Counterfactual Equivalence in Macroeconomics[*]

Martin Beraja[†]

February 7, 2018

Abstract

When studying counterfactual policy rules using structural models, researchers are often uncertain about features of the economy that are difficult to distinguish with available data. If the counterfactual is not robust to variation in such features across models, its credibility is undermined. Relatedly, model-builders have little formal guidance regarding which features are likely to be quantitatively relevant when evaluating particular policy rules. I propose a novel method to tackle these and other problems in macroeconomics. It rests on the insight that many models which are well approximated by a linear representation are both observationally equivalent under a benchmark policy and yield an identical counterfactual equilibrium under an alternative policy. Characterizing a set of models that satisfy this principle of Counterfactual Equivalence simply requires describing whether their equilibrium equations satisfy a number of linear restrictions. I illustrate the principle's usefulness in several applications. First, I present a methodology for constructing counterfactual SVARs with respect to policy rule changes in a way that is both immune to Lucas Critique and robust across a set of models. I use it to quantify how US fiscal integration contributes to regional stabilization. Second, in a set of international business cycle models, I argue that discount rate shocks are the key feature generating large stabilization gains from fiscal integration. Thus, they can help resolve a number of international risk-sharing puzzles. Third, I highlight features of search models of the labor market which are relevant when studying unemployment benefits rules. Finally, I show how to falsify a set of New Keynesian models using observed interest rate rule changes.

---

[*] I am indebted to my advisors Fernando Alvarez, Erik Hurst, and Veronica Guerrieri, for their encouragement, support, and guidance during my stay in Chicago, for our many insightful conversations, and for their patience and skill in distilling the important points from my presentations. I thank Harald Uhlig for useful comments and deep generosity of time. For their valuable suggestions, I owe Andy Atkeson, Ricardo Lagos, John Leahy, Martin Schneider, and Gianluca Violante as well as my classmates Jonathan Adams, Philip Barret, and Chanont Banternghansa. I thank those at the Review of Economic Studies Tour, Minneapolis Fed Junior Scholar Conference, SED, "Pizzanomics", and Macro Seminars at MIT, Harvard, Yale, UCLA, GWU, Chicago Fed, Stanford, NYU, Michigan, Northwestern, Cornell, Fed Board, Atlanta Fed, and Brown. I thank all participants of the Capital Theory Working Group, especially Nancy Stokey. Finally, I thank Juliette Caminade for her support and comments on preliminary versions of this paper.

[†] MIT and NBER. E-mail: maberaja@mit.edu. Web: economics.mit.edu/faculty/maberaja

1 Introduction

Economists commonly follow either *structural* or *reduced-form* approaches in order to answer counterfactual questions. The structural approach relies on specifying primitives of a particular model and identifying its parameters. This is a daunting task whenever researchers are uncertain about features of alternative models that are both difficult to distinguish with available data and reasonable a-priori. If counterfactuals differ under these alternative models, their credibility is undermined because they lack robustness. The reduced-form approach requires less commitment to particular model assumptions and has proven very useful to identify consequences of *observed* policy changes that are robust across models. However, the structural approach is the leading paradigm for studying *unobserved*, potential changes in policy rules because structural counterfactuals are not subject to the critique in Lucas (1976).

Similarly, when building models to evaluate policy rules, researchers have to decide which features of the economy to include and which to leave out. Beyond introspection and matching observed data, there is in general little systematic guidance for such choices. For example, when extending a canonical model to study a policy rule, how can we tell ex-ante which extensions are quantitatively promising? What are the features of models that are relevant for particular policy counterfactuals?

In this paper, I propose a novel method for addressing these and other questions that are common in macroeconomics. The method hinges on an insight about dynamic stochastic models with equilibria that are well approximated by a linear, stable recursive representation. The insight is that models that can match an economy's recursive equilibrium under a benchmark policy—i.e., observationally equivalent models—will also generate an identical counterfactual equilibrium under an alternative policy rule provided that the equations characterizing their equilibria exactly satisfy enough identical linear restrictions. Thus, regardless of their specific micro-foundations, such models will satisfy this principle of *Counterfactual Equivalence*.

The method has two parts. The first is describing theoretical assumptions that lead to linear restrictions on structural equilibrium equations (e.g., exclusion restrictions). This defines a set of models. The second is identifying a counterfactual after a policy change. This requires imposing enough linear restrictions such that all combinations of structural parameters that appear in equilibrium equations are identified given the recursive equilibrium representation under the benchmark policy. Taken together, these identify a subset of observationally equivalent models under a benchmark policy that are also counterfactually equivalent.

I illustrate the principle of Counterfactual Equivalence in a variety of applications. These are chosen both to demonstrate its different uses as well as how to implement its associated method. I discuss them in detail in what follows.

In the first application, I show how to conduct counterfactual analysis with respect to policy rule changes, without having to fully specify a structural model but in a way that is both immune to Lucas critique and robust to specific assumptions about primitives and parameterizations of models in a counterfactually equivalent set. This combines strengths of structural and reduced-form approaches. I call this construction a *Robust Policy Counterfactual* because it allows researchers to obtain quantitative predictions with respect to policy rule changes with minimal a-priori structural assumptions, enhancing credibility of the analysis. This application provides an alternative to fully specified DSGE models whenever researchers are uncertain about some of their features and mechanisms. As such, it echoes the motivation of a recent literature using sufficient statistics in macroeconomics (e.g., Alvarez, Le Bihan, and Lippi (2016); Auclert (2017)), trade (e.g., Arkolakis, Costinot, and Rodríguez-Clare (2012); Adao, Costinot, and Donaldson (2017)), and public finance (e.g., Chetty (2009)).

Furthermore, Robust Policy Counterfactuals are an alternative to the methodology that studies policy rules in the context of Structural Vector Autoregression (SVARs) by "zeroing-out" the response of policy variables to shocks (Bernanke, Gertler, and Watson (1997); Sims and Zha (2006); Kilian and Lewis (2011); Bachmann and Sims (2012)). While less popular in academia because it is not immune to Lucas Critique, the methodology is commonly used in many policy institutions. First, due to its flexibility. Second, because robustness is typically the researchers' main concern. Specially, when studying questions where appropriate state-of-the-art DSGE models have not yet been developed. As I show in Section 4.1, Robust Policy Counterfactuals are an alternative to such methodology because they can be implemented simply by: (i) estimating a SVAR using data from an economy under a benchmark policy, (ii) imposing the appropriate restrictions, and (iii) constructing a counterfactual SVAR for an alternative policy rule. The advantage is that such counterfactual SVAR is immune to Lucas Critique for all model variations that satisfy the imposed restrictions, i.e., those in the counterfactually equivalent set. As such, this application harkens back to the original ideas in the seminal contribution of Sims (1980).

In order to demonstrate how to construct Robust Policy Counterfactuals in practice, I use US state-level data to quantify how fiscal unions contribute to regional stabilization by redistributing resources between its members through the federal tax-and-transfer system. This question has important implications for the future of the European Monetary Union and needs robust quantitative results that go beyond existing reduced-form calculations and calibration exercises in specific models.[1] While I study a class of fiscal union models that are simple enough to illustrate the method in a transparent manner, I shall argue that this class of models is rich enough to inform discussions about the consequences of fiscal integration because it encompasses models with realistic features such as nominal rigidities, adjustment costs, and asset market incompleteness. Examples of fiscal union models that share some of these features are those in Farhi and

---

[1] There is surprisingly little quantitative and empirical research on regional stabilization in fiscal unions. On the purely empirical side, Sala-i Martin and Sachs (1991), Feyrer and Sacerdote (2013), and Bayoumi and Masson (1995) focus on reduced-form estimates of properties of the tax-and-transfer system in fiscal unions, and in some cases, present back-of-the-envelope counterfactuals. Asdrubali, Sorensen, and Yosha (1996) quantify the amount of risk sharing among states in the United States by decomposing cross-sectional variance in output. Atkeson and Bayoumi (1993) examine evidence for private insurance of regional risk in the United States and Europe. Regarding quantitative research, Evers (2015) is closest to the application in this paper, studying regional stabilization in a medium-scale DSGE model. Moreover, Oh and Reis (2012) and McKay and Reis (2013) quantify the impact of targeted transfers and automatic stabilizers in a heterogenous-agent model. However, they are largely concerned with implications for aggregate business-cycle stabilization.

Werning (2014) and Evers (2015). The distinguishing feature of the fiscal union models I study is that a federal tax-and-transfer policy rule summarizes how resources are redistributed between members in a state-contingent manner. After estimating such federal tax-and-transfer policy rule in the US, I construct a counterfactual US economy without it. My main finding is that employment differences across US states in the Great Recession would have been significantly larger in the counterfactual economy because this rule redistributed resources from well to poorly performing states, thus stabilizing regional economies.

In the second application, I show how the principle of Counterfactual Equivalence can (i) guide researchers in their modeling choices regarding which model primitives are promising for studying particular policy counterfactuals and (ii) shed light on why similar quantitative results are obtained in seemingly disparate models. In general, it is hard to gage whether certain primitives are relevant because the mapping from primitives, policies and data to a counterfactual equilibrium is highly *non-linear*. Thus, one has to proceed model by model, estimating each under the benchmark policy and then computing counterfactual equilibria under alternative policies. However, an important insight behind the construction of counterfactually equivalent sets of models is that, given the equilibrium under a benchmark policy, there is a *linear* mapping back to the combinations of structural parameters that appear in equilibrium equations of different models. Then, one can start with a benchmark model and easily identify which classes of primitives would generate counterfactually equivalent models simply by looking at the linear restrictions satisfied by the benchmark model. Irrelevant primitive variations generate models who satisfy identical restrictions. Promising and relevant primitive variations do not.

This application relates to the *business cycle accounting* literature that followed Chari, Kehoe, and McGrattan (2007). While they are concerned with guiding researchers towards developing micro-foundations that are consistent with observed patterns in the data, I am concerned with micro-foundations that are likely to be relevant for policy counterfactuals. The last section of the paper elaborates on this connection further.

To illustrate this application, I present two examples in classic topics in macroeconomics. In the first application, I consider unemployment benefits policy rules in search-theoretical models of the labor market. Starting from a canonical model (Mortensen and Pissarides (1994); Pissarides (2000)), I observe that focusing on primitives that change how wages are set is a promising avenue for building models that generate different effects of unemployment benefits policy rule changes. For example, the sticky wage model in Hall (2005) or the alternating-offer-bargaining model in Christiano, Eichenbaum, and Trabandt (2016). Alternatively, models that only change firms incentives or technologies for creating jobs are likely to be counterfactually equivalent to the canonical model with respect to changes in unemployment benefits policy rules. For example, the financial accelerator model in Wasmer and Weil (2004) or a recruiting intensity model ala Gavazza, Mongey, and Violante (2016). This implies that, if unemployment benefits policy rule changes have negligible effects in the canonical model, then this is a robust quantitative feature of models with varying micro-foundations regarding vacancy posting and job creation but not wage setting and bargaining protocols. In the second application, I consider risk-sharing arrangements in models of international business cycles. A robust result of the international business cycles literature studying the Backus-Kehoe-Kydland and the Backus-Smith puzzles is that the stabilization gains from better risk-sharing are quantitatively small, even in models with very incomplete asset markets as well a several other frictions. Relatedly, Evers (2015) finds small gains from fiscal integration—a particular risk-sharing arrangement—in a medium-scale DSGE model. In contrast, when constructing Robust Transfer Policy Counterfactuals for a set of fiscal union models, I find significant gains in terms of reductions in volatility. This implies that the models in this set are not counterfactually equivalent to those used in the international business cycles literature with respect to changes in risk-sharing rules. I show that the key difference comes from the presence of discount rate shocks in the fiscal union models I study—whereas the literature studying international risk-sharing puzzles has focused, by and large, on productivity shocks. Thus, I conclude that a potential avenue for solving a number of international risk-sharing puzzles is to consider micro-foundations that create wedges in inter-temporal equilibrium equations. In fact, Itskhoki and Mukhin (2017) shows a recent effort along these lines.

In the third application, I move beyond counterfactual analysis of unobserved policy changes. Instead, I argue that we can use observed policy *experiments* in order to falsify a set of models, as long as data before and after a policy change is available and no simultaneous structural changes occurred. Specifically, I show how to construct a simple log-likelihood ratio test in order to reject that a set of models is consistent with the observed equilibrium both before and after the policy change. The set of models is rejected whenever the Robust Policy Counterfactual and the observed equilibrium are "too far apart" in a statistical sense. This application is useful for guiding future structural modeling as well as providing information for picking linear restrictions when constructing Robust Policy Counterfactuals with respect to other unobserved policy changes. Furthermore, it provides a formal econometric framework that exploits policy experiments in order to falsify common models in macroeconomics, going beyond more informal and narrative approaches.

As an illustration, I falsify a set of New Keynesian models using the monetary policy experiment studied in the seminal paper by Clarida, Gali, and Gertler (2000). They show that the US interest rate policy rule changed after Paul Volcker's was appointed at the Fed. Then, feeding this interest rule change into a canonical New Keynesian model, they argue that such policy change stabilized inflation and output. However, I show that such canonical model belongs to a larger set of New Keynesian models that are falsified by this same policy experiment because they are not consistent with the observed behavior of output, inflation, and interest rates both before and after the policy change.

Finally, in the last part of the paper, I compare the implications of the principle of Counterfactual Equivalence to other, well established methodologies that have dealt with related issues of robustness, identification and model misspecification.

The rest of this paper is organized as follows. Section 2 introduces the principle of Counterfactual Equivalence in a simple example. Section 3 lays out the general method for constructing a set of counterfactually equivalent models. Section 4 shows how to construct Robust Policy Counterfactuals and the application to regional stabilization in fiscal unions. Section 5 discusses how to analyze micro-foundations that are relevant for particular policy counterfactuals. Section 6 shows how to falsify a set of models using policy experiments. Section 7 compares this paper to related methodologies.

2 Illustrating the Principle of Counterfactual Equivalence

The goal of this section is to illustrate the principle of counterfactual equivalence in two particular examples where models have very different primitives and economic mechanisms that are nevertheless irrelevant for the positive implications of a policy rule change. In the first example, I consider interest policy rules in the context of simple 3-equation New Keynesian models. In the second example, I consider models of a small open economy that receives lump-sum transfers from abroad according to a transfer policy rule. This example also relates to the application to fiscal unions in Section 4.1 because, to a first-order approximation, I show that one can think of this small open economy with a transfer policy rule as part of a fiscal and monetary union. The counterfactual equilibrium without a rule in place corresponds to a monetary union without fiscal integration.

2.1 Interest rate rule counterfactuals in New Keynesian models

The canonical New Keynesian model (e.g., Clarida, Gali, and Gertler (2000)) describes the equilibrium behavior of inflation π_t, output y_t and nominal interest rate r_t in log-deviations from a zero inflation steady state in terms of three equations,

$$
\begin{aligned}
0 &= -\mathbb{E}_t(y_{t+1}) - \mathbb{E}_t(\pi_{t+1}) + y_t + i_t + (\rho_\gamma - 1)\gamma_t && \text{(Euler)} \\
0 &= \beta\mathbb{E}_t(\pi_{t+1}) + \kappa y_t - \pi_t + z_t && \text{(NKPC)} \\
0 &= -i_t + \theta_y y_t + \theta_\pi \pi_t && \text{(Policy rule)}
\end{aligned}
$$

where z_t, γ_t are independent AR(1)'s and θ_p, θ_y are interest rate policy rule parameters.

The standard approach for analyzing counterfactual interest rate rules begins by estimating the structural parameters using data under a benchmark policy rule. Then, one would solve for the recursive representation of the equilibrium under alternative rules and compare properties of such alternative equilibria (e.g., output and inflation volatility). Such approach begets a natural question: would results change if one used a different model?

This concern about robustness is particularly relevant whenever researchers are uncertain about features of the economy that cannot be distinguished with data on output, inflation, and interest rates alone. Assuming no other data is available nor any a-priori reasons can be summoned up, these observationally equivalent models cannot be dismissed.

As I show in what follows, the principle of Counterfactually Equivalence can aid researchers in gaging the robustness of their counterfactual results by identifying which model features are quantitatively relevant and which ones are not. Relatedly, it is useful for thinking about promising extensions or "backward-engineering" new models that change the counterfactual conclusions obtained from a canonical model.[2]

I begin by considering three alternative New Keynesian models. The different primitives of these models change the New Keynesian Phillips Curve.

In the first model, there is a large family of agents in the households. A fraction of the agents γ_t are unemployed and consume the transfers from the family alone. Exogenous variation in γ_t over time generates variation in the inter-temporal marginal rate of substitution of consumption and the intra-temporal marginal rate of substitution between consumption and labor. This implies that that γ_t now appears in the (NKPC) equation as well.

$$0 = \beta \mathbb{E}_t(\pi_{t+1}) + \kappa y_t - \pi_t + z_t + \frac{\kappa \bar{\gamma}}{1+\bar{\gamma}} \gamma_t \tag{NKPC1}$$

The second model follows from Gabaix (2016). He derives a behavioral version of the NKPC when firms are partially inattentive to future macro conditions.

$$0 = \beta M^f E_t(\pi_{t+1}) + \kappa m^f y_t - \pi_t + z_t \tag{NKPC2}$$

The degree of inattention is parameterized by M^f and m^f. For example, the behavioral NKPC features extra discounting because $M^f < 1$ when compared to the canonical model.

In the third model, firms face a working-capital-in-advance constraint. The constraint forces them to hold a fraction χ of their sales in cash, overnight. Such cash holdings depreciate because of inflation, thus adding to the effective marginal cost.

$$0 = \beta \mathbb{E}_t(\pi_{t+1}) + \kappa y_t + (\kappa \chi - 1)\pi_t + z_t \tag{NKPC3}$$

Next, imagine the following thought experiment. We parameterize the behavioral model and solve for the recursive representation of the equilibrium $\{y_t, \pi_t\}$ under the benchmark policy θ^0_y, θ^0_π. Since the behavioral model (and all the others in this section) do not have endogenous state variables, the recursive representation of the equilibrium simply takes the form of:

$$\begin{bmatrix} y_t \\ \pi_t \end{bmatrix} = Q^0 \begin{bmatrix} \gamma_t \\ z_t \end{bmatrix} \tag{RRNK}$$

For the purposes of this section, I will assume that there always existe a recursive representation and that such representation is stable and unique.[3] Then, consider an alternative policy θ^1_y, θ^1_π.

---

[2] Section 5 shows an application of these ideas to unemployment benefits policy rules in search models of the labor market.

[3] Well known issues of multiplicity, indeterminacy, and instability in New Keynesian models are beyond the scope

Keeping all other parameters unchanged, we also construct the recursive representation of the equilibrium Q^1 under the alternative policy.

Claim 1 (Observational and Counterfactual Equivalence in New Keynesian models) *Let Q^0 and Q^1 be the recursive representation of the equilibria generated by the behavioral model under policies $\{\theta_y^0, \theta_\pi^0\}$ and $\{\theta_y^1, \theta_\pi^1\}$. Then,*

(1) (Observational Equivalence) *Both the unemployed-agents model and the working-capital model can be parameterized to match Q^0 under policy $\{\theta_y^0, \theta_\pi^0\}$.*

(2) (Counterfactual Equivalence) *Given such parameterization, the working-capital model also generates Q^1 under policy $\{\theta_y^1, \theta_\pi^1\}$. The unemployed-agents model does not.*

The first claim says that all three models satisfy a strong form of observational equivalence. These models could not be distinguished even if we observed both an infinite time-series of inflation and output, and the structural shocks that generated such time-series. This is because they have an identical recursive representation under the benchmark policy.[4]

The second claim says that the behavioral and working-capital model are counterfactually equivalent with respect to changes in interest rate policy rules, but the unemployed-agents model is not.

In Section 3, I provide formal definitions of these principles for general linear models of dynamic stochastic economies, as well as how to describe sets of models that satisfy such principle of Counterfactual Equivalence. Here, I present a simplified exposition in order to understand Claim 1.

I begin by noting that, more generally, we can describe a set of New Keynesian models—including the previous three as special cases—as being associated with the following unrestricted system of equilibrium equations,

$$0 = F\mathbb{E}_t \begin{bmatrix} y_{t+1} \\ \pi_{t+1} \end{bmatrix} + (G + \Theta) \begin{bmatrix} y_t \\ \pi_t \end{bmatrix} + M \begin{bmatrix} \gamma_t \\ z_t \end{bmatrix} \qquad \begin{matrix} \text{(Euler)} \\ \text{(NKPC)} \end{matrix}$$

$$0 = -\begin{bmatrix} \gamma_t \\ z_t \end{bmatrix} + N \begin{bmatrix} \gamma_{t-1} \\ z_{t-1} \end{bmatrix} + \Sigma \begin{bmatrix} u_t^\gamma \\ u_t^z \end{bmatrix}$$

Let $\xi \equiv \{F, G, M, N, \Sigma\}$ be a *structure*. It includes all combinations of structural, policy-invariant parameters (e.g., $\{\kappa, \beta\}$). Let Θ be a *policy*. It only includes the parameters in the interest rate policy rule (e.g., $\{\theta_y, \theta_\pi\}$). Furthermore, given (ξ, Θ), the recursive representation of the equilibrium

---

of this section.

[4] A weaker (and more typical) form of observational equivalence would only require that these models cannot be distinguished from observations of inflation and output alone. Even if they had different recursive representations, they could match the same data with alternative realizations of the structural shocks.

$Q(\xi, \Theta)$ is the solution to the *non-linear* system of equations:[5]

$$FQ(\xi, \Theta)N + (G + \Theta)Q(\xi, \Theta) + M = 0$$

Because this non-linear mapping from a structure and policy onto the recursive equilibrium is unique, the question of counterfactual equivalence is a question about the structure. Regardless of their particular micro-foundations and parameters, models with identical structures are counterfactually equivalent with respect to an interest rate policy rule change. Thus, to show Claim 1, we need to show that the behavioral and working-capital models can be parameterized to have identical structures, whereas the same is not true for the unemployed-agents model. The key insight in doing so is realizing that the reverse mapping from a recursive representation and a policy onto the structure is *linear*. That is, given Q^0, N^0 and Θ^0, we have a linear system in F, G, M

$$FQ^0N^0 + (G + \Theta^0)Q^0 + M = 0$$

This linear system has six unknowns per line in the structure but only two equations. While one of the unknowns can be normalized to 1, the system is still underdetermined. This simply means that unrestricted models cannot be identified from Q^0, N^0, Θ^0 alone. However, if particular 3-equation New Keynesian models impose enough restrictions on the structure, then the system is exactly determined. In this case, three restrictions per line is enough to uniquely identify the structure, given Q^0, N^0, Θ^0. Thus, we can conclude that all 3-equation New Keynesian models that satisfy *any* three restrictions per line (or less) are observationally equivalent under policy Θ^0 because they can match the exact same recursive representation (provided they have a sufficient number of parameters to do so). However, from this set of observationally equivalent models under policy Θ^0, if two models satisfy *identical* three restrictions per line, then they will also generate an identical counterfactual equilibrium $Q(\xi, \Theta^1)$ under the alternative policy Θ^1 because they have an identical structure, i.e., these models are counterfactually equivalent. To summarize, a counterfactually equivalent set of models is a subset of observationally equivalent models that is characterized by enough restrictions on the structure. Models that exactly satisfy those restrictions are counterfactually equivalent under alternative policies. Models that don't are not.

Armed with this characterization, we can now easily show Claim 1 simply by inspecting the New Keynesian Phillips Curve associated with the behavioral, unemployed-agents, and working-

---

[5] This is simply an application of the method of undetermined coefficients.

capital models (the Euler equation is identical in all three and can be omitted):

$$\begin{bmatrix} 0 & \beta \end{bmatrix} \mathbb{E}_t \begin{bmatrix} y_{t+1} \\ \pi_{t+1} \end{bmatrix} + \begin{bmatrix} \frac{\kappa m^f}{M^f} & -\frac{1}{M^f} \end{bmatrix} \begin{bmatrix} y_t \\ \pi_t \end{bmatrix} + \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} z_t \\ \gamma_t \end{bmatrix} = 0 \qquad \text{(Behavioral)}$$

$$\begin{bmatrix} 0 & \beta \end{bmatrix} \mathbb{E}_t \begin{bmatrix} y_{t+1} \\ \pi_{t+1} \end{bmatrix} + \begin{bmatrix} \kappa & \kappa\chi - 1 \end{bmatrix} \begin{bmatrix} y_t \\ \pi_t \end{bmatrix} + \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} z_t \\ \gamma_t \end{bmatrix} = 0 \qquad \text{(Working-capital)}$$

$$\begin{bmatrix} 0 & \beta \end{bmatrix} \mathbb{E}_t \begin{bmatrix} y_{t+1} \\ \pi_{t+1} \end{bmatrix} + \begin{bmatrix} \kappa & -1 \end{bmatrix} \begin{bmatrix} y_t \\ \pi_t \end{bmatrix} + \begin{bmatrix} 1 & \frac{\kappa\bar{\gamma}}{1+\bar{\gamma}} \end{bmatrix} \begin{bmatrix} z_t \\ \gamma_t \end{bmatrix} = 0 \qquad \text{(Unemployed-agents)}$$

Finally, note that, when compared with the behavioral model, both the unemployed-agents and working-capital models impose the same number of restrictions in the structure and have the same number of free parameters. Because of the linearity in the reverse-mapping, this implies that both models can be parameterized to match Q^0. Thus, they are observationally equivalent. This is straightforward to see, for example, by inspecting the second equation in the linear systems associated with each model. Furthermore, all models impose exactly two restrictions on the structure of the New Keynesian Phillips Curve. However, the working-capital and behavioral models impose identical restrictions—the coefficients associated with $\mathbb{E}_t[y_{t+1}]$ and γ_t are zero —whereas the unemployed agents model imposes one that is different —the coefficient associated with π_t is equal to -1. This implies that the identified structure of the working-capital to match Q^0 will be identical to that of the behavioral model whereas the unemployed-agents model's identified structure will differ. As such, the working-capital and behavioral models will generate an identical counterfactual equilibrium Q^1 under an alternative policy Θ^1 whereas the unemployed-agents model will not.

2.2 Interest rate rule counterfactuals in SVARs

Using data from an economy under a benchmark interest rate rule $\{\theta_0^y, \theta_0^\pi\}$, a researcher has estimated the following SVAR representation of output y_t and inflation π_t:[6]

$$\begin{bmatrix} y_t \\ \pi_t \end{bmatrix} = \rho^0 \begin{bmatrix} y_{t-1} \\ \pi_{t-1} \end{bmatrix} + Q^0 \begin{bmatrix} u_t^z \\ u_t^\gamma \end{bmatrix}$$

The researcher would like to evaluate how the behavior of inflation and output would change in a counterfactual economy with a more "hawkish" interest rate policy that only responds to changes in inflation (i.e., $\theta_1^y = 0$ and $\theta_1^\pi > \theta_0^\pi$). Because she is uncertain about many features of the economy (e.g., do price-setters have behavioral biases?), she is reluctant to specify a particular structural model where the counterfactual results may not be robust to variation in such features.

---

[6] While estimating ρ^0 only requires estimating the reduced-form VAR with data on inflation and output alone, identifying the impulse response matrix Q^0 is less straightforward. For the purposes of this section, we can imagine that the researcher had access to a "supply shock" that is correlated with u_t^z but is orthogonal to u_t^γ. Using this shock and the covariance matrix of the reduced form VAR errors, it is possible to identify Q^0 as is commonly done in the literature.

However, she is also worried that leading methods for performing this type of counterfactual exercise in SVARs are not appropriate because they are not immune to Lucas Critique (e.g., Bernanke, Gertler, and Watson (1997); Sims and Zha (2006); Kilian and Lewis (2011); Bachmann and Sims (2012)). Especially, since she is not concerned with the short-run equilibrium dynamics immediately after the interest rate policy rule change where one could argue that agents may not have understood that such change has occurred. Instead, she is concerned about how the behavior of inflation and output would change in the new long-run equilibrium once all agents understand this.

In Section 4, I propose a novel methodology for tackling this problem that is based on the principle of counterfactual equivalence. The methodology allows researchers to construct *Robust Policy Counterfactuals*, i.e., counterfactuals with respect to policy rule changes that are both immune to Lucas Critique and robust to variations across models that belong to the same counterfactually equivalent set. Next, I present a simplified exposition as it applies to 3-equation New Keynesian models.

I start by assuming that the estimated SVAR $\{\rho^0, Q^0\}$ was generated by some model that belongs to the class of 3-equation New Keynesian models described in the previous section. As such, its equilibrium equations can be described by a structure and a policy $\{\xi, \Theta\}$. Furthermore, if the impulse response matrix $Q(\xi, \Theta)$ is invertible, then the equilibrium has a finite SVAR representation. This is simply:

$$\begin{bmatrix} y_t \\ \pi_t \end{bmatrix} = Q(\xi,\Theta) \begin{bmatrix} z_t \\ \gamma_t \end{bmatrix} \equiv \rho(\xi,\Theta) \begin{bmatrix} y_{t-1} \\ \pi_{t-1} \end{bmatrix} + Q(\xi,\Theta) \begin{bmatrix} u_t^z \\ u_t^\gamma \end{bmatrix}$$

where $\rho(\xi,\Theta) \equiv Q(\xi,\Theta) N Q(\xi,\Theta)^{-1}$.

First, note that N^0 is simply recovered as $N^0 = (Q^0)^{-1}\rho^0 Q^0$. Second, as in the previous section, given N^0, Q^0, Θ^0, we can write the linear system of equations in the structure ξ as:

$$FQ^0N^0 + (G + \Theta^0)Q^0 + M = 0$$

As we argued before, the structure is identified by imposing any three restrictions per line. For example, consider the restrictions satisfied by the behavioral and working-capital models from the previous section:

$$F = \begin{bmatrix} f_{11} & -1 \\ 0 & \beta \end{bmatrix}; G = \begin{bmatrix} g_{11} & 0 \\ g_{21} & g_{22} \end{bmatrix}; M = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

where β is a particular value for the discount factor (e.g., $\beta = 0.95$).

Then, given the identified restricted structure ξ^s and the alternative "hawkish" interest rate policy rule $\Theta^1 = \begin{bmatrix} 0 & \theta_\pi^1 \\ 0 & 0 \end{bmatrix}$, we can solve for the counterfactual SVAR impulse response matrix

$Q(\xi^s, \Theta^1)$ by solving the non-linear system:

$$F^s Q(\xi^s, \Theta^1) N^0 + (G^s + \Theta^1) Q(\xi^s, \Theta^1) + M^s = 0$$

Finally, the counterfactual autoregressive matrix $\rho(\xi^s, \Theta^1) = Q(\xi^s, \Theta^1) N^0 Q(\xi^s, \Theta^1)^{-1}$.

To conclude, the counterfactual SVAR representation under the alternative interest rate policy rule is robust to variations across many 3-equation New Keynesian models, provided they (1) can match the SVAR representation under the benchmark policy rule and (2) satisfy identical three restrictions per line in their structure. That is, models that belong to the same counterfactually equivalent set.

3 The Principle of Counterfactual Equivalence: A General Approach

This section describes how to characterize a set of models that are *Counterfactually Equivalent* with respect to a policy change within a class of linear models of dynamic stochastic economies. First, I define this class of models, and develop notation and language that I use throughout the section. Second, I offer a formal definition of Counterfactual Equivalence. I conclude by showing how to construct a set of counterfactually equivalent models, given data from an economy under a benchmark policy rule and structural restrictions on the models that could have generated this data.

I consider a class of models whose equilibrium is characterized (or approximated) by the linear system of equations,

$$\begin{aligned}
0 &= (F + \Theta_f)\mathbb{E}_t[x_{t+1}] + (G + \Theta_c)x_t + (H + \Theta_p)x_{t-1} + L\mathbb{E}_t[z_{t+1}] + Mz_t \\
0 &= -z_t + Nz_{t-1} + u_t; \quad \text{iid } u_t \text{ with } \mathbb{E}[u_t] = 0, Var(u_t) = \Sigma \\
0 &= y_t + R_1 x_t + R_2 x_{t-1} + R_3 z_t \\
&\quad (H + \Theta_p)x_0; Mz_0 \text{ given}
\end{aligned} \tag{SME}$$

where x_t is a column vector that includes all endogenous state variables, and could include endogenous control variables too; z_t is a column vector of exogenous state variables; and y_t a column vector of other endogenous variables not included in x_t. For example, in the context of Hansen (1985) real business cycles model, x_t includes log-deviations from the steady state of capital and potentially employment, z_t includes productivity, and y_t consumption. Alternatively, in the application to fiscal unions in Section 4.1, x_t will include employment, wages and assets at the state-level in log-deviations from the aggregate union equilibrium, and z_t will include the exogenous household discount rate, productivity and wealth processes.

Given initial conditions, I say that a particular economy is characterized by the *structure* $\{\xi, \xi^a\}$ and the *policy* Θ, where $\xi \equiv \{F, G, H, L, M, N, \Sigma\}$, and $\xi^a \equiv \{R_1, R_2, R_3\}$ are policy-invariant matrices, and $\Theta \equiv \{\Theta_f, \Theta_c, \Theta_p\}$ are matrices that characterize endogenous policy rules.

Θ_f contains the policy parameters associated with future expected variables, Θ_c with contemporaneous variables, and Θ_p past variables. Elements in $\{\xi, \xi^a\}$ typically involve combinations of subsets of parameters derived from a fully specified model. They typically lack direct economic interpretation in terms of the primitives of such a model. Elements in Θ generally have a direct economic interpretation. For example, in the application to fiscal unions, the structure would include non-linear combinations of labor supply and demand elasticities, as well as parameters related to the degree of nominal wage rigidity. The policy, on the other hand, would include the transfer policy rule elasticities.

The system (SME) evaluated at $\Theta = 0$ corresponds to a benchmark economy without a policy in place.[7] Moreover, because of the linearity of the system, we can always write the policy as the difference between two economies. For example, if there is a benchmark economy with $F_0 \in \xi_0$ and and alternative economy with $F_1 \in \xi_1$, then we can define $\Theta_f = F_1 - F_0$ without loss of generality.

Next, I restrict the class of linear models to those that satisfy two assumptions.

Assumption 1 (Stability) *ξ, Θ are such that the system (SME) is stabilizable.*

Under Assumption 1, we have a stable recursive solution to (SME) that can be written as:

$$\begin{aligned} x_t &= P(\xi, \Theta)x_{t-1} + Q(\xi, \Theta)z_t \\ z_t &= Nz_{t-1} + u_t \end{aligned} \tag{RR}$$

where $P(\xi, \Theta)$ and N have all eigenvalues inside the unit circle.

Assumption 2 (Uniqueness) *ξ, Θ are such that (RR) is unique.*

These assumptions rule out, for example, models with unstable dynamics and/or multiple equilibria. These are beyond the class of models studied in this paper.

Next, I ask: what models are both observationally equivalent under a benchmark policy Θ^0 and have an identical counterfactual equilibrium under an alternative policy Θ^1? A set of Counterfactual Equivalent models is an answer to this question in models whose equilibrium is characterized (or approximated) by a system like (SME), and whose structure and policy ξ, Θ satisfy Assumptions 1 and 2. In particular, I show that the equilibrium after the policy change is identical in a set of observationally equivalent models whose structure ξ satisfies a number of linear restrictions.

The first step is realizing that, regardless of their primitives, models with an identical structure ξ are observationally equivalent under policy Θ because they have an identical recursive representation of the equilibrium $\Gamma(\xi, \Theta) \equiv \{P(\xi, \Theta), Q(\xi, \Theta), N\}$. Then, for a given benchmark policy Θ^0 and recursive representation Γ^0, we have the following definition,

---

[7] Uhlig (1995), from whom I borrow some notation, studies a very similar system of equations, and provides a computational toolkit for finding recursive solutions.

Definition 1 (Observational Equivalence) *Given* $\{\Theta^0, \Gamma^0\}$*, define* $\mathcal{O}(\Theta^0, \Gamma^0)$ *as the set of structures* ξ *that generate equilibrium representation* Γ^0 *under benchmark policy* Θ^0*.*

By definition, all models which have structures in $\mathcal{O}(\Theta^0, \Gamma^0)$ are observationally equivalent. From an econometrician's perspective, this definition of observationally equivalent models is stronger than the usual one because models with structures in $\mathcal{O}(\Theta^0, \Gamma^0)$ are not identified even if the *structural* recursive representation of the equilibrium was known, whereas the usual definition would involve knowledge of the *reduced-form* recursive representation alone.

Next, I define a subset of these observationally equivalent models that yield identical counterfactual equilibrium representation Γ^1 under an alternative policy Θ^1 and show how to identify it.

Definition 2 (Counterfactual Equivalence) *Given* $\{\Theta^0, \Gamma^0\}$ *and* $\{\Theta^1, \Gamma^1\}$*, define* $\mathcal{C}(\Theta^0, \Gamma^0, \Theta^1, \Gamma^1)$ *as the subset of structures* ξ *in* $\mathcal{O}(\Theta^0, \Gamma^0)$ *that generate identical counterfactual equilibrium recursive representation* Γ^1 *under alternative policy* Θ^1*. Formally,* $\mathcal{C}(\Theta^0, \Gamma^0, \Theta^1, \Gamma^1) = \mathcal{O}(\Theta^0, \Gamma^0) \bigcap \mathcal{O}(\Theta^1, \Gamma^1)$*.*

Definition 3 (Semi-structure) *Let* R_{n_l} *be a set of* n_l *independent linear restrictions on the coefficients in line 'l' of structure* ξ*. Then, define a semi-structure* ξ^s *as a structure* ξ *that satisfies* $R \equiv \{R_{n_l}\}_{l=1}^{L}$ *for each line* $l \in \{1, ..., L\}$ *of* ξ*.*

Proposition 1 (Identification of Counterfactually Equivalent Set of Models)

Given $\{\Theta^0, \Theta^1, \Gamma^0, R^*\}$*, a structure* ξ^* *is the unique solution to* $\xi^* = \mathcal{C}(\Theta^0, \Gamma^0, \Theta^1, \Gamma(\xi^*, \Theta^1))$ *if and only if:*

(i). ξ^* *is such that* $\Gamma(\xi^*, \Theta^0) = \Gamma^0$

(ii). ξ^* *is a semi-structure that satisfies linear restrictions* R^*

(iii). R^* *has* $n_l = 2k - (1 - rank(\Theta^0_l))$ *independent linear restrictions for each line* $l \in \{1, ..., L\}$

Thus, a model belongs to a Counterfactually Equivalent Set if and only if the model can be parameterized to generate the equilibrium recursive representation Γ^0 *under policy* Θ^0 *and its structure satisfies the linear restrictions* R^**.*

Proof. See Appendix A.

Definitions 1-3 and Proposition 1 state that we can take the set of models that are observationally equivalent under a benchmark policy and construct subsets of models by imposing linear restrictions on their structure, where models within each subset yield an identical counterfactual equilibrium under any alternative policy. In other words, these counterfactually equivalent sets include models that are observationally equivalent under both the benchmark and alternative policies.

The key to proving Proposition 1 is noticing that in order to find the equilibrium recursive representation Γ^0 under policy Θ^0, we typically solve a *non-linear* system of equations, *given* structure and policy $\{\xi^0, \Theta^0\}$. However, the system of equations is *linear* in structure ξ^0, *given* values for the recursive representation Γ^0 and policy Θ^0. Because of this linearity, we can identify the structure ξ^* that is consistent with a set of observationally equivalent models under a benchmark policy that also yield an identical counterfactual equilibrium under an alternative policy, as long as we impose enough linear restrictions on the structure in order to make the linear system exactly determined.

To summarize, the principle of Counterfactual Equivalence says that if two models are observationally equivalent under a benchmark policy rule, then they will generate an identical counterfactual equilibrium under an alternative policy rule provided that the equations characterizing equilibria in both models exactly satisfy the same number of identical restrictions. The principle has a rather surprising (and philosophical) implication. Imagine that the "true" model of the economy was only known to God. We, merely mortal researchers, have created our own model and can observe the true model's equilibrium of an economy under a benchmark policy rule. Our model can reproduce this observed equilibrium but, while observationally equivalent to the true model, it is potentially different. The principle implies that even if we analyzed a counterfactual policy rule change using our *wrong* model, we will obtain the *right* answer—the one that God would have obtained—as long as both God's and our model belong to the same counterfactually equivalent set.

4 Application I: Robust Policy Counterfactuals

In this section, I propose a methodology to conduct counterfactual analysis with respect to unobserved systematic policy changes in a way that is both immune to Lucas critique and robust to specific assumptions about primitives and parameterizations. Specifically, given data from the equilibrium under a benchmark policy, I show how to construct the counterfactual equilibrium under an alternative policy that is generated by models in a counterfactually equivalent set. I call this a *Robust Policy Counterfactual*.

Formally, using the language and notation from Section 3,

Definition 4 (Robust Policy Counterfactual) *Let ξ^* be a semi-structure satisfying linear restrictions R^* and $\xi^* = \mathcal{C}(\Theta^0, \Gamma^0, \Theta^1, \Gamma(\xi^*, \Theta^1))$. Then, a Robust Policy Counterfactual is a mapping from $\{\Gamma^0, \Theta^0, \Theta^1, R^*\}$ onto $\Gamma(\xi^*, \Theta^1)$.*

Because the proof of Proposition 1 is by construction, it implies a methodology for constructing a Robust Policy Counterfactual. Moreover, because the semi-structure ξ^* is identified, then this mapping is also unique.

Next, I illustrate the methodology and its implementation. Using state-level US data, I quantify how fiscal unions contribute to regional stabilization by redistributing resources across its members through a transfer policy rule.

4.1 Regional Stabilization in Fiscal Unions

This section shows how to identify a set of fiscal union models that are observationally equivalent under fiscal integration and yield identical counterfactual equilibrium in an economy that is not fiscally integrated. I start by describing models of fiscal unions in some generality and then apply results from Section 3 to construct a Robust Transfer Policy Counterfactual. I will focus on models that satisfy four properties. Examples of models with some of these properties include Farhi and Werning (2014), Nakamura and Steinsson (2014), Beraja, Hurst, and Ospina (2015), and Evers (2015). In particular, in Online Appendix A, I present a fiscal union model that satisfies them all. .

Assumption 3 *Models of fiscal unions satisfy the following properties:*

1. **Transfer policy rule***: Tax-and-transfer system can be summarized as federal lump-sum transfers that are a function of state-level economic variables.*
2. **Linear aggregation:** *State-level economies in log-deviations from the aggregate union behave to a first-order approximation as if they were small open economies—independent of other states.*
3. **3 by 3:** *Employment* n_t*, nominal wages* w_t*, and assets* b_t*; and exogenous processes* $\{\gamma_t, \eta_t, z_t\}$ *are sufficient variables for characterizing the state-level equilibrium in log-deviation from aggregates.*
4. **SVAR:** *The log-linearized equilibrium has a unique, finite, and stable structural vector autoregression representation.*

Properties 1-3 imply that we can characterize the equilibrium in any state in log-deviations from the aggregate with a system of equations as (SME). Property 4 requires one additional assumption over Assumptions 1 and 2. In particular, it requires that the impulse response matrix $Q(\xi, \Theta)$ be a non-singular square matrix so that the exogenous processes can be written as linear combinations of past variables and structural innovations. Such additional assumption allows us to estimate a SVAR in order to infer Γ^0 under the benchmark policy Θ^0—a necessary input for constructing a Robust Policy Counterfactual.

Property 1 excludes models where the tax-and-transfers system affects decisions at the margin, as it would be the case with distortionary taxation. In Section 4.1.2, I relax this assumption and analyze the sensitivity of the results. Property 2 excludes from the analysis models where member states are inherently different because of industrial composition or household preferences, for example, and/or models whose exogenous processes correlation structures across states are such that idiosyncratic shocks do not average out. In Property 3, assets might encompass both non-state-contingent nominal bonds and certain types of tradable physical capital. What is important for the property to hold is that no other variables that are necessary to describe the equilibrium are left out (e.g., other endogenous or exogenous state variables in the model). Property 4 excludes, for instance, models with non-fundamental equilibrium representations.

Given the description of models of fiscal unions above, the system of matrix equations that characterizes the equilibrium in a state in log-deviations from the aggregate is written below. Without loss of generality, I will say that the first equation is the (Euler) equation in these models, the second is the (Labor Market) equation, and the third is the sequential budget constraint (SB) equation. This system is analogous to the one characterizing equilibria in the models of a small open economy from Section 2.

$$
\begin{aligned}
0 &= (F+\Theta_f)\mathbb{E}_t\begin{bmatrix} n_{t+1} \\ w_{t+1} \\ b_{t+1} \end{bmatrix} + (G+\Theta_c)\begin{bmatrix} n_t \\ w_t \\ b_t \end{bmatrix} + (H+\Theta_p)\begin{bmatrix} n_{t-1} \\ w_{t-1} \\ b_{t-1} \end{bmatrix} + L\mathbb{E}_t\begin{bmatrix} \gamma_{t+1} \\ z_{t+1} \\ \eta_{t+1} \end{bmatrix} + M\begin{bmatrix} \gamma_t \\ z_t \\ \eta_t \end{bmatrix} \\
0 &= -\begin{bmatrix} \gamma_t \\ z_t \\ \eta_t \end{bmatrix} + N\begin{bmatrix} \gamma_{t-1} \\ z_{t-1} \\ \eta_{t-1} \end{bmatrix} + \Sigma\begin{bmatrix} u_t^{\gamma} \\ u_t^{z} \\ u_t^{\eta} \end{bmatrix} \qquad \text{(FiscalSME)}
\end{aligned}
$$

Given Property 1 in Assumption 3, I will assume that the transfer policy rule depends not only on state-level employment, as in the small open economy example from Section 2, but also on current wages and asset holdings at the beginning of the period. Since the third equation is the sequential budget constraint, this implies:

Assumption 4 *The policy* Θ *is:* $\Theta_c = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ \vartheta_n & \vartheta_w & 0 \end{bmatrix}$, $\Theta_p = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \vartheta_b \end{bmatrix}$, *and* $\Theta_f = 0_{3.3}$

Next, I consider linear restrictions on the structure in (FiscalSME) that take the form of exclusion restrictions. This implies that the semi-structure ξ^* consistent with a set of counterfactually equivalent models is described by certain elements in ξ being zeros. Specifically,

Assumption 5 *A set of fiscal union models that can be written as in (FiscalSME) is described by:*

$$
F = \begin{bmatrix} f_{11} & f_{12} & 0 \\ f_{21} & f_{22} & 0 \\ 0 & 0 & 0 \end{bmatrix}; G = \begin{bmatrix} g_{11} & g_{12} & 0 \\ g_{21} & g_{22} & 0 \\ g_{31} & g_{32} & g_{33} \end{bmatrix}; H = \begin{bmatrix} h_{11} & 0 & 0 \\ h_{21} & h_{22} & 0 \\ 0 & 0 & h_{33} \end{bmatrix};
$$

$$
L = \begin{bmatrix} 1 & l_{12} & l_{13} \\ 0 & 1 & l_{23} \\ 0 & 0 & 0 \end{bmatrix}; M = \begin{bmatrix} 0 & m_{12} & m_{13} \\ 0 & m_{22} & m_{23} \\ 0 & 0 & m_{33} \end{bmatrix}; N = \begin{bmatrix} n_{11} & n_{12} & n_{13} \\ 0 & n_{22} & n_{23} \\ 0 & n_{32} & n_{33} \end{bmatrix}
$$

The absence of expected and lagged terms beside assets makes the third equation consistent with most log-linearized, incomplete market models that include a sequential budget constraint (SB). Moreover, I assume that the only exogenous shifter in the sequential budget constraint is "wealth" process η_t. The other two exogenous processes do not appear in the sequential budget constraint. In terms of the (Euler) and (Labor Market) equations—the first and second equations—I assume that (i) assets (future, contemporaneous, or lagged) $\{b_{t+1}, b_t, b_{t-1}\}$ do not appear in them, (ii) lagged wages w_{t-1} does not appear in the first equation, (iii) contemporaneous "discount rate" process γ_t do not shift these equations, and (iv) future "discount rate" process γ_{t+1} do not appear in the second equation. Finally, I assume that past "discount rate" process γ_{t-1} does not cause movements in "productivity" z_t and "wealth" η_t as is evidenced by autoregressive matrix N. These assumptions are consistent with the model of fiscal unions described in Appendix A as well as many others. The key feature of models described by this semi-structure is that the (Labor Market) and (Euler) equations dependence on future and lagged variables is relatively unconstrained, as well as the exogenous processes and their correlation structure. This is important for the question of regional stabilization in fiscal unions because it means that this set encompasses many models with rich features. For example, models with varying micro-foundations for nominal rigidities that are both forward- and backward-looking, different utility functions, and so on.

4.1.1 A Counterfactual United States Economy without Fiscal Integration

Using US state-level data on employment, wages, assets, taxes and transfers, I construct a counterfactual US economy without a transfer policy rule in place. This counterfactual economy corresponds to the equilibrium of a set of counterfactually equivalent models that satisfy the assumptions from the previous section. In Appendix B, I describe in detail the data I use as well as how I estimate all inputs necessary for constructing this Robust Transfer Policy Counterfactual. In particular, (i) estimating a regional SVAR, (ii) estimating the transfer policy rule Θ, and (iii) identifying the semi-structure ξ^*.

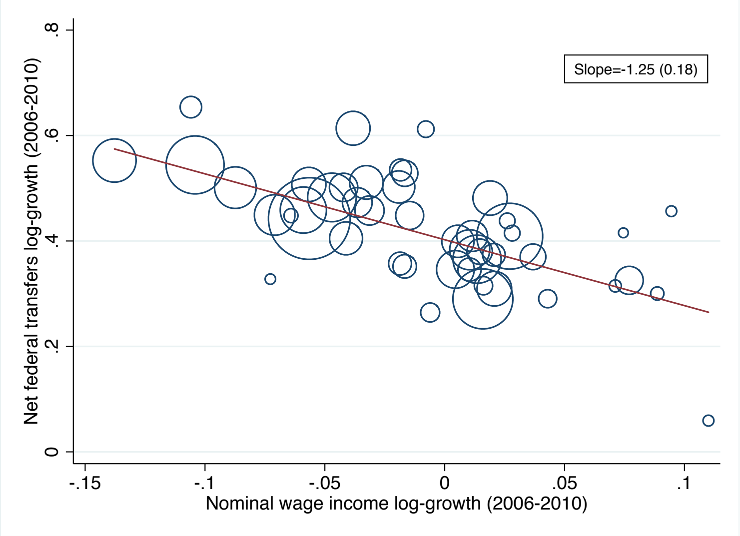
Next, I present primary findings from the counterfactual exercise. As a primer, Figure 1 shows a scatter plot of net federal transfers growth (direct federal transfers minus federal taxes growth) between 2006 and 2010 against nominal wage income growth (wage plus employment growth) between 2006 and 2010 in the United States.[8]. There is a very strong, negative relationship between the two. In particular, a one percentage point increase in nominal wage income associates with a 1.25 percentage point decrease in net transfers. If the tax-and-transfer system helps stabilize regional economies, it is because a state whose economy temporarily worsens relative to the average receives some temporary relief through transfer payments or lower tax payments to the federal government—a type of insurance against temporary negative shocks. The opposite is true for states whose economies are in a relative boom. The negative relationship is a consequence of both the progressivity of the tax system in the United States and the presence of automatic stabilizers like unemployment insurance, and particularly during this period, federal emergency unemployment compensation and food stamps.[9]

---

[8] See Section B.1 for a detailed description of the construction of these data

[9] If agents understand that federal emergency unemployment compensation would always be passed in situations of economic malaise, it is correct to include them as part of the implicit net federal transfer policy rule in place in a fiscal union.

Figure 1: Net Transfers growth v. Wage Income growth (2006-2010)



Note: Figure shows a simple scatterplot of the log-growth of net transfers in a state between 2006 and 2010 against nominal wage income log-growth during the same period. See text for details of variable construction. The population size in 2006 of the underlying state is represented by the size of the circle in the figure. The line represents a weighted regression line from the bivariate regression, where the weights are the population in 2006.

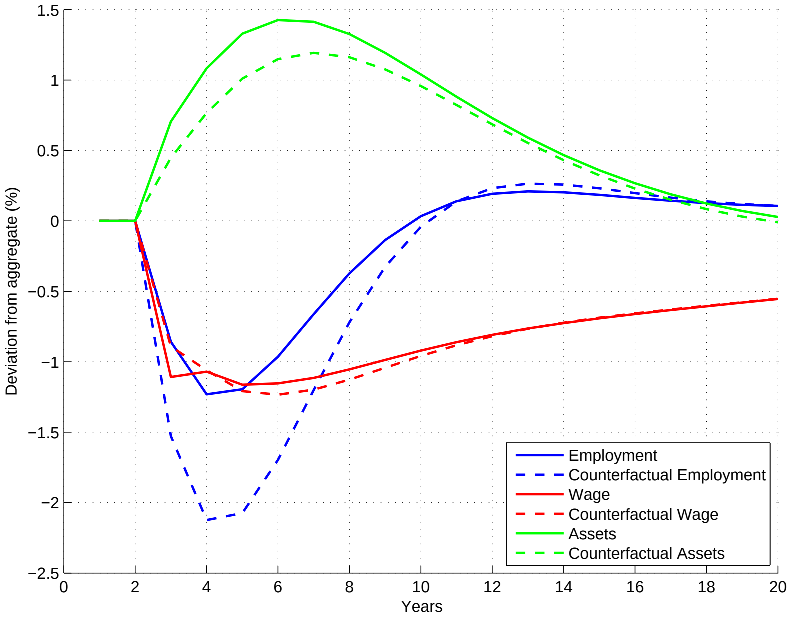
Figure 2 shows the impulse response functions of employment, wages, and assets to a one-standard-deviation "discount rate" shock γ, both in the actual and the counterfactual economy without transfers.[10] I find that employment and wages both decrease on impact, whereas assets increase in response to a "discount rate" shock. This accords with the theoretical impulse response functions in typical small open economy models such as that described in Appendix A. As for the effects of fiscal integration, I find that amplification and persistence of "discount rate" shocks are mitigated by the transfer policy rule—the employment response (after two years) is -1.2 percent in the actual economy, whereas it is -2.1 percent in the counterfactual economy without transfers.

Table 4 presents moments of the employment distribution in the actual and counterfactual economies without transfers. The cross-state employment standard deviation in the US data in 2010, σ_n^{2010}, was 2.6 percent (this corresponds to the first line and third column in the table, when all shocks are present). I then consider the following thought experiment. At the end of 2007, it is announced that from 2008 onwards the United States federal government will cease to give transfers and collect taxes to/from the states in the union. How would employment, wages, and assets have evolved were regional economies hit by the same sequence of shocks? I find that, absent a federal transfer policy rule, the standard deviation of employment in 2010 would have been 3.5 percent. The skewness of the distribution, sk_n^{2010}, would have been -0.3 instead of -0.2. To give some context to these numbers, aggregate output volatility during the pre-Volcker

---

[10] Mian and Sufi (2014) and Beraja, Hurst, and Ospina (2015), among others, argue that this type of shocks were key drivers of regional business cycles during the Great Recession in the United States.

Figure 2: Impulse Response Function to a Discount Rate Shock



Note: Figure shows the response of employment, wages and assets to a one-standard-deviation "discount rate" shock; both in the actual and counterfactual economy without transfers. The units are a state's percentage deviation from the aggregate.

period (1960:1 to 1979:2) was 2.7, whereas during the post-Volcker-disinflation period (1982:4 to 1996:4) volatility was 2.06.[11] Much literature examines causes of this "Great Moderation". The consequences of the US federal tax-and-transfer system are in the same order of magnitude.

Results above imply that the federal tax-and-transfer system helped stabilize regional economies by redistributing resources from regions that were doing relatively well to regions that were doing relatively poorly. Figure 3 elaborates on this point, showing the employment gain (or loss) from fiscal integration for each state in 2010, where states are sorted according to their employment in 2010 from lowest to highest. We observe that fiscal integration increased employment in states with the worst employment outcomes, and the opposite is true for states with the best employment outcomes.

In the first and second columns of Table 4, I calculate the same statistics if regional economies had been hit by only "discount rate" shocks or both "discount rate" and "productivity" shocks. Comparing the first and last columns in the first line, I find that most of the employment variation across states in 2010 is accounted for by "discount rate" shocks (approximately 90 percent). This is in line with findings in Beraja, Hurst, and Ospina (2015), in which data, and particularly the identification strategy, are different. Similarly, the federal transfer policy rule reduced employment dispersion primarily by stabilizing regional "discount rate" shocks. This is evidenced by comparing the first column in the second line of the table to the other columns. Of the 0.9

---

[11] These numbers come from Clarida, Gali, and Gertler (2000).

Table 4: Employment statistics: Fiscal Integration v. Fiscal Autonomy

		γ	(γ, z)	(γ, z, η)
$\boldsymbol{\sigma_n^{2010}}$	$\boldsymbol{\Theta \neq 0}$	2.3	2.5	2.6
	$\boldsymbol{\Theta = 0}$	3	3.4	3.5
$\boldsymbol{sk_n^{2010}}$	$\boldsymbol{\Theta \neq 0}$	-0.26	-0.29	-0.2
	$\boldsymbol{\Theta = 0}$	-0.39	-0.37	-0.3
$\boldsymbol{\overline{\sigma}_n}$	$\boldsymbol{\Theta \neq 0}$	2.3	2.7	3.5
	$\boldsymbol{\Theta = 0}$	3.9	4.5	4.9
$\boldsymbol{\overline{sk}_n}$	$\boldsymbol{\Theta \neq 0}$	-0.03	-0.09	0.02
	$\boldsymbol{\Theta = 0}$	-0.02	-0.07	-0.01
$\boldsymbol{\sqrt{s_n(0)}}$	$\boldsymbol{\Theta \neq 0}$	1.5	6.3	7.8
	$\boldsymbol{\Theta = 0}$	5.5	9.7	10.7

Note: σ_n^{2010} is the standard deviation of the distribution of employment n_t across states in 2010 in percentages. sk_n^{2010} is the skewness of the distribution in 2010. $\overline{\sigma}_n$ and $\overline{sk}_n$ are the standard deviation and skewness in the stationary distribution. $s_n(0)$ is the spectrum at zero frequency (the long-run variance of n_t). Line $\Theta \neq 0$ corresponds to the fiscal union economy, and line $\Theta = 0$ to results from the counterfactual. Column γ corresponds to the case with only "discount rate" shocks, and column (γ, z) to the case with "discount rate" and "supply shocks. Column (γ, z, η) corresponds to the case with all shocks.

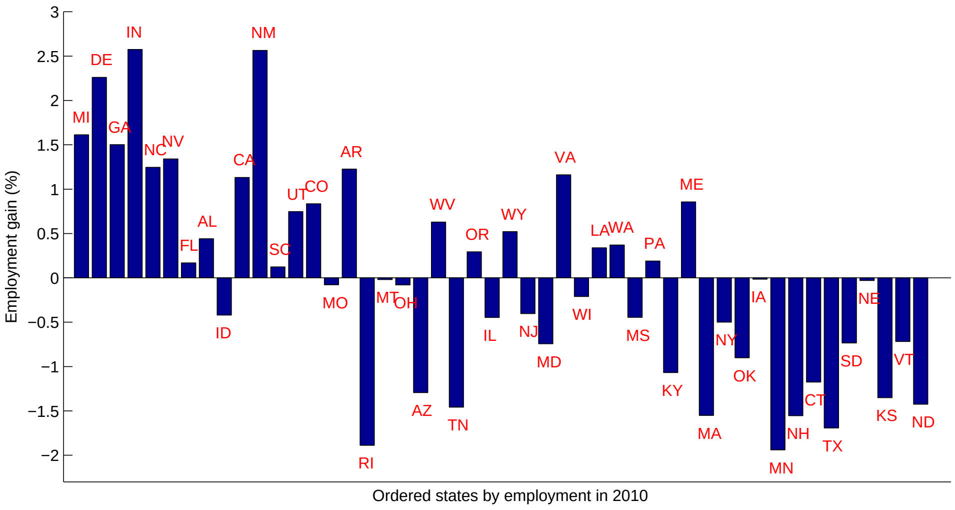
volatility reduction, 0.7 is achieved because of stabilization of "discount rate" shocks, and only 0.2 because of "productivity" shocks.[12]

In the lower half of Table 4, I present Monte Carlo estimates of the standard deviation $\overline{\sigma}_n$ and skewness $\overline{sk}_n$ of employment in the stationary distribution. I construct them by sampling with replacement 1,000,000 observations from the empirical distribution of shocks, feeding them to the SVAR and calculating the corresponding statistic. In the last line of the table, I present an estimate of the square root of the long-run variance of employment. The long-run variance is constructed as the diagonal element corresponding to employment in the spectrum at frequency zero of the multivariate SVAR. Its square root is a common measure of the persistence in the series, as proposed by Pesaran, Pierse, and Lee (1993). It associates with the speed at which the impulse response function to a given set of shocks decays, and hence their persistence. Results for the reduction in stationary employment volatility are qualitatively similar to the ones during the Great Recession. Finally, I find that the persistence of employment increases from 7.8 to 10.9 in the counterfactual economy without a federal transfer policy rule. Interestingly, most of the persistence in employment in the fiscally integrated economy is due to "productivity" shocks. In the counterfactual economy, "discount rate" shocks generate much larger persistence in employment.

---

[12] The volatility reduction of 0.9 is the difference between the actual and counterfactual volatilities of 3.5 and 2.6. Analogously, 0.7 is the difference between 3 and 2.3, and 0.2 is the difference between 3.4 minus 2.5 and 0.7.

Figure 3: Employment Gains from Fiscal Integration by State in 2010



Note: For each state, the figure shows the employment difference between the counterfactual economy without a federal transfer policy rule and the actual economy in 2010. The states were sorted according to their actual employment in 2010 in ascending order. To the left, the states with the worst employment realizations; and, to the right, the states with best.

4.1.2 Sensitivity to alternative policy specifications

Results from the previous section are based on a particular transfer policy specification where $\vartheta_n = -1.6, \vartheta_w = -0.9, \vartheta_b = 0.03$, which corresponds to the OLS estimates of the transfer policy rule in Table 3.

First, we would like to evaluate the sensitivity of results to estimates of the transfer policy rule other than the benchmark OLS estimates. Thus, I consider alternative initial policy parameterizations corresponding to the instrumental variable estimates in Table 3. Results are similar to those reported in the previous section for the benchmark policy rule. Although quantitatively reduced somewhat for some of the parameterizations, the reduction of dispersion in employment across states due to fiscal integration remains large. The largest difference is for the case in which I restrict coefficients on employment and wages in the policy rule to be identical ($\vartheta_n = \vartheta_w = -1.1$). In this case, the counterfactual employment standard deviation in 2010 would have been 3.1 percent (instead of 2.6 percent in the data), and the counterfactual standard deviation in the stationary distribution is 4.5 percent (instead of 3.5). The counterfactual using the benchmark policy estimates instead resulted in counterfactual standard deviations of 3.5 and 4.9 percent respectively.

Second, we would like to evaluate an alternative policy that accounts for the distortionary effects of taxation. The benchmark policy specification is consistent with households in each state receiving lump-sum transfers from the federal government because the only affected equation that characterizes the equilibrium is the sequential budget constraint (the third equation in this case). In practice, lump-sum transfers are rare and, instead, the federal government uses distortionary taxes.

For simplicity, consider the transfer policy without lagged assets in log-deviations from the aggregate.[13]

$$s_t = \vartheta_n n_t + \vartheta_w w_t$$

This implies that tax rate τ_t per unit of nominal labor income $w_t + n_t$ in log-deviations from the aggregate can be written as:

$$\tau_t = -(1 + \vartheta_n)n_t - (1 + \vartheta_w)w_t$$

The potential labor supply (or wage-setting) tax distortion relates to τ_t, not total transfers s_t for which we estimated elasticities ϑ_w, ϑ_n. If the federal tax-and-transfer system affects equilibrium equations beyond the sequential budget constraint, $(1 + \theta_w), (1 + \theta_n)$ would appear in these equations, not θ_w, θ_n.

I consider the case in which the second equation in (FiscalSME) is a wage-setting equation, as in Appendix A. If the federal tax-and-transfer system is distortionary, we can write it as:

$$\begin{aligned} 0 =& f_{21}\mathbb{E}_t[n_{t+1}] + f_{22}\mathbb{E}_t[w_{t+1}] + \left(g_{21} + \frac{\bar{\tau}\vartheta_n}{1 - \bar{\tau}\vartheta_n}(1 + \vartheta_n)\right) n_t + \left(g_{22} + \frac{\bar{\tau}\vartheta_n}{1 - \bar{\tau}\vartheta_n}(1 + \vartheta_w)\right) w_t \\ &+ h_{21}n_{t-1} + h_{22}w_{t-1} + \mathbb{E}_t[z_{t+1}] + l_{23}\mathbb{E}_t[\eta_{t+1}] + m_{22}z_t + m_{23}\eta_t \end{aligned}$$

The equation above accords with the tax rate that affects the target wage in the wage-setting equation by distorting the marginal rate of substitution. The distortion is given by terms $1 + \theta_n$ and $1 + \theta_w$. For example, the case $\theta_w = \theta_n = -1$ is such that the tax schedule is flat (i.e., a proportional labor income tax). Due to the lack of curvature, it would not affect the island's log-deviations of the marginal rate of substitution from the aggregate.

I construct a Robust Transfer Policy Counterfactual using this alternative policy specification, in which I set $\bar{\tau} = 0.17$ to match the average tax rate in the US economy. Results from the previous section are unchanged because estimates of transfer policy rule ϑ_n, ϑ_w are close to -1. Thus, policy-related terms that distort this equation are very small in absolute magnitude, in comparison with terms in the policy-invariant structure. To see this, consider the case in which the second equation is interpreted as a static labor supply equation, and the policy rule depends on employment alone: $w_t = \left(-g_{21} - \frac{\bar{\tau}\vartheta_n}{1-\bar{\tau}\vartheta_n}(1 + \vartheta_n)\right) n_t$. Plausible calibrations of labor supply Frisch elasticity $-g_{21}$ are in the range 0.5 to 4.[14] The policy-related term for the case when $\bar{\tau} = 0.17, \vartheta_n = -1.6$ is -0.08, which is an order of magnitude smaller than the Frisch elasticity.

---

[13] Since the estimated coefficient on lagged assets is so small, results are identical whether we include it or not.

[14] See Chetty, Guren, Manoli, and Weber (2011).

5 Application II: Which Micro-Foundations Matter?

When building models to evaluate policy rules, economists have to make difficult choices regarding which primitives and mechanisms matter for these policies' equilibrium consequences. In this section, I show how to use the principle of Counterfactual Equivalence to guide such choices—either by identifying classes of primitives that are irrelevant because they generate counterfactually equivalent models, or alternatively, by sheding light on classes of primitives that are quantitatively promising because they don't.

I present two particular applications of these ideas to classic topics in macroeconomics. The first concerns search-theoretic models of the labor market, where I analyze which primitives are relevant for the consequences of unemployment benefits policy rules. The second concerns models of international business cycles, where I discuss how the Backus-Kehoe-Kydland puzzle relates to fiscal union models as well as a simple resolution to such and other puzzles international macroeconomics.

5.1 Unemployment Benefits in Search-Theoretic Models of the Labor Market

I start by reviewing the key equations characterizing the equilibrium wage (w_t) and vacancy-to-unemployment ratio (ϑ_t) in the discrete-time version of a canonical search-and-matching model with Nash Bargaining and free entry of firms (Mortensen and Pissarides (1994); Pissarides (2000)).

$$0 = -\frac{1}{1-\phi} w_t + \frac{\phi}{1-\phi} y_t + \frac{\phi c}{1-\phi} \vartheta_t + z + b_t \tag{Wage Setting (1)}$$

$$0 = -\frac{c}{q(\vartheta_t)} + \beta \mathbb{E}_t \left[y_{t+1} - w_{t+1} + \frac{(1-s)c}{q(\vartheta_{t+1})} \right] \tag{Job Creation (1)}$$

The (Wage Setting (1)) equation determines how the surplus of a match is split between the worker and the firm. Thus, it relates the wage w_t to the (exogenous) productivity of the match y_t, the cost of posting a vacancy c, the vacancy-unemployment ratio ϑ_t, the bargaining power of the worker ϕ, the utility while unemployed z and, importantly for our purposes, unemployment benefits b_t. The (Job Creation (1)) equation determines firms' incentives to post vacancies and, because of free-entry, equates the cost of posting a vacancy to the expected benefit of a match which depends on the discount factor β and the vacancy-filling probability $q(\theta_t)$.

Next, imagine we are interested in evaluating the policy rule $b_t = \tilde{b}_t + \left(\frac{\vartheta_t}{\bar{\vartheta}}\right)^{\Theta}$, where $\tilde{b}_t$ is an exogenous policy shock and $\Theta > 0$ governs how generous benefits are when labor market slackness is above or below its long run level.[15] Moreover, suppose that after estimating the canonical model with data on wages and the vacancy-unemployment ratio alone, we conducted a series of counterfactual exercises and found that the policy Θ have negligible effects on the equilibrium behavior of these variables. How should we proceed if either (i) we wished to know

---

[15] As a motivation for this policy rule, unemployment benefits were substantially extended during the Great Recession—evidencing that its generosity may depend on the state of the business cycle.

how robust this quantitative result is to variation in model primitives or, relatedly, (ii) we wished to construct models that generated non-negligible effects?

I argue that the principle of Counterfactual Equivalence offers us some guidance. To see this, we can log-linearize the system of equilibrium equations around the steady-state, which gives rise to the following semi-structure ξ_1^*,[16]

$$F = \begin{bmatrix} 0 & 0 \\ f_{21} & f_{22} \end{bmatrix}; G = \begin{bmatrix} g_{11} & g_{12} \\ 0 & g_{22} \end{bmatrix}; H = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}; L = \begin{bmatrix} 0 & 0 \\ 0 & l_{22} \end{bmatrix}; M = \begin{bmatrix} m_{12} & 0 \\ 0 & 0 \end{bmatrix}$$

By noting the exclusion restrictions that this semi-structure satisfies, we can readily observe that the (Wage Setting (1)) equation has a rather restricted semi-structure compared to the (Job Creation (1)) equation. This observation, when combined with Proposition 1, implies that focusing on primitives or mechanisms that change how wages are set is a promising avenue for building models that could generate non-negligible effects of unemployment benefits policy (in the hypothetical case that the canonical model generated negligible effects). Analogously, models that do not significantly alter the (Job Creation (1)) equation (i.e., by changing the exclusion restrictions) will be counterfactually equivalent to the canonical model, thus implying that the negligible effects of unemployment benefits policy is a robust quantitative feature of models with varying micro-foundations regarding how firms post vacancies.

For example, if we replaced the assumption of Nash Bargaining with an *ad-hoc* wage rule ala Hall (2005), we would obtain a wage setting equation of the form,

$$0 = -w_t + z_t + \lambda y_t + (1-\lambda) w_{t-1} \qquad \text{(Wage Setting (2))}$$

where λ governs how "sticky" wages are.

Or, if we replaced it with the alternating-offer-wage-bargaining protocol in Christiano, Eichenbaum, and Trabandt (2016), we would obtain the wage setting equation,

$$\begin{aligned} 0 = &-\omega_2 \left(1 - \beta \left(1 - s - \vartheta_t q(\vartheta_t)\right)\right) \gamma + (1+\omega_1)(w_t - z_t) - (\omega_1 + \omega_3)(y_t - z_t) - c\omega_1 \vartheta_t \\ &+ \beta \left(1 - s - \vartheta_t q(\vartheta_t)\right) \omega_3 \mathbb{E}_t \left[y_{t+1} - z_{t+1}\right] \qquad \text{(Wage Setting (3))} \end{aligned}$$

where γ is the cost of delay in bargaining and the ω_i's are combinations of structural parameters.

Both semi-structures generated by these models satisfy different exclusion restrictions than the canonical model because they include backward- and forward- looking terms. Thus, they belong to different Counterfactually Equivalent Sets with respect to changes in unemployment benefits policy rule.

However, for instance, the seemingly richer model of a financial accelerator in Wasmer and Weil (2004) generates an identical semi-structure to the canonical model. They assume that

---

[16] The wage and vacancy-unemployment ratio in log-deviations from the steady-state correspond to the first and second columns of matrices F, G, H. Analogously, the exogenous policy and productivity shocks correspond to the first and second columns of matrices L, M.

matching in a credit market between firms and creditors is subject to search frictions analogous to the ones in the labor market. Then, the presence of frictional credit market adds to the cost of posting vacancies because firms have to be matched to a creditor before they can enter the labor market. In equilibrium, as opposed to the canonical model, the value of a vacancy in the labor market is given by a positive constant K that depends on the search costs in the credit market and the matching probabilities of firms and creditors. Following the the dynamic extension derivation in Petrosky-Nadeau and Wasmer (2013), we obtain equilibrium equations,[17]

$$0 = -\frac{1}{1-\phi}w_t + \frac{\phi}{1-\phi}y_t + \frac{\phi}{1-\phi}(c+(1-\beta(1-q(\vartheta_t)))K)\vartheta_t + z + b_t \tag{Wage Setting (4)}$$

$$0 = -\frac{1}{q(\vartheta_t)}\left(c+(1-\beta(1-q(\vartheta_t)))K\right) + \beta\mathbb{E}_t\left[y_{t+1} - w_{t+1} + \frac{1-s}{q(\vartheta_{t+1})}\left(c+(1-\beta(1-q(\vartheta_{t+1})))K\right)\right] \tag{Job Creation (4)}$$

Note that, compared to the canonical model, the effective cost to posting a vacancy is augmented by the additional term $(1-\beta(1-q(\theta_{t+1})))K$ which encodes the search frictions in the credit market. However, it is easy to verify that this system of equations satisfies identical exclusion restrictions than the canonical model. Thus, if the financial accelerator model and the canonical model can match the equilibrium behavior of wages and labor market slackness, they are also counterfactually equivalent with respect to changes in the unemployment benefits policy rule.

The same holds in a model where firms choose recruiting intensity e. In the spirit of Gavazza, Mongey, and Violante (2016), assume that the cost of posting a vacancy is a well-behaved function $c(e,\vartheta)$ and the probability of filling the vacancy is $q(\vartheta\bar{e})e$, where $\bar{e}$ is the average recruiting intensity in the economy. Then, we obtain the following equilibrium equations when firms optimally choose identical recruiting intensities,

$$0 = -\frac{1}{1-\phi}w_t + \frac{\phi}{1-\phi}y_t + \frac{\phi}{1-\phi}c_e(e(\vartheta_t),\vartheta_t)\vartheta_t + z_t \tag{Wage Setting (5)}$$

$$0 = -\frac{c_e(e(\vartheta_t),\vartheta_t)}{q(\vartheta_t e(\vartheta_t))} + \beta\mathbb{E}_t\left[y_{t+1} - w_{t+1} + \frac{(1-s)c_e(e(\vartheta_{t+1}),\vartheta_{t+1})}{q(\vartheta_{t+1}e(\vartheta_{t+1}))}\right] \tag{Job Creation (5)}$$

Again, while this model behaves as if it had a matching and vacancy posting cost with extra curvature, it has an identical semi-structure to the canonical model. Thus, I conclude that if the canonical model generates negligible effects of alternative unemployment benefits policy rules, then this quantitative result is robust to variation in primitives regarding certain forms of financial frictions and endogenous recruiting intensity.

---

[17] They assume that firms and creditors Nash-bargain (together) with workers in the labor market and firms and creditors Nash-bargain with each other in the credit market.

5.2 Risk-Sharing in Models of International Business Cycles

The theory behind the stabilization benefits of fiscal integration, or more generally aggregate risk sharing arrangements, is well developed. Examples include Farhi and Werning (2014), Bucovetsky (1998)), Persson and Tabellini (1996a,b), Lockwood (1999). However, there is surprisingly little quantitative research on regional stabilization in fiscal unions.[18]

An important exception is Evers (2015). He provides a quantitative analysis of federal fiscal rules in monetary unions by using a medium-scale DSGE model. The model builds on Chari and Kehoe (2007) and Kollmann (2001). It features two countries, nominal wage and price rigidities, and incomplete markets. It includes productivity and government spending shocks alone. He calibrates the model and finds very small differences in nearly every relevant business cycle statistic when comparing a monetary union with and without fiscal integration.

Because transfer rules in fiscal unions are a form of risk-sharing, the quantitative results in Evers (2015) are less surprising once we connect them to the literature studying the Backus-Kehoe-Kydland consumption correlation puzzle (see Backus, Kehoe, and Kydland (1992)). Even in leading quantitative models with very incomplete asset markets (e.g., when only a one-period bond is traded) and where many other frictions are present (e.g., nominal rigidities, habits), consumption is more highly correlated across countries than output whereas the opposite holds in the data. This implies that the consumption and employment volatility reduction gains from better risk-sharing arrangements are small in such models because the equilibrium is rather close to the complete-markets allocation.

In the language of this paper, the almost irrelevance of changes in risk-sharing rules (e.g., transfer rules in fiscal unions) is a robust result in the set of international business cycle models studied in the literature following Backus, Kehoe, and Kydland (1992) and Evers (2015). In other words, while the models in such set are not, "strictly speaking", counterfactually equivalent with respect to changes in risk-sharing policy rules, they are rather close in practice.

In contrast, for the set of models I study in Section 4.1, there are rather large reductions in the volatility of employment from fiscal integration—thus implying substantial reductions in the correlation of consumption and output since I assume that consumption is a non-tradable in these set of models.

Then, it follows that the set of models studied in Section 4.1 are not counterfactually equivalent to those in the international business cycles literature with respect to changes in risk-sharing rules. In particular, I argue that the key difference in micro-foundations comes from the presence of shifters of the Euler equation (e.g., discount rate or financial shocks). This is because, as shown in Table 4 and the discussion thereafter, I find that most of the reduction in employment volatility from fiscal integration follows from stabilizing "discount rate" shocks, and not "pro-

---

[18] On the purely empirical side, Sala-i Martin and Sachs (1991), Feyrer and Sacerdote (2013), and Bayoumi and Masson (1995) focus on reduced form estimates of properties of the tax-and-transfer system in fiscal unions, and in some cases, present back-of-the-envelope counterfactuals. Asdrubali, Sorensen, and Yosha (1996) quantify the amount of risk sharing among states in the United States by decomposing cross-sectional variance in output. Atkeson and Bayoumi (1993) examine evidence for private insurance of regional risk in the United States and Europe.

ductivity" shocks. However, to the best of my knowledge, the literature that has looked at the Backus-Kehoe-Kydland puzzle does not consider shifters in the Euler equation and focuses, by and large, on productivity shocks alone.

Thus, I conclude that the principle of counterfactually equivalence offers guidance for researchers interested in building models that can resolve the Backus-Kehoe-Kydland puzzle (and, perhaps, related ones like the Backus-Smith puzzle). A promising avenue for developing such models is to consider micro-foundations or shocks that create wedges in inter-temporal equilibrium equations. In fact, a recent paper by Itskhoki and Mukhin (2017), follows this avenue precisely and shows that it can resolve several puzzle international macroeconomics.

6 Application III: Falsifying a Set of Models

In Section 4, I discussed how to construct a counterfactual economy after a policy change given observations of the economy before the policy change. There are many instances, however, where observations before and after a policy change are available. For example, in the context of the application to stabilization in fiscal unions, we could entertain the possibility that in the future Catalunya decided to secede from the fiscal union that is Spain or that Europe would fiscally integrate. Alternatively, when the European Monetary Union was formed, country-level monetary policy rules went from being country-specific to union-wide determined. Or, when Paul Volcker started his tenure as the Federal Reserve president, the nominal interest rate policy rule changed. I argue that we could exploit this policy "experiments" in order to falsify a set of models, as long as only the policy changed and not the structure of the economy. Specifically, by constructing the a Robust Policy Counterfactual using data before a policy change and comparing it to the *observed* equilibrium after the policy change. If these are "too far apart" in a statistical sense, we would reject the null hypothesis that this set of models generated the observed equilibrium.

This application could prove useful in two ways. First, it could be used as an input for constructing Robust Policy Counterfactuals. For example, if we had several potential linear restrictions that are reasonable a-priori, we could discard some of them by falsifying their respective set of models. Second, it could be used as a guidance for future structural modeling. For example, if a particular policy "experiment" proved to be inconsistent with restrictions implied by certain models, then researchers building new structural models should take this evidence into consideration.

I formalize this idea in what follows and then illustrate it by revisiting the monetary policy experiment in Clarida, Gali, and Gertler (2000) in order to falsify a set of New Keynesian models.

As in Section B.3, suppose there is a VAR representation of the recursive equilibrium $x_t = \rho_1(\xi,\Theta)x_{t-1} + \rho_2(\xi,\Theta)x_{t-2} + e_t$, with e_t iid over time with $\mathbb{E}[e_t] = 0, Var(e_t) = V(\xi,\Theta)$. Furthermore, let $g(e_t|\xi,\Theta)$ be the theoretical distribution of the reduced-form shocks, let $\{x_t\}_{t=1}^{T_0}$ be observations from an economy under policy Θ^0, and let $\{x_t\}_{t=1}^{T_1}$ be observations from an economy under policy Θ_1. The following proposition describes how to construct a log-likelihood ratio test in order to falsify a set of models.

Proposition 2 (Falsifying a set of models via a log-likelihood ratio test)

1. *Using the observations under policy* Θ^0*, obtain maximum-likelihood estimates* $\{\hat{\rho}_1^0, \hat{\rho}_2^0, \hat{V}^0\}$.
2. *Imposing a set of restrictions* R^* *on the structure* ξ*, construct a Robust Policy Counterfactual associated with* Θ_1*. In particular, obtain maximum-likelihood estimates* $\{\xi^*, \rho_1(\xi^*, \Theta_1), \rho_2(\xi^*, \Theta_1)\}$ *and, thus, a predicted distribution* $g(e_t|\xi^*, \Theta_1)$ *from the restricted set of models.*
3. *Using the observations under policy* Θ_1*, estimate via maximum likelihood the distribution* $\hat{g}_1(e_t)$ *of the reduced-form shocks corresponding to the unrestricted set of models.*
4. *Construct the log-likelihood ratio* $LR = -2\sum_{t=1}^{T_1} log\left(\frac{g(e_t|\xi^*,\Theta_1)}{\hat{g}_1(e_t)}\right)$

We wish to test the null hypothesis $H_0 : \xi = \xi^*$*. When* H_0 *is true and* $T_1 \to \infty$*, LR converges in distribution to* χ^2_r *distribution with degrees of freedom r equal to the number of restrictions in* R^**. Thus, reject* H_0 *at significance level* α *whenever* $P(\chi^2_r > LR) \leq \alpha$.

Proof. A direct application of Wilk's Theorem in Wilks (1938).

The null hypothesis is that the restricted set of models are consistent with the observations before and after the policy change. By construction, the restricted set of models are always consistent with the observations before the policy change. Then, the log-likelihood ratio test will reject the null hypothesis that the observations after the policy change were generated by the restricted set of models whenever the predicted distribution of the reduced form shocks is far, in a statistical sense, from the observed distribution after the policy change. Intuitively, we falsify a set of models whenever its Robust Policy Counterfactual disagrees with the observed equilibrium after the policy change.

6.1 Falsifying New Keynesian Models with Paul Volcker's Policy Change

In a seminal paper, Clarida, Gali, and Gertler (2000) show how the change in US interest rate policy rule following Volcker's appointment as Fed Chairman stabilized equilibrium inflation and output. To do so, they estimate these interest rate rules and compare the equilibria under both rules in a canonical New Keynesian model. In this section, I show that this policy experiment actually falsifies a larger set of New Keynesian models (that includes such canonical model) because they are inconsistent with the behavior of output, inflation, and interest rates both before and after the policy change.

The canonical New Keynesian model in Clarida, Gali, and Gertler (2000) describes the equilibrium behavior of inflation π_t, output y_t and nominal interest rate r_t in log-deviations from a zero inflation steady state in terms of three equations,

$$
\begin{aligned}
\mathbf{0} = & \begin{bmatrix} 1 & \frac{1}{\sigma} & 0 \\ 0 & -\frac{\delta}{\lambda} & 0 \\ 0 & 0 & 0 \end{bmatrix} \mathbb{E}_t \begin{bmatrix} y_{t+1} \\ \pi_{t+1} \\ r_{t+1} \end{bmatrix} + \begin{bmatrix} -1 & 0 & -\frac{1}{\sigma} \\ -1 & \frac{1}{\lambda} & 0 \\ \gamma & \beta & -\frac{1}{1-\rho} \end{bmatrix} \begin{bmatrix} y_t \\ \pi_t \\ r_t \end{bmatrix} \\
+ & \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \frac{\rho}{1-\rho} \end{bmatrix} \begin{bmatrix} y_{t-1} \\ \pi_{t-1} \\ r_{t-1} \end{bmatrix} + \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \mathbb{E}_t \begin{bmatrix} z_t \\ g_t \\ \epsilon_t \end{bmatrix} \qquad \begin{array}{r} \text{(Euler)} \\ \text{(NKPC)} \\ \text{(Policy)} \end{array}
\end{aligned}
$$

In the (Euler) equation, g_t is an exogenous AR(1) "demand" shifter. The (NKPC) equation is the New Keynesian Phillips Curve, where z_t is an exogenous AR(1) process for natural output. Finally, the (Policy) equation specifies the interest rate policy rule, where ϵ_t is an exogenous AR(1) process encompassing misforecasts of actual and natural output, inflation, and explicit monetary policy shocks.

This model belongs to a larger set of New Keynesian models that can be described by the following policy Θ and semi-structure ζ_0^*,

$$
\begin{aligned}
\Theta_f &= \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}; \Theta_c = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ \gamma & \beta & -\frac{1}{1-\rho} \end{bmatrix}; \Theta_p = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \frac{\rho}{1-\rho} \end{bmatrix} \\
F &= \begin{bmatrix} f_{11} & f_{12} & 0 \\ 0 & f_{22} & 0 \\ 0 & 0 & 0 \end{bmatrix}; G = \begin{bmatrix} g_{11} & 0 & g_{13} \\ g_{21} & g_{22} & 0 \\ 0 & 0 & 0 \end{bmatrix}; H = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}; \\
L &= \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}; M = \begin{bmatrix} 0 & 1 & 0 \\ 1 & m_{22} & 0 \\ 0 & 0 & m_{33} \end{bmatrix}; N = \begin{bmatrix} n_{11} & n_{12} & n_{13} \\ n_{21} & n_{22} & n_{23} \\ 0 & n_{32} & n_{33} \end{bmatrix}
\end{aligned}
$$

For example, the canonical New Keynesian model satisfies one extra exclusion restriction $m_{22} = 0$ compared to this larger set of models, i.e., the "demand" shock g_t does not enter directly in the New Keynesian Philipps Curve. A different New Keynesian model that satisfies exactly the same exclusion restrictions as ζ_0^* is one with a large family of agents in the households, where a fraction of the agents g_t are non-employed and consume the transfers from the family alone. Exogenous variation in g_t over time generates variation in the intertemporal marginal rate of substitution of consumption and intratemporal marginal rate of substitution between consumption and labor, thus implying that g_t appears both in the (Euler) and (NKPC) equations.

Next, I test the null hypothesis that the set of New Keynesian models associated with semi-structure ζ_0^* are consistent with the observed equilibrium both before and after the change in policy Θ following Volcker's appointment. I use quarterly data on real GDP, inflation and nominal interest rates for the periods 1960:1 to 1979:2 (i.e., pre-Volcker) and 1982:4 to 1996:4 (i.e., pos-Volcker).[19] The policy parameters before and after Volcker come from Clarida, Gali, and Gertler (2000). In the pre-Volcker period, I set $\beta = 1.01, \gamma = 0.27$.[20] In the pos-Volcker period, I set $\beta = 1.58, \gamma = 0.14$. Both in the pre- and pos-Volcker periods, I set ρ such that the implied coefficient on lagged interest rate in the policy rule (h_{33}) is zero. Finally, I assume that the structural shocks are normally distributed and follow the procedure in Proposition 2. Under normality of the shocks, the log-likelihood ratio test essentially compares the covariance of the reduced form shocks in the pos-Volcker period to the predicted covariance by the set of models with semi-structure ξ_0^*. I find a log-likelihood ratio of 94, thus rejecting the null hypothesis at all standard levels of significance.

7 Comparison to Related Methodologies

This paper relates to much literature who has tackled issues of robustness, identification and model misspecification. The original motivation of Sims (1980) seminal contribution on structural vector autoregressions (SVARs) was to do policy analysis in a way that was robust across many models. SVARs have been very successful for doing analysis with respect to the exogenous component of policy (i.e., policy shocks) in a relatively model-free way. However, because of Lucas critique, it has not been possible to use SVARs to do analysis with respect to unobserved systematic components of policy (i.e., policy rules).[21] The methodology in this paper for constructing Robust Policy Counterfactuals shows how to restrict the set of structural models that generate SVARs in order to conduct this type of analysis in a way that is immune to Lucas critique and robust to model selection within the restricted, counterfactually equivalent set.

Furthermore, since seminal research by Harberger (1964), sufficient statistics have been used to evaluate the consequences of policy changes, without requiring full knowledge of the underlying structural model. Chetty (2009) provides an excellent survey. More recent examples in trade literature include Arkolakis, Costinot, and Rodríguez-Clare (2012) and Blaum, Lelarge, and Peters (2015), and in macroeconomics, Alvarez, Le Bihan, and Lippi (2016), Auclert (2017), and Atkeson and Burstein (2015). The non-parametric counterfactuals in Adao, Costinot, and Donaldson (2017) are in the same spirit. This paper shares their motivation and philosophy, using reduced-form statistics to characterize counterfactually equivalent models and constructing counterfactuals.

---

[19] All data comes from FRED. y_t is hp-filtered, seasonally adjusted real GDP in chained 2009 dollars (GDPC1). π_t is hp-filtered, log-change in the implicit GDP deflator (GDPDEF). r_t is the hp-filtered, effective federal funds rate (FEDFUNDS). Furthermore, I remove the first three years of Volcker's tenure from the pos-Volcker sample because several features of how monetary policy was conducted only stabilized after 1982:4. For example, as discussed by Clarida, Gali, and Gertler (2000), the Federal Reserve targeted nonborrowed reserves instead of the Federal Funds rate over this period.

[20] I use $\beta = 1.01$ instead of the point-estimate 0.86 in Clarida, Gali, and Gertler (2000) to ensure that the recursive equilibrium is stable, as they do in Section IV.C. of the paper.

[21] In the application to fiscal unions, policy shocks would be exogenous, unanticipated transfers to a state. On the other hand, I am concerned with the systematic, anticipated component of tax-and-transfer policy as a known function of state-level variables—a transfer policy rule.

Del Negro and Schorfheide (2009) are concerned with doing policy analysis in misspecified models. In particular, they study changes in the nominal interest rate policy rule. They consider a benchmark DSGE model with sticky prices and a SVAR representation of the equilibrium in this model. Then, they consider an alternative "true" model that might be different from the benchmark and is unknown. They write the SVAR representation of the equilibrium in this "true" model as a sum of the benchmark SVAR and perturbation terms. They consider various assumptions about the perturbation terms concerning how they are affected by changes in policy. Taking a stance on this is necessary to do counterfactuals that are immune to Lucas critique within their framework. For example, they conduct the analysis assuming that the perturbation terms are invariant to changes in the policy rule and that they can be estimated jointly with the benchmark DSGE model parameters from past data. This approach is useful for considering the sensitivity of results to model misspecification around the benchmark model. The methodology in this paper is rather different. While I do not describe alternative models in terms of deviations from a benchmark model in the paper, it is possible to do so and useful for comparison with their paper. In this case, we would write the structural equations characterizing the equilibrium in the "true" model as the sum of a benchmark model and a perturbation term. Then, we would state which perturbation terms are policy invariant and which ones are not. The key difference is that my perturbation would be done around a benchmark structure of the model—e.g., Euler equations, labor demand, etc.—whereas Del Negro and Schorfheide (2009) do perturbations around the SVAR representation of the equilibrium of such benchmark model. Perturbing the structure of the model makes it easy to see which models are consistent with the perturbation terms being policy invariant. Perturbing the equilibrium SVAR representation implies that the perturbation terms are in general not policy invariant.

Lastly, the accounting procedure in Chari, Kehoe, and McGrattan (2007) is useful for understanding which mechanisms to include in quantitative models in order to allow them to match observed business cycle patterns. As I have shown, this paper offers guidance regarding which mechanism are relevant in order to analyze policy rule changes. Again, in this paper, it is also possible to describe models with alternative primitives by adding stochastic *wedges* to equilibrium equations of a benchmark model. However, these wedges are endogenous objects and in general not invariant to policy changes. Thus, it would be hard to see which primitives generate models that are counterfactually equivalent.

References

Adao, R., A. Costinot, and D. Donaldson (2017): "Nonparametric counterfactual predictions in neoclassical models of international trade," *The American Economic Review*, 107, 633–689.

Alvarez, F., H. Le Bihan, and F. Lippi (2016): "The real effects of monetary shocks in sticky price models: a sufficient statistic approach," *The American Economic Review*, 106, 2817–2851.

Arkolakis, C., A. Costinot, and A. Rodríguez-Clare (2012): "New Trade Models, Same Old Gains?" *American Economic Review*, 102, 94–130.

Asdrubali, P., B. E. Sorensen, and O. Yosha (1996): "Channels of interstate risk sharing: United States 1963-1990," *The Quarterly Journal of Economics*, 1081–1110.

Atkeson, A. and T. Bayoumi (1993): "Do private capital markets insure regional risk? Evidence from the United States and Europe," *Open Economies Review*, 4, 303–324.

Atkeson, A. and A. Burstein (2015): "Aggregate Implications of Innovation Policy," .

Auclert, A. (2017): "Monetary policy and the redistribution channel," Tech. rep., National Bureau of Economic Research.

Bachmann, R. and E. R. Sims (2012): "Confidence and the transmission of government spending shocks," *Journal of Monetary Economics*, 59, 235 – 249.

Backus, D. K., P. J. Kehoe, and F. E. Kydland (1992): "International real business cycles," *Journal of political Economy*, 100, 745–775.

Baumeister, C. and J. D. Hamilton (2015): "Sign restrictions, structural vector autoregressions, and useful prior information," *Econometrica*, 1963–1999.

Bayoumi, T. and P. R. Masson (1995): "Fiscal flows in the United States and Canada: lessons for monetary union in Europe," *European Economic Review*, 39, 253–274.

Beraja, M., E. Hurst, and J. Ospina (2015): "The Aggregate Implications of Regional Business Cycles," *Unpublished Manuscript, University of Chicago*.

Bernanke, B. S., M. Gertler, and M. Watson (1997): "Systematic monetary policy and the effects of oil price shocks," *Brookings papers on economic activity*, 1997, 91–157.

Blaum, J., C. Lelarge, and M. Peters (2015): "The Gains From Trade in Firm-Based Models of Importing," .

Bucovetsky, S. (1998): "Federalism, equalization and risk aversion," *Journal of Public Economics*, 67, 301–328.

Chari, V. V. and P. J. Kehoe (2007): "On the need for fiscal constraints in a monetary union," *Journal of Monetary Economics*, 54, 2399–2408.

Chari, V. V., P. J. Kehoe, and E. R. McGrattan (2007): "Business cycle accounting," *Econometrica*, 75, 781–836.

Chetty, R. (2009): "Sufficient statistics for welfare analysis: A bridge between structural and reduced-form methods," *Annual Review of Economics*, 1, 451–488.

Chetty, R., A. Guren, D. Manoli, and A. Weber (2011): "Are micro and macro labor supply elasticities consistent? A review of evidence on the intensive and extensive margins," *The American Economic Review*, 101, 471–475.

Christiano, L. J., M. S. Eichenbaum, and M. Trabandt (2016): "Unemployment and business cycles," *Econometrica*, 84, 1523–1569.

Clarida, R., J. Gali, and M. Gertler (2000): "Monetary Policy Rules and Macroeconomic Stability: Evidence and Some Theory," *Quarterly Journal of Economics*, 147–180.

Del Negro, M. and F. Schorfheide (2009): "Monetary Policy Analysis with Potentially Misspecified Models," *The American Economic Review*, 99, 1415.

Evers, M. P. (2015): "Fiscal Federalism and Monetary Unions: A Quantitative Assessment," *Journal of International Economics*.

Farhi, E. and I. Werning (2014): "Fiscal unions," *Unpublished Manuscript, Massachusetts Institute of Technology*.

Feyrer, J. and B. Sacerdote (2013): "How Much Would US Style Fiscal Integration Buffer European Unemployment and Income Shocks?(A Comparative Empirical Analysis)," *The American Economic Review*, 103, 125–128.

Gabaix, X. (2016): "A behavioral New Keynesian model," Tech. rep., National Bureau of Economic Research.

Gavazza, A., S. Mongey, and G. L. Violante (2016): "Aggregate Recruiting Intensity," Tech. rep., National Bureau of Economic Research.

Hall, R. E. (2005): "Employment fluctuations with equilibrium wage stickiness," *The American Economic Review*, 95, 50–65.

Hansen, G. D. (1985): "Indivisible labor and the business cycle," *Journal of monetary Economics*, 16, 309–327.

Harberger, A. C. (1964): "The measurement of waste," *The American Economic Review*, 58–76.

Itskhoki, O. and D. Mukhin (2017): "Exchange rate disconnect in general equilibrium," Tech. rep., National Bureau of Economic Research.

Kilian, L. and L. T. Lewis (2011): "Does the Fed respond to oil price shocks?" *The Economic Journal*, 121, 1047–1072.

Kollmann, R. (2001): "Explaining international comovements of output and asset returns: the role of money and nominal rigidities," *Journal of Economic Dynamics and Control*, 25, 1547–1583.

Lockwood, B. (1999): "Inter-regional insurance," *Journal of Public Economics*, 72, 1–37.

Lucas, R. E. (1976): "Econometric policy evaluation: A critique," in *Carnegie-Rochester conference series on public policy*, Elsevier, vol. 1, 19–46.

McKay, A. and R. Reis (2013): "The role of automatic stabilizers in the US business cycle," Tech. rep., CEPR Discussion Papers.

Mian, A., K. Rao, A. Sufi, et al. (2013): "Household Balance Sheets, Consumption, and the Economic Slump," *The Quarterly Journal of Economics*, 128, 1687–1726.

Mian, A. and A. Sufi (2014): "What Explains the 2007–2009 Drop in Employment?" *Econometrica*, 82, 2197–2223.

Mortensen, D. T. and C. A. Pissarides (1994): "Job creation and job destruction in the theory of unemployment," *The review of economic studies*, 61, 397–415.

Nakamura, E. and J. Steinsson (2014): "Fiscal Stimulus in a Monetary Union: Evidence from US Regions," *The American Economic Review*, 104, 753–792.

Oh, H. and R. Reis (2012): "Targeted transfers and the fiscal response to the great recession," *Journal of Monetary Economics*, 59, S50–S64.

Persson, T. and G. Tabellini (1996a): "Federal fiscal constitutions: Risk sharing and moral hazard," *Econometrica: Journal of the Econometric Society*, 623–646.

——— (1996b): "Federal fiscal constitutions: risk sharing and redistribution," *Journal of Political Economy*, 979–1009.

Pesaran, M. H., R. G. Pierse, and K. C. Lee (1993): "Persistence, cointegration, and aggregation: A disaggregated analysis of output fluctuations in the US economy," *Journal of Econometrics*, 56, 57–88.

Petrosky-Nadeau, N. and E. Wasmer (2013): "The cyclical volatility of labor markets under frictional financial markets," *American Economic Journal: Macroeconomics*, 5, 193–221.

Pissarides, C. A. (2000): *Equilibrium unemployment theory*, MIT press.

Ravenna, F. (2007): "Vector autoregressions and reduced form representations of DSGE models," *Journal of monetary economics*, 54, 2048–2064.

Ruggles, S., M. Sobek, C. A. Fitch, P. K. Hall, and C. Ronnander (1997): *Integrated public use microdata series: Version 2.0*, Historical Census Projects, Department of History, University of Minnesota.

Sala-i Martin, X. and J. D. Sachs (1991): "Fiscal Federalism and Optimum Currency Areas: Evidence for Europe from the United States," *NBER Working Paper*.

Sims, C. A. (1980): "Macroeconomics and reality," *Econometrica: Journal of the Econometric Society*, 1–48.

Sims, C. A. and T. Zha (2006): "Does monetary policy generate recessions?" *Macroeconomic Dynamics*, 10, 231–272.

Uhlig, H. (1995): "A toolkit for analyzing nonlinear dynamic stochastic models easily," .

Wasmer, E. and P. Weil (2004): "The macroeconomics of labor and credit market imperfections," *The American Economic Review*, 94, 944–963.

Wilks, S. S. (1938): "The large-sample distribution of the likelihood ratio for testing composite hypotheses," *The Annals of Mathematical Statistics*, 9, 60–62.

A Proof of Proposition 1

The proof is by construction. Applying the method of undetermined coefficients to (SME) implies that under policy Θ^0, the recursive representation of the equilibrium $\{P^0, Q^0, N^0\}$ satisfies,

$$((F + \Theta^0_f)P^0 + G + \Theta^0_c)P^0 + H + \Theta^0_p = 0$$
$$(F + \Theta^0_f)(Q^0N^0 + P^0Q^0) + (G + \Theta^0_c)Q^0 + (LN^0 + M) = 0$$

Then, take equations in this system involving the first line of structure ξ. This is without loss of generality because the proof is identical for all other lines. Given $\{P^0, Q^0, N^0\}$, this can be written as

$$\begin{bmatrix} F_{1.} + \Theta^0_{f,1.} & G_{1.} + \Theta^0_{c,1.} & H_{1.} + \Theta^0_{p,1.} & (LN^0 + M)_{1.} \end{bmatrix}_{\mathbf{1.(3k+s)}} \begin{bmatrix} (P^0)^2 & Q^0N^0 + P^0Q^0 \\ P^0 & Q^0 \\ I & 0 \\ 0 & I \end{bmatrix} = \mathbf{0_{1.(k+s)}}$$

There are $3k + s$ unknowns in this system involving elements of structure ξ. There are $k + s$ equations. One of the unknown elements can be normalized to 1 if $rank(\Theta_{1.}) = 0$. This implies that if we specify a set R_{n_1} of $n_1 = 2k - 1 + rank(\Theta_{1.})$ independent linear restrictions on the elements of structure ξ, we obtain a linear system of equations that is exactly determined (as long as the $(3k + s) \times (k + s)$ matrix containing $\{P^0, Q^0, N^0\}$ has full rank). In particular, from this exactly determined linear system, we identify semi-structure ξ^* that is both consistent with the imposed linear restrictions and the recursive representation under the benchmark policy.

Finally, we solve the following system for the counterfactual equilibrium under the alternative policy Θ^1. That is, we solve for $\{P(\xi^*, \Theta^1), Q(\xi^*, \Theta^1\})$ given $\{P^0, Q^0, N^0, \Theta^0, \Theta^1\}$ and $\{F^*, G^*\} \in \xi^*$. The system results from subtracting (SME) for policy Θ^0 and Θ^1.

$$\begin{aligned}
0 =& (F^* + \Theta^1_f)(P(\xi^*, \Theta^1)^2 - (P^0)^2) + (G^* + \Theta^1_c)(P(\xi^*, \Theta^1) - P^0) \\
&+ (\Theta^1_f - \Theta^0_f)(P^0)^2 + (\Theta^1_c - \Theta^0_c)P^0 + \Theta^1_p - \Theta^0_p \\
0 =& \left(F^* + \Theta^1_f)P(\xi^*, \Theta^1) + G^* + \Theta^1_c\right)(Q(\xi^*, \Theta^1) - Q^0) + (F^* + \Theta^1_f)(Q(\xi^*, \Theta^1) - Q^0)N^0 \\
&+ (F^* + \Theta^1_f)\left(P(\xi^*, \Theta^1) - P^0\right)Q^0 + (\Theta^1_f - \Theta^0_f)\left(Q^0N^0 + P^0Q^0\right) + (\Theta^1_c - \Theta^0_c)Q^0
\end{aligned}$$

This concludes the proof because the counterfactual recursive representation under the alternative policy Θ^1 is unique and identical for all models that have semi-structure ξ^*. ■

B Robust Transfer Policy Counterfactual in Fiscal Union Models

B.1 Data description

I exclude Alaska, District of Columbia, and Hawaii from analysis, leaving 48 observations (one for each remaining state) per year, and 6 years (2006-2011) of data.

To make state-level nominal wages indices, I use data from the 2000 US Census and the 2001-2012 American Community Surveys (ACS).[22]

The 2000 Census includes 5 percent of the US population. The 2001-2012 ACS's include approximately 600,000 respondents between 2001-2004, and about 2 million after 2004. The large sample sizes allow detailed labor market information at the state level. I begin by using the data to make individual hourly nominal wages. I restrict the sample to only individuals who are employed, who report usually working at least 30 hours per week, and who worked at least 48 weeks during the prior 12 months. For each individual, I divide total labor income earned during the prior 12 months by a measure of annual hours worked during the prior 12 months.[23] The composition of workers differs across states and within a state over time, which might explain some variation in nominal wages across states over time. To account for this, I run the following regression:

$$\ln(w_{itk}) = K_t + \Gamma_t X_{itk} + u_{itk}$$

where $\ln(w_{kt})$ is log-nominal wages for household i in period t residing in state k, and X_{kt} is a vector of household specific controls. The vector of controls include a series of dummy variables for usual hours worked (30-39, 50-59, and 60+), a series of five-year age dummies (with 40-44 being the omitted group), 4 educational attainment dummies (with some college being the omitted group), three citizenship dummies (with native born being the omitted group), and a series of race dummies (with white being the omitted group). I run these regressions separately for each year such that both constant K_t and the vector of coefficients on the controls, Γ_t, can differ for each year. I then take the residuals from these regressions for each individual, u_{itk}, and add back constant K_t. Adding back the constant from the regression preserves differences over time in average log wages. To compute average log wages within a state, holding composition fixed, I average $u_{itk} + K_t$ across all individuals in state k. I refer to this measure as the demographically adjusted, log-nominal wage in time t in state k.

The measure of employment at the state level is the employment rate for each state, calculated using data from the US Bureau of Labor Statistics. The BLS reports annual employment counts

---

[22] I access the data through the IPUMS-USA website https://usa.ipums.org/usa/. See Ruggles, Sobek, Fitch, Hall, and Ronnander (1997).

[23] Total labor income during the prior 12 months is the sum of both wage and salary earnings and business earnings. Total hours worked during the previous 12 months is a multiple of total weeks worked during the prior 12 months and the respondents' reports of their usual hours worked per week. For some years, bracketed reports are provided for weeks worked during the prior 12 months, and the usual hours per week worked. In those cases, I take the midpoint of the brackets.

and population numbers for each state and year. I divide employment counts by population to make an annual employment rate measure for each state.

Data on federal transfers net of taxes paid come from the Bureau of Economic Analysis.[24] Transfers include retirement and disability insurance benefits, medical benefits, income maintenance benefits, unemployment insurance compensation, veterans benefits, federal education and training assistance, and other transfer receipts of individuals from governments. Federal taxes are the sum of personal income taxes that are withheld, usually by employers, from wages and salaries, quarterly payments of estimated taxes on income that is usually not subject to withholding, and final settlements, which are additional tax payments made when tax returns for a year are filed, or as a result of audits by the Federal Government.[25]

Given the unavailability of official state-level data on asset positions, I construct a measure of state-level assets as the sum of physical and financial assets. From national account identities, we can derive the law of motion for assets B_t in a given state as:

$$B_t = B_{t-1}(1 + r_t) + Y_t - C_t + S_t - G_t^{local} + v_t$$

where Y_t is nominal gross domestic product, C_t is private consumption expenditures, S_t are net transfers (i.e., expenditures minus taxes) from the federal government, G_t^{local} are expenditures from the local government, and r_{t-1} captures the change in asset valuation between $t-1$ and t. Finally, error term v_t includes income receipts from abroad minus income payments to foreigners, federal government expenditures not counted as federal transfers (e.g., salaries and wages), and differences in returns between physical and financial assets for which no data are available.[26] I obtain Y_t and C_t directly from the Bureau of Economic Analysis website. $S_t - G_t^{local}$ also comes from several variables in the BEA. I calculate it as (personal current transfers receipts) - (personal current taxes paid + taxes on production and imports net of subsidies).[27] The revaluation of assets term r_t is obtained residually to ensure that the growth rate of the sum of local assets across states is consistent with the growth rate of aggregate net worth in the US economy. Having all components in the law of motion for B_t, I calculate assets at each point in time for each state simply by iterating forward with 2006 as the initial observation. I obtain initial assets in 2006 by aggregating at the state level, the zip code total net worth data from Mian, Rao, Sufi et al. (2013). In order to construct financial assets at the zip code level Mian, Rao, Sufi et al. (2013) they use data on dividends and interest income from the IRS Statistics of Income (SOI). They assume that households hold identical shares of stocks and bonds (they hold the market index portfolio). Given the share of total dividends and interest income received by a zip code they can construct

---

[24] I access the data through the BEA website on regional GDP and personal income: http://www.bea.gov/iTable/index_regional.cfm

[25] Excise, Medicare and Social security federal taxes are not included in this measure.

[26] Error term v_t accounts for most of the "wealth" exogenous process. The remainder is the error term e_t in the difference between observed net transfers and estimated policy rule in equation Section B.2.

[27] As long as local government expenditures plus transfers are close enough to local tax revenues (i.e., local governments have a nearly balanced budget), the calculation is accurate. If not, the difference is absorbed by error term e_t.

the share of total stocks and bonds held by that zip code. Then, they total financial assets from the Federal Reserve's Flow of Funds data to zip codes based on these shares. For the value of nominal debt owed by households they use data based on information from Equifax Predictive Services. Then they match the Federal Reserve Flow of Funds data by using the share of Equifax total debt in a zip code to allocate Flow of Funds debt. The final component of the asset measure is the value of housing wealth which they estimate using the 2000 Decennial Census data. They construct total home value as of 2000 in a zip code as the product of the number of home owners and the median home value. Then, they project it forward into later years using the CoreLogic zip code level house price index and an aggregate estimate of the change in homeownership and population growth.

B.2 Estimating the transfer policy rule Θ

As a reminder, the transfer policy rule is:

$$s_t = \vartheta_n n_t + \vartheta_w w_t + \vartheta_b b_{t-1} + e_t$$

For regional data to be used to estimate Θ, one of the following must hold: (1) the innovations to the policy rule have no regional component ($e_t = 0$)—in which case, a simple OLS regression produces consistent estimates—or (2) valid instruments can be found that isolate movements in n_t, w_t, b_{t-1} that are orthogonal to e_t. The issue of endogeneity arises because of reverse causality. When the innovation in the policy rule is part of the "wealth" shock u_t^{η}, employment and wages both cause and are caused by net transfers in the equation above. To deal with the endogeneity of n_t, w_t, b_{t-1}, I proceed variously. First, I estimate a regression of net transfers onto nominal wage income alone (assuming $\theta_w = \theta_n$) using house price growth between 2006 and 2010 as an instrument. This accords with many recent papers, including Mian and Sufi (2014). Contemporaneous housing price growth strongly predicts contemporaneous nominal wage income growth. The instrument is valid as long as local housing prices are orthogonal to the transfer policy rule shock, which appears plausible. In the second approach, I use "discount rate" and "productivity" shocks in 2008, estimated from (FiscalSVAR), as instruments for wages and employment. They are linear combinations of wages, employment, and assets in 2008 that are orthogonal to the "wealth" shock, and hence e_t, by construction.

Table 3 presents results for several specifications. The dependent variable is the log-growth rate of transfers minus the growth rate of taxes between 2006 and 2010 for each state. The independent variables are the log-growth rate of nominal wages between 2006 and 2010 and the log-growth rate of employment between 2006 and 2010 in the first two columns, and the log-growth rate of assets between 2006 and 2009 in the third column. In the fourth column, the independent variable is the sum of wage and employment growth. The first line is a simple OLS regression. The second presents two-stage, least-squares results using the "discount rate" and "productivity" shocks in 2008 $u_{2008}^{\gamma}, u_{2008}^{z}$. The third uses house price log-growth between 2006 and 2008 as an instrument instead. The fourth uses all three instruments. For all specifications, and when possible, I consider case (1) when b_{t-1} is not endogenous, and case (2) when b_{t-1} might be endogenous.

Table 3: Policy rule baseline estimates

	$\boldsymbol{\vartheta_n}$	$\boldsymbol{\vartheta_w}$	$\boldsymbol{\vartheta_b}$	$\boldsymbol{\vartheta_{w+n}}$	$\boldsymbol{R^2}$
OLS	-1.6^{**} (0.5)	-0.9^{*} (0.7)	-0.03 (0.02)	.	0.41
IV w/ shocks (1)	-1.3^{*} (0.7)	-1.4^{*} (0.8)	-0.02 (0.02)	.	0.41
IV w/ house prices (1)	.	.	-0.03 (0.02)	-1.1^{**} (0.4)	0.42
IV w/ house prices and shocks (1)	-1.4^{*} (0.6)	-1.2^{*} (0.6)	-0.02 (0.02)	.	0.43
(2)	-1.3^{*} (0.7)	-1.4^{*} (0.7)	0.01 (0.08)	.	0.38

Note: Numbers in parenthesis are OLS (or second stage) standard errors. Variables with '*' are significant at a 5% level. Variables with '**' are significant at 1%. All variables are state log-growth rates between 2006 and 2010. b_{t-1} is exogenous in (1) and endogenous in (2)

I find that the policy rule estimates have the expected sign and are significant in all specifications. They are also similar in magnitude, ranging from -1.3 to -1.6 for ϑ_n and -0.9 to -1.4 for ϑ_w. Lagged assets have nearly no independent explanatory power for net transfers across all specifications. To give a sense of the magnitudes involved, when net transfers increase by 30 percent for every 1 percent decrease in nominal wage income, and the average income tax rate is 0.17, for every 1 dollar decrease in nominal wage income, a state receives 0.22 dollars in federal transfers. This result is similar to findings by Feyrer and Sacerdote (2013), who find a 0.25 decrease, and Bayoumi and Masson (1995), who find a 0.31 decrease.

B.3 SVAR identification

A necessary input in the construction of Robust Policy Counterfactuals is the impulse response matrix $Q(\xi, \Theta)$. The literature proposes myriad ways to identify it, ranging from simple ordering assumptions to more sophisticated sign and long-run restrictions. These represent several routes that could be followed as long as their implied linear restrictions on the structure ξ are consistent with the restrictions R^* imposed on semi-structure ξ^*. Alternatively, in this section I show how to use those equilibrium equations of structural models that we feel more confident about in order to derive linear restrictions on the structure ξ that are sufficient to identify Q. Specifically, I use the sequential budget constraint (SB) to generate these theoretical restrictions because, as I argued in Section 2, many fiscal union models are consistent with it. These theoretical restrictions imply a series of particular linear restrictions linking the reduced form errors to the structural shocks. Hence, this identification scheme fits nicely with the philosophy in this paper and makes it easy to verify that the restrictions in R^* are not violated. Beraja, Hurst, and Ospina (2015) shows another application of this scheme. Baumeister and Hamilton (2015) present a scheme that is very similar in spirit.

Following Ravenna (2007), if Assumptions 1, 2 hold and $Q(\xi,\Theta)$ is a non-singular square matrix, then there is a structural vector autoregression (SVAR) representation of the solution (RR) of the form:

$$x_t = \rho_1(\xi,\Theta)x_{t-1} + \rho_2(\xi,\Theta)x_{t-2} + Q(\xi,\Theta)u_t \tag{SVAR}$$

where $\rho_1(\xi,\Theta) \equiv P(\xi,\Theta) + Q(\xi,\Theta)NQ(\xi,\Theta)^{-1}$; $\rho_2(\xi,\Theta) \equiv (P(\xi,\Theta) - \rho_1(\xi,\Theta))P(\xi,\Theta)$ and $V(\xi,\Theta) \equiv Var(Q(\xi,\Theta)u_t) = Q(\xi,\Theta)\Sigma\Sigma' Q(\xi,\Theta)'$.

To see this, note that we can write $z_{t-1} = Q(\xi,\Theta)^{-1}(x_{t-1} - P(\xi,\Theta)x_{t-2})$ and replace it and the law of motion for the exogenous states into the law of motion for the endogenous variables to obtain the SVAR(2) representation.

Without loss of generality, I normalize the covariance matrix of structural shocks Σ to the identity matrix in what follows. The first step in the procedure consists of estimating the reduced form VAR to obtain the autoregressive matrices $\{\rho_1, \rho_2\}$, and the reduced form errors covariance matrix V. The second step is deriving identification restrictions that will allow us to infer Q and the shocks.

Applying the conditional expectation operator $\mathbb{E}_{t-1}(.)$ on both sides of the (SB) and constructing the reduced form expectational errors, we obtain:

$$0 = \begin{bmatrix} G_{31} + \vartheta_n & G_{32} + \vartheta_w & G_{33} \end{bmatrix} Q \begin{bmatrix} u_t^{\gamma} \\ u_t^{z} \\ u_t^{\eta} \end{bmatrix} + M_{33}u_t^{\eta} \tag{Id1}$$

This equation must hold for all realizations of the shocks. Whenever there is an innovation to u_t^{γ} or u_t^{z} and $u_t^{\eta} = 0$, employment, wages, and debt must co-move on impact in a way that satisfies this linear relationship. Hence, it gives us two linear restrictions in the second and third columns' elements of Q for a given parameterization of (SB) when there are either contemporaneous u_t^{γ} or u_t^{z} shocks.

Similarly, constructing $\mathbb{E}_{t-1}(.) - \mathbb{E}_{t-2}(.)$, we obtain:

$$0 = \left(\begin{bmatrix} G_{31} + \vartheta_n & G_{32} + \vartheta_w & G_{33} \end{bmatrix} \rho_1 + \begin{bmatrix} 0 & 0 & H_{33} + \vartheta_b \end{bmatrix} \right) Q \begin{bmatrix} u^{\gamma}_{t-1} \\ u^{z}_{t-1} \\ u^{\eta}_{t-1} \end{bmatrix}$$
$$+ M_{33} n_{33} u^{\eta}_{t-1} + M_{33} n_{32} u^{z}_{t-1} \tag{Id2}$$

This gives us one extra linear restriction in the second column's elements of Q for a given parameterization of (SB) when there are u^{γ}_{t-1} shocks. These three linear restrictions, combined with six non-linear restrictions coming from the orthogonalization of the shocks, are sufficient to identify all nine elements in Q. Intuitively, equation (Id1) separates the "wealth" shock from the other two shocks. If the unexpected component of employment, wages, and assets does not co-move in the linear way implied by equation (Id1), when $u^{\eta}_{t} = 0$, a "wealth" shock must have occurred. Analogously, equation (Id2) separates "discount rate" and "productivity" shocks. If the unexpected component of employment, wages, and assets does not co-move in the linear way implied by equation (Id2), when $u^{z}_{t-1} = u^{\eta}_{t-1} = 0$, a "discount rate" shock occurred. For completeness, matrix Q solves the system:

$$\begin{bmatrix} G_{31} + \vartheta_n & G_{32} + \vartheta_w & G_{33} \end{bmatrix} Q \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 \end{bmatrix}$$

$$\left(\begin{bmatrix} G_{31} + \vartheta_n & G_{32} + \vartheta_w & G_{33} \end{bmatrix} \rho_1 + \begin{bmatrix} 0 & 0 & H_{33} + \vartheta_b \end{bmatrix} \right) Q \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} = 0$$

$$V = QQ'$$

B.4 Estimating the semi-structure ξ^* and the impulse response matrix Q^0

I estimate a vector autoregression on employment, wages, and assets via weighted OLS where the weights are the 2006 population in the state, using data described in subsection B.1. For each variable and year, I take the cumulative log-growth between 2006 and 2011 and express it in log-deviations from the average across states. I pool all data between 2006 and 2011, leaving 240 observations (5 years*48 states), and estimate common autoregressive coefficients ρ^0_1, ρ^0_2 and reduced form errors U covariance matrix for all states $V^0 = \frac{UU'}{240-3*2}$.

Given ρ^0_1, ρ^0_2, we find solutions with all eigenvalues inside the unit circle to the quadratic equation $\rho^0_2 = (P - \rho^0_1)P$. Under Assumptions 1-2 and Property 4, there are only two such solutions. The first corresponds to P^0 in the unique stable recursive representation of the equilibrium under fiscal integration. The second corresponds to $Q^0 N (Q^0)^{-1}$. I identify P^0 as the solution that results in an implied $N = \left(Q^0\right)^{-1} (P^0 - \rho^0_1) Q^0$ that satisfies the exclusion restrictions in ξ^*.

From the restrictions implied by the third line of the semi-structure in Assumption 5, together with results in Proposition 1, we have,

$$\left[\begin{array}{ccc} G_{31}+\vartheta_n^0 & G_{32}+\vartheta_w^0 & G_{33} \end{array}\right] P^0 + \left[\begin{array}{ccc} 0 & 0 & H_{33}+\vartheta_b^0 \end{array}\right] = \mathbf{0}_{1.3}$$

Then, G_{31}, G_{32}, H_{33} are identified from the above system of equations, given P^0 and setting $G_{33} \equiv \frac{\bar{B}}{\bar{S}} = 2.25$, to match the median net worth to revenues ratio across states in the United States in 2006, and $\vartheta_n^0 = -1.6, \vartheta_w^0 = -0.9, \vartheta_b^0 = -0.03$, which correspond to the OLS policy rule estimates in Table 3.

Then, with the above inputs, I follow B.3 to identify Q^0. Furthermore, the rest of the semi-structure ξ^* is identified by following the results in Proposition 1 using the restrictions implied by Assumption 5 and the estimated P^0, Q^0, Θ^0 and $N^* = \left(Q^0\right)^{-1}(P^0 - \rho_1^0)Q^0$.

Online Appendix

A Fiscal Unions as a Collection of Small Open Economies

Consider an economy comprised of many islands, inhabited by a representative household and firm. The only other agent in the economy is a federal government. Households consume, work, and save/borrow in a non-state-contingent asset—a nominal bond in zero net supply. Firms produce final consumption goods using labor and intermediate goods. By assumption, the final consumption good is non-tradable, intermediate goods are tradable, and labor is not mobile across islands. Finally, each island has an exogenous endowment of intermediate goods. The federal government sets the nominal interest rate on the nominal bond, and gives lump-sum transfers to the islands. Assume that the nominal interest rate follows an endogenous rule that is a function of only aggregate variables (together with a fixed nominal exchange rate, this implies that the islands are part of a monetary union). Also, assume that federal transfers are a function of island-level variables alone. Throughout, I assume that parameters governing preferences and production are identical across islands and the islands only differ, potentially, in the shocks that hit them—these shocks include a shifter of the households discount rate, a productivity shifter in the production function of final goods, and the exogenous endowment of tradable intermediate goods. Finally, I assume that all labor, goods and asset markets are competitive.

A.1 Firms and Households

Final goods producers use labor N^y_{kt} and intermediates X_{kt} in island k at time t and face prices P_{kt}, wages W_{kt}, and intermediate prices Q_t (equalized across all islands because of assumed tradability). Their profits are

$$\max_{N^y_{kt}, X_{kt}} P_{kt} e^{z_{kt}} (N^y_{kt})^{\alpha} (X_{kt})^{\beta} - W_{kt} N^y_{kt} - Q_t X_{kt}$$

where z_{kt} is a productivity shock and $(\alpha, \beta) : \alpha + \beta < 1$ are the labor and intermediates shares. Unlike the tradable goods prices, final good prices (P_{kt}) vary across islands.[28]

Households preferences are given by

$$\mathbb{E}_0 \left[\sum_{t=0}^{\infty} e^{-\rho_{kt} - \delta_{kt}} \frac{(C_{kt} - \frac{\phi}{1+\phi} N_{kt}^{\frac{1+\phi}{\phi}})^{1-\sigma}}{1-\sigma} \right]$$

where C_{kt} is consumption of the final good, N_{kt} is labor, δ_{kt} is an exogenous processes driving the household's discount rate.

Households are able to spend their labor income $W_{kt} N_{kt}$ plus profits accruing from firms Π_{kt} and exogenous endowment of tradable goods $Q_t e^{\eta_{kt}}$, financial income $B_{kt} i_t$ and transfers from

---

[28] It is worth noting that all model shocks will generate endogenous variation in markups given assumed decreasing returns to scale. Additionally, what I call a "productivity shock" is isomorphic to any shifter of unit labor costs and, hence, labor demand schedules. I will not attempt to distinguish between the different interpretations of this shock in this paper.

the government S_{kt}, where B_{kt} are nominal bond holdings at the beginning of the period and i_t is the nominal interest (equalized across islands given our assumption of a monetary union where the bonds are freely traded) on consumption goods (C_{kt}) and savings $(B_{kt+1} - B_{kt})$. Thus, they face the period-by-period budget constraint

$$P_{kt}C_{kt} + B_{kt+1} \leq B_{kt}(1 + i_t) + W_{kt}N_{kt} + \Pi_{kt} + S_{kt} + Q_t e^{\eta_{kt}}$$

A well known issue in the international macroeconomics literature is that under market incompleteness of the type we just described there is no stationary distribution for bond holdings across islands in the log-linearized economy, and all other island variables in the model have unit roots. This is problematic for reasons both theoretical (I will like to study log-deviations from a deterministic steady state) and empirical (regional data for the US does not suggest the presence of such unit roots). I follow Schmitt-Grohe and Uribe (2003) and let ρ_{kt} be the endogenous component of the discount factor that satisfies $\rho_{kt+1} = \rho_{kt} + \Phi(.)$ for some function $\Phi(.)$ of the average per capita variables in an island. As such, agents do not internalize this dependence when making their choices. This modification induces stationarity for an appropriately chosen function $\Phi(.)$. Schmitt-Grohe and Uribe (2003) show that alternative stationary inducing modifications (a specification with internalization, a debt-elastic interest rate or convex portfolio adjustment costs) all deliver similar quantitative results in the context of a small open economy real business cycle model.

A.2 Federal government

The federal government budget constraint is

$$B_t^g + \sum_k S_{kt} + Q_t G = B_{t-1}^g(1 + i_t)$$

where G is some exogenous level of government spending in intermediate goods. The key feature of a fiscally integrated economy is that the federal government has the ability to redistribute resources across islands via transfers S_{kt}. If the islands where fiscally independent such transfers would not be possible.

I assume that the federal government announces a nominal interest rate rule $i_t = i(.)$ as a function of aggregate variables in the economy alone. Moreover, it announces a transfer policy rule as a function of per-capita employment, wages and assets in an island

$$S_{kt} = \bar{S}(\tilde{W}_{kt})^{\vartheta_w}(\tilde{N}_{kt})^{\vartheta_n}(\tilde{B}_{kt-1})^{\vartheta_b}$$

Again, agents do not internalize this dependence when making their choices.

A.3 Exogenous shocks and processes

I assume the exogenous processes are AR(1) processes, with an identical autoregressive coefficient across islands, and that the innovations are iid, mean zero, random variables with an aggregate and island specific component. First, define $\gamma_{kt} \equiv \delta_{kt} - \delta_{kt-1}$. Then,

$$\begin{aligned} z_{kt} &= \rho_z z_{kt-1} + \tilde{\sigma}_z v_t^z + \sigma_z u_{kt}^z \\ \gamma_{kt} &= \rho_\gamma \gamma_{kt-1} + \tilde{\sigma}_\gamma v_t^\gamma + \sigma_\gamma u_{kt}^\gamma \\ \eta_{kt} &= \rho_\eta \eta_{kt-1} + \tilde{\sigma}_\eta v_t^\eta + \sigma_\eta u_{kt}^\eta \end{aligned}$$

with $\sum_k u_{kt}^z = \sum_k u_{kt}^\gamma = \sum_k u_{kt}^\eta = 0$. By assumption, I assume the average of the regional shocks sum to zero in all periods.

The "discount rate" process γ_{kt} is a shifter of a household's discount rate, but it can be viewed as a proxy for the tightening of household borrowing limits. The "productivity" process z_{kt} can be interpreted as actual productivity, or a shifter of firm's demand for labor or firm's markups. Finally, "wealth" process η_{kt} is modeled as an endowment of intermediate goods but can be interpreted as shifters of the budget constraint that agents face such as exogenous changes in household wealth.

A.4 Equilibrium

An equilibrium is a collection of prices $\{P_{kt}, W_{kt}, Q_t\}$ and quantities $\{C_{kt}, N_{kt}, B_{kt}, N_{kt}^y, X_{kt}\}$ for each island k and time t such that, for an interest rate rule $i_t = i(.)$ and given exogenous processes $\{z_{kt}, \eta_{kt}, \gamma_{kt}\}$, they are consistent with household utility maximization and firm profit maximization and such that the following market clearing conditions hold:

$$\begin{aligned} C_{kt} &= e^{z_{kt}} (N_{kt}^y)^\alpha (X_{kt})^\beta \\ N_{kt} &= N_{kt}^y \\ G + \sum_k X_{kt} &= \sum_k e^{\eta_{kt}} \\ 0 &= \sum_k B_{kt} + B_t^g \end{aligned}$$

A.5 Aggregation

The first important assumption for aggregation is that all islands are identical with respect to their underlying production and utility parameters.[29] The second assumption is that the joint distribution of island-specific shocks is such that its cross-sectional summation is zero. If K, the number of islands, is large this holds in the limit because of the law of large numbers. I log-linearize the model around this steady state and show that it aggregates up to a representative

---

[29] Given that the broad industrial composition at the state level does not differ much across states, the assumption that productivity parameters are roughly similar across states is not dramatically at odds with the data.

economy where all aggregate variables are independent of any cross-sectional considerations to a first order approximation.[30] I denote with lowercase letters an island variable's log-deviation from the aggregate union equilibrium. Lowercase letters with a tilde denote deviations from the steady state. For example, $n_{kt} \equiv \tilde{n}_{kt} - \tilde{n}_t$ and $\tilde{n}_t \equiv \sum_k \frac{1}{K}\tilde{n}_{kt} = \sum_k \frac{1}{K} log\left(N_{kt}/\bar{N}\right)$. I assume that the monetary authority announces the nominal interest rate rule in log-linearized form: $\tilde{i}_{t+1} = \varphi_\pi \mathbb{E}_t[\tilde{\pi}_{t+1}]$ where $\tilde{\pi}_t$ is the aggregate inflation rate. Finally, I assume that the endogenous component of the discount factor is $\Phi(.) = \Phi_0 n_{kt} + \Phi_1 n_{kt-1}$.[31]

The following lemma present the aggregation result and shows that we can write the island level equilibrium in deviations from these aggregates.

Lemma 1 *For given $\{z_{kt}, \gamma_{kt}, \eta_{kt}\}$, the behavior of $\{w_{kt}, n_{kt}, b_{kt}, p_{kt}, c_{kt}, x_{kt}\}$ in the log-linearized economy for each island in log-deviations from aggregates is identical to that of a small open economy where the price of intermediates and the nominal interest rate are at their steady state levels, i.e. $\tilde{q}_t = \tilde{i}_t = 0\ \forall t$.*

Proof.

The following equations characterize the log-linearized equilibrium

$$\begin{aligned}
\tilde{w}_{kt} - \tilde{p}_{kt} &= \frac{1}{\phi}\tilde{n}_{kt} \\
\tilde{w}_{kt} - \tilde{p}_{kt} &= (\alpha - 1)\tilde{n}_{kt} + \beta\tilde{x}_{kt} + \tilde{z}_{kt} \\
\tilde{q}_t - \tilde{p}_{kt} &= \alpha\tilde{n}_{kt} + (\beta - 1)\tilde{x}_{kt} + \tilde{z}_{kt} \\
0 &= \mathbb{E}_t\left(\tilde{m}u_{kt+1} - \tilde{m}u_{kt+1} - (\tilde{p}_{kt+1} - \tilde{p}_{kt}) - \gamma_{kt+1} - \Phi_0(\tilde{c}_{kt} - \tilde{c}_t) + \tilde{i}_{t+1}\right) \\
\tilde{m}u_{kt+1} &= -\frac{\sigma}{C - \frac{\phi}{1+\phi}N^{\frac{1+\phi}{\phi}}}\left(C\tilde{c}_{kt+1} - N^{\frac{1+\phi}{\phi}}\tilde{n}_{kt+1}\right) \\
\tilde{c}_{kt} &= \tilde{w}_{kt} - \tilde{p}_{kt} + \tilde{n}_{kt} \\
B\tilde{b}_{kt} &= B(1+r)(\tilde{b}_{kt-1} + \tilde{i}_t) + \eta_{kt} - X(\tilde{q}_t + \tilde{x}_{kt}) + S\tilde{s}_{kt} \\
\sum_k \tilde{x}_{kt} &= \sum_k \tilde{\eta}_{kt} \\
B^g\tilde{b}^g_t + S\sum_k \tilde{s}_{kt} + G\tilde{q}_t &= B^g(1+r)(\tilde{b}^g_{t-1} + \tilde{i}_t) \\
\tilde{s}_{kt} &= \vartheta_w\tilde{w}_{kt} + \vartheta_n\tilde{n}_{kt} + \vartheta_b\tilde{b}_{kt-1} \\
\tilde{i}_{t+1} &= \phi_p\mathbb{E}_t[\tilde{p}_{t+1} - \tilde{p}_t]
\end{aligned}$$

---

[30] The model we presented has many islands subject to idiosyncratic shocks that cannot be fully hedged because asset markets are incomplete. By log-linearizing the equilibrium we gain in tractability, but ignore these considerations and the aggregate consequences of heterogeneity. As usual, the approximation will be a good one as long as the underlying volatility of the idiosyncratic shocks is not too large. If our unit of study was an individual, as for example in the precautionary savings literature with incomplete markets, the use of linear approximations would likely not be appropriate. However, since our unit of study is an island the size of a state I believe this is not too egregious of an assumption. The volatilities of key economic variables of interest at the state level are orders of magnitude smaller than the corresponding variables at the individual level.

[31] When $\Phi_0 > 0$ this will be enough to induce stationary of island level variables in log-deviations from the aggregate. At the same time, since $\Phi(.)$ depends only on these deviations, the aggregate equilibrium will feature a constant endogenous discount factor ρ.

After adding up, the aggregate log-linearized equilibrium evolution of $\{\tilde{w}_t - \tilde{p}_t, \tilde{n}_t\}$ is characterized by

$$\begin{aligned}
0 &= \mathbb{E}_t(\widetilde{mu}_{t+1} - \widetilde{mu}_t + (\phi_p - 1)(\tilde{p}_{t+1} - \tilde{p}_t) + \tilde{\gamma}_{t+1}) \\
0 &= \frac{1}{\phi}\tilde{n}_t - (\tilde{w}_t - \tilde{p}_t) \\
\tilde{w}_t - \tilde{p}_t &= (\alpha - 1)\tilde{n}_t + \tilde{z}_t + \beta\tilde{\eta}_t \\
\widetilde{mu}_{t+1} &\equiv -\frac{\sigma}{C - \frac{\phi}{1+\phi}N^{\frac{1+\phi}{\phi}}}\left(C(\tilde{w}_{t+1} - \tilde{p}_{t+1} + \tilde{n}_{t+1}) - N^{\frac{1+\phi}{\phi}}(\tilde{n}_{t+1})\right)
\end{aligned}$$

which is equivalent to the system of equations characterizing the log-linearized equilibrium in a representative agent economy with a production technology that utilizes labor alone with an elasticity of α, no endogenous discounting and only 2 exogenous processes $\{\tilde{z}_t + \beta\tilde{\eta}_t, \tilde{\gamma}_t\}$.

Next, take log-deviations from the aggregate in the original system. This results in the system characterizing the evolution of $\{p_{kt}, w_{kt}, n_{kt}, b_{kt}, c_{kt}\}$ for given $\{z_{kt}, \eta_{kt}, \gamma_{kt}\}$

$$\begin{aligned}
0 = \mathbb{E}_t &\Bigg[\frac{\sigma\beta}{\frac{\alpha\phi}{1+\phi} - 1}(w_{kt+1} + n_{t+1} - w_{kt} - n_{kt}) + (\alpha + \beta - 1)(n_{kt+1} - n_{kt}) \\
&+(\beta - 1)(w_{kt+1} - w_{kt}) + (1 + \frac{\sigma}{\frac{\alpha\phi}{1+\phi} - 1})(z_{kt+1} - z_{kt}) + \Phi_0 c_{kt} - \gamma_{kt+1}\Bigg] \\
\beta w_{kt} &= \left(\frac{1+\phi}{\phi} - (\alpha + \beta)\right) n_{kt} - z_{kt} \\
\frac{B}{S} b_{kt} &= \frac{B}{S}(1 + r) b_{kt-1} - \frac{X}{S}(w_{kt} + n_{kt}) + \vartheta_n n_{kt} + \vartheta_w w_{kt} + \vartheta_b b_{kt-1} + \frac{1}{S}\eta_{kt} \\
c_{kt} &= w_{kt} - p_{kt} + n_{kt} \\
p_{kt} &= (1 - \beta) w_{kt} - (z_{kt} + (\alpha + \beta - 1) n_{kt})
\end{aligned}$$

This system is independent of all aggregate variables and is analogous to the system characterizing the equilibrium in a small open economy without movements in the terms of trade and nominal interest rate. ■