

# A Semi-structural Methodology for Policy Counterfactuals\*

Martin Beraja<sup>†</sup>

*Massachusetts Institute of Technology and NBER*

March 14, 2022

## Abstract

I propose a methodology for constructing counterfactuals with respect to changes in policy rules which does not require fully specifying a particular model, yet is not subject to Lucas Critique. It applies to a class of dynamic stochastic models whose equilibria are well approximated by a linear representation. It rests on the insight that many such models satisfy a principle of counterfactual equivalence: they are observationally equivalent under a benchmark policy and yield an identical counterfactual equilibrium under an alternative one.

---

\*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. For their valuable suggestions, I owe George-Marios Angeletos, Andy Atkeson, Philip Haile, Ricardo Lagos, John Leahy, Martin Schneider, Gianluca Violante, and Ivan Werning as well as my classmates Jonathan Adams, Philip Barret, and Chanont Banterngghansa. I thank those at the Review of Economic Studies Tour, Minneapolis Fed Junior Scholar Conference, SED, "Pizzanomics," NBER SI Dynamic Equilibrium Models, and Macro Seminars at MIT, Harvard, Yale, Wharton, UCLA, GWU, Stanford, Princeton, NYU, Michigan, Northwestern, Cornell, Brown, LSE, BU, Berkeley, Fed Board, Atlanta Fed, Chicago Fed, BCRA, and Norges Bank. I thank all participants of the Capital Theory Working Group, especially Harald Uhlig and Nancy Stokey. Finally, I thank Juliette Caminade for her support and comments on preliminary versions of this paper.

<sup>†</sup>E-mail: maberaja@mit.edu. Web: [economics.mit.edu/faculty/maberaja](http://economics.mit.edu/faculty/maberaja)

# 1 Introduction

Economists follow either structural or reduced-form (and semi-structural) approaches to answer counterfactual questions. The former relies on specifying primitives of a particular model. This is a daunting task whenever researchers are uncertain about features of alternative models that are hard to distinguish with available data and reasonable a-priori. If counterfactuals differ under these alternative models, their credibility is undermined. The latter approaches require less commitment to a particular model and have proven useful to evaluate observed policy changes, as in Sims (1980). However, the structural approach is the leading paradigm for studying unobserved, potential changes in policy rules because the analysis is not subject to the critique in Lucas (1976). Are there then reasonable circumstances when we can analyze the effects of a counterfactual policy rule change without having to commit to a particular model?

In this paper, I propose a semi-structural methodology for constructing counterfactuals with respect to policy rule changes. It is useful in circumstances when a linearization approach is reasonable and we do not wish to fully specify a model's microfoundations, yet we are worried about using approaches that contain too little information for the analysis to be immune to Lucas Critique — like those based on structural VARs.<sup>1</sup> The method hinges on an insight about dynamic stochastic models with equilibria that are well approximated by a linear representation, like the ones in Uhlig (1995). The insight is that many models that can match an economy's reduced-form equilibrium under a benchmark policy rule — observationally equivalent models — will also generate an identical counterfactual equilibrium under an alternative policy rule. Regardless of their microfoundations, these models thus satisfy a *principle of counterfactual equivalence*.

The method has three steps. The first is using data from an economy under a benchmark policy rule to estimate a recursive law of motion of the equilibrium of dynamic stochastic models — a *reduced form model*. The second is imposing restrictions on these models' equilibrium equations. They need to be enough to identify all coefficients in such equations — the *structural model* — given the reduced form model. The last step is solving for the counterfactual equilibrium under an alternative policy rule, given the identified structural model. As such, the method is in the spirit of ideas advanced at the Cowles Commission (e.g., Hurwicz, 1962; Marschak, 1974) as well as more recently by the literature on sufficient statistics (e.g., Chetty, 2009).<sup>2</sup>

## 2 Illustrative example: Interest rate rules in New Keynesian models

The goal of this section is to illustrate the semi-structural methodology in a simple example. I consider counterfactuals with respect to changes in interest rate policy rules in the context of 3-equation New Keynesian models. I begin by describing notation and language that will be used

---

<sup>1</sup>Bernanke, Gertler, and Watson (1997) and Sims and Zha (2006) evaluate counterfactual policy rules by “zeroing-out” policy responses in SVARs. Policy institutions use this method as well as macroeconometric models like FRB/US.

<sup>2</sup>Marschak's Maxim says that for many policy questions it may not be necessary to fully identify all model parameters, but only combinations of subsets of parameters. Hurwicz (1962) noted that these should be policy invariant.

throughout. I then show that different models are counterfactually equivalent. Finally, I show how to construct counterfactuals without having to specify a particular New Keynesian model.

The canonical New Keynesian model describes the equilibrium behavior of output  $y_t$ , inflation  $\pi_t$ , and the nominal interest rate  $i_t$  in log-deviations from a zero inflation steady state in terms of three equations. In matrix form, the equilibrium conditions include the Euler equation and the New Keynesian Phillips Curve (NKPC)

$$0 = \begin{bmatrix} 1 & \frac{1}{\sigma} & 0 \\ 0 & \beta & 0 \end{bmatrix} \mathbb{E}_t \begin{bmatrix} y_{t+1} \\ \pi_{t+1} \\ i_{t+1} \end{bmatrix} + \begin{bmatrix} -1 & 0 & -\frac{1}{\sigma} \\ \kappa & -1 & 0 \end{bmatrix} \begin{bmatrix} y_t \\ \pi_t \\ i_t \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} y_{t-1} \\ \pi_{t-1} \\ i_{t-1} \end{bmatrix} + \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} b_t \\ a_t \\ m_t \end{bmatrix} \quad \begin{array}{l} \text{(Euler)} \\ \text{(NKPC)} \end{array}$$

and the interest rate policy rule

$$0 = \begin{bmatrix} 0 & \theta_\pi & 0 \end{bmatrix} \mathbb{E}_t \begin{bmatrix} y_{t+1} \\ \pi_{t+1} \\ i_{t+1} \end{bmatrix} + \begin{bmatrix} \theta_y & 0 & -1 \end{bmatrix} \begin{bmatrix} y_t \\ \pi_t \\ i_t \end{bmatrix} + \begin{bmatrix} 0 & 0 & \theta_i \end{bmatrix} \begin{bmatrix} y_{t-1} \\ \pi_{t-1} \\ i_{t-1} \end{bmatrix} + \begin{bmatrix} 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} b_t \\ a_t \\ m_t \end{bmatrix}, \quad \text{(Rule)}$$

where  $1/\sigma$  is the intertemporal elasticity of substitution,  $\beta$  is the discount factor,  $\kappa$  is a combination of subsets of parameters (e.g., the frequency of price changes), and  $\{\theta_\pi, \theta_y, \theta_i\}$  are policy rule parameters. Moreover,  $\{b_t, a_t, m_t\}$  are assumed to be independent AR(1) processes described by

$$0 = - \begin{bmatrix} b_{t+1} \\ a_{t+1} \\ m_{t+1} \end{bmatrix} + N \begin{bmatrix} b_t \\ a_t \\ m_t \end{bmatrix} + \begin{bmatrix} \epsilon_{t+1}^b \\ \epsilon_{t+1}^a \\ \epsilon_{t+1}^m \end{bmatrix}, \quad \text{(Shocks)}$$

where  $N$  is a diagonal matrix of parameters governing persistence, and  $\{\epsilon_t^b, \epsilon_t^a, \epsilon_t^m\}$  are respectively a demand shock, a cost-push shock, and a monetary policy shock.

From now on, let  $\zeta$  denote the collection of matrices with structural parameters in (Euler) and (NKPC), and  $\Theta$  denote the collection of matrices with policy parameters in (Rule). I will say that  $\mu \equiv \{\zeta, \Theta\}$  is the *structural model* described by the *structure*  $\zeta$  and the *policy*  $\Theta$ .

Suppose that there is a unique and stable recursive law of motion of the equilibrium

$$\begin{bmatrix} y_t \\ \pi_t \\ i_t \end{bmatrix} = P \begin{bmatrix} y_{t-1} \\ \pi_{t-1} \\ i_{t-1} \end{bmatrix} + Q \begin{bmatrix} b_t \\ a_t \\ m_t \end{bmatrix}. \quad \text{(RLM-NK)}$$

From here on, I will say that  $\Gamma \equiv \{P, Q, N\}$  is the *reduced form model*, where  $P$  and  $Q$  are the (stable) solution to the non-linear system of matrix equations

$$\begin{bmatrix} 1 & \frac{1}{\sigma} & 0 & -1 & 0 & -\frac{1}{\sigma} & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & \beta & 0 & \kappa & -1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & \theta_\pi & 0 & \theta_y & 0 & -1 & 0 & 0 & \theta_i & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} (P)^2 & QN + PQ \\ P & Q \\ 0_{3.3} & 0_{3.3} \\ 0_{3.3} & I_{3.3} \end{bmatrix} = 0_{3.6}. \quad \text{(SMERLM-NK)}$$

For example, the structural model  $\mu^0$  described by structure  $\zeta^0$  and policy  $\Theta^0$  below

$$\zeta^0 = \begin{bmatrix} 1 & \frac{1}{2} & 0 & -1 & 0 & -\frac{1}{2} & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0.9 & 0 & 0.3 & -1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$

$$\Theta^0 = \begin{bmatrix} 0 & 0.7 & 0 & 0.5 & 0 & -1 & 0 & 0 & 0.6 & 0 & 0 & 1 \end{bmatrix}$$

has the reduced form model  $\Gamma^0$

$$P^0 = \begin{bmatrix} 0 & 0 & -0.35 \\ 0 & 0 & -0.16 \\ 0 & 0 & 0.38 \end{bmatrix} \quad Q^0 = \begin{bmatrix} 1.70 & -0.50 & -0.59 \\ 1.46 & 3.24 & -0.27 \\ 1.59 & 1.61 & 0.63 \end{bmatrix} \quad N^0 = \begin{bmatrix} 0.9 & 0 & 0 \\ 0 & 0.9 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

**Observational equivalence.** Many models have the same reduced form under policy  $\Theta^0$ : they are observationally equivalent. To see this, consider two models with a different NKPC. The first one is from Gabaix (2020) where firms are inattentive. The NKPC in this “behavioral model” is

$$0 = \begin{bmatrix} 0 & \beta M^f & 0 \end{bmatrix} \mathbb{E}_t \begin{bmatrix} y_{t+1} \\ \pi_{t+1} \\ i_{t+1} \end{bmatrix} + \begin{bmatrix} \kappa m^f - 1 & 0 \end{bmatrix} \begin{bmatrix} y_t \\ \pi_t \\ i_t \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} y_{t-1} \\ \pi_{t-1} \\ i_{t-1} \end{bmatrix} + \begin{bmatrix} 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} b_t \\ a_t \\ m_t \end{bmatrix} \quad (\text{NKPC1})$$

where  $M^f$  and  $m^f$  govern inattention to future economic conditions. The second model is from Christiano, Trabandt, and Walentin (2010) where firms need to borrow short term to finance materials and labor inputs. The NKPC in this “working-capital model” is

$$0 = \begin{bmatrix} 0 & \beta & 0 \end{bmatrix} \mathbb{E}_t \begin{bmatrix} y_{t+1} \\ \pi_{t+1} \\ i_{t+1} \end{bmatrix} + \begin{bmatrix} \kappa & -1 & \chi \end{bmatrix} \begin{bmatrix} y_t \\ \pi_t \\ i_t \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} y_{t-1} \\ \pi_{t-1} \\ i_{t-1} \end{bmatrix} + \begin{bmatrix} 0 & \gamma & 0 \end{bmatrix} \begin{bmatrix} b_t \\ a_t \\ m_t \end{bmatrix} \quad (\text{NKPC2})$$

where  $\chi$  and  $\gamma$  are combinations of parameters, and account for financing and materials needs.

The behavioral model has reduced form  $\Gamma^0$  because it has structure  $\zeta^0$  when  $\beta M^f = 0.9$  and  $\kappa m^f = 0.3$ . The working-capital model has reduced form  $\Gamma^0$  even if it has a different structure  $\zeta^1$

$$\zeta^1 = \begin{bmatrix} 1 & \frac{1}{2} & 0 & -1 & 0 & -\frac{1}{2} & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0.70 & 0 & 0.38 & -1 & 0.04 & 0 & 0 & 0 & 0 & 1.5 & 0 \end{bmatrix}.$$

The common feature of these models is that their structures satisfy exactly 6 restrictions per line — see the structure in (Restrictions) below. As Lemma A.1 in the Online Appendix shows in the general case, this feature makes them observationally equivalent.

**Counterfactual equivalence.** Consider a policy change from  $\Theta^0$  to  $\Theta^1$ . There are models that have the same reduced form  $\Gamma^0$  under policy  $\Theta^0$  that *also have the same reduced form*  $\Gamma^1$  under policy  $\Theta^1$ : they are counterfactually equivalent. But not all observationally equivalent models

satisfy such principle of counterfactually equivalence.

For instance, consider a more “hawkish” interest rate rule with  $\theta_y = 0$  instead of  $\theta_y = 0.4$ . Then, both the canonical model and behavioral models associated with structural model  $\{\zeta^0, \Theta^1\}$  have the same reduced form  $\Gamma^1$ , but the working-capital model associated with structural model  $\{\zeta^1, \Theta^1\}$  has a different reduced form  $\tilde{\Gamma}^1$ . In particular, the two counterfactuals are

$$P^1 = \begin{bmatrix} 0 & 0 & -0.63 \\ 0 & 0 & -0.34 \\ 0 & 0 & 0.48 \end{bmatrix} \quad Q^1 = \begin{bmatrix} 5.32 & -1.56 & -0.84 \\ 5.86 & 1.95 & -0.38 \\ 3.32 & 1.11 & 0.90 \end{bmatrix}; \quad \tilde{P}^1 = \begin{bmatrix} 0 & 0 & -0.63 \\ 0 & 0 & -0.33 \\ 0 & 0 & 0.49 \end{bmatrix} \quad \tilde{Q}^1 = \begin{bmatrix} 5.72 & -1.67 & -0.84 \\ 5.36 & 2.10 & -0.38 \\ 3.03 & 1.19 & 0.90 \end{bmatrix}.$$

The example shows that the principle of counterfactual equivalence is intimately related to Lucas Critique. If the data generating process is the structural model  $\mu^0$ , a model with structure  $\zeta^1$  would match the reduced form  $\Gamma^0$  under policy  $\Theta^0$  but would lead to the “wrong” counterfactual  $\Gamma^1$  under policy  $\Theta^1$ . Such counterfactual analysis would not be immune to Lucas Critique.

**The semi-structural methodology.** Suppose that a researcher has estimated the reduced form  $\Gamma^0$  using data from an economy under the interest rate policy rule  $\Theta^0$ .<sup>3</sup> Knowledge of this reduced form is not sufficient to identify a counterfactual with respect to a policy rule change. However, assume that the researcher also has some a-priori knowledge about the structural model that generated it. In particular, she knows that the structure  $\zeta$  satisfies the following 6 restrictions in each of its lines

$$\zeta = \begin{bmatrix} \zeta_{11} & \zeta_{12} & 0 & \zeta_{14} & 0 & \zeta_{16} & \zeta_{17} & \zeta_{18} & 0 & 1 & 0 & 0 \\ \zeta_{12} & \zeta_{22} & 0 & \zeta_{24} & \zeta_{25} & 0 & \zeta_{27} & \zeta_{28} & 0 & 0 & 1 & 0 \end{bmatrix} \quad (\text{Restrictions})$$

The restrictions imply, for example, that only demand shocks shift the Euler equation — the coefficients  $\{\zeta_{111}, \zeta_{112}\}$  associated with  $\{a_t, m_t\}$  are set to zero — or that the interest rate does not appear in the NKPC — the coefficients  $\{\zeta_{23}, \zeta_{26}, \zeta_{29}\}$  associated with  $\{i_{t+1}, i_t, i_{t-1}\}$  are set to zero.

Imposing these restrictions is enough to identify the unknown coefficients in the structure  $\zeta$ , given the reduced form model  $\Gamma^0$  and the policy  $\Theta^0$ . The reason is that, as Theorems 1 shows in the general case, knowledge of  $\{\Gamma^0, \Theta^0\}$  only imposes 6 restrictions per line in the structure, whereas there are 12 unknown structural coefficients per line. Then, imposing 6 additional restrictions per line identifies the full structure  $\zeta^0$ .

Finally, to construct a counterfactual equilibrium, the researcher proceeds as usual: solve a system like (SMERLM-NK) for  $\{P^1, Q^1\}$  given the structure  $\zeta^0$  and a counterfactual policy  $\Theta^1$  of interest. The counterfactual equilibrium is identical for all models described by the same structure  $\zeta^0$  regardless of their microfoundations — like the canonical and behavioral New Keynesian models in the example above. It is also robust to variations across models with different structures  $\zeta$  that are nevertheless counterfactually equivalent — these are described by Lemma A.2 and Propositions A.1 and A.2 in the Online Appendix.

<sup>3</sup>Online Appendix A.3 shows how to recover the reduced form when the equilibrium has a SVAR representation.

### 3 General case: Policy rules in dynamic stochastic models

In this section, I present the semi-structural methodology for policy counterfactuals. I first describe the class of dynamic stochastic models that it applies to. Then, I introduce the principle of counterfactual equivalence and show the main result that underpins the methodology.

Consider dynamic stochastic models with equilibria that are characterized by a linear system

$$\begin{aligned} 0 &= F\mathbb{E}_t[x_{t+1}] + Gx_t + Hx_{t-1} + L\mathbb{E}_t[z_{t+1}] + Mz_t \\ 0 &= \Theta_f\mathbb{E}_t[x_{t+1}] + \Theta_c x_t + \Theta_p x_{t-1} + \Theta_z z_t \\ 0 &= -z_{t+1} + Nz_t + \epsilon_{t+1}, \end{aligned} \tag{SME}$$

where  $x_t$  is a column vector of length 'k' that includes all observed endogenous variables in deviations from the steady state — a number 'p' of which are policy variables — and  $z_t$  is a column vector of length 's' that includes exogenous unobserved state variables. As Uhlig (1995) and the Online Appendix show, many models' equilibria can be approximated by such system.

**Definition 1 (Structural model).** *A structural model  $\mu \equiv \{\zeta, \Theta\}$  is described by a structure  $\zeta$  collecting policy-invariant matrices  $\zeta \equiv \begin{bmatrix} F & G & H & (LN + M) \end{bmatrix}$  and a policy  $\Theta$  collecting matrices of endogenous policy rule parameters  $\Theta \equiv \begin{bmatrix} \Theta_f & \Theta_c & \Theta_p & \Theta_z \end{bmatrix}$ .*

The coefficients in structure  $\zeta$  are combinations of parameters derived from fully specifying a particular model — e.g., a New Keynesian model. They often lack a direct economic interpretation. Coefficients in policy  $\Theta$  do have such interpretation — e.g., interest rate rule parameters. Note that the definition implies that policy does not affect steady state values potentially in  $\zeta$ .

Next, I restrict the class of models in (SME) to those that satisfy two assumptions.

**Assumption 1 (Stability).**  *$\{\zeta, N, \Theta\}$  are such that the system (SME) is stabilizable.*

Under Assumption 1, there is a stable recursive law of motion of the equilibrium

$$\begin{aligned} x_t &= Px_{t-1} + Qz_t \\ z_t &= Nz_{t-1} + \epsilon_t \end{aligned} \tag{RLM}$$

where  $P$  and  $N$  have all eigenvalues inside the unit circle. In particular, using the method of undetermined coefficients,  $\{P, Q\}$  solve the non-linear system of matrix equations

$$\begin{bmatrix} F & G & H & LN + M \\ \Theta_f & \Theta_c & \Theta_p & \Theta_z \end{bmatrix}_{k.(3k+s)} \begin{bmatrix} (P)^2 & QN + PQ \\ P & Q \\ I_k & 0_k \\ 0_s & I_s \end{bmatrix}_{(3k+s).(k+s)} = 0_{k.(k+s)}. \tag{SMERLM}$$

**Assumption 2 (Uniqueness).**  *$\{\zeta, N, \Theta\}$  are such that  $\{P, Q\}$  are unique.*

The assumptions ensure that the reduced form model defined below is stable and unique.

**Definition 2 (Reduced form).** *The structural model  $\mu$  has the reduced form model  $\Gamma \equiv \{P, Q, N\}$ .*

**The principle of counterfactual equivalence.** The examples in Section 2 and Online Appendix A.2 illustrate that many models are observationally equivalent: they have the same reduced form under a given policy  $\Theta$ . Lemma A.1 in the Online Appendix formally characterizes a set of observationally equivalent models in the general case. It shows that the structural model cannot be uniquely recovered from the reduced form model. There are  $3k + s$  coefficients in each line of the structure  $\zeta$ , yet the reduced form  $\Gamma$  only imposes  $k + s$  linear restrictions per line.

Next consider a policy change from  $\Theta^0$  to  $\Theta^1$ . Many models that have the same reduced form under  $\Theta^0$  would typically have a different counterfactual equilibrium under  $\Theta^1$  — the Lucas Critique. But are there models that have reduced form  $\Gamma^0$  under policy  $\Theta^0$  that also have reduced form  $\Gamma^1$  under policy  $\Theta^1$ ? I will say that such models satisfy a *principle of counterfactual equivalence*: they are observationally equivalent under both a benchmark and a counterfactual policy.

Section 2 and Online Appendix A.2 show examples of models that satisfy the principle (and some that do not). Lemma A.2 together with Propositions A.1 and A.2 in the Online Appendix characterize a set of counterfactually equivalent models in the general case. They show that there are two “types” of counterfactually equivalent models. The first are all models with the same structure  $\zeta$ . These models are nevertheless different from each other in that they have different microfoundations. The second type are models with different structures, but which satisfy the conditions in the Lemma. For example, the Propositions show that a sufficient but not necessary condition is that the lines in a model’s structure are linear combinations of the lines in another’s.

**The semi-structural methodology.** The central question of how to construct a counterfactual without having to commit to a particular microfounded model can now be stated as: What additional restrictions need to be imposed on the structural model  $\mu$  so that knowledge of the reduced form  $\Gamma^0$  under the observed policy  $\Theta^0$  uniquely identifies a structure  $\zeta^0$  and, therefore, a unique counterfactual  $\Gamma^1$  under a counterfactual policy  $\Theta^1$ ?

The theorem shows the answer: it suffices to impose  $2k$  independent linear restrictions in each line  $l = \{1, \dots, k - p\}$  of the structure, where  $k$  is the number of endogenous variables in (SME).

**Theorem 1. (Counterfactual Identification)** *Let  $\{R_l, r_l\}$  describe a set of  $2k$  independent linear restrictions on the coefficients in line ‘ $l$ ’ of a structure  $\zeta$ . Given observed policy  $\Theta^0$  and reduced form  $\Gamma^0$ , there is a unique structural model  $\mu^0 = \{\zeta^0, \Theta^0\}$  that:*

- (i) *has reduced form  $\Gamma^0$  under policy  $\Theta^0$ ,*
- (ii) *has a structure  $\zeta^0$  that satisfies the restrictions  $R_l(\zeta_l^0)' = r_l$  in each of its lines  $l = \{1, \dots, k - p\}$ .*

*Then, given  $\zeta^0$  and counterfactual policy  $\Theta^1$ , there is a unique counterfactual reduced form  $\Gamma^1$ .*

The proof is by construction. As such, it describes a semi-structural methodology for policy counterfactuals. We noticed before that solving for the reduced form  $\Gamma^0$  entailed solving the *non-linear* system of equations (SMERLM), *given* the structural model  $\mu^0 = \{\zeta^0, \Theta^0\}$  and  $N^0$ . However, consider the reverse mapping instead. *Given* the reduced form  $\Gamma^0$ , (SMERLM) implies a *linear* system of equations that any structure  $\zeta$  must satisfy in order to match this reduced form.

In particular, each line  $l = \{1, \dots, k - p\}$  in a structure solves

$$\begin{bmatrix} (P^0)^{2l} & P^{0l} & I_k & 0_{k,k} \\ (Q^0 N^0 + P^0 Q^0)^l & Q^{0l} & 0_{s,k} & I_s \end{bmatrix}_{(k+s).(3k+s)} \begin{bmatrix} F'_l \\ G'_l \\ H'_l \\ (LN^0 + M)_{l'} \end{bmatrix}_{(3k+s).1} = 0_{(k+s).1} .$$

This system is generally underdetermined — many structural models have the same reduced form. There are  $3k + s$  unknown coefficients but only  $k + s$  equations. Then, adding a set  $\{R_l, r_l\}$  of  $2k$  independent linear restrictions for line  $l'$ , we obtain the exactly determined system

$$\begin{bmatrix} (P^0)^{2l} & P^{0l} & I_k & 0_{k,k} \\ (Q^0 N^0 + P^0 Q^0)^l & Q^{0l} & 0_{s,k} & I_s \\ & R_l & & \end{bmatrix}_{(3k+s).(3k+s)} \begin{bmatrix} F'_l \\ G'_l \\ H'_l \\ (LN^0 + M)_{l'} \end{bmatrix}_{(3k+s).1} = \begin{bmatrix} 0_{(k+s).1} \\ r_l \end{bmatrix}_{(3k+s).1} .$$

Solving this system for every line  $l'$  uniquely identifies a structure  $\zeta^0$  that (i) generates  $\Gamma^0$  under  $\Theta^0$  and (ii) satisfies  $\{R_l, r_l\}$ . To construct the unique counterfactual  $\Gamma^1$ , we then solve the usual non-linear system (SMERLM) for  $\{P^1, Q^1\}$  given structure  $\zeta^0$  and counterfactual policy  $\Theta^1$ .

**Discussion.** Knowledge of the reduced form alone is not sufficient to identify a counterfactual with respect to a policy rule change. Since Lucas (1976) it has been thought that fully specifying a model's microfoundations was needed to fill this gap and to ensure that the counterfactual accounts for changes in the behavior of agents that understand that policy has changed. The semi-structural methodology just described shows that this is not necessarily the case: we can construct a counterfactual with sufficient knowledge about the structural model  $\mu$  alone.

There are three non-trivial parts to the methodology. The first is realizing that a counterfactual depends only on the structural model  $\mu$  and not the particular microfoundations that led to it (as Section 2 illustrates). This requires momentarily abandoning how we typically think about models and instead thinking about them as a collection of equilibrium equations and matrices. The second is to realize that imposing restrictions directly on equilibrium equations can identify a structure  $\zeta$  and thus a counterfactual. This requires noting that the reverse mapping from the reduced form to the structure is linear (as the proof of Theorem 1 shows). The last is realizing that the counterfactual is robust to variation in primitives across models with identical structures but also more generally across all that belong to a counterfactually equivalent set. This set is characterized by Lemma A.2 and Propositions A.1 and A.2 in the Online Appendix.

The required restrictions can be obtained in two ways in practice. The first is from *inspecting* the equilibrium equations of known models. If an equation satisfies a restriction in many models, then this is a good candidate to be part of the  $2k$  required ones. The examples in Section 2 and in the Online Appendix as well as Beraja (2021) follow this approach. The second way is *ad-hoc*. Try different restrictions and check how the counterfactual changes — a form of sensitivity analysis.



## 4 Conclusions

I have presented a semi-structural methodology to construct counterfactuals with respect to policy rule changes which does not require committing to a particular model. It rests on an insight about widely used linear dynamic stochastic models which I have called the principle of counterfactual equivalence. The principle says that many models which are observationally equivalent under a given policy also yield an identical counterfactual equilibrium under an alternative one.

The examples here served to illustrate the methodology but are not enough to assess its practical relevance. A richer application is Beraja (2021). It shows that US fiscal integration helps stabilize regional business cycles. That paper and Section A.2.3 in the Online Appendix also show two other uses. First, while it may not be computationally hard to compute counterfactuals for several models, the methodology boils them down to their core essence. This can help clarify which assumptions are important and organize previous results coming from disparate models in a literature. Second, the methodology can guide building new models that would match evidence from a policy change. One begins by trying different *ad-hoc* structural restrictions and solving for the equilibrium after the policy change. If some restrictions were able to reproduce the evidence, one can then think about microfoundations for a model that would satisfy them.

## References

- Beraja, Martin. 2021. "Does US fiscal integration stabilize regional business cycles?" *MIT mimeo* .
- Bernanke, Ben S, Mark Gertler, and Mark Watson. 1997. "Systematic monetary policy and the effects of oil price shocks." *Brookings papers on economic activity* 1997 (1):91–157.
- Chetty, Raj. 2009. "Sufficient statistics for welfare analysis: A bridge between structural and reduced-form methods." *Annual Review of Economics* 1 (1):451–488.
- Christiano, Lawrence J, Mathias Trabandt, and Karl Walentin. 2010. "DSGE models for monetary policy analysis." In *Handbook of monetary economics*, vol. 3. Elsevier, 285–367.
- Gabaix, Xavier. 2020. "A behavioral New Keynesian model." *American Economic Review* 110 (8):2271–2327.
- Hurwicz, Leonid. 1962. "On the structural form of interdependent systems, Logic, Methodology and Philosophy of Science, edited by E. Nagel, P. Suppes & A. Tarski."
- Lucas, Robert E. 1976. "Econometric policy evaluation: A critique." In *Carnegie-Rochester conference series on public policy*, vol. 1. Elsevier, 19–46.
- Marschak, Jacob. 1974. "Economic measurements for policy and prediction." In *Economic Information, Decision, and Prediction*. Springer, 293–322.
- Sims, Christopher A. 1980. "Macroeconomics and reality." *Econometrica* :1–48.
- Sims, Christopher A and Tao Zha. 2006. "Does monetary policy generate recessions?" *Macroeconomic Dynamics* 10 (2):231–272.
- Uhlig, Harald. 1995. "A toolkit for analyzing nonlinear dynamic stochastic models easily." .

ONLINE APPENDIX (NOT FOR  
PUBLICATION)

## A.1 The Principle of Counterfactual Equivalence

This section describes how to characterize a set of models that are counterfactually equivalent with respect to a policy rule change within a class of linear models of dynamic stochastic economies.

I begin by defining a class of models, and develop notation and language that I use throughout. Consider a model  $j$  whose equilibrium is generically described by the non-linear conditions:

$$\begin{aligned} 0 &= \mathbb{E}_t [f_j(x_{t+1}, x_t, x_{t-1}, z_{t+1}, z_t)] \\ 0 &= \mathbb{E}_t [\theta(x_{t+1}, x_t, x_{t-1}, z_{t+1}, z_t)] \\ 0 &= g(z_{t+1}, z_t) + \epsilon_{t+1} \quad \text{iid } \epsilon_{t+1} \text{ with } \mathbb{E}[\epsilon_{t+1}] = 0, \text{Var}(\epsilon_{t+1}) = \Sigma, \end{aligned}$$

where  $x_t$  is a column vector of length ' $k$ ' that includes all observed endogenous state variables and could include observed endogenous control variables as well,  $z_t$  is a column vector of length ' $s$ ' that includes exogenous unobserved state variables,<sup>1</sup>  $f_j(\cdot)$  is a vector function with codomain  $\mathbb{R}^k$ , and  $\theta(\cdot)$  is a policy rule with codomain  $\mathbb{R}^p$ , where  $p$  is the number of policy variables.<sup>2</sup>

A first order approximation around the non-stochastic steady state of model  $j$  results in the following system of equations:

$$\begin{aligned} 0 &= F^j \mathbb{E}_t[x_{t+1}] + G^j x_t + H^j x_{t-1} + L^j \mathbb{E}_t[z_{t+1}] + M^j z_t \\ 0 &= \Theta_f \mathbb{E}_t[x_{t+1}] + \Theta_c x_t + \Theta_p x_{t-1} + \Theta_z z_t \\ 0 &= -z_{t+1} + N z_t + \epsilon_{t+1}, \end{aligned} \tag{SME}$$

where, with some abuse of notation,  $x_t$  and  $z_t$  now represent deviations from the steady state.<sup>3</sup>

As in the paper, I next define a structural and reduced form models, and impose the assumptions of stability and uniqueness.

**Definition A.1 (Structural model).** A structural model  $\mu \equiv \{\zeta, \Theta\}$  is described by a structure  $\zeta$  collecting policy-invariant matrices  $\zeta \equiv \begin{bmatrix} F & G & H & (LN + M) \end{bmatrix}$  and a policy  $\Theta$  collecting matrices of endogenous policy rule parameters  $\Theta \equiv \begin{bmatrix} \Theta_f & \Theta_c & \Theta_p & \Theta_z \end{bmatrix}$ .

**Assumption A.1 (Stability).**  $\{\zeta, N, \Theta\}$  are such that the system (SME) is stabilizable.

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

$$\begin{aligned} x_t &= P x_{t-1} + Q z_t \\ z_t &= N z_{t-1} + \epsilon_t \end{aligned} \tag{RLM}$$

<sup>1</sup>Note that restricting the system to only one lag is without loss of generality because we can always "stack" equilibrium variables in models with more than one lag.

<sup>2</sup>Formally, the full description of the equilibrium conditions also includes initial conditions for endogenous and exogenous state variables, as well as terminal conditions. Moreover, note that neither the policy nor the laws of motion for exogenous variables are indexed by  $j$ .

<sup>3</sup>Uhlig (1995), from whom I borrow some notation, studies a very similar system of equations.

where  $P$  and  $N$  have all eigenvalues inside the unit circle. In particular, using the method of undetermined coefficients,  $\{P, Q\}$  solve the non-linear system of matrix equations

$$\begin{bmatrix} F & G & H & LN + M \\ \Theta_f & \Theta_c & \Theta_p & \Theta_z \end{bmatrix}_{k.(3k+s)} \begin{bmatrix} (P)^2 & QN + PQ \\ P & Q \\ I_k & 0_k \\ 0_s & I_s \end{bmatrix}_{(3k+s).(k+s)} = 0_{k.(k+s)}. \quad (\text{SMERLM})$$

**Assumption A.2 (Uniqueness).**  $\{\zeta, N, \Theta\}$  are such that  $\{P, Q\}$  are unique.

The assumptions ensure that the reduced form model defined below is stable and unique.

**Definition A.2 (Reduced form).** The structural model  $\mu$  has the reduced form model  $\Gamma \equiv \{P, Q, N\}$ .

### A.1.1 Observationally equivalent models under a benchmark policy $\Theta^0$

Imagine that we parameterize a model  $j^*$ , resulting in structure  $\zeta^*$ . We consider a benchmark policy  $\Theta^0$  and obtain the reduced form  $\Gamma^0$ . We now ask: what are the set of models  $j$  that can generate  $\Gamma^0$ ? Or, in other words, which models are *observationally equivalent* to  $j^*$  under the benchmark policy  $\Theta^0$ ?<sup>4</sup>

The first step in answering this question is realizing that, regardless of their primitives and particular parameters, any two models  $j$  and  $i$  with an identical structure  $\zeta = \zeta^j = \zeta^i$  will generate an identical reduced form  $\Gamma$  for any  $\Theta$ . Thus, as the following definition states, the question of observational equivalence can be understood as one about model structures.

**Definition A.3. (Observational equivalence)** Given  $\{\Theta^0, \Gamma^0\}$ , define  $\mathcal{O}(\Theta^0, \Gamma^0)$  as the set of structures  $\zeta$  that generate reduced form  $\Gamma^0$  under benchmark policy  $\Theta^0$ . Model  $j$  and  $j^*$  are observationally equivalent if and only if model  $j$  can be parameterized so that  $\zeta^j \in \mathcal{O}(\Theta^0, \Gamma^0)$ .

Because, given policy  $\Theta^0$  and model  $j^*$ 's structure  $\zeta^*$ , the recursive representation  $\Gamma^0$  solves (SMERLM), we immediately obtain the following lemma characterizing  $\mathcal{O}(\Theta^0, \Gamma^0)$ .

**Lemma A.1. (Necessary and sufficient conditions for observational equivalence)** A structure  $\zeta$  belongs to  $\mathcal{O}(\Theta^0, \Gamma^0)$  if and only if every line  $\zeta_l$  for  $l = \{1, 2, \dots, k - p\}$  satisfies the following linear system, given  $\Gamma^0$ :

$$\begin{bmatrix} (P^0)^2 & P^0 & I_k & 0_{k.s} \\ (Q^0 N^0 + P^0 Q^0)' & Q^0 & 0_{k.k} & I_s \end{bmatrix}_{(k+s).(3k+s)} (\zeta_l)' = 0_{(k+s).1} \quad (\text{Null OE})$$

A number of comments are in order. First, notice that a researcher solving for  $\Gamma^0$  needs to solve a *non-linear* system (SMERLM), given a structure and policy  $\{\zeta^*, \Theta^0\}$  and  $N^0$ . However, consider the reverse mapping instead. Given values for the reduced form  $\Gamma^0$ , (SMERLM) implies a

<sup>4</sup>Note that this notion of observationally equivalence is stronger than the usual one because two models cannot be told apart even if the impulse response matrix  $Q_0$  was known and the structural shocks could be recovered.

linear system of equations that any structure  $\xi$  must satisfy in order to match  $\Gamma^0$  under policy  $\Theta^0$ . This linear mapping then characterizes a set of observationally equivalent models, as stated in the lemma above. Formally, a structure belongs to  $\mathcal{O}(\Theta^0, \Gamma^0)$  if and only if each of its lines is in the null space of the  $(k+s) \cdot (3k+s)$  matrix above and thus solve (Null OE). Second, we can readily see that, because  $3k+s > k+s$ , the linear system (Null OE) is in general underdetermined. This implies that  $\mathcal{O}(\Theta^0, \Gamma^0)$  includes multiple structures  $\xi$ . In other words, knowledge of  $\Gamma^0, \Theta^0$  alone are not enough to uniquely identify the model (described by  $\xi^*$ ) which generated such  $\Gamma^0$  under  $\Theta^0$ .

### A.1.2 Counterfactually equivalent models when policy changes from $\Theta^0$ to $\Theta^1$

Consider an alternative policy  $\Theta^1$ . We again solve model  $j^*$  and obtain the reduced form  $\Gamma^1$ . We now ask: from those models that were observationally equivalent to  $j^*$  under policy  $\Theta^0$ , which models would also generate the reduced form  $\Gamma^1$  if the policy were to change from  $\Theta^0$  to  $\Theta^1$ ? Or in other words, which models are observationally equivalent to  $j^*$  both under policies  $\Theta^0$  and  $\Theta^1$ ? As the following definition states, we will call such models *counterfactually equivalent*.

**Definition A.4. (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  $\xi$  in  $\mathcal{O}(\Theta^0, \Gamma^0)$  that generate reduced form  $\Gamma^1$  under alternative policy  $\Theta^1$ . Formally,  $\mathcal{C}(\Theta^0, \Gamma^0, \Theta^1, \Gamma^1) = \mathcal{O}(\Theta^0, \Gamma^0) \cap \mathcal{O}(\Theta^1, \Gamma^1)$ . Model  $j$  and  $j^*$  are counterfactually equivalent if and only if model  $j$  can be parameterized so that  $\xi^j \in \mathcal{C}(\Theta^0, \Gamma^0, \Theta^1, \Gamma^1)$ .

Given the above definition and Lemma A.1, we obtain the following lemma characterizing the counterfactually equivalent set  $\mathcal{C}(\Theta^0, \Gamma^0, \Theta^1, \Gamma^1)$ .<sup>5</sup>

**Lemma A.2. (Necessary and sufficient conditions for counterfactual equivalence)** A structure  $\xi$  belongs to  $\mathcal{C}(\Theta^0, \Gamma^0, \Theta^1, \Gamma^1)$  if and only if every line  $\xi_l$  for  $l = \{1, 2, \dots, k-p\}$  satisfies both (Null OE) and the following linear system, given  $\Gamma^0$  and  $\Gamma^1$ :

$$\begin{bmatrix} (P^1)^2' - (P^0)^2' & P^1' - P^0' \\ (Q^1 N^0 + P^1 Q^1)' - (Q^0 N^0 + P^0 Q^0)' & Q^1' - Q^0' \end{bmatrix}_{(k+s), 2k} \begin{bmatrix} F_l' \\ G_l' \end{bmatrix}_{2k, 1} = 0_{(k+s), 1}. \quad (\text{Null CE})$$

We have seen before that  $\mathcal{O}(\Theta^0, \Gamma^0)$  is not a singleton because (Null OE) was underdetermined. Is it the case now that (Null CE) is determined and, therefore,  $\mathcal{C}(\Theta^0, \Gamma^0, \Theta^1, \Gamma^1)$  is a singleton? Propositions A.1 and A.2 show that the answer is again no, and further characterize this counterfactually equivalent set.

**Proposition A.1. (Sufficient condition for counterfactual equivalence)** If  $(\xi_l)' = \sum_{n=1}^{k-p} c_n^l (\xi_n^*)'$  for all  $l = \{1, 2, \dots, k-p\}$  and some constants  $c_n^l$ , then  $\xi \in \mathcal{C}(\Theta^0, \Gamma^0, \Theta^1, \Gamma^1)$ .

**Proof.** For any line  $n = \{1, 2, \dots, k-p\}$ , we know that  $\begin{bmatrix} F_n^{*'} \\ G_n^{*'} \end{bmatrix}$  from the structure of model  $j^*$  is a solution to (Null CE), by construction. This implies that, given constants  $c_n^l$ , a linear combination

<sup>5</sup>(Null CE) is obtained by subtracting (Null OE) under policy  $\Theta^0$  from the corresponding system under policy  $\Theta^1$ .

$\begin{bmatrix} F_{l.}' \\ G_{l.}' \end{bmatrix} \equiv \sum_{n=1}^{k-p} c_n^l \begin{bmatrix} F_{n.}' \\ G_{n.}' \end{bmatrix}$  is also a solution. Replacing any such solution into (Null OE), we obtain the remaining elements  $\begin{bmatrix} H_l' \\ (LN^0 + M)'_l \end{bmatrix}$  in this proposed structure  $\zeta$ . This shows that, if every line in the structure  $\zeta$  of a model can be written as  $(\zeta_l)' = \sum_{n=1}^{k-p} c_n^l (\zeta_n^*)'$ , then  $\zeta$  solves both (Null OE) and (Null CE) and, therefore, belongs to  $\mathcal{C}(\Theta^0, \Gamma^0, \Theta^1, \Gamma^1)$ . ■

The proposition shows that if a model  $j$  can be parameterized so that the lines in its structure  $\zeta^j$  can be written as a linear combination of the lines in  $\zeta^*$ , then this is a sufficient condition for  $j$  and  $j^*$  to be counterfactually equivalent when the policy changes from  $\Theta^0$  to  $\Theta^1$ . This is because every line in  $\zeta^*$  is a solution to (Null OE) and (Null CE) and thus so are any linear combination of them. Therefore, (Null CE) is underdetermined with rank less than  $2k - (k - p)$  and the counterfactually equivalent set includes more structures than  $\zeta^*$  alone.

Are these structures the only ones in  $\mathcal{C}(\Theta^0, \Gamma^0, \Theta^1, \Gamma^1)$ ? The next proposition shows that this is not the case when the number of endogenous and exogenous state variables ( $rank(P^1 - P^0) + s$ ) is strictly smaller than  $k + p$ , where, as reminder,  $k$  is the total number of endogenous variables (policy or not) and  $p$  is the number of policy variables. Otherwise, the sufficient condition in Proposition A.1 may or may not be necessary as well, depending on the application.

**Proposition A.2. (Sufficient condition is not necessary)** *If  $rank(P^1 - P^0) + s < k + p$ , then  $\mathcal{C}(\Theta^0, \Gamma^0, \Theta^1, \Gamma^1)$  includes structures  $\zeta$  which do not satisfy the sufficient condition in Proposition A.1.*

**Proof.** We begin by defining  $A(\Gamma^0, \Gamma^1)$  in system (Null CE)

$$A(\Gamma^0, \Gamma^1) \equiv \begin{bmatrix} (P^1)^{2'} - (P^0)^{2'} & P^1' - P^0' \\ (Q^1 N^0 + P^1 Q^1)' - (Q^0 N^0 + P^0 Q^0)' & Q^1' - Q^0' \end{bmatrix}_{(k+s), 2k}$$

We have shown above that  $A(\Gamma^0, \Gamma^1)$  must have incomplete rank since the span of the  $k - p$  columns  $\begin{bmatrix} F_{n.}' \\ G_{n.}' \end{bmatrix}$  are solutions to (Null CE). In particular, we know that it has at least  $k - p$  linearly dependent columns and thus  $rank(A(\Gamma^0, \Gamma^1)) \leq \min\{rank(P^1 - P^0) + s, 2k - (k - p)\}$ . Therefore, if the total number of endogenous and exogenous variables  $rank(P^1 - P^0) + s$  is strictly less than  $k + p$ , then  $rank(A(\Gamma^0, \Gamma^1)) < 2k - (k - p)$ . As a result, there must be other solutions to (Null CE) than the ones in the span of the  $k - p$  columns  $\begin{bmatrix} F_{n.}' \\ G_{n.}' \end{bmatrix}$ . Taking any such solution and replacing it into (Null OE), we again obtain the remaining elements  $\begin{bmatrix} H_l' \\ (LN^0 + M)'_l \end{bmatrix}$  in a structure that solves both (Null OE) and (Null CE). This shows that, if  $rank(P^1 - P^0) + s < k + p$ , then  $\mathcal{C}(\Theta^0, \Gamma^0, \Theta^1, \Gamma^1)$  includes more structures beyond the ones whose lines can be written as  $(\zeta_l)' = \sum_{n=1}^{k-p} c_n^l (\zeta_n^*)'$ . ■

Taken together, the propositions imply that there are three “types” of counterfactually equivalent models. The first type are those models  $j$  which can be parameterized so that  $\zeta^j = \zeta^*$ ,

which is a special case of the sufficient condition in Proposition A.1. While their structures are identical, these counterfactually models are nevertheless different to each other in the sense that they have different microfoundations and, therefore, different nonlinear conditions  $0 = \mathbb{E}_t [f_j(x_{t+1}, x_t, x_{t-1}, z_{t+1}, z_t)]$  characterizing their equilibrium. The second type are those models whose structures  $\bar{\zeta}^j$  differ from  $\zeta^*$ , but can nevertheless be parameterized so that the lines in  $\bar{\zeta}^j$  are linear combinations of those in  $\zeta^*$ . The last type are models whose structures do not satisfy this sufficient condition, but can still be parameterized to satisfy the necessary and sufficient conditions in Lemma A.2.

## A.2 Additional Illustrative Examples

This section presents two additional illustrative examples beyond the one in Section 2 of the paper. The first example is on federal transfers policy rules in models of a fiscal and monetary union. The second example is on unemployment benefits policy rules in search-theoretic models of labor market.

### A.2.1 Federal transfer rules in models of a fiscal and monetary union

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.

**Firms and Households.** Final goods producers use labor  $N_{kt}^y$  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

$$\max_{N_{kt}^y, X_{kt}} P_{kt} e^{a_{kt}} (N_{kt}^y)^\alpha (X_{kt})^{1-\alpha} - W_{kt} N_{kt}^y - Q_t X_{kt}$$

where  $a_{kt}$  is a productivity shock and  $\alpha : \alpha < 1$  is the labor share. Unlike the tradable goods prices, final good prices ( $P_{kt}$ ) vary across islands.

Households preferences are given by

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

where  $C_{kt}$  is consumption of the final good,  $N_{kt}$  is labor,  $\delta_{kt}$  is an exogenous processes driving the household's discount rate. Moreover, I follow Schmitt-Grohé and Uribe (2003) and let  $\rho_{kt}$  be the endogenous component of the discount factor that satisfies  $\rho_{kt+1} = \rho_{kt} + \Phi(\cdot)$  for some function  $\Phi(\cdot)$  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  $\Phi(\cdot)$  when assets markets are incomplete (as we assume below).

Households are able to spend their labor income  $W_{kt}N_{kt}$  plus profits accruing from firms  $\Pi_{kt}$  and exogenous endowment of tradable goods  $Q_t e^{\eta_{kt}}$ , financial income  $B_{kt-1}i_{t-1}$  and transfers from the government  $\tau_{kt}$ , where  $B_{kt-1}$  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 ( $B_{kt} - B_{kt-1}$ ). Thus, they face the period-by-period budget constraint

$$P_{kt}C_{kt} + B_{kt} \leq B_{kt-1}(1 + i_{t-1}) + W_{kt}N_{kt} + \Pi_{kt} + \tau_{kt} + Q_t \bar{\eta} e^{\eta_{kt}}$$

**Federal government.** The federal government budget constraint is

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

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  $\tau_{kt}$ . If the islands were fiscally independent such transfers would not be possible.

I assume that the federal government announces a nominal interest rate rule  $i_t = i(\cdot)$  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

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

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

**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,

$$\begin{aligned} a_{kt} &= \rho_a a_{kt-1} + \tilde{\sigma}_a v_t^a + \sigma_a \epsilon_{kt}^a \\ \gamma_{kt} &= \rho_\gamma \gamma_{kt-1} + \tilde{\sigma}_\gamma v_t^\gamma + \sigma_\gamma \epsilon_{kt}^\gamma \\ \eta_{kt} &= \rho_\eta \eta_{kt-1} + \tilde{\sigma}_\eta v_t^\eta + \sigma_\eta \epsilon_{kt}^\eta \end{aligned}$$

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

The "discount rate" process  $\gamma_{kt}$  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  $a_{kt}$  can be interpreted as actual productivity, or a shifter of firm's demand for labor or firm's mark-ups. Finally, "wealth" process  $\eta_{kt}$  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.

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

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

**Aggregation.** The first important assumption for aggregation is that all islands are identical with respect to their underlying production and utility parameters.<sup>6</sup> 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 economy where all aggregate variables are independent of any cross-sectional considerations to a first order approximation.<sup>7</sup> I denote with lowercase letters an island variable's

<sup>6</sup>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.

<sup>7</sup>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

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  $\tilde{n}_t \equiv \sum_k \frac{1}{\bar{K}} \tilde{n}_{kt} = \sum_k \frac{1}{\bar{K}} \log(N_{kt}/\bar{N})$ . 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 such that  $\Phi(\cdot) = \phi n_{kt}$ .

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

**Lemma A.3.** *For given  $\{a_{kt}, \gamma_{kt}, \eta_{kt}\}$ , the behavior of  $\{w_{kt}, n_{kt}, b_{kt}, \tau_{kt}, p_{kt}, c_{kt}, x_{kt}\}$  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

$$\begin{aligned}
\tilde{w}_{kt} - \tilde{p}_{kt} &= \frac{1}{\nu} \tilde{n}_{kt} + \sigma c_{kt} \\
\tilde{w}_{kt} - \tilde{p}_{kt} &= (\alpha - 1)(\tilde{n}_{kt} - \tilde{x}_{kt}) + \tilde{a}_{kt} \\
\tilde{q}_t - \tilde{p}_{kt} &= \alpha(\tilde{n}_{kt} - \tilde{x}_{kt}) + \tilde{a}_{kt} \\
0 &= \mathbb{E}_t(-\tilde{m}u_{kt+1} - \tilde{m}u_{kt+1}) + (\tilde{p}_{kt+1} - \tilde{p}_{kt}) + \phi(\tilde{n}_{kt} - \tilde{n}_t) + \gamma_{kt+1} - \tilde{i}_t \\
\tilde{m}u_{kt} &= -\sigma \tilde{c}_{kt} \\
\tilde{c}_{kt} &= \tilde{w}_{kt} - \tilde{p}_{kt} + \tilde{n}_{kt} \\
\tilde{B}\tilde{b}_{kt} &= \bar{B}(1+r)(\tilde{b}_{kt-1} + \tilde{i}_t) + \tilde{\eta}\eta_{kt} - \tilde{\eta}(\tilde{q}_t + \tilde{x}_{kt}) + \tilde{\tau}\tilde{\tau}_{kt} \\
\sum_k \tilde{x}_{kt} &= \sum_k \tilde{\eta}_{kt} \\
\bar{B}^s \tilde{b}_t^s + \tilde{\tau} \sum_k \tilde{\tau}_{kt} + \bar{G}\tilde{q}_t &= \bar{B}^s(1+r)(\tilde{b}_{t-1}^s + \tilde{i}_t) \\
\tilde{\tau}_{kt} &= \vartheta_w \tilde{w}_{kt} + \vartheta_n \tilde{n}_{kt} + \vartheta_b \tilde{b}_{kt-1} \\
\tilde{i}_{t+1} &= \phi_p \mathbb{E}_t[\tilde{p}_{t+1} - \tilde{p}_t]
\end{aligned}$$

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

$$\begin{aligned}
0 &= \mathbb{E}_t(-\tilde{m}u_{t+1} - \tilde{m}u_t) + (1 - \phi_p)(\tilde{p}_{t+1} - \tilde{p}_t) + \tilde{\gamma}_{t+1} \\
0 &= \sigma(\tilde{w}_t - \tilde{p}_t + \tilde{n}_t) + \frac{1}{\nu} \tilde{n}_t - (\tilde{w}_t - \tilde{p}_t) \\
\tilde{w}_t - \tilde{p}_t &= (\alpha - 1)\tilde{n}_t + \tilde{a}_t + (1 - \alpha)\tilde{\eta}_t \\
\tilde{m}u_t &\equiv -\sigma(\tilde{w}_t - \tilde{p}_t + \tilde{n}_t)
\end{aligned}$$

which is equivalent to the system of equations characterizing the log-linearized equilibrium in 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.

a representative agent economy with a production technology that utilizes labor alone with an elasticity of  $\alpha$ , no endogenous discounting and only 2 exogenous processes  $\{\tilde{a}_t + (1 - \alpha)\tilde{\eta}_t, \tilde{\gamma}_t\}$ .

Next, take log-deviations from the aggregate in the original system and replace  $c_{kt}, p_{kt}, mu_{kt}$  for their corresponding expressions. When we set  $\rho_\gamma = \rho_a = \rho_\eta = 0$  and  $\theta_w = \theta_b = 0$ , this results in the system characterizing the equilibrium of  $\{n_{kt}, w_{kt}, b_{kt}, \tau_{kt}\}$  (where we drop the 'k' index for convenience).

$$0 = \mathbb{E}_t(n_{t+1} - n_t) + \left(\alpha + \frac{1}{\sigma}(1 - \alpha)\right) \mathbb{E}_t(w_{t+1} - w_t) + \left(\frac{1}{\sigma} - 1\right)a_t + \frac{\phi}{\sigma}n_t \quad (\text{Euler})$$

$$0 = -\alpha w_t + \left(\frac{(1 + v)}{(1 - \sigma)} - 1\right) n_t - a_t \quad (\text{Labor market})$$

$$0 = -\frac{\bar{B}}{\bar{\tau}}b_t + (1 + r)\frac{\bar{B}}{\bar{\tau}}b_{t-1} + \frac{\bar{\eta}}{\bar{\tau}}(\eta_t - (w_t + n_t)) + \tau_t \quad (\text{Budget Constraint})$$

$$0 = -\tau_t + \theta_n n_t \quad (\text{Policy})$$

$$0 = -a_{t+1} + \epsilon_{t+1}^a; \quad 0 = -\eta_{t+1} + \epsilon_{t+1}^\eta \quad (\text{Shocks})$$

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. ■

Then, to connect to the general case described by (SME), let  $x_t \equiv \left[ n_t \quad w_t \quad b_t \quad \tau_t \right]'$  and  $z_t \equiv \left[ a_t \quad \eta_t \right]'$ . We have that the first line  $\xi_1$  in the structure (corresponding to the Euler equation) is:

$$\xi_1 \equiv \left[ \overbrace{\begin{matrix} F_1 \\ 1 & \alpha + \frac{1}{\sigma}(1 - \alpha) & 0 & 0 \end{matrix}} \quad \overbrace{\begin{matrix} G_1 \\ -1 + \frac{\phi}{\sigma} & -\left(\alpha + \frac{1}{\sigma}(1 - \alpha)\right) & 0 & 0 \end{matrix}} \quad \overbrace{\begin{matrix} H_1 \\ 0_{1,4} \end{matrix}} \quad \overbrace{\begin{matrix} L_1, N + M_1 \\ \frac{1}{\sigma} - 1 & 0 \end{matrix}} \right] \quad (\text{Euler})$$

Next, consider a different small open economy model with a "discounted" Euler equation, which we call model '1'. Gabaix (2020) presents a "discounted Euler equation" that arises when households have behavioral biases (e.g., inattention to macroeconomic variables or cognitive discounting). McKay, Nakamura, and Steinsson (2017) offers an alternative micro-foundation in a model with idiosyncratic income risk and borrowing constraints. In both cases, given a parameter  $\delta$  governing discounting, the Euler equation takes the form

$$0 = \delta \mathbb{E}_t(\tilde{c}_{kt+1}) - \tilde{c}_{kt} + \frac{1}{\sigma} \mathbb{E}_t(\tilde{p}_{kt+1} - \tilde{p}_{kt}) - \frac{1}{\sigma} \tilde{i}_t.$$

When writing in terms of deviations from the aggregate union equilibrium and replacing  $c_{kt}, p_{kt}$  (and dropping the 'k' subscript for convenience), we obtain

$$0 = \delta \mathbb{E}_t(n_{t+1}) - n_t + \left(\alpha \delta + \frac{1}{\sigma}(1 - \alpha)\right) \mathbb{E}_t(w_{t+1}) - \left(\alpha + \frac{1}{\sigma}(1 - \alpha)\right) w_t + \left(\frac{1}{\sigma} - 1\right)a_t. \quad (\text{Euler 1})$$

The first line in the structure  $\zeta^1$  becomes

$$\zeta_1^1 \equiv \left[ \delta \quad \alpha\delta + \frac{1}{\sigma}(1-\alpha) \quad 0 \quad 0 \quad -1 \quad -\left(\alpha + \frac{1}{\sigma}(1-\alpha)\right) \quad 0 \quad 0 \quad 0_{1,4} \quad \frac{1}{\sigma} - 1 \quad 0 \right]. \quad (\text{Euler 1})$$

Alternatively, consider a model with a quadratic portfolio adjustment cost, as in Schmitt-Grohé and Uribe (2003) — which we call model ‘2’. Given a parameter  $\psi$  governing the cost, the Euler equation becomes

$$0 = \mathbb{E}_t(\tilde{c}_{kt+1}) - \tilde{c}_{kt} + \frac{1}{\sigma}\mathbb{E}_t(\tilde{p}_{kt+1} - \tilde{p}_{kt}) - \frac{1}{\sigma}\tilde{i}_t + \frac{\psi\bar{B}}{\sigma}\tilde{b}_{kt}.$$

And, therefore, we obtain,

$$0 = \mathbb{E}_t(n_{t+1}) - n_t + \left(\alpha + \frac{1}{\sigma}(1-\alpha)\right)\mathbb{E}_t(w_{t+1} - w_t) + \left(\frac{1}{\sigma} - 1\right)a_t + \frac{\psi\bar{B}}{\sigma}b_t. \quad (\text{Euler 2})$$

The first line in the structure  $\zeta^2$  becomes

$$\zeta_1^2 \equiv \left[ 1 \quad \alpha + \frac{1}{\sigma}(1-\alpha) \quad 0 \quad 0 \quad -1 \quad -\left(\alpha + \frac{1}{\sigma}(1-\alpha)\right) \quad \frac{\psi\bar{B}}{\sigma} \quad 0 \quad 0_{1,4} \quad \frac{1}{\sigma} - 1 \quad 0 \right]. \quad (\text{Euler 2})$$

### A.2.1.1 Observational Equivalence

Imagine that we have parameterized the baseline small open economy model, resulting in  $\zeta^*$  and  $\Gamma^0$  under policy  $\Theta^0$ . We next show that, even when models ‘1’ and ‘2’ do not nest the baseline model, there are parameterization of these models such that both  $\zeta^1, \zeta^2 \in \mathcal{O}(\Gamma^0, \Theta^0)$  for any parameterization of the baseline, and, therefore, all models are observationally equivalent under policy  $\Theta^0$ .<sup>8</sup>

Consider a parameterization of such baseline  $\sigma^*, \alpha^*, \nu^*, \phi^*, \bar{B}, \bar{\tau}, \bar{\eta}$  which gives rise to structure  $\zeta^*$  and generates recursive representation  $\Gamma^0$  under policy  $\theta_n^0$ . Then, by construction, we have that the first line in the structure corresponding to the Euler equation ( $\zeta_1^*$ ) satisfies (Null OE). Since

---

<sup>8</sup>Notice that, even when  $\delta = 1$ , (Euler 1) is different than (Euler) because of the endogenous discount factor (governed by  $\phi$ ). Therefore, model ‘1’ does not nest the baseline. The same holds for model ‘2’ even when  $\psi = 0$ .

the only state variable is  $b_t$ , this is:

$$\begin{bmatrix}
(P^0)^2 & P^0 & \mathbb{I}_k & 0_{k,s} \\
(Q^0 N^0 + P^0 Q^0)' & Q^0 & 0_{s,k} & \mathbb{I}_s
\end{bmatrix}
\begin{bmatrix}
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
p_{13}^0 p_{33}^0 & p_{23}^0 p_{33}^0 & (p_{33}^0)^2 & p_{43}^0 p_{33}^0 & p_{13}^0 & p_{23}^0 & p_{33}^0 & p_{43}^0 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
q_{31}^0 p_{13}^0 & q_{31}^0 p_{23}^0 & q_{31}^0 p_{33}^0 & q_{31}^0 p_{43}^0 & q_{11}^0 & q_{21}^0 & q_{31}^0 & q_{41}^0 & 0 & 0 & 0 & 0 & 1 & 0 \\
q_{32}^0 p_{13}^0 & q_{32}^0 p_{23}^0 & q_{32}^0 p_{33}^0 & q_{32}^0 p_{43}^0 & q_{12}^0 & q_{22}^0 & q_{32}^0 & q_{42}^0 & 0 & 0 & 0 & 0 & 0 & 1
\end{bmatrix}
\begin{bmatrix}
\zeta_1^* \\
1 \\
\alpha^* + \frac{1}{\sigma^*}(1 - \alpha^*) \\
0 \\
0 \\
-1 + \frac{\phi^*}{\sigma^*} \\
-(\alpha^* + \frac{1}{\sigma^*}(1 - \alpha^*)) \\
0 \\
0 \\
0_{4,1} \\
\frac{1}{\sigma^*} - 1 \\
0
\end{bmatrix} =$$

$$\begin{bmatrix}
0 \\
0 \\
p_{13}^0 (p_{33}^0 - 1 + \frac{\phi^*}{\sigma^*}) + p_{23}^0 (p_{33}^0 - 1) (\alpha^* + \frac{1}{\sigma^*}(1 - \alpha^*)) \\
0 \\
q_{31}^0 p_{13}^0 + (-1 + \frac{\phi^*}{\sigma^*}) q_{11}^0 + (q_{31}^0 p_{23}^0 - q_{21}^0) (\alpha^* + \frac{1}{\sigma^*}(1 - \alpha^*)) + \frac{1}{\sigma^*} - 1 \\
q_{32}^0 p_{13}^0 + (-1 + \frac{\phi^*}{\sigma^*}) q_{12}^0 + (q_{32}^0 p_{23}^0 - q_{22}^0) (\alpha^* + \frac{1}{\sigma^*}(1 - \alpha^*))
\end{bmatrix} = 0$$

Analogously, for the second line  $\zeta_2^*$ , we have that (Null OE) is,

$$\begin{bmatrix}
0 \\
0 \\
p_{13}^0 \left( \frac{1+\nu^*}{1-\sigma^*} - 1 \right) - p_{23}^0 \alpha^* \\
0 \\
q_{11}^0 \left( \frac{1+\nu^*}{1-\sigma^*} - 1 \right) - q_{21}^0 \alpha^* - 1 \\
q_{12}^0 \left( \frac{1+\nu^*}{1-\sigma^*} - 1 \right) - q_{22}^0 \alpha^*
\end{bmatrix} = 0$$

**Model '1' is observationally equivalent to the baseline.** Consider a parametrization  $\sigma^1, \alpha^1, \nu^1, \delta^1, \bar{B}, \bar{\tau}, \bar{\eta}$  for model '1' with a "discounted Euler equation." First, notice that budget constraint is identical to the baseline model. This directly implies that (Null OE) is satisfied for the third line in the model's structure  $\zeta_3^1$ . Second, guess that  $\alpha^1 = \alpha^*$  and  $\frac{1+\nu^1}{1-\sigma^1} = \frac{1+\nu^*}{1-\sigma^*}$ . This directly implies that (Null OE) is satisfied for the second line in the model's structure  $\zeta_2^1$  as well. Then, to show that model '1' is observationally equivalent to the baseline, we just need to find

$\sigma^1, \delta^1$  such that  $\zeta_1^1$  satisfies (Null OE). That is,

$$\begin{bmatrix} 0 \\ 0 \\ p_{13}^0 p_{33}^0 \delta^1 - 1 + p_{23}^0 p_{33}^0 (\alpha^1 \delta^1 + \frac{1}{\sigma^1} (1 - \alpha^1)) - p_{23}^0 (\alpha^1 + \frac{1}{\sigma^1} (1 - \alpha^1)) \\ 0 \\ q_{31}^0 p_{13}^0 \delta^1 - q_{11}^0 + q_{31}^0 p_{23}^0 (\alpha^1 \delta^1 + \frac{1}{\sigma^1} (1 - \alpha^1)) - q_{21}^0 (\alpha^1 + \frac{1}{\sigma^1} (1 - \alpha^1)) + \frac{1}{\sigma^1} - 1 \\ q_{32}^0 p_{13}^0 \delta^1 - q_{12}^0 + q_{32}^0 p_{23}^0 (\alpha^1 \delta^1 + \frac{1}{\sigma^1} (1 - \alpha^1)) - q_{22}^0 (\alpha^1 + \frac{1}{\sigma^1} (1 - \alpha^1)) \end{bmatrix} = 0$$

Guessing that  $\delta^1 = \frac{1}{1 + \frac{\sigma^* - 1}{1 + \nu^*} \frac{\phi^*}{\sigma^*}}$  and  $\sigma^1 = \frac{\sigma^*}{\delta^1}$ , and replacing above shows that the system is satisfied. To conclude, we have shown that whenever the lines in  $\zeta^*$  satisfy (Null OE) — which was by construction since  $\Gamma^0$  was generated by the baseline model — then there is a parameterization of model '1' such that the lines in  $\zeta^1$  satisfy it as well.

**Model '2' is observationally equivalent to the baseline.** Consider a parametrization  $\sigma^2, \alpha^2, \nu^2, \psi^2, \bar{B}, \bar{\tau}, \bar{\eta}$  for model '2' with "portfolio adjustment costs." Again, both  $\zeta_3^2$  and  $\zeta_2^2$  satisfy (Null OE) when guessing that  $\alpha^2 = \alpha^*$  and  $\frac{1 + \nu^2}{1 - \sigma^2} = \frac{1 + \nu^*}{1 - \sigma^*}$ . Thus, we just need to find  $\sigma^2, \psi^2$  such that  $\zeta_1^2$  satisfies (Null OE) as well. That is,

$$\begin{bmatrix} 0 \\ 0 \\ p_{13}^0 (p_{33}^0 - 1) + p_{23}^0 (p_{33}^0 - 1) (\alpha^2 + \frac{1}{\sigma^2} (1 - \alpha^*)) + \frac{\psi^2 \bar{B}}{\sigma^2} p_{33}^0 \\ 0 \\ q_{31}^0 p_{13}^0 - q_{11}^0 + (q_{31}^0 p_{23}^0 - q_{21}^0) (\alpha^2 + \frac{1}{\sigma^2} (1 - \alpha^2)) + \frac{1}{\sigma^*} - 1 + \frac{\psi^2 \bar{B}}{\sigma^2} q_{31}^0 \\ q_{32}^0 p_{13}^0 - q_{12}^0 + (q_{32}^0 p_{23}^0 - q_{22}^0) (\alpha^2 + \frac{1}{\sigma^2} (1 - \alpha^2)) + \frac{\psi^2 \bar{B}}{\sigma^2} q_{32}^0 \end{bmatrix}$$

Guessing that  $\sigma^2 = \sigma^* + \phi^* \frac{\sigma^* - 1}{1 + \nu^*}$  and  $\psi^2 \bar{B} = \phi^* p_{13}^0$ , and replacing above shows that the system is satisfied. To conclude, we have shown that whenever the lines in  $\zeta^*$  satisfy (Null OE), then there is a parameterization of model '2' such that the lines in  $\zeta^2$  satisfy it as well.

## A.2.2 Counterfactual Equivalence

We now show that, while both models '1' and '2' are observationally equivalent to the baseline model under policy  $\Theta^0$ , only model '1' is counterfactually equivalent to the baseline model when policy changes from  $\Theta^0$  to  $\Theta^1$ . Formally, for the parameterizations that make  $\zeta^1, \zeta^2$  belong to  $\mathcal{O}(\Theta^0, \Gamma^0)$ , only  $\zeta^1$  is also in  $\mathcal{C}(\Theta^0, \Gamma^0, \Theta^1, \Gamma^1)$  because  $\zeta_1^1$  can be written as  $\delta \zeta_1^* + (1 - \delta) \zeta_2^*$ .

**Model '1' is counterfactually equivalent to the baseline.** For the parameterization above, we have that  $\zeta_2^1 = \zeta_2^*$  and  $\zeta_3^1 = \zeta_3^*$ . We next show that, for such parameterization, there are constants  $c_1^1, c_2^1$  such that  $\zeta_1^1 = c_1^1 \zeta_1^* + c_2^1 \zeta_2^*$  so that  $\zeta_1^1$  is a solution to (Null CE). In particular, guessing that

$c_1^1 = \delta^1$  and  $c_2^1 = 1 - \delta^1$ , we have that

$$\begin{aligned}
\delta^1 \bar{\zeta}_1^* + (1 - \delta^1) \bar{\zeta}_2^* &= \delta^1 \begin{bmatrix} 1 \\ \alpha^* + \frac{1}{\sigma^*} (1 - \alpha^*) \\ 0 \\ 0 \\ -1 + \frac{\phi^*}{\sigma^*} \\ -\left(\alpha^* + \frac{1}{\sigma^*} (1 - \alpha^*)\right) \\ 0 \\ 0 \\ 0_{4.1} \\ \frac{1}{\sigma^*} - 1 \\ 0 \end{bmatrix} + (1 - \delta^1) \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ \frac{1+\nu^*}{1-\sigma^*} - 1 \\ -\alpha^* \\ 0 \\ 0 \\ 0_{4.1} \\ -1 \\ 0 \end{bmatrix} \\
&= \begin{bmatrix} \delta^1 \\ \delta^1 \alpha^* + \frac{\delta^1}{\sigma^*} (1 - \alpha^*) \\ 0 \\ 0 \\ -1 + \delta^1 \frac{\phi^*}{\sigma^*} + (1 - \delta^1) \frac{1+\nu^*}{1-\sigma^*} \\ -\left(\alpha^* + \frac{\delta^1}{\sigma^*} (1 - \alpha^*)\right) \\ 0 \\ 0 \\ 0_{4.1} \\ \frac{\delta^1}{\sigma^*} - 1 \\ 0 \end{bmatrix} = \begin{bmatrix} \delta^1 \\ \delta^1 \alpha^1 + \frac{1}{\sigma^1} (1 - \alpha^1) \\ 0 \\ 0 \\ -1 \\ -\left(\alpha^1 + \frac{1}{\sigma^1} (1 - \alpha^1)\right) \\ 0 \\ 0 \\ 0_{4.1} \\ \frac{1}{\sigma^1} - 1 \\ 0 \end{bmatrix} = \bar{\zeta}_1^1
\end{aligned}$$

where the last line follows from the fact that, for model '1' to be observationally equivalent to the baseline, we have required  $\delta^1 = \frac{1}{1 + \frac{\sigma^* - 1}{1 + \nu^*} \frac{\phi^*}{\sigma^*}}$  and  $\sigma^1 = \frac{\sigma^*}{\delta^1}$  and  $\alpha^1 = \alpha^*$ . Therefore, the above shows that for, this same parameterization, model '1' is not only observationally but also counterfactually equivalent to the baseline.

**Model '2' is not counterfactually equivalent to the baseline.** The coefficient  $\frac{\psi^2}{\sigma^2}$  in (Euler 2) associated with  $b_t$  makes it such that, in order to write  $\bar{\zeta}_1^2$  as a linear combination of the lines in  $\bar{\zeta}^*$ , we would have to combine both the (Euler) and (Budget Constraint) equations ( $\bar{\zeta}_1^*$  and  $\bar{\zeta}_3^*$ ) because  $b_t$  does not show up in the (Labor Market) equation ( $\bar{\zeta}_2^*$ ). However, when doing so, such linear combination would also have non-zero coefficients associated with  $b_{t-1}$ ,  $\tau_t$  and  $\eta_t$ , whereas (Euler 2) restricts those coefficients to be zero. This implies that it is not possible to write  $\bar{\zeta}_1^2$  as a linear combination of the lines in  $\bar{\zeta}^*$ .

Is it still the case there is some parameterization of model '2' that makes it counterfactually equivalent to the baseline, even when it does not satisfy the above sufficient condition? The

answer is no. To see this, note that the parameterization necessary so that model '2' is observationally equivalent to the baseline has  $\psi^2 \bar{B} = \phi^* p_{13}^0$ . When policy changes from  $\Theta^0$  to  $\Theta^1$ , so does  $p_{13}^0$  change to  $p_{13}^1$ . Thus, model '2' cannot be parameterized to be observationally equivalent under both  $\Theta^0$  and  $\Theta^1$ , and, by definition, is not counterfactually equivalent to the baseline.

### A.2.3 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 ( $\vartheta_t$ ) 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)).

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

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

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  $\vartheta_t$ , the bargaining power of the worker  $\phi$ , 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  $\beta$  and the vacancy-filling probability  $q(\vartheta_t)$ .

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.<sup>9</sup> 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  $\Theta$  have negligible effects on the equilibrium behavior of these variables. How should we proceed if either (i) we wished to know 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 a structure with the following restrictions,<sup>10</sup>

$$F = \begin{bmatrix} 0 & 0 & 0 \\ f_{21} & f_{22} & 0 \end{bmatrix}; G = \begin{bmatrix} g_{11} & g_{12} & g_{13} \\ 0 & g_{22} & 0 \end{bmatrix}; H = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}; L = \begin{bmatrix} 0 & 0 \\ 0 & l_{22} \end{bmatrix}; M = \begin{bmatrix} m_{12} & 0 \\ 0 & 0 \end{bmatrix}$$

<sup>9</sup>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.

<sup>10</sup>The wage, vacancy-unemployment ratio, and benefits in log-deviations from the steady-state correspond to the first, second, and third columns of matrices  $F, G, H$ . Analogously, the exogenous policy and productivity shocks correspond to the first and second columns of matrices  $L, M$ . The first line corresponds to (Wage Setting (1)) and the second line to (Job Creation (1))



By noting the exclusion restrictions that this structure satisfies, we can readily observe that the (Wage Setting (1)) equation has a rather restricted structure compared to the (Job Creation (1)) equation. This observation, when combined with the results from Section 2 in the paper, 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 likely be counterfactually equivalent to the canonical model, thus implying that the negligible effect of changes in the unemployment benefits policy rule would be a robust 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,

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

where  $\lambda$  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,

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

where  $\gamma$  is the cost of delay in bargaining and the  $\omega_i$ 's are combinations of structural parameters.

Both structures generated by these models satisfy different exclusion restrictions than the canonical model because they include backward- and forward- looking terms. As a result, it is easy to show that they will 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 structure to the canonical model. They assume that 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,<sup>11</sup>

---

<sup>11</sup>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.

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

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

Note that, compared to the canonical model, the effective cost to posting a vacancy is augmented by the additional term  $(1-\beta(1-q(\vartheta_{t+1})))K$  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,

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

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

Again, while this model behaves as if it had a matching and vacancy posting cost with extra curvature, it has an identical 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.

### A.3 Estimating the reduced form $\Gamma$ from a SVAR

A necessary input in the construction of the semi-structural counterfactuals is the reduced form  $\Gamma^0 = \{P^0, Q^0, N^0\}$  under policy  $\Theta^0$ . I next show how to recover this reduced-form when the equilibrium has a SVAR representation.

Following Ravenna (2007), if Assumptions A.1 and A.2 hold and  $Q$  is a non-singular square matrix, then there is a SVAR representation of the recursive law of motion (RLM) of the form:

$$x_t = \rho_1 x_{t-1} + \rho_2 x_{t-2} + Q\epsilon_t \quad (\text{SVAR})$$

where  $\rho_1 \equiv P + QNQ^{-1}$ ;  $\rho_2 \equiv (P - \rho_1)P$  and  $V \equiv \text{Var}(Q\epsilon_t) = Q\Sigma\Sigma'Q'$ . To see this, note that we can write  $z_{t-1} = Q^{-1}(x_{t-1} - Px_{t-2})$  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.

Next, suppose that we have estimated the reduced-form VAR — e.g., via OLS equation by equation — and obtained  $\{\rho_1^0, \rho_2^0, V^0\}$  in an economy under policy  $\Theta^0$ . In addition, assume that we have imposed restrictions on the structural model so that we have identified the impulse response matrix  $Q^0$ .<sup>12</sup> These restrictions are generally in addition to the  $2k$  linear restrictions in Theorem 1 of the paper which are needed to identify a counterfactual.

We then find solutions  $X$  with all eigenvalues inside the unit circle to the quadratic equation  $\rho_2^0 = (X - \rho_1^0)X$ . Under Assumptions A.1 and A.2, there are only two such solutions. The first corresponds to  $P^0$  in the unique stable recursive law of motion. The second corresponds to  $Q^0N^0(Q^0)^{-1}$ . Then,  $P^0$  is identified as the solution that results in an implied  $N^0 = (Q^0)^{-1}(P^0 - \rho_1^0)Q^0$  that satisfies the restrictions on the structure  $\xi^0$  in Theorem 1.

## References

- Christiano, Lawrence J, Martin S Eichenbaum, and Mathias Trabandt. 2016. “Unemployment and business cycles.” *Econometrica* 84 (4):1523–1569.
- Gabaix, Xavier. 2020. “A behavioral New Keynesian model.” *American Economic Review* 110 (8):2271–2327.
- Gavazza, Alessandro, Simon Mongey, and Giovanni L Violante. 2016. “Aggregate Recruiting Intensity.” Tech. rep., National Bureau of Economic Research.
- Hall, Robert E. 2005. “Employment fluctuations with equilibrium wage stickiness.” *The American Economic Review* 95 (1):50–65.
- McKay, Alisdair, Emi Nakamura, and Jón Steinsson. 2017. “The discounted euler equation: A note.” *Economica* 84 (336):820–831.
- Mortensen, Dale T and Christopher A Pissarides. 1994. “Job creation and job destruction in the theory of unemployment.” *The review of economic studies* 61 (3):397–415.
- Petrosky-Nadeau, Nicolas and Etienne Wasmer. 2013. “The cyclical volatility of labor markets under frictional financial markets.” *American Economic Journal: Macroeconomics* 5 (1):193–221.
- Pissarides, Christopher A. 2000. *Equilibrium unemployment theory*. MIT press.
- Ravenna, Federico. 2007. “Vector autoregressions and reduced form representations of DSGE models.” *Journal of monetary economics* 54 (7):2048–2064.
- Schmitt-Grohé, Stephanie and Martin Uribe. 2003. “Closing small open economy models.” *Journal of international Economics* 61 (1):163–185.

<sup>12</sup>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  $\xi$  are consistent with the linear restrictions  $\{R_l, r_l\}$  used for constructing the semi-structural counterfactual. In particular, one can use the equilibrium equations of structural models to derive linear restrictions on the structure  $\xi$  that are sufficient to identify  $Q$ .

Uhlig, Harald. 1995. "A toolkit for analyzing nonlinear dynamic stochastic models easily." .

Wasmer, Etienne and Philippe Weil. 2004. "The macroeconomics of labor and credit market imperfections." *The American Economic Review* 94 (4):944–963.