* $s_t^j = \hat{s}_t^j$ *and* $W_t^j = \hat{W}_t^j$.

3. Likelihood approximation:

$$\ln \hat{p}\left(Y^T|\Theta\right) = \sum_{t=1}^{T} \ln\left(M^{-1}\sum_{j=1}^{M}\hat{w}_t^j W_{t-1}^j\right)$$

Finally, we use a Metropolis-Hastings algorithm to construct the posterior distribution:

$$p\left(\Theta|Y_T\right) \propto p\left(Y_T|\Theta\right) \times p\left(\Theta\right)$$

where $p(\Theta)$ is the prior distribution and $(\Theta|Y_T)$ is the object of interest. Our algorithm samples from the posterior distribution through an adapted Random-Walk Metropolis Hastings (RWMH) MCMC technique, with the particle-filter based estimate of the likelihood function. Convergence properties of the algorithm are discussed in Herbst and Schorfheide (2015). One challenge for the RWMH is identifying a good candidate distribution to draw from, and to be sure that the chain samples from the entire support of the distribution. We do this in two ways. First, we use an adaptive approach to recursively update the candidate distribution combined with a long transient burn-in period. Second, in 99% of the time we draw from a mixture distribution, each proportional to the recursively updated candidate that is centered on the previously accepted draw from the RWMH. The remaining time the draws come naively from the candidate distribution. This ensures that the algorithm is not concentrated near a local maxima and that convergence occurs relatively quickly. We also used a tempering procedure in the early stages of the burn-in period.

We can now describe the Metropolis-Hastings algorithm.

Algorithm 14 *Metropolis-Hastings. For* $i = 1, ..., N$

1. *Draw a candidate* θ *from* $q\left(\theta|\Theta^{i-1}\right)$
2. *Set* $\Theta^i = \theta$ *with probability*

$$\alpha\left(\theta|\Theta^{i-1}\right) = min\left\{1, \frac{\hat{p}(Y^T|\theta)p(\theta)/q(\theta|\Theta^{i-1})}{\hat{p}(Y^T|\Theta^{i-1})p(\Theta^{i-1})/q(\Theta^{i-1}|\theta)}\right\}$$

where $\hat{p}(\cdot|\cdot)$ *is computed using Algorithm 1, and* $p(\cdot)$ *is the prior.*

For the candidate density we use a variant on a random-walk Metropolis Hastings. With probability $\delta \approx 1$ we set

$$q(\cdot|\Theta^{i-1}) = N(\Theta^{i-1}, c\hat{\Sigma}_i)$$

where we use a recursive adaptive algorithm to compute $\hat{\Sigma}$:

$$\begin{aligned}
\bar{\Theta}_i &= \frac{i+1}{100+i+1}\left[\frac{i}{1+i}\bar{\Theta}_{i-1} + \frac{1}{1+i}\Theta^{i-1}\right] + \frac{100}{100+i+1}\Theta^0 \\
\hat{\Sigma}_i &= \frac{i}{100+i}\left[\frac{i-1}{i}\hat{\Sigma}_{i-1} + \left(i\bar{\Theta}_{i-1}\bar{\Theta}'_{i-1} - (i+1)\bar{\Theta}_i\bar{\Theta}'_i\Theta^{i-1}\Theta^{(i-1)'}\right)/i\right] + \frac{100}{100+i}\hat{\Sigma}_0
\end{aligned}$$

and with complementary probability

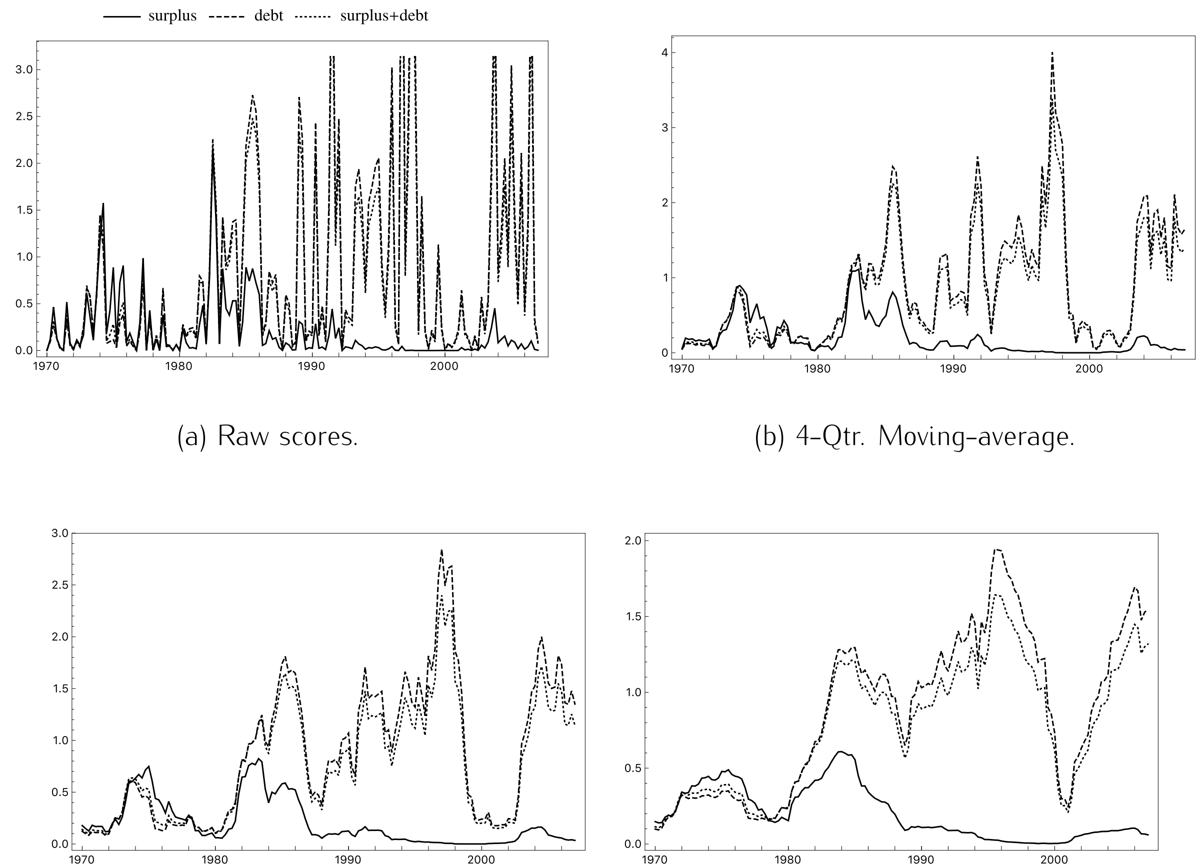
$$q(\cdot|\Theta^{i-1}) = N(\bar{\Theta}_i, c\hat{\Sigma}_i)$$

In the estimation, we set $M = 120,000$ and $N = 200,000$. We use a burn-in period of $80,000$ draws. We ran a tuning run to initialize $\hat{\Sigma}$ during which we also used a tempering schedule. The priors are specified below.

Many of our priors are the same as in Eusepi and Preston (2018) and Herbst and Schorfheide (2015). We set the prior for the intensity of choice in line with survey estimates provided in Branch (2004). Our prior for the gain parameters are informed by previous estimates provided in the literature. The observation equation also includes measurement errros, as discussed in the main text. We set these, following Herbst and Schorfheide (2015), as $\Sigma_u = 0.25 \times diag\left[Var\left(Y^T\right)\right]$.

F. Restricted perceptions vs. rational expectations

Since $n = 0$ is equivalent to the rational expectations model our estimate the estimation results present an econometric test of our model of restricted perceptions against rational expectations. The results tell us that the data prefer the specification of the model with a non-Ricardian equilibrium over most of the sample. 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 of providing some external evidence, 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,



(a) Raw scores.

(b) 4-Qtr. Moving-average.

(c) 8-Qtr. Moving-average.

(d) 5-Yr. Moving-average.

Figure A2: 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.

we compute moving averages of the statistical score $Ex_{j,t-1}\left(\pi_t - \pi^e_{t-1,t}\right)$ where $x_{j,t} \in \{s_t, b_t\}$, π_t is the PCE inflation rate, and $\pi^e_{t-1,t}$ is the one-step ahead median SPF survey forecast. The results are plotted in Figure A2. 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 the scores-Figure 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.

Table A1: Prior distribution of parameters

Parameter	Dist.	Para(1)	Para(2)
Structural parameters			
κ	Beta	0.1	0.2
α	Beta	0.6	0.2
Policy parameters			
ϕ_π	Normal	1.50	0.25
ϕ_y	Normal	0.125	0.05
ϕ_b	Inv. Gamma	0.15	0.05
Exogenous shocks			
ρ_g	Beta	0.50	0.20
ρ_u	Beta	0.50	0.20
ρ_w	Beta	0.50	0.20
ρ_z	Beta	0.50	0.20
$100\sigma_g$	Inv. Gamma	0.01	2.00
$100\sigma_u$	Inv. Gamma	0.01	2.00
$100\sigma_w$	Inv. Gamma	0.01	2.00
$100\sigma_z$	Inv. Gamma	0.01	2.00
Learning parameters			
ω	Gamma	5.00	2.00
γ_1	Beta	0.015	0.015
γ_2	Beta	0.03	0.03