Recovering Macro Elasticities from Regional Data*

Andrés Pablo Sarto

MIT

December 2018

Job Market Paper

Click here: http://economics.mit.edu/grad/apsarto for the most recent version.

Abstract

I propose a new methodology to estimate macro elasticities in linear economies by exploiting regional data. The key identification assumption is that regions are heterogeneous in their sensitivities to aggregate macro shocks and policies. This assumption is satisfied if regions differ in their fundamentals, such as their technology or their intertemporal elasticity of substitution. First, I show that regardless of the heterogeneity assumption, the macro elasticity is a function of the micro-global elasticities, which measure how regions react to aggregate policies or shocks. Then, I combine typical structural VAR approaches with various panel data methods, such as asymptotic principal components, to show that heterogeneity makes it possible to recover the micro-global elasticities. These are then used to construct an estimate of the macro elasticity. Moreover, I show that the estimates are robust: they are consistent for a wide variety of data-generating processes, including models with incomplete and complete markets, sticky and flexible prices, and different market structures. Compared to existing approaches, I show that the methodology allows for weaker identification assumptions and greater robustness. Finally, I present an empirical application to fiscal multipliers in the U.S. Using state-level data for the period 1971 to 2008, I find a fiscal multiplier of total spending (federal, state and local) that falls in the range 0.7-1.2.

*I am deeply indebted to Iván Werning, Daron Acemoglu, Isaiah Andrews and Anna Mikusheva for their continuous guidance and support throughout this project. I also thank George-Marios Angeletos, Martin Beraja, Ricardo Caballero, Nicolas Caramp, Gabriel Chodorow-Reich, Arnaud Costinot, Leonardo Elias, Joel Flynn, Jerry Hausman, Jonathon Hazell, Juan Passadore, Mikkel Plagborg-Møller, Alp Simsek, Mariano Spector, Ludwig Straub, Marco Tabellini, Robert Townsend, Emil Verner, Olivier Wang, Nathan Zorzi, and the seminar participants at the MIT Macro Seminar, MIT Macro Lunch and MIT Econometrics Lunch for their helpful comments and suggestions. All errors are my own. *Email:* apsarto@mit.edu.

1 Introduction

Macro elasticities measure how an aggregate outcome, for example gross domestic product (GDP), changes when there is a shift in an aggregate policy or a macro shock, such as aggregate government spending. They are crucial to macroeconomics because, among other uses, they can help forecast the depth of a crisis when a negative shock hits the economy, or they can help inform policy makers about the usefulness of a given policy to offset such a shock. Currently, there are three approaches to estimating macro elasticities. The first is a pure aggregate time series method that is typically implemented either as a structural VAR or as narrative techniques that use historical documents to identify a particular change in the variables of interest.[1] The second is a panel data approach that estimates local elasticities and uses them as restrictions in a macroeconomic model to infer the macro elasticity.[2] The third one relies exclusively on a model to obtain the macro elasticity through simulated method of moments, calibration or indirect inference.[3] Each of these has its own problems, which are well known: the first, when it takes the form of a structural VAR, typically relies on assumptions about the ordering of different shocks. When the narrative approach is taken instead, it is difficult to isolate the change in the variables of interest from other confounding influences that might be happening simultaneously. The second and the third approaches work only as long as we impose the correct model on the data.

This paper offers a new alternative for estimating macro elasticities. Although it draws on ideas present in both the aggregate time series and the fully structural approaches, it is closest in spirit to the panel data approach because of its reliance on regional variation.

The main reason to resort to a macroeconomic model in the panel data approach is that the elasticities recovered in the regressions are local, in the sense that they do not account for the interactions among regions. For example, in a monetary union, local demand in region n might depend on aggregate demand, and thus policies that affect income or employment in other regions will have spillover effects in region n through this channel. These spillovers across regions make it difficult to think about how the empirical results obtained in partial equilibrium frameworks, or in small sub-populations, would change if the policy were applied at the union level. The effects could disappear or be dramatically attenuated. Consequently, it is very hard to estimate the macro elasticity using only these local elasticities. Macroeconomic models address this problem because they incorporate all the ingredients necessary to think about these interactions, and thus offer a guide to track these general equilibrium effects.[4] However, resorting to a macroeconomic model comes at a cost: different models have different channels built in for the spillovers, and thus deliver different macro elasticities. Hence, by choosing one model (or a small set), we are left with model-specific estimates that are valid only if the correct model was chosen.

[1] See Ramey (2016) for a review of many examples of these methods. For fiscal multipliers in particular, Blanchard and Perotti (2002) is an example of the structural VAR approach, and Ramey and Shapiro (1998) is an example of the narrative approach.

[2] Sometimes the interest lies exclusively in the local elasticities, so there is no second step. Recent papers that deal with regional interactions and the macroeconomy include, among many others: Hagedorn, Manovskii and Mitman (2016), Serrato and Wingender (2016), Nakamura and Steinsson (2014), Beraja, Hurst and Ospina (2016), Mian and Sufi (2014), Chodorow-Reich (2017), Wilson (2012), Acemoglu and Restrepo (2017), Beraja, Fuster, Hurst and Vavra (2018).

[3] See, for example, Smets and Wouters (2007) and Uhlig (2010).

[4] Sometimes the macro elasticity is derived by assuming no spillovers.

This paper proposes a new approach to estimating the macro elasticity that avoids the problem of model-specific estimates. The idea is to bypass the problem by estimating a different type of micro elasticity than those just described. Instead of focusing on how a region reacts to a policy that is applied regionally, this different type of elasticity measures how a region reacts to a policy that is applied nationally. It is still a micro elasticity, but because it measures the reaction to a change that happens "globally" in the system, I will refer to it, for convenience, as a "micro-global" elasticity. In a similar fashion, I will refer to the local elasticities described previously as "micro-local" elasticities, because they measure a change against a local policy. Since the micro-global elasticities measure a change against a national policy, they incorporate all of the general equilibrium effects caused by the spillovers across regions. These effects are absent from the micro-local elasticities, and this is why, in general, the two can differ substantially. Based on this property, one can intuitively guess that the macro elasticity should be tightly connected to the micro-global elasticities, since they effectively capture the general equilibrium effects from the spillovers. I show this is correct: the macro elasticity is a function of the micro-global elasticities. Hence, estimation of the micro-global elasticities gives the building blocks of the macro elasticity.

To completely bypass the problem of model-specific estimates, the estimation of the micro-global elasticities must not depend on a single model or small set of models. To achieve this, the method imposes no restrictions on the micro-global elasticities. Instead, it uses the key structural equations in which the micro-global elasticities appear and imposes restrictions on the way the variables can enter the equations. For example, one of the key requirements is that the structural equations should display sufficient heterogeneity in their coefficients. To illustrate, this implies that if there are K macro shocks, their heterogeneous impact on different regions cannot display a linear relationship. The restrictions are weak enough that there is a large class of macro models that satisfy them. As a consequence, the estimates are shown to be robust: they are consistent for a wide variety of data-generating processes, including models with incomplete and complete markets, sticky and flexible prices, and different market structures. In particular, they are robust to the channel of the spillovers. To continue with the example, it does not matter whether the spillovers arise because the local demand of region n depends on aggregate demand, on an aggregate price index, or on a combination of the two. The estimates obtained encompass all of them; in fact, it is not possible, at least without further assumptions, to disentangle the effects coming from the different channels.

Let me illustrate these ideas by focusing on the fiscal multiplier example, although what I say here holds for other policies or macro shocks in general. In this case, the macro elasticity captures how aggregate government spending affects aggregate GDP.[5] Let $\widetilde{Y}_{nt}$ denote the growth rate of output in region n in period t, $\widetilde{G}_t$ the growth rate of aggregate government spending, $\varepsilon_{\widetilde{G}_{nt}}$ the growth specific to region n's government spending, and $\widetilde{a}_t$ the growth rate of an unobserved aggregate TFP shock. Then, a wide array of regional macro models display the following type of equilibrium equation for regional output:[6]

$$\widetilde{Y}_{nt} = \beta_{1n}\widetilde{G}_t + \beta_{2n}\varepsilon_{\widetilde{G}_{nt}} + \lambda_n\widetilde{a}_t + \epsilon_{nt}. \tag{1}$$

[5] Usually we measure this impact in dollar units instead, but this is inessential for the argument.

[6] The same ideas apply if we interpret $\widetilde{Y}_{nt}$ as the deviation of output in region n in period t from its non-stochastic steady state, $\widetilde{G}_t$ the equivalent of aggregate government spending, etc.

In equation (1), β_{1n} is the micro-global elasticity, since it measures how $\widetilde{G}_t$ affects $\widetilde{Y}_{nt}$. In contrast, β_{2n} is the micro-local elasticity, because it measures how regional output reacts to regional spending. In these regional macroeconomic models, and in most macroeconomic models commonly used, the macro-elasticity for aggregate spending, which is given by $\eta_{macro} = \frac{\partial \widetilde{Y}_t}{\partial \widetilde{G}_t}$, is related to the β_{1n}'s by means of an aggregator, which is a rule that relates the regional variables to their aggregate counterparts. In the particular case of fiscal multipliers, the aggregator comes from the GDP accounting rules, which translates into a weighted average of the regions' β_{1n}: $\eta_{macro} = \frac{1}{N}\sum_{n=1}^{N}(Y_n/Y)^{*}\beta_{1n}$, where $\frac{1}{N}\sum_{n=1}^{N}(Y_n/Y)^{*} = 1$, and $(Y_n/Y)^{*}$ represents the long-run average of the relative size of region n's real per capita GDP to the national real per capita GDP.

Hence, even if we could consistently estimate the micro-local elasticities, the β_{2n}'s, they would not answer the relevant macroeconomic questions we are asking, those for which the answer lies in the micro-global elasticities, the β_{1n}'s. As I show in a series of simple examples, the micro-local elasticities can differ substantially from the micro-global ones.[7] The difference is due to the spillovers across regions, which generate general equilibrium effects that only the micro-global elasticities capture. Moreover, I show that equations like (1) arise for models in which markets are complete or incomplete, with flexible or sticky prices, etc. The difference in all those cases only appears in the specific form of $\beta_{1n}, \beta_{2n}, \lambda_n$. These elasticities are functions of the underlying parameters of the fundamentals of each region (technology, preferences, etc.), and changing those fundamentals changes the expression for $\beta_{1n}, \beta_{2n}, \lambda_n$, but not equation (1). Hence, recovering the β_{1n}'s from (1) without using any information on how they relate to the fundamentals gives an estimate of the macro elasticity that is valid under any model with (1) as the equilibrium equation for regional output.

One of the main concerns with the estimation of (1) is that the aggregate TFP shock, $\widetilde{a}_t$, is unobservable, and both $\widetilde{G}_t$ and $\varepsilon_{\widetilde{G}_{nt}}$ might be correlated with it. To fix ideas, suppose first that ϵ_{nt} is independent of the right-hand side variables and, hence, $\widetilde{a}_t$ is the only concern. The approach in this paper treats $\widetilde{\lambda}_n$ and $\widetilde{a}_t$ as fixed-effects parameters to be estimated, imposing no restrictions on them. As a result, the strategy can be labeled a fixed-effects one, both because it treats the unobserved macro shocks and sensitivities as parameters to be estimated and because it delivers consistent estimates for the β_{1n}'s without imposing any restriction on the distribution of those shocks conditional on the observables of the model. The strategy incorporates more information to equation (1) in the form of structural equations for $\widetilde{G}_t$ and $\varepsilon_{\widetilde{G}_{nt}}$.[8] Since the models that feature (1) as the equilibrium equation for regional output are obtained by log-linearizing the equilibrium equations around a non-stochastic steady state, whatever structural equations $\widetilde{G}_t$ or $\varepsilon_{\widetilde{G}_{nt}}$ might have, these will be linear. This, in turn, means we are working in the universe of simultaneous equations models (SEMs), or in a regional structural VAR (RSVAR), when we consider dynamics along the time dimension.

The procedure uses the whole system of equations (every equation for every variable, period and region) to build estimates of $\widetilde{a}_t$; call them $\hat{\widetilde{a}}_t$. For this, I rely on a combination of time series methods used in the structural VAR literature along with different panel data methods, such as interactive-effects estimators and asymptotic principal components. The estimators of the micro-global and micro-local

[7] Needless to say, this does not mean the micro-local elasticities cannot be interesting per se.

[8] Nevertheless, some of the results rely only on properties of the reduced forms of $\widetilde{G}_t$ and $\varepsilon_{\widetilde{G}_{nt}}$.

elasticities are then obtained by running (1) as a time series regression for every region separately, using $\widetilde{G}_t$, $\varepsilon_{\widetilde{G}_{nt}}$ and $\hat{\widetilde{a}}_t$ as regressors. If the $\hat{\widetilde{a}}_t$ is precise enough, then controlling for it in (1) handles the concern that $\widetilde{G}_t$ or $\varepsilon_{\widetilde{G}_{nt}}$ may be correlated with the aggregate TFP shock. The inclusion of $\hat{\widetilde{a}}_t$ as a control follows the same logic as that of many studies using factor-augmented regressions; see for example Stock and Watson (2002). Since all sensitivities are region-varying, all of the results are derived under $N, T \to \infty$.

Let me now try to put the strategy in perspective by relating it to two well-known approaches. The first is the control function approach: in this paper $\hat{\widetilde{a}}_t$ works as an extra regressor that breaks the correlation between $\widetilde{G}_t$ or $\varepsilon_{\widetilde{G}_{nt}}$ and the unobservable shocks affecting GDP in (1). The difference lies in the way $\hat{\widetilde{a}}_t$ is obtained which, as I mentioned before, relies on the whole system of equations. Nonetheless, using $\hat{\widetilde{a}}_t$ as a control is not the only way to think about the strategy. There is an equivalent way that is related to instrumental variables (IV). We can readily tell that if we project the policies, $\widetilde{G}_t$ and $\varepsilon_{\widetilde{G}_{nt}}$, on $\hat{\widetilde{a}}_t$, the residuals can be used as IVs in (1). And in principle, the estimation can be implemented as the usual IV strategy, albeit with estimated instruments. However, the connection is deeper than that. In some of the cases I will discuss, it is necessary to first estimate these IVs to get at $\hat{\widetilde{a}}_t$. Hence, in those cases the system estimation of $\widetilde{a}_t$ is reversed: first we estimate the IVs and then we use them to estimate $\widetilde{a}_t$. Thus, we see that, in general, the system estimation of $\widetilde{a}_t$ can be equivalently viewed as a system estimation of IVs for the policy variables. And controlling for $\widetilde{a}_t$ or using the IVs are equivalent ways of thinking about the strategy. The main difference from the usual IV strategy is that in this case the IVs are not observed and must be estimated. Indeed, I show examples in which there is no observable variable that could serve as an IV.

The key to the strategy is getting at estimates of $\widetilde{a}_t$. The crucial assumption that allows this is the heterogeneous effects assumption. When regions are heterogeneous and $\widetilde{a}_t$ has a different impact on different regions, we can combine results from the interactive-effects estimators and asymptotic principal components to extract a flexible set of factors that capture the variation in $\widetilde{a}_t$. Moreover, we can do so without imposing any restriction on the unobserved macro shocks. It is useful to compare these results to the case of homogeneity. Suppose in (1) that $\beta_{1n} = \beta_1$, $\beta_{2n} = \beta_2$ and $\lambda_n = \lambda$ for every region, and that the whole system of equations is also homogeneous. In this case, it is impossible to estimate $\widetilde{a}_t$ with the method in this paper, because it was precisely the different impact of the same shock in different regions that allowed its estimation. With an homogeneous impact, there is an infinity of equivalent ways to attribute the changes in the observed variables to $\widetilde{a}_t$, and all of them have different implications for the elasticities. Furthermore, note that if one includes a time fixed effect in (1), the micro-global elasticities are not identified, unless one is able to apply the approach in Hausman and Taylor (1981).[9] In contrast, this is not a problem under heterogeneity: the IV implementation of the approach in this paper can be viewed as a generalized version of Hausman and Taylor (1981) where the instrument is an unobservable variable that can be estimated from within the system of equations. It is similar in that the instrument comes from within the system, but it differs in that it is not observable (although it can be estimated with the observables of the model). Hence, taking the IV view, the crucial difference from the homogeneity case is that heterogeneity allows the estimation of unobservable instruments that can

[9] Notice that the micro-local elasticities are still identified regardless of these issues. But note that if the true dgp features heterogeneity instead, the usual within estimator (least square dummy variable) is inconsistent, as the interactive effects are no longer removed from the equation or effectively controlled for.

serve identify the micro-global elasticities.

Even though the heterogeneous effects assumption is at the core of the strategy, the exclusion restrictions that are imposed in the system of equations are also relevant, because the estimation of the unobserved macro shocks relies on the whole system of equations. As usual, how reasonable these restrictions are should be evaluated on a case by case basis. In this paper, I focus on the fiscal multiplier application and, consequently, argue for restrictions that seem reasonable for it. This gives a clear sense of the approach in action, which can then easily be replicated in relation to other questions. Moreover, I show that the approach in this paper needs fewer exclusion restrictions in comparison to the pure aggregate time series approach described previously, and thus has the potential to be more robust than both the structural VAR approach or the narrative approach. Of course, this greater robustness comes at a cost. The data requirements for the approach in this paper are much higher, since it requires regional variation for the estimation.

The other concern one might have in (1) is the correlation of the regressors with ϵ_{nt}, something I have put on hold thus far. I will argue that the concern for $\widetilde{G}_t$ is not justified in the usual questions addressed by the typical macro models, but it is indeed valid for $\varepsilon_{\widetilde{G}_{nt}}$. To address this concern, I will assume that there is an observable instrument for $\varepsilon_{\widetilde{G}_{nt}}$ that, in spite of being orthogonal to ϵ_{nt}, might be correlated with $\widetilde{a}_t$. The system estimation described previously takes into account this observable instrument and incorporates it into the strategy.

Finally, I offer a detailed application of this methodology to the case of fiscal multipliers in the US. The fiscal multiplier is a key input to many policy decisions, for example, decisions about using fiscal policy to combat a recession. Moreover, this is a natural setting for the methods in this paper because the response to government spending stimulus is likely heterogeneous across states, depending on their industrial composition, geography, demographics, etc. I apply the method to a panel of the 50 US states and Washington DC, from 1971 to 2008, using total spending (federal, plus state, plus local) as the policy variable. My results suggest a very precisely estimated fiscal multiplier of around 1, depending on the specification used. Thus, it is not possible to rule out the possibility that government spending crowds out/in private spending. However, given that the lower estimates of the multiplier fall in the range of $0.7 - 0.9$, the results do suggest that if there is crowding out, it is not severe. Moreover, I also get multipliers in the range of $1.1 - 1.2$, so, similarly, if there is crowding in, it is likely small.

The application also illustrates the importance of looking at the micro-global multipliers instead of the micro-local ones. The micro-global multipliers for the different states all lie in the range $0.15 - 1.92$.[10] In contrast, approximately 80% of states display a micro-local elasticity that falls below the micro-global one, and a high fraction of those is statistically indistinguishable from zero. To illustrate, the state of Nevada has a micro-global multiplier of 1.39 (significant at the 1% level) and a micro-local multiplier of 0.09 (statistically indistinguishable from zero). Even more, if we were to instead use the micro-local multipliers to compute the fiscal multiplier, we would get an estimate of 0.11 (statistically indistinguishable from zero).

Related Literature This paper contributes to and benefits from several strands of the macroeconomics and econometrics literature. First and foremost, it is closely related to the papers addressing local

[10] These numbers change slightly depending on the specification, but the general picture remains.

versus national multipliers. During the past several years, many papers have studied such multipliers in relation to different shocks and policies from both theoretical and empirical perspectives. A non-exhaustive list includes Nakamura and Steinsson (2014), Farhi and Werning (2016), Chodorow-Reich, Feiveson, Liscow and Woolston (2012), Chodorow-Reich (2017), Mian and Sufi (2011), Mian, Rao and Sufi (2013), Mian and Sufi (2014), Wilson (2012), Beraja, Hurst and Ospina (2016), Hagedorn, Manovskii and Mitman (2016), Acemoglu and Restrepo (2017), Beraja, Fuster, Hurst and Vavra (2018) and Serrato and Wingender (2016). On the more theoretical side, Farhi and Werning (2016) provide solutions of fiscal multipliers under a liquidity trap and under fixed exchange regimes (confirming the potential for large multipliers during a liquidity trap) and provide formulas for local multipliers. In the same vein, Beraja, Hurst and Ospina (2016) highlight the difference between local and aggregate elasticities, and Nakamura and Steinsson (2014) distinguish between local government spending multipliers and the usual closed economy aggregate fiscal multiplier. In this paper, I provide a framework that makes a clear distinction between three types of elasticities, which I refer to as micro-local, micro-global and macro. I argue that the macro elasticity we are usually interested in is a function of the micro-global elasticities, something that, as a consequence, make them a crucial object of interest. This framework also helps clarify some of the contributions of this literature, in which some of the distinctions were made in terms of, as I call them here, micro-global elasticities versus the macro elasticity, or micro-local elasticities versus the macro elasticity.

Among the more empirical papers, many estimate what I call micro-local elasticities and then use a theoretical framework to fill in the general equilibrium effects that come from the spillovers in order to arrive at the macro elasticity. Both the empirical strategies used to to recover the micro-local elasticities and the theoretical tools used to estimate the macro elasticity depend on the particular question at hand. Thus there is significant variation across these strategies. For example, Nakamura and Steinsson (2014) use an IV approach related to the difference in the response of regional spending to national military buildups, whereas Serrato and Wingender (2016) exploit population revisions in census years due to accumulated measurement error. Abstracting from these differences, in this paper I show the benefits of departing from the usual homogeneous effects assumptions. I state assumptions under which we can recover the micro-global elasticities that make up the macro elasticity. This allows us to overcome the model-specific estimates problem by providing estimates that are consistent for a large class of models, including many commonly used in the macroeconomics literature. A recent paper dealing with robustness in macroeconomics, although for counterfactuals, is Beraja (2017).

The empirical strategy relies heavily on the literature on interactive fixed effects and dynamic factor models. There is a large literature on these topics, including, among many others, Ahn, Hoon Lee and Schmidt (2001), Ahn, Lee and Schmidt (2013), Bai (2009), Ando and Bai (2015), Stock and Watson (2002), Stock and Watson (2005), Connor and Korajczyk (1986), Holtz-Eakin, Newey and Rosen (1988) and Pesaran (2006). In particular, in this paper the strategy builds on Bai (2009), Ando and Bai (2015) and Pesaran (2006) and on typical structural VAR approaches. These papers, and the strategy in this paper as well, are also closely related to dynamic factor models as in Stock and Watson (2002). The contribution of this paper on this front is twofold. First, by analyzing various regional versions of canonical macroeconomic models, it shows with clear emphasis which dimensions of this literature are more relevant for the empirical macroeconomics questions analyzed here. Second, it shows how different approaches present in them can be combined to obtain an empirical strategy that addresses those dimensions. In term of inference, results for many of these papers are developed in Mikusheva and Anatolyev (2018). Gonçalves and Perron (2014) and Djogbenou, Gonçalves and Perron (2015) provide valid bootstrap methods for many approaches that rely on factor augmented regressions. In this paper, I adapt their wild bootstrap to my setting. I also apply results from Bai and Ng (2002) to determine the number of unobserved macro shocks.

The empirical application presented is closely related to the (macro) fiscal multipliers literature of, among others, Ramey and Shapiro (1998), Ramey (2011), Ramey (2016), Blanchard and Perotti (2002), Barro and Redlick (2011), Hall (2009), Ilzetzki, Mendoza and Végh (2013). See Ramey (2016) for a detailed review of this literature. Most of these papers rely predominantly on (different) pure time series approaches. Blanchard and Perotti (2002) illustrate the structural VAR approach, whereas Ramey and Shapiro (1998) illustrate the narrative approach. In this paper, I show how regional variation can be combined with time series variation to weaken the identification assumptions. In terms of the empirical results obtained, and with respect to the papers that are more directly comparable to this paper, I get higher effects for government spending, with multipliers that are very precisely estimated. Smets and Wouters (2007) and Uhlig (2010), among others, offer a more structural approach, and hence the same ideas related to the robustness of the approach proposed here also apply. Farhi and Werning (2016) also discuss different measures of summary fiscal multipliers and the connection among them, a framework that I adopt to discuss my empirical results in Section 6. Parker (2011)'s discussion of the lack of a good measure of the effects of fiscal policy during a recession also applies to the results in this paper, and a future extension should address these issues.

Layout The remainder of the paper is organized as follows. In Section 2, I provide a minimal example that gives an overview of all of the relevant results in the paper. Section 3 provides various canonical regional models and shows that they give rise to regional equations of the form previously discussed. Building on these insights, in Section 4 I define a large class of models in which estimation will take place and provide a detailed description of all of the assumptions. In Section 5, I detail the empirical strategy to recover the macro elasticities from the system of equations in a way that is valid under all of the models in the class defined. Moreover, I show that the estimators proposed are consistent, and I discuss the main identification assumptions. Section 6 presents an application to the case of fiscal multipliers in the US. Section 7 concludes. All the proofs can be found in Appendix A. Appendix B contains the figures and tables missing from the main text.

2 A Simple Regional Model

The purpose of this section is to survey the main results of the paper. To that end, I use a reduced form model that lets me discuss them in a straightforward and simplified manner. Section 3 shows that the intuitions discussed here carry over to more commonly used macro models.

A Minimal Model

Suppose we have a continuum of regions $i \in [0,1]$ each with a supply and demand of a consumption good, in logs, of the following sort:

$$\begin{aligned}
(Demand) \quad & p_{it}^{d} = -\rho_i^d y_{it} + \alpha_1 Y_t + \alpha_2 \xi_{1t} + u_{it}^{d} \\
(Supply) \quad & p_{it}^{s} = \rho_i^s y_{it} + \beta_1 s_t + \beta_1 \varepsilon_{s_{it}} + \beta_2 \xi_{2t} + u_{it}^{s} \\
(Market\ Clearing) \quad & p_{it}^{d} = p_{it}^{s} = p_{it}
\end{aligned}$$

where x_{it} stands for $\log(X_{it})$, p_{it} is the (log of) price of the consumption good in region i in period t, y_{it} is the quantity of the consumption good, $Y_t := \int_0^1 y_{it} di$, s_t is the policy variable of interest (for example, a subsidy to suppliers of the good) that is common to all regions, $\varepsilon_{s_{it}}$ is the policy variable particular to region i, ξ_{1t} is an aggregate taste shock to demand and ξ_{2t} is the aggregate TFP of suppliers. Neither ξ_{1t} nor ξ_{2t} is observable. For convenience, I will refer to this model as the regional supply and demand model.

The following proposition characterizes the equilibrium:

Proposition 1. *Suppose that in the regional supply and demand model the $u_{it}^{s/d}$ have zero mean, are independent and have finite variances ($\sigma_{ii} \leq M < \infty$) for all i,*[11] *and that $\rho_i^s + \rho_i^d \in \mathbb{R}_{\neq 0}$ $(\forall i)$ and are Riemann integrable. Then in equilibrium the following regional equation holds:*

$$y_{it} = \eta_{MG}^{i} s_t + \eta_{ML}^{i} \varepsilon_{s_{it}} + \delta_1^i \xi_{1t} + \delta_2^i \xi_{2t} + \epsilon_{it}, \tag{2}$$

where the ϵ_{it}'s have zero mean, are independent and have finite variances ($\sigma_{ii} \leq M < \infty$) for all i and:

$$\eta_{macro} := \frac{\partial Y_t}{\partial s_t} = \int_0^1 \eta_{MG}^{i} di = -\frac{\int_0^1 \left(\frac{\beta_1}{\rho_i^s+\rho_i^d}\right) di}{\left[1 - \int_0^1 \left(\frac{\alpha_1}{\rho_i^s+\rho_i^d}\right) di\right]} \tag{3}$$

$$\eta_{MG}^{i} := \frac{\partial y_{it}}{\partial s_t} = -\left\{\left(\frac{\beta_1}{\rho_i^s + \rho_i^d}\right) + \left(\frac{\alpha_1}{\rho_i^s + \rho_i^d}\right)\frac{\int_0^1 \left(\frac{\beta_1}{\rho_i^s+\rho_i^d}\right) di}{\left[1 - \int_0^1 \left(\frac{\alpha_1}{\rho_i^s+\rho_i^d}\right) di\right]}\right\} \tag{4}$$

$$\eta_{ML}^{i} := \frac{\partial y_{it}}{\partial \varepsilon_{s_{it}}} = -\left(\frac{\beta_1}{\rho_i^s + \rho_i^d}\right). \tag{5}$$

Three Elasticities at Play As Proposition 1 illustrates, there are three elasticities at play in the models we use for policy analysis. The first is the "micro-local" elasticity, η_{ML}^{i}. As (5) shows, this measures how the regional dependent variable changes with the policy variable particular to region i, i.e. $\frac{\partial y_{it}}{\partial \varepsilon_{s_{it}}}$. The second is the "micro-global," which measures how the regional variables change with the aggregate policy variable, $\frac{\partial y_{it}}{\partial s_t}$. And the last, the macro elasticity, measures how the aggregate responds to the

[11] The independence requirement in this condition could be weakened, but the main idea would be the same.

change in the aggregate policy variable, $\frac{\partial Y_t}{\partial s_t}$. That is, we have:

$$\frac{\partial y_{it}}{\partial \varepsilon_{s_{it}}} = micro - local\ elasticity$$
$$\frac{\partial y_{it}}{\partial s_t} = micro - global\ elasticity$$
$$\frac{\partial Y_t}{\partial s_t} = macro\ elasticity.$$

There are some important distinctions between these elasticities. First, as (3) makes clear, when our interest lies in macro elasticities, what we want to get at is η^i_{MG}, not η^i_{ML}. The reason is that the "building blocks" of the macro elasticity are the η^i_{MG} of every region, not the η^i_{ML}. The link between η^i_{MG} and η_{macro} of (3) is a consequence of the aggregator in this model, which is given by $Y_t := \int_0^1 y_{it} di$.

Second, even if what we want is η^i_{MG}, it might be the case that $\eta^i_{MG} = \eta^i_{ML}$. So why and when are $\eta^i_{MG} \neq \eta^i_{ML}$? From (4) and (5), we see that when $\alpha_1 = 0$, the two coincide. That is, when there is no interaction across regions, because Y_t drops from the demand equation, the two elasticities coincide. By contrast, when $\alpha_1 \neq 0$, the two elasticities differ. The reason is that the regional demand in this case depends on Y_t, and thus, for a given region i, conditions affecting markets from other regions have spillover effects on its own market through this channel. When other markets expand, aggregate demand increases, which increases demand in region i. As a consequence, we can conceptualize the effect of an increase in s_t on y_{it} as operating through two channels. There is a direct impact of the change in region i, causing an increase in output, because suppliers in that region get a higher subsidy. But there is also an indirect impact coming from the spillover effects through aggregate demand, because, for the same reasons, other regions also see their own markets expand. The micro-global elasticity captures both channels, the micro-local only the direct one.

In more general models, even in the absence of a direct aggregate like the one in this example, the spillovers could operate through a common price, and in those cases we would also have $\eta^i_{MG} \neq \eta^i_{ML}$. In this minimal model, the difference between the two elasticities amounts only to the second term in (4) because there is only one such dependence. But if the regional demand (and/or supply) equations had more dependencies on common prices or aggregates, the number of terms in the difference of η^i_{MG} and η^i_{ML} would increase. For example, adding a dependence on the supply curve of an aggregate price index, like $P_t := \int_0^1 p_{it} di$, would add a third term to (4). In this case η^i_{MG} would differ both because of the spillovers transmitted through aggregate demand, and because of those transmitted through the aggregate price index. These observations imply that the difference between η^i_{MG} and η^i_{ML} could potentially be very large, and there is little guidance on how this difference would relate to η^i_{ML}. This is the reason there is always great caution in extrapolating η_{macro} from η^i_{ML} in studies that estimate η^i_{ML}.

Finally, note that if we recover η^i_{MG}, although this captures the direct and indirect effects coming from the spillovers, it is impossible, without further assumptions, to attribute parts of it to the different channels. They are all grouped together, and we can distinguish neither the number nor the nature of the channels through which the spillovers operate.

Robustness Proposition 1 has important implications for robustness. It says that the specific form of η^i_{MG} depends on all the "structural" parameters of the regional supply and demand model, such as, for example, ρ_i^d. It is also easy to see that other similar models that allow for more general patterns will also display an equation similar to (2). For example, under the following variations, which except for the last are generalizations, (2) would still hold:

1. $Y_t, s_t, \xi_{1t}, \xi_{2t}$ could enter both the supply and demand equations.

2. $\alpha_1, \alpha_2, \beta_1, \beta_2$ need not be region-invariant.

3. The coefficients on s_t and $\varepsilon_{s_{it}}$ need not be the same.

4. Other aggregates like $P_t := \int_0^1 p_{it} di$ could enter both equations.

5. Other prices being, in equilibrium, functions of the aggregates could enter both equations.

6. Y_t could be absent from both equations.

Hence, the idea in this paper is to get an estimate of η_{MG}^i without relying on the structural form of (4). To do this, we will use only the information contained in (2), where we must keep in mind that ξ_{1t} and ξ_{2t} are unobserved. This will allow us to get an estimate of η_{MG}^i that is robust to any model generating the data we observe, conditional on them having an equilibrium equation like (2). For example, in this particular case, the estimate will be valid if the regional supply and demand model is the true data-generating process, but also if the true process is one in which Y_t enters both the supply and demand equations, etc. Finally, with this estimate we can use (3), which holds across all these models, to construct an estimate of η_{macro}. Thus, in this manner, the estimate of η_{macro} is robust to any of the models that have equilibrium equations like (2) that share the same aggregator.

Therefore, to get an idea of the robustness of the procedure, the key is to understand which models have an equilibrium equation like (2) that we can exploit. This is the subject of Section 4, where I build on some of the more widely used macroeconomic models.

Empirical Strategy Once we have understood which models share an equilibrium equation like (2), the second question is how to estimate η_{MG}^i using (2) without a model-specific structure. Clearly, if we ran a time series OLS regression for every region of y_{it} on s_t and $\varepsilon_{s_{it}}$, we would likely get biased estimates of the elasticities, because it is reasonable to think that s_t, and/or $\varepsilon_{s_{it}}$, is correlated with ξ_{1t} and ξ_{2t}. On top of this, we could also have both regressors correlated with ϵ_{it}.

The first step in the empirical strategy is to think about the cause of the endogeneity of s_t and/or $\varepsilon_{s_{it}}$ in (2), in order to be able to address it. Of course, for general setups, the reasons could be varied. Nevertheless, in this paper, I will focus on situations in which the estimating equation of interest, like (2), represents an equilibrium relationship from a macroeconomic model, in which forward looking agents optimize their choices subject to resource constraints. And I will also think of s_t and $\varepsilon_{s_{it}}$ as policy variables set by a policy maker, or decision process more generally (a legislature, for example).[12] The endogeneity causes I will address are those in which the policy maker is reacting either directly to y_{it} (or some function of y_{it}'s) or to the unobserved macro shocks themselves (or to a signal of them, if they are unobservable to the policy maker[13]).

[12] However, I am going to refer to the policy maker from now on for convenience.

[13] I am always assuming they are unobserved to the econometrician.

In the interest of clarity, let me show the simplest example in terms of estimation. And let me assume as well that ϵ_{it} is $i.i.d.$ To motivate the key equation, the cause for the endogeneity in this example is that the policy maker is reacting to a signal of the unobserved macro shocks. Suppose the policy maker setting $\varepsilon_{s_{it}}$ chooses the spending to maximize $\mathbb{E}\left[q\left(\varepsilon_{s_{it}}, \xi_{1t}, \xi_{2t}\right) \middle| \varphi^1_{it}, \varphi^2_{it}\right] - c\left(\varepsilon_{s_{it}}\right)$, where $q(.)$ and $c(.)$ are a benefit and cost function, respectively, and φ^k_{it} is a signal of ξ_{kt}, $k = 1, 2$. This problem captures the idea that the benefit to the policy maker depends on the shocks hitting her region (which hit other regions as well), but she does not observe them directly. Instead she observes a pair of signals. In choosing $\varepsilon_{s_{it}}$ there is also a cost that can represent resources, lobbying, etc. In Section 5, I show that there are assumptions on the signals and the functions such that the optimal choice satisfies:

$$\varepsilon_{s_{it}} = \theta^{\varepsilon}_{1i}\xi_{1t} + \theta^{\varepsilon}_{2i}\xi_{2t} + \epsilon^s_{it}, \tag{6}$$

where ϵ^s_{it} is i.i.d. across i and t and independent of the rest of the variables. Since the key for the estimation is equation (6), reasons other than the one I used to motivate that equation are compatible with the strategy as well.

The strategy can be implemented as a two-step procedure. First, using the method of asymptotic principal components, it estimates ξ_{1t}, ξ_{2t} from (6) as the solution to the following problem:[14]

$$\left(\hat{\Xi}, \left\{\hat{\theta}^{\varepsilon}_i\right\}_{i=1}^N\right) = \underset{\left(\Xi, \left\{\theta^{\varepsilon}_i\right\}_{i=1}^N\right)}{\arg\min} \; \frac{1}{NT}\sum_{i=1}^N \left(\varepsilon_{s_i} - \Xi\theta^{\varepsilon}_i\right)'\left(\varepsilon_{s_i} - \Xi\theta^{\varepsilon}_i\right) \tag{7}$$

$$s.t.: \begin{cases} \frac{\Xi'\Xi}{T} = I_2 \\ \Theta^{\varepsilon\prime}\Theta^{\varepsilon} \; diagonal \end{cases}$$

where $\varepsilon_{s_i} = \left(\varepsilon_{s_{i1}}, ..., \varepsilon_{s_{iT}}\right)'$, $\Theta^{\varepsilon} = \left(\theta^{\varepsilon}_1, ..., \theta^{\varepsilon}_N\right)'$, $\theta^{\varepsilon}_i = \left(\theta^{\varepsilon}_{1i}, \theta^{\varepsilon}_{2i}\right)$, $\Xi = (\xi_1, \xi_2)$, $\xi_k = (\xi_{k1}, ..., \xi_{kT})'$. The second step involves running a time series OLS regression of (2) for every region, using $\hat{\xi}_{1t}$ and $\hat{\xi}_{2t}$ to control for ξ_{1t} and ξ_{2t}. Therefore, if we let $\beta_i = \left(\eta^i_{MG}, \eta^i_{ML}\right)'$, $Y_i = \left(y_{i1}, ..., y_{iT}\right)'$, $X_i = \left(s, \varepsilon_{s_i}\right)$, $s = (s_1, ..., s_T)$ the estimator is given by:

$$\beta_i\left(\hat{\Xi}\right) = \left(X_i' M_{\hat{\Xi}} X_i\right)^{-1} X_i' M_{\hat{\Xi}} Y_i, \tag{8}$$

where $M_{\Xi} = I_T - \Xi\left(\Xi'\Xi\right)^{-1}\Xi'$. In Section 5 I give conditions under which (8) is consistent.

As this simple example illustrates, the key that enables the estimation of η^i_{MG} is the heterogeneity assumption, which allows precise enough estimates of $\hat{\xi}_{1t}$ and $\hat{\xi}_{2t}$. Note that in the regional supply and demand model, the only source of heterogeneity in the regional supply and demand system comes from the own-price elasticity of both curves. The rest of the structural parameters are homogenous across regions. This suffices, though, to have an equilibrium equation like (2), in which all of the parameters are heterogeneous. If more parameters in the regional supply and demand system were heterogeneous, this feature would only be reinforced. And, as I show in Section 5, to generate enough heterogeneity in (6), it suffices that the policy makers from different regions have different precisions in their signals.

[14] The constraints in (7) are estimation restrictions, not assumptions about the underlying true processes for Ξ and Θ^{ε}. See Section 5 for further details.

Thus, this heterogeneity in the structural equations that is key for the empirical strategy can arise with only a "tiny" amount of heterogeneity in the fundamentals of the economy, and this is something that Section 3 will show is true as well for the most widely used macro models.

Finally, note that, in principle, the linearity might seem arbitrary in the regional supply and demand model, and indeed it is to some extent. However, I will apply these ideas in contexts where the equations come from a log-linear approximation of the equilibrium around a non-stochastic steady state. Hence, the linearity in this context is less of a concern.

3 Regional Versions of Canonical Macroeconomic Models

In this section, I present two canonical examples of models widely used in the macroeconomics literature. The first model is a regional New Keynesian model, and the second one is a regional RBC model. Since the application in Section 6 deals with government spending, the policy variable of interest here is government spending. However, I also present results for variations of these and also look at another policy, a subsidy to the production of consumption goods. Thus, the results are seen to be robust to a very wide range of models and policies.

The goal is twofold. First, these examples show that the results of Section 2, which might have seemed particular to a reduced form model, are actually very general. The examples in this section feature much more complex dynamics, with expectations of how the economy will evolve in the future playing a crucial role in determining the equilibrium today, and still behave, from an estimation perspective, in a fashion similar to the regional supply and demand model of Section 2.

Second, these models help us build intuition about how the class of models used for estimation should be defined. In particular, they offer a clear idea of what elements the regional structural system needs to include. This last motive will become much more clear in the next section.

Thus, in a sense, this section works like a bridge between the regional supply and demand model and the general setups considered in Section 4.

3.1 A Regional New Keynesian Model

Households Suppose there is a continuum of consumption goods c_{it} with $i \in [0,1]$ and a representative agent who solves the following problem:

$$\max_{(c_{it})_{i\in[0,1]}, L_t, B_t} \mathbb{E}_0 \left[\sum_{t=0}^{\infty} \beta^t \left(\frac{C_t^{1-\sigma}}{1-\sigma} - \frac{L_t^{1+\varphi}}{1+\varphi} \right) \right]$$

$$s.t.: \begin{cases} \int_0^1 p_{it} c_{it} di + B_t = B_{t-1}\left(1+i_{t-1}\right) + \int_0^1 \Omega_{it} di + W_t L_t - T_t \\ B_t\left(1+i_t\right) \geq -\sum_{T=t+1}^{\infty} \mathbb{E}_{t+1}\left\{ Q_{t+1,T}\left(\Omega_T + W_T L_T - T_T\right)\right\} \\ B_{-1} = 0, C_t = \left[\int_0^1 c_{it}^{\frac{\varepsilon-1}{\varepsilon}} di\right]^{\frac{\varepsilon}{\varepsilon-1}}, \Omega_t = \int_0^1 \Omega_{it} di \end{cases}$$

where i_t is the nominal interest rate, B_t are the nominal bond holdings, p_{it} is the price of good i in period t, L_t is labor time, W_t is the nominal wage, T_t are lump sum taxes paid to the government and

Ω_{it} are the profits of firm producing good i in period t. Markets are complete but I omit in the notation the arrow securities for clarity.

Firms Suppose there are $N \in \mathbb{N}$ regions ordered in the real line such that $\forall i \in [\overline{\omega}_{n-1}, \overline{\omega}_n)$ for $n = 1, 2..., N$, with $\overline{\omega}_0 = 0$ and $\overline{\omega}_N = 1$ and $\overline{\omega}_n - \overline{\omega}_{n-1} = \frac{1}{N}$, $\forall n$, lie in the same region. Let n denote an arbitrary region. The only difference across regions lies in the production function that firms operating in each region have access to. In particular, they produce according to:[15]

$$y_{int} = a_{nt} l_{int}^{1-\alpha_n}, \ \alpha_n \in (0,1) .$$

Government The government demands:

$$g_{it} = G_t \left(\frac{p_{it}}{P_t} \right)^{-\varepsilon}$$

and finances $\int_0^1 p_{it} g_{it} di$ with the lump-sum taxes T_t. Since markets are complete, the timing does not matter.

Price Dynamics Prices may be sticky in this economy (as in Calvo (1983)). At any given period t and in every region n, every firm faces a probability $1 - \theta$ of being able to adjust its price. Thus, following Uhlig (1996), in every region n a measure $[1 - \theta] \frac{1}{N}$ of producers change their prices and $\theta \frac{1}{N}$ keep them fixed. Also, the expected duration of a price is $\frac{1}{1-\theta}$. Finally, note that for the economy as a whole, in every period a fraction $1 - \theta$ of firms is resetting its price.

Aggregator The natural data counterpart of Y_{nt} in these type of models is real per capita GDP. Thus, the aggregator that links these regional per capita variables with the aggregate per capita one is:

$$Y_t = \sum_n \frac{1}{N} Y_{nt}. \tag{9}$$

Equilibrium The next proposition characterizes the equilibrium in this economy. The whole derivation is given in Appendix A. Before looking into the statement, let us define for $n = 1, ..., N$:

$$P_{n,t} := \left(N \int_{\overline{\omega}_{n-1}}^{\overline{\omega}_n} p_{it}^{-(\varepsilon-1)} di \right)^{-\frac{1}{\varepsilon-1}}$$

and let us denote by $P^*_{n+1,t}$ the index that would arise in this economy if prices were completely flexible. With these definitions:

[15] In contrast to the RBC model of the next section, I impose here that all i in $[\overline{\omega}_{n-1}, \overline{\omega}_n)$ share the same technology and interpret the set $[\overline{\omega}_{n-1}, \overline{\omega}_n)$ as the region. This allows me to use well-known machinery to solve these kind of models (see, for example, Woodford (2003)). Otherwise, however, the economic interpretation is the same.

Proposition 2. *The log-linearized equilibrium around a non-stochastic steady state with:*

$$\frac{P_{nt}}{P_{nt-1}} = \overline{\Pi}_n = 1 = \Pi_n^* = \frac{P_{nt}^*}{P_{nt}}, \frac{W_t}{P_{nt}} = \overline{\left(\frac{W}{P_n}\right)}$$
$$\forall n \in \{1, \dots, N\}, \; p_i = p_j =: \overline{p_n} \; \forall i, j \in [\overline{\omega}_{n-1}, \overline{\omega}_n)$$

in the regional new Keynesian model with a central bank that implements monetary policy using a Taylor rule of the form:

$$\widetilde{\iota}_t = v_t + \phi_\pi \pi_t + \phi_y \widetilde{Y}_t^C \tag{10}$$

is characterized by the following equations:

$$\begin{aligned}
\pi_{nt} &= \lambda_n \left(\frac{\sigma}{(1-\mathcal{G})} - \frac{1}{\varepsilon}\right) \widetilde{Y}_t^C - \lambda_n \frac{\sigma \mathcal{G}}{(1-\mathcal{G})} \widetilde{G}_t + \lambda_n \varphi \left(\sum_{n=1}^{N} \frac{\overline{L}_n}{\overline{L}_t} \left(\frac{1}{1-\alpha_n}\right) \left(\widetilde{Y}_{n,t} - \widetilde{a}_{n,t}\right)\right) \\
&+ \lambda_n \left(\frac{\alpha_n}{1-\alpha_n} + \frac{1}{\varepsilon}\right) \widetilde{Y}_{nt} - \frac{\lambda_n}{1-\alpha_n} \widetilde{a_{nt}} + \beta \mathbb{E}_t \left\{\pi_{nt+1}\right\} \\
0 &= -\frac{\sigma}{(1-\mathcal{G})} \mathbb{E}_t \left[\widetilde{Y}_{t+1}^C\right] + \frac{\sigma \mathcal{G}}{(1-\mathcal{G})} \mathbb{E}_t \left[\widetilde{G}_{t+1}\right] \\
&+ \frac{\sigma}{(1-\mathcal{G})} \widetilde{Y}_t^C - \frac{\sigma \mathcal{G}}{(1-\mathcal{G})} \widetilde{G}_t - \mathbb{E}_t \left[\pi_{t+1}\right] + v_t + \phi_\pi \pi_t + \phi_y \widetilde{Y}_t^C \\
\widetilde{Y}_{n,t} - \widetilde{Y}_{n,t-1} &= \widetilde{Y}_t^C - \widetilde{Y}_{t-1}^C - \varepsilon \left(\pi_{nt} - \pi_t\right)
\end{aligned}$$

where $\pi_{nt} = ln\left(P_{nt}\right) - ln\left(P_{nt-1}\right)$, $\lambda_n := \left(\frac{1-\theta}{\theta}\right)\left(1-\theta\beta\right)\left(\frac{1-\alpha_n}{1-\alpha_n+\varepsilon\alpha_n}\right)$, *for any variable* x_t, $\widetilde{x}_t := ln\left(\frac{x_t}{\overline{x}}\right)$, *where* $\overline{x}$ *is the value of* x_t *in the non-stochastic steady state and* $\widetilde{Y}_t^C = \sum_{n=1}^{N} \frac{1}{N} \left(\frac{\overline{Y}_n}{\overline{Y}}\right)^{\frac{\varepsilon-1}{\varepsilon}} \widetilde{Y}_{n,t}$.

The following result, which is a corollary of Proposition 12 in Section 4, gives sufficient conditions under which this model will feature an equilibrium regional equation like the equation we encountered in Section 2:

Corollary 1. *Consider the stacked version of the equilibrium equations in the regional new Keynesian model:*

$$\begin{pmatrix} \vec{K}_{t+1} \\ \mathbb{E}_t\left[\vec{P}_{t+1}\right] \end{pmatrix} = A \begin{pmatrix} \vec{K}_t \\ \vec{P}_t \end{pmatrix} + \Theta \Xi_t$$

where $\vec{K}_t := \left(\widetilde{Y}_{1t-1}, \dots, \widetilde{Y}_{Nt-1}\right)'$, $\vec{P}_t := \left(\pi_{1t}, \dots, \pi_{Nt}, \widetilde{Y}_{1t}, \dots, \widetilde{Y}_{Nt}\right)'$, $\Xi_t = \left(\widetilde{G}_t, v_t, \widetilde{a}_t, \varepsilon_{\widetilde{a}_{nt}}\right)'$, *with* $\widetilde{a}_{nt} = \widetilde{a}_t + \varepsilon_{\widetilde{a}_{nt}}$ *and* $\vec{K}_0$ *given. Suppose the following conditions hold:*

1. *A and Θ are bounded,*
2. *$\mathbb{E}_t\left[\Xi_{t+i}\right]$ does not explode,*[16] *and $\mathbb{E}_t\left[\Xi_{t+i}\right]$ is a linear function of Ξ_t,*
3. *A has exactly $2N$ eigenvalues outside the unit circle.*

[16] Condition $(1c)$ in Blanchard and Kahn (1980).

Then, the solution for $\widetilde{Y}_{nt}$ *is given by:*[17]

$$\widetilde{Y}_{nt} = \eta^n_{MG}\widetilde{G}_t + c_n'\vec{K}_t + F_t'\lambda_n + \epsilon_{nt}, \tag{11}$$

where $F_t = (v_t, \widetilde{a}_t)$, $\epsilon_{nt} = e_{nn}\varepsilon_{\widetilde{a}_{nt}} + (e_n'\mathbf{1})\left(\sum_{j\neq n}^N \frac{e_{nj}}{(e_n'\mathbf{1})}\varepsilon_{\widetilde{a}_{jt}}\right)$*, and*

$$\eta_{macro} := \frac{\partial \widetilde{Y}_t}{\partial \widetilde{G}_t} = \sum_{n=1}^N \omega_i \eta^n_{MG}, \tag{12}$$

with $\omega_i := \frac{1}{N}\left(\frac{\overline{Y_n}}{\overline{Y}}\right)$ *and* $\sum_{n=1}^N \omega_i = 1$.

Corollary 1 is a nice result because, if we compare equations (11) and (12) with (2) and (3), we see that although there are some important differences, we obtain an equilibrium regional equation with a structure very similar to that of the cases analyzed in Section 2. I will discuss the differences in detail in Sections 4 and 5 in light of how they matter for estimation and the definition of the class of models, so I postpone the discussion for the moment and instead turn to another important example.

3.2 A Regional RBC Model

Households Suppose there is a continuum of regions $i \in [0,1]$ and a representative agent who solves the following problem:

$$\max_{(c_{it},l_{it})_{i\in[0,1]},B_t} \mathbb{E}_0\left[\sum_{t=0}^{\infty}\beta^t\left(\frac{C_t^{1-\sigma}}{1-\sigma} - \int_0^1 \frac{l_{it}^{1+\varphi_i}}{1+\varphi_i}di\right)\right]$$

$$s.t.: \begin{cases} \int_0^1 p_{it}c_{it}di + B_t = B_{t-1}(1+i_{t-1}) + \int_0^1 \Omega_{it}di + \int_0^1 W_{it}l_{it}di - T_t \\ B_t(1+i_t) \geq -\sum_{T=t+1}^{\infty}\mathbb{E}_{t+1}\left\{Q_{t+1,T}\left(\Omega_T + \int_0^1 W_{iT}l_{iT}di - T_T\right)\right\} \\ B_{-1} = 0, C_t = \left[\int_0^1 c_{it}^{\frac{\varepsilon-1}{\varepsilon}}di\right]^{\frac{\varepsilon}{\varepsilon-1}}, \Omega_t = \int_0^1 \Omega_{it}di \end{cases}$$

where i_t is the nominal interest rate, B_t are the nominal bond holdings, p_{it} is the price of good i in period t, l_{it} is labor time employed in region i, W_{it} is the nominal wage of region i, T_t are lump sum taxes paid to the government and Ω_{it} are the profits of firm producing good i in period t. Markets are complete, but I omit in the notation the arrow securities for clarity.

Firms There is a monopolist firm in each region i that has technology:

$$y_{it} = a_{it}l_{it}^{1-\alpha_i},\ \alpha_i \in (0,1).$$

Government The government demands:

$$g_{it} = G_t\left(\frac{p_{it}}{P_t}\right)^{-\varepsilon}$$

and finances $\int_0^1 p_{it}g_{it}di$ with the lump-sum taxes T_t. Since markets are complete, the timing does not matter.

[17] The bold $\mathbf{1}$ denotes a column of ones.

Aggregator Note that here the equivalent of (9) is given by:

$$Y_t = \int_0^1 y_{it} di.$$

Equilibrium Before diving into the characterization of the equilibrium, let me provide a definition that will prove useful in future results:

Definition 1. *For any collection of random variables $x_t(i)$, $i \in [0,1]$, the process $x_t(.)$ is* purely regional *if $x_t(i)$ have zero mean, are independent and have finite variances ($\sigma_{ii} \leq M < \infty$) for all i.*

The next proposition characterizes the equilibrium in this economy (as before, for any variable x_t, $\widetilde{x}_t := ln\left(\frac{x_t}{\overline{x}}\right)$ where $\overline{x}$ is the value of x_t in the non-stochastic steady state):

Proposition 3. *Suppose that $\widetilde{a}_{it} = \widetilde{a}_t + \varepsilon_{\widetilde{a}_{it}}$ where $\varepsilon_{\widetilde{a}_{it}}$, $i \in [0,1]$, form a purely regional process. Then, in the log-linearized equilibrium, around a non-stochastic steady state with no inflation, regional output behaves according to:*

$$\widetilde{Y}_{it} = \eta^i_{MG}\widetilde{G}_t + \lambda_i\widetilde{a}_t + \lambda_i^{\varepsilon_a}\varepsilon_{\widetilde{a}_{it}} \tag{13}$$

where

$$\eta_{macro} := \frac{\partial\widetilde{Y}_t}{\partial\widetilde{G}_t} = \int_0^1 \omega_i\eta^i_{MG} di \tag{14}$$

with $\omega_i := \left(\frac{\overline{Y}_i}{\overline{Y}}\right)$ and $\int_0^1 \omega_i di = 1$.

Hence, as with the regional New Keynesian model, Proposition 3 shows that we also obtain an equilibrium regional equation for the RBC model with a structure very similar to that of the cases analyzed in Section 2.

3.3 Extensions and Discussion

The previous two examples show that some of the most commonly used models in macroeconomics, when we allow for heterogeneity in the fundamentals of different regions, have equilibrium regional equations that are very similar to those encountered in Section 2.

There are some differences, though; some of them will have consequences for the estimation framework and will be discussed in later sections, but there is one that deserves to be treated now because it is mainly an extension of the previous results. Both models in the previous subsections lack a region-specific policy variable, i.e. they have no $\varepsilon_{\widetilde{G}_{it}}$ in them. The reason for this is that the models, as presented, are the minimal deviations from the usual employed versions, and the main takeaway was to see how they would change when adding some heterogeneity. Now I want to go a step further and include, for the case of government spending as well, the $\varepsilon_{\widetilde{G}_{it}}$. I show the result for the RBC model:

Proposition 4. *Suppose the economy is given by the one in Proposition 3 with the difference that $\widetilde{G}_{it} = \widetilde{G}_t + \varepsilon_{\widetilde{G}_{it}}$, where $\varepsilon_{\widetilde{G}_{it}}$, $i \in [0,1]$, form a purely regional process and the technology of each firm is now given by $y_{it} = a_{it} l_{it}$ under perfect competition. Then, in the log-linearized equilibrium, regional output behaves according to:*

$$\widetilde{Y}_{it} = \eta^i_{MG} \widetilde{G}_t + \eta^i_{ML} \varepsilon_{\widetilde{G}_{it}} + \lambda_i \widetilde{a}_t + \lambda_i^{\varepsilon_a} \varepsilon_{\widetilde{a}_{it}} \tag{15}$$

where

$$\eta_{macro} := \frac{\partial \widetilde{Y}_t}{\partial \widetilde{G}_t} = \int_0^1 \omega_i \eta^i_{MG} di \tag{16}$$

with $\omega_i := \left(\frac{\overline{Y}_i}{\overline{Y}}\right)$ and $\int_0^1 \omega_i di = 1$.

Hence, note that Proposition 4 has the same structure as the simple model we went over in Section 2. The purpose of the next proposition is again to speak to the robustness of these equations and show that if government spending also serves a purpose in the production of consumption goods, the same results go through. We can think of it as government spending capturing some kind of infrastructure spending that enhances the productivity of firms:

Proposition 5. *Suppose the economy is given by the one in Proposition 4 with the difference that the technology is now given by $y_{it} = a_{it} G_{it}^{\rho_i} l_{it}$. Then, in the log-linearized equilibrium, regional output behaves according to:*

$$\widetilde{Y}_{it} = \eta^i_{MG} \widetilde{G}_t + \eta^i_{ML} \varepsilon_{\widetilde{G}_{it}} + \lambda_i \widetilde{a}_t + \lambda_i^{\varepsilon_a} \varepsilon_{\widetilde{a}_{it}} \tag{17}$$

where

$$\eta_{macro} := \frac{\partial \widetilde{Y}_t}{\partial \widetilde{G}_t} = \int_0^1 \omega_i \eta^i_{MG} di \tag{18}$$

with $\omega_i := \left(\frac{\overline{Y}_i}{\overline{Y}}\right)$ and $\int_0^1 \omega_i di = 1$.

Now I turn to other extensions. These show that the qualitative results obtained do not depend on the exact policy analyzed, i.e., the policy of government spending, although Proposition 5 already hinted at this. The next proposition shows that the same qualitative results arise if we instead look at a subsidy to the production of goods in the regional new Keynesian model:

Proposition 6. *Suppose that in the regional new Keynesian model there is no government spending on consumption goods, but the government subsidizes their production. In particular, assume that each firm in region n receives a subsidy, which is common to every firm in region n, proportional to its payroll of the form:*

$$[1 - s_{nt}] W_t l_{it}.$$

Then, Proposition 2 holds with:

$$\begin{aligned}
\pi_{n,t} = \lambda_n \left(\sigma - \frac{1}{\varepsilon}\right) \widetilde{Y}^C_t + \lambda_n \widetilde{s}_{n,t} + \lambda_n \varphi \left(\sum_{n=1}^N \frac{\overline{L}_n}{\overline{L}} \left(\frac{1}{1-\alpha_n}\right) \left(\widetilde{Y}_{n,t} - \widetilde{a}_{n,t}\right) \right) \\
+ \lambda_n \left(\frac{\alpha_n}{1-\alpha_n} + \frac{1}{\varepsilon}\right) \widetilde{Y}_{n,t} - \frac{\lambda_n}{1-\alpha_n} \widetilde{a}_{n,t} + \beta \mathbb{E}_t \left\{\pi_{n,t+1}\right\}
\end{aligned}$$

$$\sigma \mathbb{E}_t \left[\widetilde{Y}^C_{t+1}\right] = \sigma \widetilde{Y}^C_t - \mathbb{E}_t \left[\pi_{t+1}\right] + v_t + \phi_\pi \pi_t + \phi_y \widetilde{Y}^C_t$$

$$\widetilde{Y}_{n,t} - \widetilde{Y}_{n,t-1} = \widetilde{Y}^C_t - \widetilde{Y}^C_{t-1} - \varepsilon \left(\pi_{nt} - \pi_t\right).$$

I omit here the parallel result of Corollary 1 because it is straightforward to see that it still holds in this case. With the subsidy instead of government spending, we would obtain equations like (11) and (12).

The next proposition analyzes what happens in the regional RBC model:

Proposition 7. *Suppose that in the regional RBC model, there is no government spending on consumption goods, but the government subsidizes their production. In particular, assume that each firm receives a subsidy proportional to its payroll of the form:*

$$[1 - s_{it}]\, W_{it} l_{it}$$

where $\widetilde{s}_{it} = \widetilde{s}_t + \varepsilon_{\widetilde{s}_{it}}$*, and the* $\varepsilon_{\widetilde{s}_{it}}$*,* $i \in [0,1]$ *form a purely regional process. Then, Proposition 3 holds with:*

$$\widetilde{Y}_{it} = \eta^i_{MG}\widetilde{s}_t + \eta^i_{ML}\varepsilon_{\widetilde{s}_{it}} + \lambda_i \widetilde{a}_t + \lambda_i^{\varepsilon_a}\varepsilon_{\widetilde{a}_{it}} \tag{19}$$

where

$$\eta_{macro} := \frac{\partial \widetilde{Y}_t}{\partial \widetilde{s}_t} = \int_0^1 \omega_i \eta^i_{MG} di,$$

with $\omega_i := \left(\frac{\overline{Y}_i}{\overline{Y}}\right)$ *and* $\int_0^1 \omega_i di = 1$.

Finally, the following result shows that equations of these type are unrelated to the market completeness of the models analyzed thus far:

Proposition 8. *Suppose there are N regions, each with a representative agent with preferences over a tradable consumption good,* C_{nt}*, and labor,* L_{nt}*, given by:*

$$\mathbb{E}_0\left[\sum_{t=0}^{\infty}\beta^t\left(\frac{C_{nt}^{1-\sigma_n^c}}{1-\sigma_n^c} - \frac{L_{nt}^{1+\varphi_n}}{1+\varphi_n}\right)\right].$$

There is a representative firm in every region n that produces, under constant returns to scale, the tradable consumption good according to $Y_{nt} = a_{nt}L_{nt}$ *where* $\widetilde{a}_{it} = \widetilde{a}_t + \varepsilon_{\widetilde{a}_{it}}$*. Suppose the government spends* G_t *and finances this spending with lump sum taxes of* $\frac{1}{N}G_t$ *to every region, so that it holds a balanced budget in every state and period. Although there is a common market across regions for the tradable consumption good, labor markets are region-specific, and thus each representative agent earns* W_{nt} *for every unit of* L_{nt} *supplied. Moreover, assume the market incompleteness takes the extreme form of financial autarky, i.e., each representative agent has the budget constraint of* $C_{nt} = W_{nt}L_{nt} - T_{nt}$*.*

Then, in equilibrium:

$$\widetilde{Y}_{nt} = \eta^n_{MG}\widetilde{G}_t + \lambda_n \widetilde{a}_t + \epsilon_{nt},$$

where

$$\eta_{macro} := \frac{\partial \widetilde{Y}_t}{\partial \widetilde{G}_t} = \sum_n \omega_n \eta^n_{MG},$$

with $\omega_n := \frac{1}{N}\frac{\overline{Y}_n}{\overline{Y}}$ *and* $\sum_n \omega_n = 1$.

4 General Data Generating Processes

In this section, I present two general classes of models under which estimation will be carried out in the next section. Note that the two main models from Section 3 have an important difference in terms of their state variables. The regional RBC model had no state variables in the equilibrium equation, whereas the regional New Keynesian model had the vector of $\left(\widetilde{Y}_{1t-1}, \ldots, \widetilde{Y}_{Nt-1}\right)$ as state variables. The idea in this section is to encompass all of these cases. However, allowing for region-specific state variables is much more complicated in terms of estimation, and requires specific assumptions. In particular, since the number of regressors grows with the number of observations, some constraints on how this can happen are required in order to be able to prove the consistency of the estimators proposed. Because of this, I split the exposition in two. First, I define a class that includes general economies in which regions are only allowed to have common state variables (which could be lagged aggregates, for instance). For example, the RBC model has this structure. After this, I focus on the more involved class that allows the regional state variables. I first present the assumptions needed to specify the classes.

4.1 Assumptions

In this subsection, I will state some assumptions that will prove useful in the definitions of the classes and in the estimation results of next. However, I rarely apply all of these assumptions together, so it is best to look at this subsection as a collection of assumptions that will be used repeatedly in the upcoming results, but with different results based in different combinations. I present the main assumptions in this manner not only for the sake of organization, but also to allow readers to skip this section if they wish without losing conceptual clarity.

In this paper, all asymptotic results are obtained under $T, N \to \infty$. In the first class of models, this takes the form of a simultaneous limit, so all paths are allowed. In the second class, I consider paths of the form $\frac{N}{T} \underset{N,T\to\infty}{\longrightarrow} \rho \in (0,1)$. See the discussion in the proof of Proposition 13 for further details.

Suppose we have a model:

$$Y_n = X_n \beta_n^{(N)} + L_n^{(N)} \gamma_n^{(N)} + F \lambda_n^{(N)} + \epsilon_n \tag{20}$$

where $Y_n = (Y_{n1}, \ldots, Y_{nT})'$, $X_n = (X_{n1}, \ldots, X_{nP})$ with $X_{nj} = \left(X_{nj1}, \ldots, X_{njT}\right)'$, $L_n = (L_{n1}, \ldots, L_{nN})$ with $L_{nj} = \left(L_{nj1}, \ldots, L_{njT}\right)$, $F = (F_1, \ldots, F_K) = \left(F_1', \ldots, F_T'\right)'$ with $F_k = (F_{k1}, \ldots, F_{kT})'$ and $F_t = (F_{1t}, \ldots, F_{Kt})'$, and $\epsilon_n = (\epsilon_{n1}, \ldots, \epsilon_{nT})'$. Let me also define $Z_n^{(N)} := \left(X_n, L_n^{(N)}\right)$, $\delta_n^{(N)} := \left(\beta_n^{(N)\prime}, \gamma_n^{(N)\prime}\right)'$ and $\mathbb{X}_n := \left(X_n, L_n^{(N)}, F\right)$. To fix ideas, think of X_n as a fixed number of regressors (where typically those of interest, like the policy variable s_t, lie), $L_n^{(N)}$ as an extra set of regressors that are allowed to grow in number with N (the specific role of these will become clear when I define the last class of models), and F as the unobserved macro shocks or macro factors, like v_t in (10) (the monetary policy shock in the context of a new Keynesian model). The pair $\{F, \epsilon_n\}$ is unobserved for every n while the tuple $\left\{Y_n, X_n, L_n^{(N)}\right\}$ is observed for every region.

Then we have:

Assumption 1. $\exists\, u_{nt}, \theta_{nt}$ *such that:*

(1.1) $\epsilon_n = u_n + \sum_{j=1}^{J} \omega_{jn}$ *where* $\omega_{jn} = \vec{\omega}_{jn} \left(\check{\omega}_{j1}, ..., \check{\omega}_{jT}\right)'$ *with* $\vec{\omega}_{jn}$ *a scalar and* $\check{\omega}_{jt} = \frac{1}{N} \sum_{k=1}^{N} \phi_k w_{jkt}$*. Moreover,* $\exists C < \infty$ *such that for all* N *and* T*:*

(1.2) $\mathbb{E}\left[u_{nt}\right] = 0, \mathbb{E}\left[\left|u_{nt}\right|^{16}\right] < C, \ \forall n, t$

(1.3) $\mathbb{E}\left[u_{nt} u_{js}\right] = \tau_{nj,ts}$ *with* $\left|\tau_{nj,ts}\right| \leq \left|\tau_{nj}\right|$ *for some* τ_{nj} *for all* (t, s) *and* $\frac{1}{N} \sum_{n,j=1}^{N} \left|\tau_{nj}\right| < C$*; and* $\left|\tau_{nj,ts}\right| \leq \left|\kappa_{ts}\right|$ *for some* κ_{ts} *for all* (n, j) *and* $\frac{1}{T} \sum_{t,s=1}^{T} \left|\kappa_{ts}\right| < C$*; and* $\frac{1}{NT} \sum_{n,j,t,s=1} \left|\tau_{nj,ts}\right| < C$*.*

(1.4) For every (t, s)*,* $\mathbb{E}\left[\left|\frac{1}{\sqrt{N}} \sum_{n=1}^{N} \left(u_{ns} u_{nt} - \mathbb{E}\left[u_{ns} u_{nt}\right]\right)\right|^4\right] < C$*.*

(1.5) $T^{-2} N^{-1} \sum_{t,s,u,v} \sum_{n,j} \left|cov\left(u_{ns} u_{nt}, u_{ju} u_{jv}\right)\right| < C$ *and* $T^{-1} N^{-2} \sum_{t,s} \sum_{n,j,k,l} \left|cov\left(u_{nt} u_{jt}, u_{ksu} u_{ls}\right)\right| < C$*.*

(1.6) u_{nt} *is independent of* $X_{js}, L_{js}^{(N)}, F_s, \lambda_j$ *for all* n, t, j, s*.*

(1.7) $\mathbb{E}\left[w_{jnt}\right] = 0, \mathbb{E}\left[\left|w_{jnt}\right|^{16}\right] < C, \ \forall n, t$ *and* w_{jnt} *independent of* $X_{js}, L_{js}^{(N)}, F_s, \lambda_j$ *for all* n, t, j, s*.*

Assumption 1 allows the error term u_{nt} to be heteroskedastic and weakly correlated across time and in the cross-section. It also allows for heteroskedasticities. A completely cross-sectional and time independent u_{nt} satisfies conditions (1.3) – (1.5).

Assumption 2. *There are* K *macro shocks that satisfy* $\mathbb{E}\left[\|F_t\|^{16}\right] < \infty$ *and:*

$$\frac{1}{T} \sum_{t=1}^{T} F_t F_t' \xrightarrow{p} \Sigma_F$$

as $T \to \infty$ *with* Σ_F *a positive definite matrix.*

Assumption 2, like Assumption 3, is necessary for the number of unobserved macro shocks to be K.

Assumption 3. *The macro shocks matrix of loadings,* $\Lambda = \left[\lambda_1^{(N)}, ..., \lambda_N^{(N)}\right]'$*, satisfies* $\mathbb{E}\left[\left\|\lambda_n^{(N)}\right\|^{16}\right] < \infty$ *and:*

$$\frac{1}{N} \Lambda' \Lambda \xrightarrow{p} \Sigma_\Lambda$$

as $N \to \infty$ *with* Σ_Λ *a positive definite matrix.*

The next assumptions concern the observable regressors. In them, I denote as F^0 the true unobserved macro shocks, which generate the data in (20), and by F the parameters matrix used to estimate F^0. Thus, equation (20) now holds with F^0 instead of F.

Assumption 4. *The regressors satisfy* $\mathbb{E}\left[\|X_{nt}\|^{16}\right] < C, \mathbb{E}\left[\left|L_{ntp}\right|^{16}\right] < C$*,* $\{\mathbb{X}_{nt}, u_{nt}\}$ *is a stationary ergodic sequence,* $\frac{P+N+K}{T} \underset{N,T\to\infty}{\longrightarrow} \rho \in (0, 1)$ *and:*

$$\left[\frac{Z_n^{(N)'} M_{F^0} Z_n^{(N)}}{T}\right] \tag{21}$$

is a positive definite matrix and $O_p(1)$*, where* $M_F := I_T - F\left(F'F\right)^{-1} F'$*, and* M_{F^0} *equals* M_F *evaluated at the true macro shocks* F^0*. Moreover,* $\mathbb{E}\left\{\left[tr\left(D^{-2}\right)\right]^2\right\} = \mathbb{E}\left\{\left(\sum_u \left[\iota_{n,u}^T\right]^{-2}\right)^2\right\} \leq M < \infty$*,* $\forall T, n$ *where* $\iota_{n,u}^T$ *are the eigenvalues of* $(1/T) \mathbb{X}_n' \mathbb{X}_n$*.*

The assumption that (21) is positive definite (p.d.) is just the usual invertibility condition one would assume in model (20) if the unobserved macro shocks in F^0 were observable. This assumption means that when F is evaluated at the true macro shocks, the regressors in our model are not spanned by them. Some of the results in the following sections will treat cases in which regressors $L_n^{(N)}$ are absent. In those cases, it is understood that Assumption 4 holds for X_{nt} alone and there is no restriction on the path of $N, T \to \infty$.

Moreover, I will always assume:

Assumption 5. *The slopes in model (20) satisfy:*

$$\beta_n^{0(N)} \in \mathcal{B}_n \subseteq B(r, x), r < \infty, \ \forall n \text{ and } \forall N \text{ for some } x$$
$$\gamma_n^{(N)0} \in \mathcal{G}_n^{(N)} \subseteq B(r(N), h), \text{ with } r(N) = O\left(N^{-\frac{1}{2}}\right) \text{ for some } h,$$

where $B(l, m)$ is an open ball with center m and radius l. Moreover, $\beta_{np}^{0(N)}$ and $\frac{1}{N}\sum_{n=1}^{N}\beta_{np}^{0(N)}$ converge as $N \to \infty$ for every p.

Also, I will sometimes assume:

Assumption 6. *The matrix:*

$$\inf_{F \in \mathcal{F}} \frac{1}{N}\sum_{n=1}^{N} E_n \tag{22}$$

is positive definite, where:

$$\begin{aligned}
\mathcal{F} &:= \left\{F : \frac{F'F}{T} = I_K\right\} \\
E_n &:= B_n - C_n' A_n^g C_n \\
B_n &:= \left(\lambda_n \lambda_n'\right)' \otimes I_T \\
C_n &:= \left[\lambda_n' \otimes \left(\frac{(M_F Z_n)'}{\sqrt{T}}\right)\right] \\
A_n &:= \left(\frac{Z_n' M_F Z_n}{T}\right)
\end{aligned}$$

and A_n^g is a generalized inverse of A_n.

Assumption 6 is part of a set of sufficient conditions to consistently estimate the space spanned by the true unobserved macro shocks in single-equation interactive-effects estimators. See, for example, Bai (2009) and Ando and Bai (2015).

4.2 Common State Variables Class (CSVC), Regional State Variables Class (RSVC), Subclasses and Examples

In this subsection, I introduce the classes of models that we will use for estimation in Section 5. Let me start with the CSVC. As the name makes clear, in this class, only models that feature common state variables across regions are allowed. For the definition, I will denote as M an arbitrary model. That is, M stands for the market structure, preferences, technologies of production, trading rules, equilibrium concepts, etc. Making explicit all these characteristics would not only be very tedious, but also would likely restrict the applicability of the results in unexpected ways, so I choose to leave them implicit.

The definition of the class is the following:

Definition 2. *A model M belongs to the $CSVC$ iff:*

1. *In equilibrium, regional output (its first log-difference, or its deviation from the non-stochastic steady state) can be written as* (20)*, and Assumptions 1 through 5 are satisfied.*

2. *The regional state variables are common to all regions.*

As we saw in Section 3, there are many commonly used models in which regional output (its deviation from the non-stochastic steady state) can be written as (20) in equilibrium. And this is true for different policies as well, such as government spending or regional subsidies. The first condition in the definition makes sure that Assumptions 1 through 5 are satisfied in the models considered in the class. The second condition implies that the size of the vector $L_n^{(N)}$ does not depend on N.

The following proposition is useful to fix ideas. It shows that the discrete region version of the regional RBC model of Subsection 3.2 belongs to the CSVC.[18]

Proposition 9. *Suppose that in the regional RBC model of Subsection 3.2 there are N equidistant regions, obtained by imposing $\varphi_i = \varphi_n, \alpha_i = \alpha_n, a_{it} = a_{nt}$ for all $i \in \left[\frac{n-1}{N}, \frac{n}{N}\right)$, and that the following conditions hold for all N and T:*

1. *$\varphi_n, \alpha_n, \varepsilon, \sigma, \mathcal{G}$ are bounded,*

2. *$\Delta \log\left(\varepsilon_{a_{it}}\right)$ are $i.i.d.$ with mean zero and variance $\sigma^2 < \infty$,*

3. *$\left(\frac{\varepsilon\left(1+\varphi_i\right)}{\left(\alpha_i+\varphi_i\right)\varepsilon+1-\alpha_i}\right) \Delta \log\left(\varepsilon_{a_{it}}\right)$ satisfies conditions* (1.2) *and* (1.6) *of Assumption 1,*

4. *The sequences $\frac{1}{N}\sum_{i=1}^{N}\left(\frac{\overline{Y}_i}{Y^C}\right)^{\frac{\varepsilon-1}{\varepsilon}}\left(\frac{(1-\alpha_i)\{(1-\mathcal{G})-\varepsilon\sigma\}}{\left\{\left(\alpha_i+\varphi_i\right)\varepsilon+1-\alpha_i\right\}(1-\mathcal{G})}\right)$, $\frac{1}{N}\sum_{i=1}^{N}\left(\frac{\overline{Y}_i}{Y^C}\right)^{\frac{\varepsilon-1}{\varepsilon}}\frac{(1-\alpha_i)\varepsilon\sigma\mathcal{G}}{(1-\mathcal{G})\left\{\left(\alpha_i+\varphi_i\right)\varepsilon+1-\alpha_i\right\}}$, $\frac{1}{N}\sum_{i=1}^{N}\left(\frac{\overline{Y}_i}{Y^C}\right)^{2\left(\frac{\varepsilon-1}{\varepsilon}\right)}\left(\frac{\varepsilon\left(1+\varphi_i\right)}{\left(\alpha_i+\varphi_i\right)\varepsilon+1-\alpha_i}\right)^2$, $\frac{1}{N}\sum_{i=1}^{N}\left(\frac{\overline{Y}_i}{Y^C}\right)^{\frac{\varepsilon-1}{\varepsilon}}\left(\frac{\varepsilon\left(1+\varphi_i\right)}{\left(\alpha_i+\varphi_i\right)\varepsilon+1-\alpha_i}\right)$, $\frac{1}{N}\sum_{i=1}^{N}\left(\frac{(1-\alpha_i)\{(1-\mathcal{G})-\varepsilon\sigma\}}{\left\{\left(\alpha_i+\varphi_i\right)\varepsilon+1-\alpha_i\right\}(1-\mathcal{G})}\right) \neq 1$,*
$\frac{1}{N}\sum_{i=1}^{N}\frac{(1-\alpha_i)\varepsilon\sigma\mathcal{G}}{(1-\mathcal{G})\left\{\left(\alpha_i+\varphi_i\right)\varepsilon+1-\alpha_i\right\}}$, converge,

5. *$\Delta \log\left(a_t\right)$ satisfies Assumption 2,*

6. *$\Delta \log\left(G_t\right)$ satisfies Assumption 4.*

Then, the regional RBC model of Subsection 3.2 belongs to the CSVC.

[18] Condition 2 in Proposition 9 could be weakened to allow for weak cross-sectional dependence, but this would come at the expense of a less clean statement (and proof). Since we are interested only in illustrating examples of specific models included under the classes in consideration, I stick with the stronger assumptions. Moreover, for the same reason I rarely focus on giving a minimal, non-redundant, set of assumptions.

Using the same machinery we have used thus far, it is easy to see that versions of the regional RBC model with additional macro shocks, such as the production subsidy shock discussed earlier, will also fit in the CSVC. This suggests that indeed the CSVC is a very large class, and that we can go further in characterizing it. Of course, for estimation purposes, as long as the dgp belongs to the class, our estimator will be robust to any model in it. But having a clearer idea of what belongs to the class and what does not is useful. The next result, instead of showing that a particular model (or a number of particular models) belongs to the CSVC, characterizes an appealing subclass of the CSVC.

Before stating the result, let me define:

Definition 3. *$\widetilde{X}_t$ is a* generalized symmetric aggregator *(gsa) of the vector $\left(\widetilde{X}_{1t}, \ldots, \widetilde{X}_{Nt}\right)$ iff:*

$$\widetilde{X}_t = \sum_{n=1}^{N} f_n \left(\overline{X}_{1t}, \ldots, \overline{X}_{Nt}\right) \widetilde{X}_{nt},$$

with $\sum_{n=1}^{N} f_n \left(\overline{X}_{1t}, \ldots, \overline{X}_{Nt}\right) = 1$.

Note that the CES aggregator that appeared in both the regional new Keynesian model and the regional RBC is a gsa.[19] Let me also define:

Definition 4. *$\widetilde{X}_t$ is a* simple aggregator *of the vector $\left(\widetilde{X}_{1t}, \ldots, \widetilde{X}_{Nt}\right)$ iff $\widetilde{X}_t = \frac{1}{N} \sum_{n=1}^{N} \widetilde{X}_{nt}$.*

The result then is the following:

Proposition 10. *Suppose we have a regional model M with equilibrium equations given by:*

$$\mathbb{E}_t \left[\vec{Y}_{t+1}\right] = A\vec{Y}_t + \Theta\vec{G}_t$$

where $\vec{X}_t := \left(\widetilde{X}_{1t}, \ldots, \widetilde{X}_{Nt}\right)'$ for an arbitrary vector $\vec{X}_t$, and interactions across regions come only through simple aggregators, either directly or through common prices, i.e.:

$$A = \begin{pmatrix} a_{11} + b_1 \frac{1}{N} & \cdots & b_1 \frac{1}{N} \\ \vdots & \ddots & \vdots \\ b_N \frac{1}{N} & \cdots & a_{NN} + b_N \frac{1}{N} \end{pmatrix}, \Theta = \begin{pmatrix} \theta_{11} + \nu_1 \frac{1}{N} & \cdots & \nu_1 \frac{1}{N} \\ \vdots & \ddots & \vdots \\ \nu_N \frac{1}{N} & \cdots & \theta_{NN} + \nu_N \frac{1}{N} \end{pmatrix}.$$

Suppose as well that for all N:[20]

1. *$a_{ii}, b_i, \theta_{ii}, v_i$ are bounded for all i, $a_{ii} \neq 0$ for all i,*

[19] In general, an aggregator of the form:

$$Y_t = f^{-1}\left(\sum_{n=1}^{N} f\left(Y_{nt}\right)\right) \implies \widetilde{Y}_t = \sum_{n=1}^{N} \frac{f'\left(\overline{Y}_{nt}\right)\overline{Y}_{nt}}{f'\left(\overline{Y}_t\right)\overline{Y}_t} \widetilde{Y_{nt}}$$

is a gsa if $\sum_{n=1}^{N} \frac{f'\left(\overline{Y}_{nt}\right)\overline{Y}_{nt}}{f'\left(\overline{Y}_t\right)\overline{Y}_t} = 1$.

[20] Note that we can easily choose parameters such that A is diagonally dominant both by rows and by columns; hence, condition 3 is easily satisfiable for many matrices with a structure like A. See, for example, Varah (1975).

2. $1+\frac{1}{N}\sum_{h=1}^{N}a_{hh}^{-1}b_h\neq 0$,

3. *All eigenvalues of A lie outside the unit circle,*

4. $\vec{G}_t$ *is a vector such that* $\widetilde{G}_{nt}=\widetilde{G}_t+\varepsilon_{\widetilde{G}_{nt}}$ *with* $\widetilde{G}_t$ *and* $\varepsilon_{\widetilde{G}_{nt}}$ *martingale difference sequences with* $\overline{\varepsilon}_{G_n}=1$,

5. $\varepsilon_{\widetilde{G}_{nt}}$ *are* $i.i.d.$ *with mean zero and variance* $\sigma^2<\infty$,

6. $\frac{1}{N}\sum_{i=1}^{N}\left[a_{nn}^{-1}\nu_n-a_{jj}^{-1}\theta_{jj}\frac{a_{nn}^{-1}b_n}{1+\frac{1}{N}\sum_{h=1}^{N}a_{hh}^{-1}b_h}-\left(\frac{a_{nn}^{-1}b_n}{1+\frac{1}{N}\sum_{h=1}^{N}a_{hh}^{-1}b_h}\right)\left(\frac{1}{N}\sum_{h=1}^{N}a_{hh}^{-1}\nu_h\right)\right]^2$, $\frac{1}{N}\sum_{h=1}^{N}a_{hh}^{-1}b_h$, $\frac{1}{N}\sum_{h=1}^{N}a_{hh}^{-1}\nu_h$, *and* $\frac{1}{N}\sum_{h=1}^{N}a_{hh}^{-1}\theta_{hh}$ *converge,*

7. $\widetilde{G}_t$ *and* $\varepsilon_{\widetilde{G}_{nt}}$ *satisfy Assumption 4.*

Then, M belongs to the CSVC.

The following is a straightforward corollary. It says that in the subclass we just analyzed, the equilibrium equation for regional output takes a very simple form:

Corollary 2. *Suppose the conditions for Proposition 10 hold, then:*

$$\widetilde{Y}_{nt}=\gamma_n\widetilde{G}_t+\beta_n\varepsilon_{\widetilde{G}_{Nt}}+o_p(1)$$
$$\frac{1}{N}\sum_{n=1}^{N}\gamma_n=\eta_{macro}.$$

In what follows, I will denote as **1** the column vector of ones. The following result characterizes an even bigger subclass:[21]

Proposition 11. *Part (1): Suppose we have a regional model M with equilibrium equations given by:*

$$\begin{pmatrix}\vec{K}_{t+1}\\ \mathbb{E}_t\left[\vec{P}_{t+1}\right]\end{pmatrix}=A\begin{pmatrix}\vec{K}_t\\ \vec{P}_t\end{pmatrix}+\Theta\Xi_t$$

where $\vec{K}_t:=\left(\widetilde{K}_{1t},...,\widetilde{K}_{Jt}\right)'$, $\vec{P}_t:=\left(\widetilde{P}_{11t},...,\widetilde{P}_{1Nt},...,\widetilde{P}_{R1t},...,\widetilde{P}_{RNt}\right)'$, $\left(\widetilde{P}_{11t},...,\widetilde{P}_{1Nt}\right)=\left(\widetilde{Y}_{1t},...,\widetilde{Y}_{Nt}\right)$, $\Xi_t=\left(\widetilde{s}_t,...,\widetilde{s}_{t-q},\varepsilon_{\widetilde{s}_{nt}},\varepsilon_{\widetilde{s}_{nt-2}},\xi_{1t},...,\xi_{1t-q},\xi_{2t},...,\xi_{2t-q},\varepsilon_{\xi_{1t}^n}\right)'$ *and* $\vec{K}_0$ *given. Suppose for all N the following conditions hold:*

1. *A and* Θ *are bounded,*

2. $\mathbb{E}_t\left[\Xi_{t+i}\right]$ *does not explode,*[22] *and* $\mathbb{E}_t\left[\Xi_{t+i}\right]$ *is a linear function of* Ξ_t,

3. *A has exactly* $N*R$ *eigenvalues outside the unit circle,*

[21] I restrict the number of macro shocks to 2 and the number of lags on some variables. This is all without loss of generality and just for notational simplicity.

[22] Condition (1c) in Blanchard and Kahn (1980).

Then the solution for $\widetilde{Y}_{nt}$ is given by:

$$
\begin{aligned}
\widetilde{Y}_{nt} &= a_n^{(N)\prime}\vec{s}_t + b_n^{(N)\prime}\varepsilon_{\vec{s}_{nt}} + c_n^{(N)\prime}\vec{K}_t + F_t'\lambda_n^{(N)} + u_{nt} \\
F_t &= \left(\xi_{1t}, \ldots, \xi_{1t-q}, \xi_{2t}, \ldots, \xi_{2t-q}\right) \\
\vec{s}_t &= \left(\widetilde{s}_t, \widetilde{s}_{t-1}, \ldots, \widetilde{s}_{t-q}\right) \\
\varepsilon_{\vec{s}_{nt}} &= \left(\varepsilon_{s_{nt}}, \varepsilon_{s_{nt-1}}\right) \\
u_{nt} &= e_{nn}^{(N)}\varepsilon_{\xi_{1t}^n} + \left(c_n^{(N)\prime}, d_n^{(N)\prime}, e_n^{(N)\prime}\right)\mathbf{1}\left\{\begin{array}{c}
\sum_{j\neq n}^N\left(\frac{c_{nj}^{(N)}}{\left(c_n^{(N)\prime}, d_n^{(N)\prime}, e_n^{(N)\prime}\right)\mathbf{1}}\right)\varepsilon_{s_{jt}} \\
+\sum_{j\neq n}^N\left(\frac{d_{nj}^{(N)}}{\left(c_n^{(N)\prime}, d_n^{(N)\prime}, e_n^{(N)\prime}\right)\mathbf{1}}\right)\varepsilon_{s_{jt-1}} \\
+\sum_{j\neq n}^N\left(\frac{e_{nj}^{(N)}}{\left(c_n^{(N)\prime}, d_n^{(N)\prime}, e_n^{(N)\prime}\right)\mathbf{1}}\right)\varepsilon_{\xi_{1t}^j}
\end{array}\right\}
\end{aligned}
$$

Part (2)*: Suppose as well that for all N:*[23]

1. $\Delta\log\left(\varepsilon_{s_{nt}}\right)$ *and* $\Delta\log\left(\varepsilon_{\xi_{1t}^j}\right)$ *are $i.i.d.$ with mean zero and variance $\sigma^2 < \infty$,*
2. $a_n^{(N)}, b_n^{(N)}, c_n^{(N)}, \lambda_n^{(N)}$ *and their cross-sectional averages converge, and* $\left(c_n^{(N)\prime}, d_n^{(N)\prime}, e_n^{(N)\prime}\right)\mathbf{1}$ *is bounded,*
3. *F_t satisfies Assumption 2 and $\lambda_n^{(N)}$ satisfies Assumption 3,*
4. *$\vec{s}_t$, $\vec{\varepsilon}_{s_{nt}}$ and $\vec{K}_t$ satisfy Assumption 4.*

Then, M belongs to the CSVC.

Something worth pointing out is that models with $K_t = Y_{t-1}$, i.e., having lagged aggregate variables as the state variables common to every region, are included in Proposition 11.

Now I turn to the RSVC. The main difference is that now regional state variables are allowed:

Definition 5. *A model M belongs to the $RSVC$ iff:*

1. *In equilibrium, regional output (its first log-difference, or its deviation from the non-stochastic steady state) can be written as* (20)*, and Assumptions 1 through 5 are satisfied.*
2. *The number of regional state variables is N.*

As next section will make evident, all the results regarding estimation and related to the RSVC would still hold under the assumption that the number of regional state variables is $O(N)$, adjusting T accordingly. But since our examples deal mainly with the case in which this number is N, that is the case I analyze. The following proposition is the generalization of Proposition 11 to the RSVC:

[23] As before, the first condition could be weakened but at the expense of less clean notation. The same applies to the second condition. Conceptually there is nothing lost in specifying the result like this.

Proposition 12. *Part* (1)*: Suppose we have a regional model M with equilibrium equations given by:*

$$\begin{pmatrix} \vec{K}_{t+1} \\ \mathbb{E}_t\left[\vec{P}_{t+1}\right] \end{pmatrix} = A \begin{pmatrix} \vec{K}_t \\ \vec{P}_t \end{pmatrix} + \Theta \Xi_t$$

where $\vec{K}_t := \left(\widetilde{K}_{1t}, ..., \widetilde{K}_{Nt}\right)'$, $\vec{P}_t := \left(\widetilde{P}_{11t}, ..., \widetilde{P}_{1Nt}, ..., \widetilde{P}_{R1t}, ..., \widetilde{P}_{RNt}\right)'$, $\left(\widetilde{P}_{11t}, ..., \widetilde{P}_{1Nt}\right) = \left(\widetilde{Y}_{1t}, ..., \widetilde{Y}_{Nt}\right)$, $\Xi_t = \left(\widetilde{s}_t, ..., \widetilde{s}_{t-q}, \varepsilon_{\widetilde{s}_{nt}}, \varepsilon_{\widetilde{s}_{nt-2}}, \xi_{1t}, ..., \xi_{1t-q}, \xi_{2t}, ..., \xi_{2t-q}, \varepsilon_{\xi^n_{1t}}\right)'$ *and* $\vec{K}_0$ *given. Suppose for all N the following conditions hold:*

1. *A and* Θ *are bounded,*
2. $\mathbb{E}_t\left[\Xi_{t+i}\right]$ *does not explode,*[24] *and* $\mathbb{E}_t\left[\Xi_{t+i}\right]$ *is a linear function of* Ξ_t*,*
3. *A has exactly* $N * R$ *eigenvalues outside the unit circle,*

Then the solution for $\widetilde{Y}_{nt}$ *is given by:*

$$\begin{aligned}
\widetilde{Y}_{nt} &= a_n^{(N)'} \vec{s}_t + b_n^{(N)'} \varepsilon_{\vec{s}_{nt}} + c_n^{(N)'} \vec{K}_t + F_t' \lambda_n^{(N)} + u_{nt} \\
F_t &= \left(\xi_{1t}, ..., \xi_{1t-q}, \xi_{2t}, ..., \xi_{2t-q}\right) \\
\vec{s}_t &= \left(\widetilde{s}_t, \widetilde{s}_{t-1}, ..., \widetilde{s}_{t-q}\right) \\
\varepsilon_{\vec{s}_{nt}} &= \left(\varepsilon_{s_{nt}}, \varepsilon_{s_{nt-1}}\right) \\
u_{nt} &= e_{nn}^{(N)} \varepsilon_{\xi^n_{1t}} + \left(c_n^{(N)'}, d_n^{(N)'}, e_n^{(N)'}\right)\mathbf{1} \left\{ \begin{array}{c} \sum_{j\neq n}^N \left(\frac{c_{nj}^{(N)}}{\left(c_n^{(N)'}, d_n^{(N)'}, e_n^{(N)'}\right)\mathbf{1}}\right) \varepsilon_{s_{jt}} \\ + \sum_{j\neq n}^N \left(\frac{d_{nj}^{(N)}}{\left(c_n^{(N)'}, d_n^{(N)'}, e_n^{(N)'}\right)\mathbf{1}}\right) \varepsilon_{s_{jt-1}} \\ + \sum_{j\neq n}^N \left(\frac{e_{nj}^{(N)}}{\left(c_n^{(N)'}, d_n^{(N)'}, e_n^{(N)'}\right)\mathbf{1}}\right) \varepsilon_{\xi^j_{1t}} \end{array} \right\}
\end{aligned}$$

Part (2)*: Suppose as well that for all N:*[25]

1. $\Delta \log\left(\varepsilon_{s_{nt}}\right)$ *and* $\Delta \log\left(\varepsilon_{\xi^j_{1t}}\right)$ *are* $i.i.d.$ *with mean zero and variance* $\sigma^2 < \infty$*,*
2. $a_n^{(N)}, b_n^{(N)}, \lambda_n^{(N)}$ *and their cross-sectional averages converge,* $\left\|c_n^{(N)}\right\| = O\left(N^{-\frac{1}{2}}\right)$ *and* $\left(c_n^{(N)'}, d_n^{(N)'}, e_n^{(N)'}\right)\mathbf{1}$ *is bounded,*
3. F_t *satisfies Assumption 2 and* $\lambda_n^{(N)}$ *satisfies Assumption 3,*
4. $\vec{s}_t$*,* $\vec{\varepsilon}_{s_{nt}}$ *and* $\vec{K}_t$ *satisfy Assumption 4.*

Then, M belongs to the RSVC.

[24] Condition ($1c$) in Blanchard and Kahn (1980).

[25] As before, the first condition could be weakened but at the expense of less clean notation. The same applies to the second condition. Conceptually, there is nothing lost in specifying the result like this.

At first glance, it might seem annoying to make a clear distinction between the CSVC and the RSVC. The reason for this, which is the topic of next section, is that when the dgp belongs to the RSVC, we need to be very careful when proving the consistency of the estimators considered, because the number of regressors increases with the cross-sectional dimension.

Proposition 12 implies that if the regional new Keynesian model of subsection 3.1 satisfies the assumptions of the proposition, then that model belongs to the RSVC.

Moreover, note that in Proposition 12, and as in the new Keynesian model of subsection 3.1, we can have $\widetilde{K}_{nt} = \widetilde{Y}_{nt-1}$ as the state variable for region n. As we mentioned before, we can also allow for more than one state variable per region on average, but I stick with this version here for conceptual clarity.

Finally, note that making the trivial adjustments needed on the statements, we have that $CSVC \subseteq RSVC$.

5 Estimation and Asymptotic Results

The purpose of this section is to describe estimators that allow the recovery of the micro-global elasticities from the regional equations, and with them the estimation of the macro elasticities. The estimation framework combines the equilibrium equations from the previous section with a structural equation for each of the endogenous variables.

In any of the different specifications, the procedure is always to first obtain estimators of the unobserved macro shocks from the whole system of equations, and then to use these estimates as controls in the regional equations of interest. Throughout these results, I assume that the number of unobserved macro shocks is known. In the last subsection, I discuss what to do when the number is unknown. Moreover, for every specification I show the consistency of the estimators. I discuss inference in Subsection 5.6.

The ultimate goal, then, is to recover the micro-global elasticities in a way that is consistent with the class of models specified in the previous section. Although one could pursue the following theory for general classes as the classes presented, it is worthwhile to be specific and to stay close to the examples we have seen so far. The reason is that the specification of the remaining equations of the system is naturally linked to the identification concerns we have in these contexts, and by allowing more generality, we distance ourselves from these concerns. However, the definition of the classes helps show clearly where we stand in terms of generality and how the methods can be extended.

Hence, I now focus particularly on models in which the key regressors represent policy variables set by a policy maker,[26] examples of which were covered in Section 3. To that end, let me start by classifying the different variables in (20) in terms of our needs and the role they play. As part of the regressors X_n, we have the common policy variable, s_t, and the region-specific policy variable, $\varepsilon_{s_{nt}}$. In the context of the fiscal multipliers examples, like those in Section 3, these were $\widetilde{G}_t$ and $\varepsilon_{\widetilde{G}_{nt}}$. The regressors in $L_n^{(N)}$ capture different specifications for the state variables, like the complete vector of lagged regional output growth from the regional New Keynesian model. I begin without these variables. In F_t we have

[26] The policy variable could be the result of the decision of a policy maker, legislature, etc. For convenience, I will refer to the policy as being set by a policy maker.

the unobserved macro shocks, and ϵ_{nt} captures a region-specific idiosyncratic shock, which in some of the models we reviewed represented region-specific TFP shocks. Remember that all these variables represented growth rates. If rewrite (20) with this simpler notation, we get:

$$Y_{nt} = \eta^{n}_{MG} s_t + \eta^{n}_{ML} \varepsilon_{s_{nt}} + F_t \lambda_n + \epsilon^{Y}_{nt}. \tag{23}$$

Context of (23) Because of the interaction between λ_n and F_t, models like (23) belong to the realm of interactive-effects models. Let me briefly put these models in context to clarify the scope of the results, and let me assume, for clarity, that F_t and λ_n are scalars. The usual fixed-effects model is a particular case of interactive-effects models, and would, instead, display those terms additively, as in $\lambda_n + F_t$. The interactive-effects model, in turn, is a particular case of a more general linear panel data model in which all parameters are allowed to vary with n and t, as in:

$$Y_{nt} = \eta^{nt}_{MG} s_t + \eta^{nt}_{ML} \varepsilon_{s_{nt}} + \alpha_{nt} + \epsilon^{Y}_{nt}. \tag{24}$$

Unfortunately, model (24) is not estimable since the number of parameters to estimate is greater than the number of observations, and consequently, some restrictions need to be imposed to estimate the model. In this sense, model (24) is too general. In principle, it's not clear where to get these restrictions. Nonetheless, the results discussed in previous sections say that we can use macroeconomic theory to guide this choice: the most commonly used macro models lead to model (23), in which $\alpha_{nt} = \lambda_n F_t$. In particular, theory, at least for macroeconomics, seems to disfavor the usual additive-effects model, in which $\alpha_{nt} = \lambda_n + F_t$. This is because, as the results in Section 3 show, there seem to be two reasonable cases. We either have economies with heterogeneity as in (23), in which $\alpha_{nt} = \lambda_n F_t$ and the elasticities of the observable variables are region-varying. Alternatively, we have economies with homogeneity, for which these elasticities are region-invariant but $\alpha_{nt} = F_t$, as in:[27]

$$Y_{nt} = \eta_{MG} s_t + \eta_{ML} \varepsilon_{s_{nt}} + \alpha_t + \epsilon^{Y}_{nt}. \tag{25}$$

The intuition behind this is that one cannot choose where the model displays the heterogeneity: if the aggregate unobserved macro shock, F_t, has an heterogeneous impact on different regions, then so too must s_t.

A Menu of Empirical Strategies In what follows, I present four different ways of adding the remaining equations to (23) in order to have a complete simultaneous equation model. Even though I present the results progressively, adding more threats to identification as we move forward, each stage could by itself represent the right assumptions in different applications. Hence, one way to look at the different specifications in this section is as a menu of different empirical strategies, with some of them being better suited for a particular application than the rest.[28] In Section 6, I offer an application of these methods to the case of fiscal multipliers, and there I discuss the strategies in this section in light of the particular problems we typically think of when analyzing government spending. Thus, although I

[27] In (25) a grand-mean can be included in α_t.

[28] Moreover, the application need not be, in principle, restricted to measuring the effects of policy variables, and could, for example, be used to analyze the effects of aggregate macro shocks as well.

sometimes use the example of fiscal multipliers to motivate the various strategies, as well as examples from Section 3, the intention in this section is to show many possible strategies and assumptions that make it possible to recover the micro-global elasticities, without regarding any of them as the most reasonable.

Roadmap The ultimate goal of our estimation procedure is to recover η^n_{MG} from (23). The problems we have to address to do so are that the regressors, s_t and $\varepsilon_{s_{nt}}$, might be correlated with F_t and/or ϵ^Y_{nt}. The following points summarize the main ideas:

- The four specifications will allow arbitrary correlations of s_t and $\varepsilon_{s_{nt}}$ with F_t.
- The four specifications will treat cases in which s_t is uncorrelated, at least asymptotically (when $N \to \infty$), with ϵ_{nt}. The reason for this is that the most likely cause for the econometric endogeneity of s_t is that the policy maker might be responding to some index of economic performance in choosing s_t. As long as this index is "well-diversified" in the sense that there is no particular region that monopolizes this response, s_t will become uncorrelated with ϵ^Y_{nt} as $N \to \infty$.
- For $\varepsilon_{s_{nt}}$, correlation with ϵ^Y_{nt} needs to be taken into account. Hence, the three strategies will move progressively on this front, starting from an assumption of uncorrelatedness and finishing with assumptions that resemble those of s_t, in which the structural equation for $\varepsilon_{s_{nt}}$ depends directly on Y_{nt}.

My purpose here is to show all of the results as simply as possible. This means, in particular, that my purpose is not to show the results with the minimal necessary set of assumptions for them to go through. Instead, I often choose to make stronger assumptions in order to be able to use a simpler notation in the statements and proofs.

5.1 Case 0: $\varepsilon_{s_{nt}} \perp \epsilon_{nt}$ Without Model Selection

Let me start with a case that leads to the easiest scenario, in terms of getting at the estimator. The title of this subsection points out that it won't be necessary to apply a model selection technique under these assumptions, although I will do so in the next subsection.

Denoting as F^0_t the unobserved macro shocks entering (23), the assumption in terms of reduced forms is:

Assumption 7. *The reduced forms of s_t and $\varepsilon_{s_{nt}}$ satisfy:*

$$s_t = \theta^{s\prime} F^0_t + u^s_t \tag{26}$$

$$\varepsilon_{s_{nt}} = \theta^{\varepsilon\prime}_n F^0_t + \epsilon^s_{nt} \tag{27}$$

$$\epsilon^s_{nt} \perp \epsilon^Y_{nt}. \tag{28}$$

In the next subsection, I will present a structural system that gives rise to these reduced forms and makes them appealing. First, however, I will explain how the estimation works.

The general idea for estimation under the system given by (23), (26), (27) and (28) is straightforward. Under some normalization restrictions, we can estimate the unobserved macro shocks entering (27) with the method of asymptotic principal components. With these estimates then we can run a time series OLS regression of (23), controlling for the unobserved macro shocks, to obtain the regional elasticities. Although our setup has particular features that are different, this general idea is the same as in many strategies that rely on factor-augmented regressions; see for example Stock and Watson (2002).

Now I present the estimation in detail. For the moment, the total number of unobserved macro shocks, K, is treated as known. Later in this section, I will discuss what to do when this is not the case. The first step is getting an estimate of F^0. This is the solution to the following problem:

$$\left(\hat{F},\left\{\hat{\theta}_n^{\varepsilon}\right\}_{n=1}^{N}\right)=\underset{\left(F,\{\theta_n^{\varepsilon}\}_{n=1}^{N}\right)}{\arg\min}\ \frac{1}{NT}\sum_{n=1}^{N}\left(\varepsilon_{s_n}-F\theta_n^{\varepsilon}\right)'\left(\varepsilon_{s_n}-F\theta_n^{\varepsilon}\right) \tag{29}$$
$$s.t.:\begin{cases}\frac{F'F}{T}=I_K\\ \Theta^{\varepsilon\prime}\Theta^{\varepsilon}\ diagonal\end{cases}$$

where $\varepsilon_{s_n}=\left(\varepsilon_{s_{n1}},...,\varepsilon_{s_{nT}}\right)'$ and $\Theta^{\varepsilon}=\left(\theta_1^{\varepsilon},....,\theta_N^{\varepsilon}\right)'$. The constraints in (29) are estimation restrictions, not assumptions about the underlying true processes for F^0 and Θ^{ε}; see, for example, Stock and Watson (2002), Bai and Ng (2002) and Bai (2009). With these estimates, the regional elasticities are obtained as:

$$\beta_n\left(\hat{F}\right)=\left(X_n'M_{\hat{F}}X_n\right)^{-1}X_n'M_{\hat{F}}Y_n, \tag{30}$$

where $M_A=I_T-A\left(A'A\right)^{-1}A'$, for an arbitrary $T\times h$ matrix A.

Identification in Detail and General Discussion As the proof of the consistency of the estimators makes clear, (29) will in general recover only a rotation of F^0. This means that we will not recover the exact unobserved macro shocks, but just K linear combinations of them. This is acceptable because we only need to control for their variation in (23) to recover the regional elasticities.

As something that will arise repeatedly in the following subsections, identification is essentially achieved because of the system estimation coupled with the heterogeneity assumption that allows the recovery of the unobserved macro shocks. In this case we have (28), which also makes things easier, but, as we will see in the following cases, in the absence of this assumption, and if we observe an instrument for $\varepsilon_{s_{nt}}$, we can still combine these ideas with the instrument to estimate the regional elasticities.

I now give an example of a model that delivers the reduced forms of the previous subsection. To start, suppose F_t represents economic variables that are unobserved to the policy maker, but to which she wants to react nonetheless. In particular, suppose F_t is comprised of the monetary policy shock, v_t, and the TFP shock, a_t, as the models of Section 3 displayed. Suppose as well, for simplicity, that these are independent shocks. Let me focus first in $\varepsilon_{s_{nt}}$. Suppose $\varepsilon_{s_{nt}}$ is set by the policy maker as the solution to:

$$\max_{\varepsilon_{s_{nt}}}\mathbb{E}\left[q\left(\varepsilon_{s_{nt}},a_t,v_t\right)\middle|\,\varphi_{nt}^{a},\varphi_{nt}^{v}\right]-c\left(\varepsilon_{s_{nt}}\right) \tag{31}$$

Here, $q\left(\varepsilon_{s_{nt}}, a_t, v_t\right)$ represents the benefit to the policy maker of choosing $\varepsilon_{s_{nt}}$ for given shocks a_t and v_t. Since the shocks are unobserved by her, she uses the signals:

$$\varphi^a_{nt} = a_t + u^a_{nt} \tag{32}$$

$$\varphi^v_{nt} = v_t + u^v_{nt}, \tag{33}$$

where u^a_{nt} and u^v_{nt} are independent noise shocks, to compute the expected benefit $\mathbb{E}\left[q\left(\varepsilon_{s_{nt}}, a_t, v_t\right) \middle| \varphi^a_{nt}, \varphi^v_{nt}\right]$. $c\left(\varepsilon_{s_{nt}}\right)$ captures the cost for the policy maker of choosing $\varepsilon_{s_{nt}}$. The following result characterizes the solution to (31) for a case that is simple to compute:

Lemma 1. *Suppose in problem* (31) *we have:*

$$q\left(\varepsilon_{s_{nt}}, a_t, v_t\right) = \varepsilon_{s_{nt}} a_t + \varepsilon_{s_{nt}} v_t$$

$$c\left(\varepsilon_{s_{nt}}\right) = \frac{1}{2}\varepsilon^2_{s_{nt}}$$

and also:

$$a_t \sim N\left(\mu_{a,n}, \sigma^2_{a,n}\right)$$

$$v_t \sim N\left(\mu_{v,n}, \sigma^2_{v,n}\right)$$

$$u^a_{nt} \sim N\left(0, \sigma^2_{u^a,n}\right)$$

$$u^v_{nt} \sim N\left(0, \sigma^2_{u^v,n}\right)$$

Then, if, for simplicity, $\mu_{a,n} = \mu_{v,n} = 0$ *, the solution is given by:*

$$\varepsilon_{s_{nt}} = \frac{\tau_{u^a,n}}{\tau_{u^a,n} + \tau_{a,n}} a_t + \frac{\tau_{u^v,n}}{\tau_{u^v,n} + \tau_{v,n}} v_t + \epsilon^s_{nt} \tag{34}$$

where:

$$\tau^{-1}_{x,n} = \sigma^2_{x,n}, \textit{ for } x = a, v, u^a, u^v$$

$$\epsilon^s_{nt} = \frac{\tau_{u^a,n}}{\tau_{u^a,n} + \tau_{a,n}} u^a_{nt} + \frac{\tau_{u^v,n}}{\tau_{u^v,n} + \tau_{v,n}} u^v_{nt}.$$

Suppose as well that s_t is set in the same fashion; a very simple corollary to Lemma 1 derives the same result for s_t:

Corollary 3. *Suppose, analogously to problem* (31)*,* s_t *is the solution to:*

$$\max_{s_t} \mathbb{E}\left[q\left(s_t, a_t, v_t\right) \middle| \varphi^a_{st}, \varphi^v_{st}\right] - c\left(s_t\right)$$

where:

$$q\left(s_t, a_t, v_t\right) = s_t a_t + s_t v_t$$

$$c\left(s_t\right) = \frac{1}{2}s^2_t$$

$$\varphi^a_{st} = a_t + u^a_{st}$$

$$\varphi^v_{st} = v_t + u^v_{st}$$

and also:

$$\begin{aligned} a_t &\sim N\left(\mu_{a,s}, \sigma^2_{a,s}\right) \\ v_t &\sim N\left(\mu_{v,s}, \sigma^2_{v,s}\right) \\ u^a_{st} &\sim N\left(0, \sigma^2_{u^a,s}\right) \\ u^v_{st} &\sim N\left(0, \sigma^2_{u^v,s}\right) \end{aligned}$$

Then, if, for simplicity, $\mu_{a,s} = \mu_{v,s} = 0$, the solution is given by:

$$s_t = \frac{\tau_{u^a,s}}{\tau_{u^a,s} + \tau_{a,s}} a_t + \frac{\tau_{u^v,s}}{\tau_{u^v,s} + \tau_{v,s}} v_t + u^s_t \tag{35}$$

where:

$$\begin{aligned} \tau^{-1}_{x,s} &= \sigma^2_{x,s},\ \textit{for } x = a, v, u^a, u^v \\ u^s_t &= \frac{\tau_{u^a,s}}{\tau_{u^a,s} + \tau_{a,s}} u^a_{st} + \frac{\tau_{u^v,s}}{\tau_{u^v,s} + \tau_{v,s}} u^v_{st}. \end{aligned}$$

As the previous results show, the reduced forms in this model fit into the framework of the previous subsection.

Moreover, the estimation described makes clear that variations of this model also fit the purpose of recovering the regional elasticities. For example, if we added a cost-shifter common to all regions in the cost functions, all the results would still hold. We would be recovering, in (29), more unobserved shocks than needed, but this would not be a problem because we would still be controlling for the variation of F^0_t, which is what we need.

Asymptotic Results The following proposition shows that (30) is consistent under different specifications of (23). Almost all of the results in this paper will be obtained under the simultaneous limit of $N, T \to \infty$; when this is not the case, I will make it explicit. Also, I will use the notation in (20) because it is more compact. For the same reason, I will denote the true loadings in (34) as $\theta^{\varepsilon 0}_k = \left(\theta^{\varepsilon 0}_{1k}, ..., \theta^{\varepsilon 0}_{Nk}\right)'$.

Proposition 13. *Suppose the system of equations is given by (23), (26) and (27), and $\hat{F}$ is the solution in (29). Suppose as well that:*

1. *the errors in (23) and in (27) satisfy Assumption 1 with w_{jnt} independent for all n and t and identically distributed across t,*
2. *F^0 satisfies Assumption 2,*
3. *their loadings in (27) satisfy Assumption 3 and $\theta^{\varepsilon\prime}_k \mathbf{1} = 0\ \forall k$,*
4. *all the regressors and the errors satisfy Assumption 4.*

Then:

$$\left\| \hat{\beta}^{(N)}_n \left(\hat{F}\right) - \beta^{(N)}_n \right\| = o_p\left(1\right). \tag{36}$$

Moreover, if (23) is replaced by (20) for a model that belongs to the RSVC, we also have:

$$\left\| \begin{pmatrix} \hat{\beta}_n^{(N)}(\hat{F}) \\ \hat{\gamma}_n^{(N)}(\hat{F}) \end{pmatrix} - \begin{pmatrix} \beta_n^{(N)} \\ \gamma_n^{(N)} \end{pmatrix} \right\| = o_p(1). \tag{37}$$

5.2 Case 1: $\varepsilon_{s_{nt}} \perp \epsilon_{nt}$ With Model Selection

The previous case is nice because it has the property that the unobserved macro shocks extracted from (34) have two key features. The first is that they include all of the unobserved macro shocks we want to control for in (23). The second is that they do not include all of the unobserved macro shocks entering (35). As soon as we start looking at systems in which the policy maker is reacting to the outcome variable of interest, output in this case, this will no longer be true. This, in turn, means that if (34) has the same unobserved macro shocks as (35), we will not be able to include all of them in (23) because of perfect multicollinearity. The case in this subsection will show precisely this, and it will also show how we can work around this inconvenience.

Denote by F_t^0 the vector of unobserved macro shocks entering equation (23), by F_t^S the vector of unobserved macro shocks present in the entire system of equations, with K being the total number of them, and by $F_t^{S\backslash 0}$ the vector of unobserved macro shocks that belong to F_t^S but not to F_t^0.

Let me first state the assumption we will use in terms of reduced forms, and then I will discuss its main points and what structural equations can lead to them:

Assumption 8. *The reduced forms of s_t and $\varepsilon_{s_{nt}}$ satisfy:*

$$s_t = \theta^{s\prime} F_t^S \tag{38}$$

$$\varepsilon_{s_{nt}} = \theta_n^{\varepsilon\prime} F_t^S + \epsilon_{nt}^s \tag{39}$$

$$\epsilon_{nt}^s \perp \epsilon_{nt}^Y \tag{40}$$

To describe how the estimation works under (23), (38), (39) and (40), first note that again, under (40), if we could get precise estimates of F_t^0, we could run (23) as a time series OLS regression for every region separately to obtain consistent estimates of η_{MG}^n and η_{ML}^n. As before, the problem is how to get at those estimates, and, furthermore, ensuring we have the correct restrictions for these estimates to be "precise enough," so that the estimators of η_{MG}^n and η_{ML}^n are indeed consistent. But there is an added complication here as I mentioned before. Even though we can estimate the unobserved macro shocks entering (39) by the method of asymptotic principal components (under some normalization restrictions), this procedure gives us an estimate of the whole vector $\hat{F}_t^S$. Because of (38), we cannot include all of those in (23) due to perfect multicollinearity. Thus, it might seem that we need to be able to tell apart F_t^0 from F_t^S. In fact, however, it suffices to distinguish a subset of F_t^S that contains F_t^0, because doing so controls for the variation in F_t^0, which is what we need, avoiding the perfect multicollinearity problem. We select this subset by checking which of all possible subsets gives a smaller sum of squared residuals in (23).

I now present in detail the procedure. I assume for the moment that K, the total number of unobserved macro shocks, is known. The first step is getting an estimate of $\hat{F}^S$. As in the previous subsection, this is the solution to problem (29). For the second part, denote as $\hat{F}^{(k)}$ an arbitrary subset of $K-2$ estimated macro shocks, and define:

$$\mathcal{F} := \left\{\hat{F}^{(1)}, \ldots, \hat{F}^{(K^*)}\right\}, \tag{41}$$

where:

$$K^* := \begin{pmatrix} K \\ K-2 \end{pmatrix} = \frac{K!}{(K-2)!2!}. \tag{42}$$

The estimator is the same as that in the previous section, with the difference that there is now a model selection step before the computation the estimator that selects the "best" subset from $\mathcal{F}$. Thus, we can write the estimator as:

$$\beta_n\left(\hat{F}^*\right) = \left(X_n' M_{\hat{F}^*} X_n\right)^{-1} X_n' M_{\hat{F}^*} Y_n \tag{43}$$

where $\hat{F}^*$ is obtained as:

$$\hat{F}^* := \underset{F \in \mathcal{F}}{\arg\min} \frac{1}{NT} \sum_{n=1}^{N} \left(Y_n - X_n \beta_n(F)\right)' M_F \left(Y_n - X_n \beta_n(F)\right) \tag{44}$$

and $M_A = I_T - A\left(A'A\right)^{-1} A'$, for an arbitrary $T \times h$ matrix A.

Remark 1. *This procedure could, in principle, be highly taxing, computationally, when the combinatorial possibilities for the subsets increases. In our applications, it is rare to have more than* 6 *or* 7 *unobserved macro shocks, however, so this is not a problem.*[29]

Remark 2. *Note that* (44) *puts this procedure in the realm of model selection techniques. Hence, in principle, all the problems associated with the (lack of) uniform consistency of the estimators will be present here. See Section 1 in Leeb and Pötscher (2008) for a nice discussion of these issues.*

Identification in Detail and General Discussion As the previous equations make clear, the subsets need to have $K-2$ unobserved macro shocks, and as we will see in the next subsection, this also needs to be true for (23). That is, F_t^0 can have at most $K-2$ shocks from F_t^S. The reason the subsets need to have $K-2$ is that otherwise the objective function in (44) is flat: if F_t^0 had $K-1$, then (44) would converge to the same value for any subset of $K-1$ macro shocks. Note that F_t^0 can have, at most, $K-1$ shocks for the system to be well defined in the sense that otherwise, even if all the unobserved macro shocks were observed by the econometrician, (23) would not be a proper regression equation because of perfect multicollinearity. Thus, we see that we need one more layer for this method to work.

The key identification condition, which we will require for consistency, is related to the reduced form (39). In addition to Assumptions 2 and 3, we will require:

[29] Note, further, that this is independent of the sample size. Even if one had, for example, county-level data in the US on a monthly basis for a large number of years, something that would allow estimation of a high number of unobserved macro shocks, it is unlikely that the application would require such estimation, at least theoretically. The reason is that it is usually very hard to think of more than a dozen candidates in a given application. Nonetheless, with a very large sample size, we could test more precisely the number of them present in the sample, and, ultimately, we could empirically guide the choice. I will elaborate more on these issues later in this section.

Assumption 9. *$\Sigma_F\Sigma_\theta$ is a block diagonal matrix on the included $\left(F_t^0\right)$ and excluded $\left(F_t^{S\setminus 0}\right)$ unobserved macro shocks, with each block a diagonalizable matrix.*

The importance of Assumption 9 is that it implies (44) will select, asymptotically, a subset that correctly captures the variation in F_t^0. However, it also implies that this variation will be captured by linear combinations of the elements in F_t^0, rather than by F_t^0 itself. To have a better idea of this assumption, it is useful to think about a sufficient condition for it, which is to ask for Σ_F and Σ_θ to themselves be block diagonal on included and excluded unobserved macro shocks. Suppose the loadings as well as the shocks have mean zero to make the discussion easier. This condition then says that the included shocks are allowed to be correlated among them, but should be orthogonal to the excluded ones, and vice versa. And, in terms of loadings, it says that observing how the policy maker reacts to an included shock in principle should give no information on how she would react to an excluded one, although it could give information on how she would react to another included one (and vice versa).

I now discuss which structural equations can give rise to (38), (39) and (40). Stating the assumptions in terms of reduced forms is nice in that we can capture different structural systems at once. However, knowing which systems exactly map into the assumptions made is important for understanding the scope of the method. It will also help us understand what the condition on $\Sigma_F\Sigma_\theta$ is about. The idea is that both s_t and $\varepsilon_{s_{nt}}$ are responding to an economic performance index. In particular, suppose:

$$s_t = \theta_F^{s\prime} F_t^S + \theta_y^s \overline{Y}_t \tag{45}$$

$$\varepsilon_{s_{nt}} = \theta_{nF}^{\varepsilon\prime} F_t^S + \theta_{ny}^{\varepsilon} \overline{Y}_t + \epsilon_t^{s_n} \tag{46}$$

where some of the entries in the vectors $\theta_F^{s\prime}$ or $\theta_{nF}^{\varepsilon\prime}$ could be zero so that F_t^0 could appear in the reduced forms because of the dependence of (42) and (46) on $\overline{Y}_t$. The restrictions discussed before now translate into restrictions on $\theta_{ny}^{\varepsilon}$ and $\theta_{nF}^{\varepsilon\prime}$. In general, there are many choices that satisfy them.

Asymptotic Results The following proposition shows that (43) is consistent under (23). For the following proposition, let $(k)^*$ denote a subset of unobserved macro shocks that includes F^0 and (k) an arbitrary subset.

Proposition 14. *Suppose the system of equations is given by (23), (38) and (39). Suppose as well that:*

1. *the number of unobserved macro shocks entering (23) is at most $K-2$,*
2. *the errors in (23) and in (39) satisfy Assumption 1 with w_{jnt} independent for all n and t and identically distributed across t,*
3. *the regressors $\mathbb{X}_n$ satisfy Assumption 4,*
4. *F^S satisfy Assumption 2,*
5. *θ^ε satisfy Assumption 3 and $\theta_k^{\varepsilon\prime}\mathbf{1} = \theta_k^{\varepsilon\prime}\eta_{ML} = 0\ \forall k$, $\|\lambda_n\| \leq M$,*
6. *Assumption 9 holds,*

7. *if* $(k) \neq (k)^*$, $\mathbb{E}\left[u_n^{(k)'} M_{\left[X_n : F^S H^{(k)}\right]} u_n^{(k)}\right] > O(T) > 0$,[30] *where* $u_n^{(k)} = F^0\lambda_n - \mathcal{P}\left(F^0\lambda_n \middle| \left[X_n : F^S H^{(k)}\right]\right)$ *is a projection error, and* $\hat{u}_{nt}^{(k)2}$ *stationary with absolutely summable autocovariances, where* $\hat{u}_n^{(k)} = F^0\lambda_n - \hat{\mathcal{P}}\left(F^0\lambda_n \middle| \left[X_n : F^S H^{(k)}\right]\right)$.

Then:

$$\left\|\hat{\beta}_n^{(N)}\left(\hat{F}^*\right) - \beta_n^{(N)}\right\| = o_p(1). \tag{47}$$

5.3 Case 2: Instrument for $\varepsilon_{s_{nt}}$

This case is the first to analyze a situation in which $\varepsilon_{s_{nt}}$ is not only correlated with F in (23) but also potentially correlated with ϵ_{nt}^Y. The main assumption in this case is that we observe an instrument $\varepsilon_{Z_{nt}}$ for $\varepsilon_{s_{nt}}$. In principle, one might think that $\varepsilon_{Z_{nt}}$ should be orthogonal to F on top of ϵ_{nt}^Y. As the previous cases show, this need not be true to obtain consistent estimators, since we are also controlling for $\hat{F}$. Controlling for $\hat{F}$ is necessary because here too it is crucial in order to recover the elasticities on s_t. This case also has the nice feature that there is no need for a model selection step as in Case 1.

The main assumption here will be:

Assumption 10. *The reduced forms of* s_t *and* $\varepsilon_{s_{nt}}$ *satisfy:*

$$s_t = \theta^{s'} F_t^S \tag{48}$$

$$\varepsilon_{s_{nt}} = \theta_n^{\varepsilon'} F_t^{\varepsilon} + \theta_n^{\varepsilon Z}\varepsilon_{Z_{nt}} + \epsilon_{nt}^s \tag{49}$$

$$\varepsilon_{Z_{nt}} \perp \epsilon_{nt}^Y. \tag{50}$$

Here F_t^{ε} is used to denote a proper subset of F_t^S. The first main difference from the previous case is that ϵ_{nt}^s is not required to be orthogonal to ϵ_{nt}^Y. Second, although the reduced form for $\varepsilon_{s_{nt}}$ is allowed to depend on the whole vector F_t^S, this, at least in part, has to be mediated by the instrument $\varepsilon_{Z_{nt}}$. Moreover, note that the instrument is allowed to be correlated with all the unobserved macro shocks; it just has to be orthogonal to ϵ_{nt}^Y.

The general idea here is to again use the whole system to get at the estimates $\hat{F}_t^0$ but combine these estimates with $\varepsilon_{Z_{nt}}$ in order to address the correlation of $\varepsilon_{s_{nt}}$ with ϵ_{nt}^Y. In detail, this means that the estimates of the unobserved macro shocks come from:

$$\left(\hat{F}, \left(\hat{\theta}_n^{\varepsilon'}, \hat{\theta}_n^{\varepsilon Z}\right)_{n=1}^N\right) := \underset{\left(F, \left(\theta_n^{\varepsilon'}, \theta_n^{\varepsilon Z}\right)_{n=1}^N\right)}{\arg\min} \frac{1}{NT}\sum_{n=1}^N \left(\varepsilon_{s_n} - F\theta_n^{\varepsilon} - \theta_n^{\varepsilon Z}\varepsilon_{Z_n}\right)'\left(\varepsilon_{s_n} - F\theta_n^{\varepsilon} - \theta_n^{\varepsilon Z}\varepsilon_{Z_n}\right) \tag{51}$$

$$s.t.: \begin{cases} \frac{F'F}{T} = I_K \\ \Theta^{\varepsilon'}\Theta^{\varepsilon}\ diagonal \end{cases}$$

where $\Theta^{\varepsilon} = \left(\theta_1^{\varepsilon}, ..., \theta_N^{\varepsilon}\right)'$. With the estimates $\hat{F}$ from (51), the regional elasticities are estimated as:

$$\beta_n\left(\hat{F}\right) = \left(X_n' M_{\hat{F}} X_n\right)^{-1} X_n' M_{\hat{F}} Y_n, \tag{52}$$

[30] The $O(T)$ term should be $O^{(k)}(T)$ to make clear that $O^{(k)}(T)$ is an $O(T)$ term that depends on the subset (k), but we omit the (k) in the superscript for a cleaner notation.

where, in the regressors X_n, $\varepsilon_{s_{nt}}$ is replaced by $\varepsilon_{Z_{nt}}$. That is, in this case in (23), the regressors $X_n = (s_t, \varepsilon_{Z_{nt}})$. See Ando and Bai (2015) for a detailed analysis of (51).

Identification in Detail and General Discussion As (51) makes clear, the key to identification is that F_t^0 should be a subset of F_t^{ε}. If we are in the general framework of Case 1, where a reduced form like (39) ultimately applies, if F_t^0 is not a subset of F_t^{ε} we are back in Case 1 and need to apply the ideas of the model selection technique.

Furthermore, let me note that (51) allows ε_{Z_n} to be correlated with F_t^{ε}. That is, the idea that an instrument for $\varepsilon_{s_{nt}}$ should be uncorrelated with ϵ_{nt}^Y and F_t^0 is too strict. The crucial property we need is orthogonality with respect to ϵ_{nt}^Y.

This case is suitable for scenarios in which the econometrician observes an intermediate cause for at least one of the unobserved macro shocks, provided the conditions described before are met. Let me give an illustrative example with a structural system that gives rise to (48), (49) and (50). The following result illustrates an extension of the environment in Subsection 5.1:

Corollary 4. *Suppose the environment is the same as in Lemma 1 and Corollary 3 with the difference that now in setting $\varepsilon_{s_{nt}}$ the policy maker is also trying to target the signal received for v_t by the policy maker setting s_t, so that now the problem is:*

$$\max_{\varepsilon_{s_{nt}}} \mathbb{E}\left[q\left(\varepsilon_{s_{nt}}, a_t, v_t, u_{st}^v\right) \middle| \varphi_{nt}^a, \varphi_{nt}^v, \varphi_{nt}^{u_s^v} \right] - c\left(\varepsilon_{s_{nt}}\right)$$

where:

$$\begin{aligned} q\left(\varepsilon_{s_{nt}}, a_t, v_t, u_{st}^v\right) &= \varepsilon_{s_{nt}} a_t + \varepsilon_{s_{nt}} v_t + \varepsilon_{s_{nt}} u_{st}^v \\ \varphi_{nt}^{u_s^v} &= u_{st}^v + u_{nt}^{u_s^v} \\ u_{st}^v &\sim N\left(0, \sigma_{u^v,s}^2\right) \\ u_{nt}^{u_s^v} &\sim N\left(0, \sigma_{u^{u_s^v},n}^2\right), \end{aligned}$$

where as before $u_{nt}^{u_s^v}$ is an independent noise shock. The rest of the assumptions are maintained.

Then, the solution to this problem is given by:

$$\varepsilon_{s_{nt}} = \frac{\tau_{u^a,n}}{\tau_{u^a,n} + \tau_{a,n}} a_t + \frac{\tau_{u^v,n}}{\tau_{u^v,n} + \tau_{v,n}} v_t + \frac{\tau_{u^{u_s^v},n}}{\tau_{u^{u_s^v},n} + \tau_{u^v,s}} \varphi_{nt}^{u_s^v} + \epsilon_{nt}^s \tag{53}$$

where:

$$\begin{aligned} \tau_{u^{u_s^v},n}^{-1} &= \sigma_{u^{u_s^v},n}^2, \tau_{u^v,s}^{-1} = \sigma_{u^v,s}^2 \\ \epsilon_{nt}^s &= \frac{\tau_{u^a,n}}{\tau_{u^a,n} + \tau_{a,n}} u_{nt}^a + \frac{\tau_{u^v,n}}{\tau_{u^v,n} + \tau_{v,n}} u_{nt}^v. \end{aligned}$$

As Corollary 4 shows with (53), if the signal $\varphi_{nt}^{u_s^v}$ is observed by the econometrician, we are in the territory of (48), (49) and (50).

Asymptotic Results The following proposition shows that (52) is consistent:

Proposition 15. *Suppose the system of equations is given by* (23), (48) *and* (49). *Suppose as well that:*

1. *the errors in* (23) *after replacing the reduced form* (49) *and the errors in* (49) *satisfy assumption* 1 *with* w_{jnt} *independent for all* n *and* t *and identically distributed across* t*,*

2. *the regressors* $\mathbb{X}_n$ *satisfy Assumption* 4*,*

3. F^S *satisfies Assumption* 2 *and* F_t^{ε} *spans* F_t^0*,*

4. θ^{ε} *satisfy Assumption* 3 *and* $\theta_k^{\varepsilon'}\mathbf{1} = \theta_k^{\varepsilon'}\eta_{ML} = 0\ \forall k$.

5. *Assumption* 6 *holds for* (49).

Then:

$$\left\|\hat{\beta}_n^{(N)}\left(\hat{F}\right) - \beta_n^{(N)}\right\| = o_p\left(1\right). \tag{54}$$

Moreover, if (23) *is replaced by* (20) *for a model that belongs to the RSVC, we also have:*

$$\left\|\begin{pmatrix} \hat{\beta}_n^{(N)}\left(\hat{F}\right) \\ \hat{\gamma}_n^{(N)}\left(\hat{F}\right) \end{pmatrix} - \begin{pmatrix} \beta_n^{(N)} \\ \gamma_n^{(N)} \end{pmatrix}\right\| = o_p\left(1\right). \tag{55}$$

5.4 Case 3: Instrument for $\varepsilon_{s_{nt}}$ and True SEMs

Up to this point, almost all of the scenarios analyzed have allowed s_t to suffer from simultaneous equation bias. However, the region-specific policy $\varepsilon_{s_{nt}}$ suffered only from omitted variables bias. Hence, it makes sense to ask what happens in the event that $\varepsilon_{s_{nt}}$ suffers from true simultaneous equation bias as well. As in the previous subsection, we continue to assume there is an observed instrument $\varepsilon_{Z_{nt}}$, and thus we increase the threat to identification but maintain the presence of an instrument. As we will see shortly, the complexity of the procedure is greatly increased when dealing with this type of bias as well.

The case in this subsection treats explicitly the situation in which the endogeneity of $\varepsilon_{s_{nt}}$ is due to the policy maker responding directly to Y_{nt}. As the first case illustrated, the previous methods can accommodate situations in which the response is with respect to some index of economic performance for the economy as a whole, like $\overline{Y}_t$. Responding to Y_{nt} implies that $\varepsilon_{s_{nt}}$ will necessary have ϵ_{nt}^{Y} in its reduced form for any sample size, and thus complicates matters further.

As I mentioned in the introduction to this section, different specifications of the ones seen so far might be suitable for different applications. Responding to Y_{nt} is a version of the main concern we typically have when thinking about fiscal multipliers: if the fiscal authority is responding to the economic performance of the region to which belongs, then Y_{nt} and $\varepsilon_{s_{nt}}$ are simultaneously determined, and it is difficult to understand what causal effect, if any, one is recovering by running regressions that involve spending and output.

To recover the regional elasticities in this case, we will apply both panel data methods related to the single-equation interactive-effects estimators and strategies related to the typical SVAR approaches in pure time series models. However, the requirements in terms of the time series methods will be lower than the usual applied ones, and thus one way to see the whole framework is as the advantage that regional variation brings to the usual methods, in the sense of lowering the identification assumptions we need to make in terms of ordering of shocks, short-run restrictions, long-run restrictions, etc.

Since the procedure is a bit more involved than in the previous cases, I conduct the analysis with a much more specific notation. However, once the main point is understood, it is straightforward to see how it generalizes. Suppose then that the whole system of equations now becomes:

$$Y_{nt} = \beta_{1n} s_t + \beta_{2n} \varepsilon_{s_{nt}} + \delta_{1n} F_{1t} + \delta_{2n} F_{2t} + \omega^Y_{nt} \tag{56}$$

$$s_t = \theta_{sY} \overline{Y}_t + \theta_{sZ} Z_t + u^s_t \tag{57}$$

$$\varepsilon_{s_{nt}} = \theta_{Yn} \left(Y_{nt} - \overline{Y}_t \right) + \theta_{Zn} \varepsilon_{Z_{nt}} + \theta_{u^s n} u^s_t + \omega^s_{nt} \tag{58}$$

$$Z_t = \kappa_s s_t + \kappa_Y \overline{Y}_t + u^Z_t \tag{59}$$

$$\varepsilon_{Z_{nt}} = \kappa_{0n} F_{1t} + \kappa_{1n} F_{2t} + \kappa_{u^s,n} u^s_t + \kappa_{u^Z,n} u^Z_t + \omega^Z_{nt} \tag{60}$$

where as usual $\overline{x}_t := \frac{1}{N} \sum_n x_{nt}$ for an arbitrary variable x_{nt}. In the system (56)-(60), Z_t denotes the common component of the instrument $\varepsilon_{Z_{nt}}$. The idea here is that we observe not only the region-specific instrument $\varepsilon_{Z_{nt}}$ but also the region-invariant component Z_t. However, the system (56)-(60) implies that Z_t cannot be used as an instrument for s_t. That is, the point of being explicit about Z_t in the system is to rule out this possibility. Furthermore, the presence of (60) illustrates that $\varepsilon_{Z_{nt}}$ is allowed to be correlated with all of the unobserved macro shocks. The key requirement I will impose is that $\omega^Z_{nt} \perp \omega^Y_{nt}$. Thus, although $\varepsilon_{Z_{nt}}$ may be orthogonal to ω^Y_{nt}, both the common component Z_t, and $\varepsilon_{Z_{nt}}$ may be highly endogenous with respect to the unobserved macro shocks; Z_t is allowed to depend explicitly on s_t and $\overline{Y}_t$, and s_t on $\overline{Y}_t$ and Z_t.

To further illustrate these points, note that the system in (56)-(60) displays, at the aggregate level, a familiar simultaneous equation model, a SVAR without lags (for the moment):

$$\overline{Y}_t = \overline{\beta_1} s_t + \overline{\delta_1} F_{1t} + \overline{\delta_2} F_{2t} + \overline{\beta_2 \varepsilon_{s_t}} + \overline{\omega^Y_t} \tag{61}$$

$$s_t = \theta_{sY} \overline{Y}_t + \theta_{sZ} Z_t + u^s_t \tag{62}$$

$$Z_t = \kappa_Y \overline{Y}_t + \kappa_s s_t + u^Z_t. \tag{63}$$

Note that if we wanted to identify this upper layer of (56)-(60) given by (61)-(63) with, for example, short-run restrictions, we would need two more such restrictions. The restriction we have comes from the exclusion of Z_t from (61), and such an exclusion, as usual, should be argued on the basis of economic theory. In this paper, I will argue for it in Section 6 on the basis of the models of Section 3 and the nature of the variable Z_t I use. If one could argue, for example, for $\kappa_Y = \kappa_s = 0$, the system (61)-(63) would be identified because we would have an instrument for s_t in our hands, Z_t. And with Z_t and $\varepsilon_{Z_{nt}}$ we could instrument both s_t and $\varepsilon_{s_{nt}}$ in (56) to recover the regional elasticities of interest. Hence, one of the purposes of this case is to show that even without arguing for any extra exclusion restrictions in (61)-(63), the whole system in (56)-(60) can be used to recover the regional elasticities of interest. That is, we can use every equation of every variable, for every region and period, to do so.

Moreover, let me note here that as (61) illustrates, (56) gives a microfoundation of an "aggregate GDP shock": is a mixture of the unobserved macro shocks entering (56), as in $u^Y_t := \overline{\delta_1} F_{1t} + \overline{\delta_2} F_{2t} + \overline{\beta_2 \varepsilon_{s_t}} + \overline{\omega^Y_t}$.

I present first the main assumptions used here and then I will present the procedure in detail. The assumptions are:

Assumption 11. *In the regional system given by* (56)-(60)*:*

$$u_t^s \perp u_t^Z \perp F_{1t}, F_{2t} \tag{64}$$

$$\omega_{nt}^Z \perp \omega_{nt}^Y. \tag{65}$$

Note that (64) is just the usual assumption made in a SVAR context, that is, that the unobserved macro shocks of different variables are orthogonal to each other. (65) is the assumption I already mentioned.

Like every strategy presented in this paper, the estimation works by first using the whole system of equations to obtain an estimate of the unobserved macro shocks, and then running equation (56) to obtain the regional elasticities of interest. For maximum clarity, I present the estimation in a series of steps. I then explain the intuition and general idea:

1. Using Y_{nt} and $\varepsilon_{Z_{nt}}$ we solve:

   (66)
   $$\left(\hat{F}, \left(\hat{\phi}_n^{y\prime}, \hat{\phi}_{Zn}^{y}\right)_{n=1}^{N}\right) := \underset{\left(F, \left(\phi_n^{y\prime}, \phi_{Zn}^{y}\right)_{n=1}^{N}\right)}{\arg\min} \frac{1}{NT}\sum_{n=1}^{N}\left(Y_n - F\phi_n^y - \phi_{Zn}^y \varepsilon_{Z_n}\right)^{\prime}\left(Y_n - F\phi_n^y - \phi_{Zn}^y \varepsilon_{Z_n}\right)$$
   $$s.t. : \begin{cases} \frac{F^{\prime}F}{T} = I_K \\ \Phi^{y\prime}\Phi^{y} \ diagonal \end{cases}$$

   where $\Phi^y = \left(\phi_1^y, ..., \phi_N^y\right)^{\prime}$.

2. Using $\varepsilon_{s_{nt}}$ and $\varepsilon_{Z_{nt}}$ we solve:

   (67)
   $$\left(\hat{u}^s, \left(\hat{\phi}_n^{\varepsilon}, \hat{\phi}_{Zn}^{\varepsilon}, \hat{\phi}_{\Omega n}^{\varepsilon}\right)_{n=1}^{N}\right) := \underset{\left(u^s, \left(\phi_n^{\varepsilon}, \phi_{Zn}^{\varepsilon}, \phi_{\Omega n}^{\varepsilon}\right)_{n=1}^{N}\right)}{\arg\min} \frac{1}{NT}\sum_{n=1}^{N}\sum_{t=1}^{T}\left(\varepsilon_{s_{nt}} - \phi_n^{\varepsilon}u_t^s - \phi_{Zn}^{\varepsilon}\varepsilon_{Z_{nt}} - \phi_{\Omega n}^{\varepsilon}\hat{F}_t^{\prime}\hat{\phi}_n^y\right)^2$$
   $$s.t. : \begin{cases} \frac{u^{s\prime}u^s}{T} = 1 \\ \Phi^{\varepsilon\prime}\Phi^{\varepsilon} \ diagonal \end{cases}$$

   where $\Phi^{\varepsilon} = \left(\phi_1^{\varepsilon}, ..., \phi_N^{\varepsilon}\right)^{\prime}$.

3. Using $\hat{u}_t^s$ to instrument s_t in (61) we get the estimate of the residuals $\hat{r}_t$, where $r_t = \overline{\delta_1}F_{1t} + \overline{\delta_2}F_{2t} + \overline{\beta_2\varepsilon_{s_t}} + \overline{\omega_t^Y}$.

4. Using $\hat{u}_t^s$ to instrument s_t, and $\hat{r}_t$ to instrument $\overline{Y}_t$, we get the estimate $\hat{u}_t^Z$ of the residuals from (63).

5. Regressing $\hat{F}_{kt}$ on $\hat{u}_t^Z$ for $k = 1, 2, 3, 4$, we get the estimate of the residuals, which we denote by $\hat{F}_{kt}^R$.

6. We obtain the estimates of the micro-global elasticities as the OLS coefficient on s_t of:

$$\hat{\beta}_n = \left(X_n' X_n\right)^{-1} X_n' Y_n \tag{68}$$

with $X_{nt} = \left(s_t, \varepsilon_{Z_{nt}}, \hat{F}_{1t}^{RC}, \hat{F}_{2t}^{RC}, \hat{F}_{3t}^{RC}\right)'$ where $\hat{F}_{kt}^{RC}$, $k = 1, 2, 3$, is an arbitrary linear combination with strictly positive weights of $\hat{F}_{1t}^R, \hat{F}_{2t}^R, \hat{F}_{3t}^R, \hat{F}_{4t}^R$.

Identification in Detail and General Discussion Let me first elaborate on the general idea of (68) that will make all of the steps in the estimation straightforward. One way to look at them is the following: the first step gives as an estimate of $\vec{F}_t = \left(F_t, u_t^s, u_t^Z\right)$. However, since there are no restrictions on $\Sigma_{\vec{F}}$ or on the equivalent matrix for the loadings, step 1 actually gives four linear combinations of $\vec{F}_t$. That is, it does not give us the true $\vec{F}_t$ but a rotation of itself. Using all of them in a regional OLS time series regression of Y_{nt} on s_t and $\varepsilon_{Z_{nt}}$ is again not possible because of perfect multicollinearity in the limit. Note that if this were not the case, using the rotation that controls for the variation of all of them would be fine, since this is all what we need to recover the parameters that interest us. Nonetheless, if we could separate $\left(F_t, u_t^s\right)$ from $\vec{F}_t$, we could control for that separation in the regression of Y_{nt} on s_t and $\varepsilon_{Z_{nt}}$, which is what we actually need. Thus, steps 2 to 5 are designed to achieve just that, i.e., are designed to use the information in all the equations to separate $\left(F_t, u_t^s\right)$ from $\vec{F}_t$. Finally, step 6 computes the regional elasticities using the instrument and the separated unobserved macro shocks.

I now discuss the key identification assumptions by diving into each step in great detail. This will also help clarify every part of the estimation. As I already mentioned, step 1 gives the estimate of a rotation of $\vec{F}_t$. Note once again that if we were able to pull apart the unobserved macro shocks at this point, we could jump immediately to step 6. Since this is not possible unfortunately, we must transit through 2-5. Note that in addition to this rotation, step 1 also gives a consistent estimate of $\hat{F}_t' \hat{\phi}_n^y$, which will be crucial for pulling apart the unobserved macro shocks.

The second step uses this $\hat{F}_t' \hat{\phi}_n^y$ as a regressor in another single-equation interactive effects estimation, along with the instrument $\varepsilon_{Z_{nt}}$, with the goal of getting at an estimate of u_t^s. Here is where the first key assumption comes in. For this step to work, we need, as always, Assumption (6) to hold. This means in particular that for an arbitrary candidate u^s, with $\frac{u^{s\prime} u^s}{T} = I_K$, in (67), the residuals of regressing $\hat{F}_t' \hat{\phi}_n^y$ on the candidate u^s should display sufficient regional heterogeneity. Thus, in our case this is satisfied because of a key assumption that is coded in the exclusion restrictions of system (56)-(60). We can describe it in this manner: the reason $\varepsilon_{s_{nt}}$ is endogenous in (56), and the reason its reduced form depends on all of the unobserved macro shocks, is that $\varepsilon_{s_{nt}}$ responds to Y_{nt}, and thus its correlation with the right-hand side variables of (56) comes exclusively from this interaction. That is, if we were to assume that $\varepsilon_{s_{nt}}$ responds not only to Y_{nt} but also to all of the unobserved macro shocks, this assumption would be violated. The intuition of the exclusion restriction is that $\varepsilon_{s_{nt}}$ is correlated with F_{1t} and F_{2t} not because the policy maker cares about those shocks (as in some of the examples in the previous cases), but because it wants to target Y_{nt}, and through this behavior $\varepsilon_{s_{nt}}$ ends up being correlated with F_{1t} and F_{2t}. Of course, as usual, the plausibility of this exclusion restriction should be judged in each application. Nonetheless, for the fiscal spending application of this paper, it seems to satisfy our usual main concern that the policy maker is responding to economic activity and this response drives the correlation of spending with the unobserved macro shocks.

Steps 3 and 4 are just the usual IV treatment of exclusion restrictions, short-run restrictions, in "world ordering" of SVARs. The goal of these steps is to reach $\hat{u}_t^Z$. As I mentioned before, if it weren't for steps 1 and 2, we would need more exclusion restrictions in the system to get at $\hat{u}_t^Z$.

Step 5 is just the actual separation of u_t^Z from $\vec{F}_t$. There is nothing special about this step apart from the separation. Step 6 merely computes the actual estimator.

Note that assumption 11 is also important and used in all of the steps just described: for example, in Step 5 to separate the unobserved macro shocks, and in steps 1 and 2 to compute the single-equation interactive-effects estimators. Moreover, see Ando and Bai (2015) for a detailed analysis of problems like (66) and (67).

Asymptotic Results The following proposition shows that (68) is consistent. I refer to ω_{nt}^x, for $x = Y, s, Z$, as unobserved regional shocks.

Proposition 16. *Suppose the system of equations is given by* (56)-(60). *Suppose as well that:*

1. ω_{nt}^x *is independent of all of the unobserved macro shocks, unobserved regional shocks and coefficients in the system, and identically distributed across* t *with* $\mathbb{E}\left[\omega_{nt}^x\right] = 0$ *and* $\mathbb{E}\left[\left|\omega_{nt}^x\right|^{16}\right] < C$ *for* $x = Y, s, Z$,

2. *Assumption* 11 *holds,*

3. *the regressors* $\mathbb{X}_n$ *satisfy Assumption* 4,

4. $\vec{F}$ *satisfies Assumption* 2,

5. *the loadings on the reduced form of* $Y_{nt} - \overline{Y}_t$ *satisfy Assumption* 3, $\theta_{Yn} = \theta_Y$, *and* $\sum_{n=1}^N \theta_{u^s n} = \sum_{n=1}^N \theta_{u^s n} \frac{\beta_{2n}}{1-\delta_Y \beta_{2n}} = \sum_{n=1}^N \theta_{Zn} x_n = \sum_{n=1}^N \theta_{Zn} \frac{\beta_{2n}}{1-\delta_Y \beta_{2n}} x_n = 0, x_n = \kappa_{0n}, \kappa_{1n}, \kappa_{u^s,n}, \kappa_{u^Z,n}$,

6. $\varepsilon_{Z_{nt}}$ *and the unobserved macro shocks in* $\vec{F}_t$ *satisfy Assumptions* 4 *and* 6,

7. Ω_{nt}, $\varepsilon_{Z_{nt}}$ *and the unobserved macro shock* u_t^s *satisfy Assumptions* 4 *and* 6, *where* Ω_{nt} *is the inner product of the unobserved macro shocks and their loadings in the reduced form of* $Y_{nt} - \overline{Y}_t$.

Then the estimator in (68) *satisfies:*

$$\left\|\hat{\beta}_n^{(N)} - \beta_n^{(N)}\right\| = o_p(1). \tag{69}$$

5.5 Monte Carlo Simulations

In this subsection, I present results on Monte Carlo simulations to assess the performance of the different estimators for different sample sizes. Except for case 0, I present simulations for all of them. For cases 1 and 2, I generate data according to:

$$y_{nt} = \eta_{MG}^{n} s_t + \eta_{ML}^{n} \varepsilon_{s_{nt}} + \delta_1^n \xi_{1t} + \delta_2^n \xi_{2t} + u_{nt} \tag{70}$$

$$s_t = \xi_{1t} + \xi_{2t} + \xi_{3t} + \xi_{4t} \tag{71}$$

$$\varepsilon_{Z_{nt}} = \kappa_{1n}\xi_{1t} + \kappa_{2n}\xi_{2t} + \kappa_{3n}\xi_{3t} + \kappa_{4n}\xi_{4t} + v_{nt} \tag{72}$$

and:

$$\varepsilon_{s_{nt}} = \begin{cases} \theta_{1n}\xi_{1t} + \theta_{2n}\xi_{2t} + \theta_{3n}\xi_{3t} + \theta_{4n}\xi_{4t} + \omega_{nt} & (Case\ 1) \\ \theta_{1n}\xi_{1t} + \theta_{2n}\xi_{2t} + \theta_{3n}\xi_{3t} + \theta_{4n}\varepsilon_{Z_{nt}} + 0.5u_{nt} + \omega_{nt} & (Case\ 2) \end{cases}$$

where: $\xi_{kt} \sim N(0,1)$, $\delta_{1n} \sim 0.9 + N(0,1)$, $\delta_{2n} \sim 1.3 + N(0,1)$, $\eta_{MG}^n \sim 0.8 + N(0,1)$, $\eta_{ML}^n \sim 0.3 + N(0,1)$, $u_{nt} \sim N(0,1)$, $v_{nt} \sim N(0,1)$, $\omega_{nt} \sim N(0,1)$, $\kappa's$ and θ's $\sim N(0,\sigma)$ for varying $\sigma's$ $(0.5, 1, 1.5, \ldots)$.

For case 3, I simulate the model (56)-(60) with $F_{kt} \sim N(0,1)$, $u_t^s \sim N\left(0, \frac{9}{4}\right)$, $u_t^Z \sim N(0,1)$, $\gamma \sim U_{[-0.1,0.1]}$, $\delta_{1n} \sim U_{[0,\frac{1}{4}]} + \gamma$, $\delta_{2n} \sim 0.2 + U_{[0,0.4]} + \gamma$, $\beta_{1n} \sim 0.8 + U_{[-\frac{1}{4},\frac{1}{4}]} + \gamma$, $\beta_{2n} \sim 0.1 + U_{[-\frac{1}{8},\frac{1}{8}]} + \frac{1}{4}\gamma$, $\omega_{nt}^Y \sim U_{[-\frac{1}{2},\frac{1}{2}]}$, $\omega_{nt}^Z \sim N(0,1)$, $\omega_{nt}^s \sim U_{[-\frac{1}{2},\frac{1}{2}]}$, $\kappa_n' s \sim N(0,1)$, $\theta_Y = 0.3$, $\theta_{Zn} \sim U_{[-\frac{1}{4},\frac{1}{4}]}$, $\theta_{u^s n} \sim U_{[-\frac{1}{2},\frac{1}{2}]}$, $\kappa_s = 1.54$, $\kappa_Y = 0.7$, $\theta_{sY} = 0.4$ and $\theta_{sZ} = 2.1$.

The results for the simulations at different sample sizes are presented in figures 1 through 9. All the figures show the results we would obtain if we were to instead run a simple OLS time series regression for each region to gain a sense of the magnitude of the bias.

Figures 1 through 3 show the results for the macro elasticity. As the three figures show, the estimators perform very well even in a case like this, in which, as the OLS points illustrate, there is a huge underlying bias.

Since we are also interested in the micro-global elasticities, Figures 8 and 9, in Appendix B, show the analog of Figures 1 through 3 for these elasticities in cases 1 and 2. As expected, the standard errors are much higher for the micro-global elasticities than for the macro elasticity, since in the latter we are pulling together all of the micro-global elasticities. However, Figures 8 and 9 also show a good performance of the estimators even in small samples.

5.6 Further Issues: Endogeneity Causes, Identification Assumptions, Inference and Choosing K

Endogeneity Causes and Identification Assumptions Before turning into the empirical application, I want to spend a few paragraphs explaining different aspects of the identification assumptions. Although I have not emphasized them so far, it is important to keep them in mind.

For context, let me reiterate that the first and perhaps most important assumption is the heterogeneity of regions, which is reflected, for example, in Assumption 3.

However, another important assumption is condition (1.6) in Assumption 1. Condition (1.6) says that the errors in our econometric model are independent of the regressors and the unobserved macro shocks. This assumption could easily be weakened to absence of correlation among them. In terms of the regional RBC model of Subsection 3.2, for example, this means that the deviation from the technology shock is uncorrelated with aggregate government spending and with the aggregate technology shock.

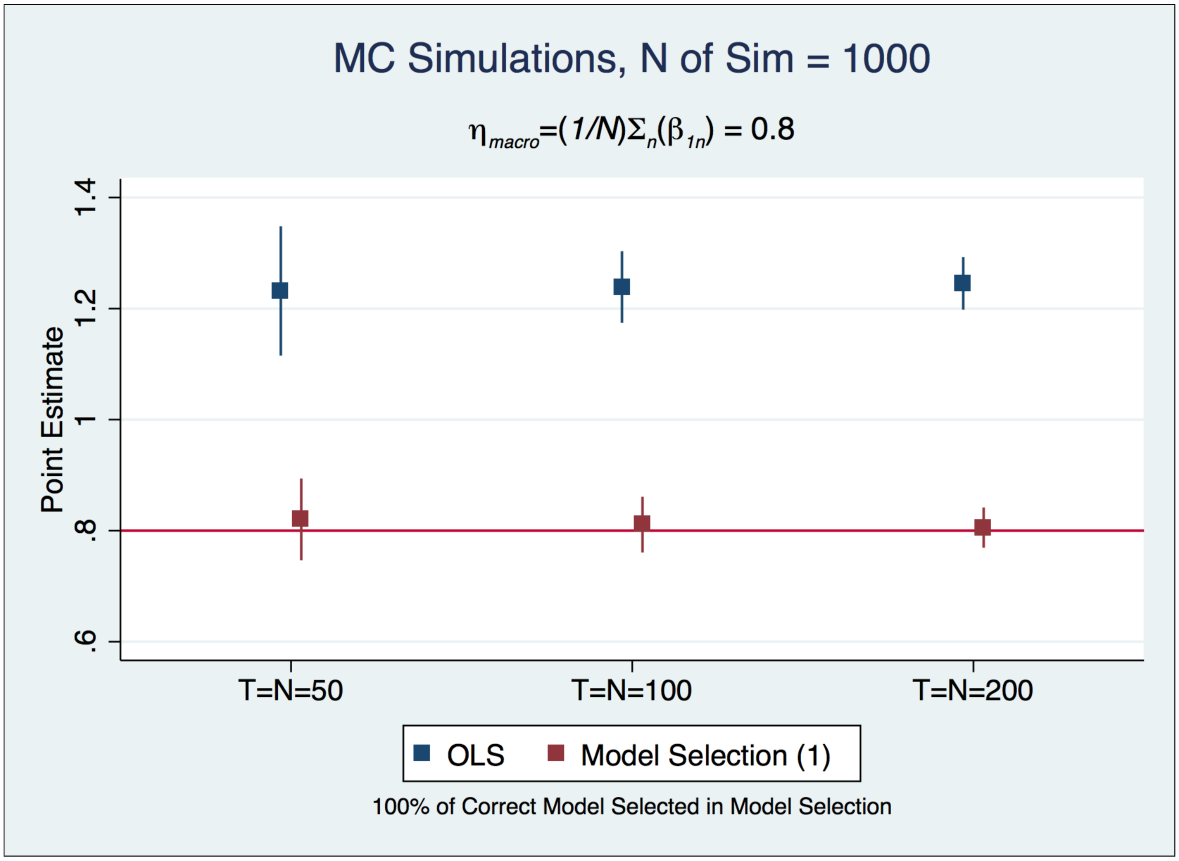

Figure 1: This figure shows the Monte Carlo simulations for the macro elasticity in Case 1.

Although there are many ways this assumption could be violated,[31] two particular concerns arise naturally in our framework. The first is an omitted variables concern, and the second is a reverse causation concern. I now discuss each in turn and give examples and intuition of when condition (1.6) holds.

With respect to the omitted variables concern, and given that we are thinking about a region-specific error in the presence of region-invariant regressors, the most natural starting point to think about this may be to consider that there is a variable, let's call it ζ_t, that is affecting both u_{nt} and u_{jt}, and also s_t. This makes u_{nt} and u_{jt} comove, and it also makes the pairs $(u_{nt}, \widetilde{s}_t)$ and $\left(u_{jt}, \widetilde{s}_t\right)$ comove as well. But then this means that we can think of ζ_t as an extra unobserved macro shock: in our setup, the macro shocks are allowed to be correlated among themselves and with the regressors, in particular the aggregate policy variable, which is the one that interests us. Thus, the most natural way to think about the identification assumption being violated suggests that with a large enough K we can guard ourselves against these type of violations.

Before turning to the second concern, let me discuss a setup under which this first concern can be addressed in an even starker way. For classes of models that assume a continuum of regions, thanks to Al-Najjar (1995), we know that for an $L_2 - bounded$ process ε_{nt}, $n \in [0, 1]$, if the process is weakly

[31] See Andrews (2005).

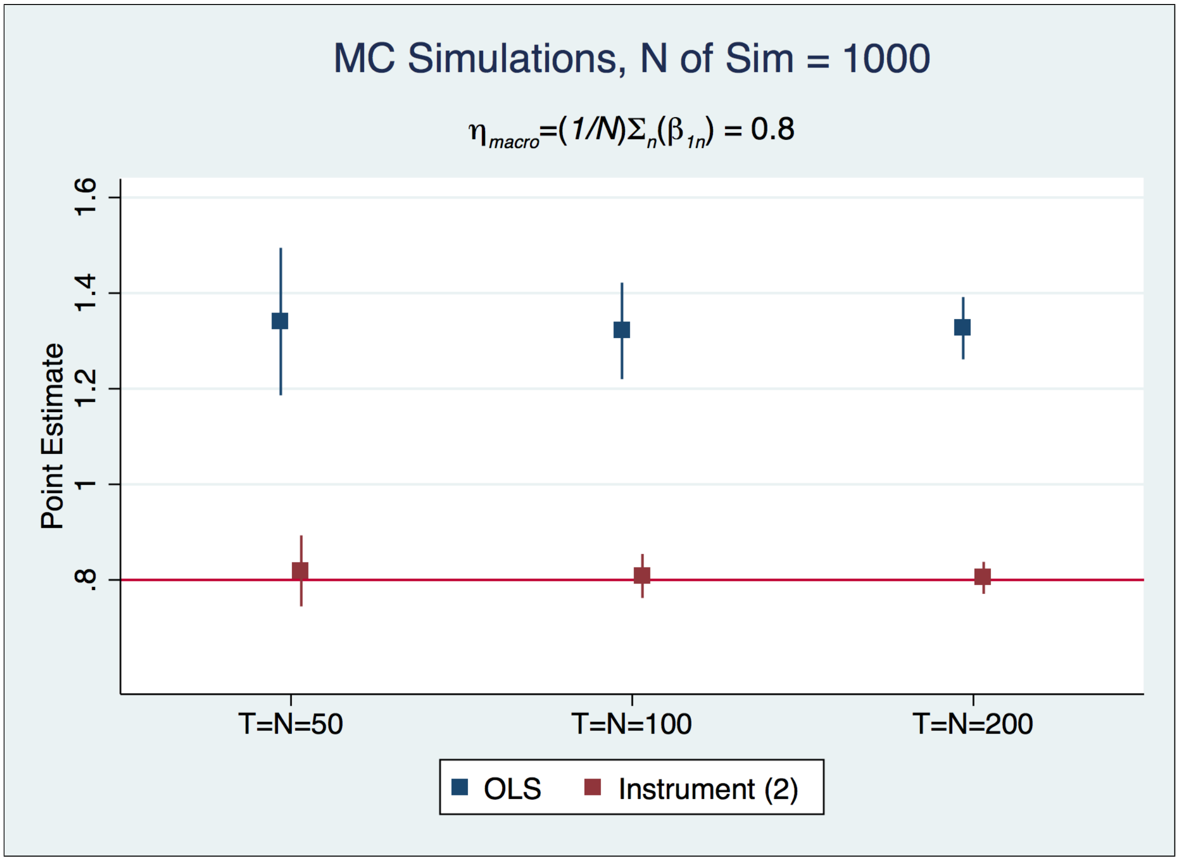

Figure 2: This figure shows the Monte Carlo simulations for the macro elasticity in Case 2.

measurable, it can be (essentially) decomposed as the sum of an aggregate process and an idiosyncratic process.[32] The aggregate process takes the form of $g_{nt} = \sum_{\kappa=1}^{\infty} \alpha_{\kappa nt}\zeta_{\kappa t}$, where the $\zeta_{\kappa t}$ form a countable set of orthonormal[33] random variables, and the idiosyncratic process is a process h_{nt}, which is orthogonal to any random variable in L_2. In particular, they are orthogonal to the $\zeta_{\kappa t}$'s. We also know that for any $\vartheta > 0$, any bounded aggregate process can be written as the sum of a finitely generated process $g_{nt}^{\vartheta} = \sum_{\kappa=1}^{K(\vartheta)} \alpha_{\kappa nt}\zeta_{\kappa t}$, and a residual process $w_{nt} = g_{nt} - g_{nt}^{\vartheta}$, with $\int_{[0,1]} \|w_{nt}\|\, dn < \vartheta$. Thus, combining these two results, we have that any weakly measurable process can be written as the sum of an aggregate process with a finite number of $\zeta_{\kappa t}$'s and a residual $\check{h}_{nt} = w_{nt} + h_{nt}$, where h_{nt} is idiosyncratic and $\int_{[0,1]} \|w_{nt}\|\, dn < \vartheta$.[34]

Thus, with a sufficiently large number of $\zeta_{\kappa t}$'s, $\check{h}_{nt}$ will be almost uncorrelated across regions. Furthermore, we can always include a constant among the regressors; thus, without loss of generality, we can assume the unobserved macro shocks have zero means and then $\check{h}_{nt}$ will satisfy condition (1.6).[35]

[32] As the author points out in the paper, the weakly measurability condition is a very weak one; see Al-Najjar (1995) for more details.

[33] Taking as the inner product of two random variables the expected value of their product.

[34] All three results can be found in Al-Najjar (1995): the decomposition result is part of the Main Theorem, the finite approximation is Proposition 1, and the third observation is given right after the Main Theorem.

[35] Of course, for this to be true, we are also assuming that the policy variables can be correlated with, at most, a finite number

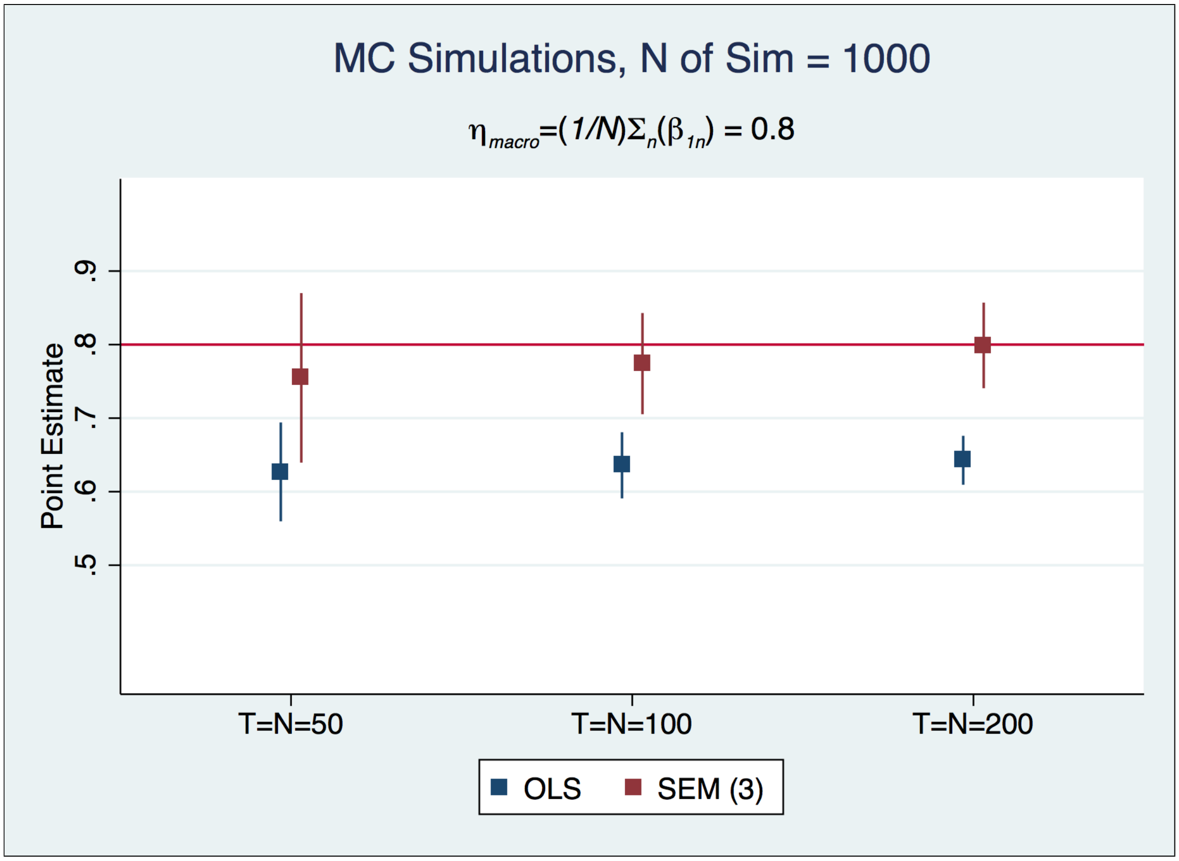

Figure 3: This figure shows the Monte Carlo simulations for the macro elasticity in Case 3.

Note also that in many of the cases reviewed in the previous subsection, the cross-correlation of the $\check{h}_{nt}$'s that might remain, even after accounting for many of the unobserved macro shocks, does not need to be addressed in terms of the single-equation interactive-effects methods. When it does, it seems that a good approximation would make this cross-sectional correlation moderately low, allowing us to safeguard the consistency of the estimators.

In the previous subsection, I also analyzed situations in which the main threat to identification comes from reverse causation, but another potential source of a violation of econometric endogeneity is current expectations of the future. That is, on top of having reverse feedback from current variables, we can have current expectations of future variables breaking the identification assumption. In principle, dealing with this concern would require again a detailed study, but a few points can be made that follow from what I have already shown and thus can be made without further investment. The following proposition shows a class of economies for which the procedure of case 3 can be applied:

of unobserved macro shocks.

Proposition 17. *Suppose we have a regional model M with equilibrium equations given by:*

$$\begin{aligned}(73)\qquad &\mathbb{E}_t\left[\left(\vec{Y}_{t+1}, G_{t+1}, m_{t+1}, \varepsilon_{G_{1t+1}}, ..., \varepsilon_{G_{Nt+1}}\right)'\right]\\ &= A\left(\vec{Y}_t, G_t, m_t, \varepsilon_{G_{1t}}, ..., \varepsilon_{G_{Nt}}\right)' + \Theta\left(\xi_t, \varepsilon_{\xi^1_t}, ..., \varepsilon_{\xi^N_t}, \xi^g_t, \epsilon^g_{1t}, ..., \epsilon^g_{Nt}, \varepsilon_{m_{1t}}, ..., \varepsilon_{m_{Nt}}, \xi^4_t\right)'\end{aligned}$$

or, with a more compact notation,

$$\mathbb{E}_t\left[\left(\vec{Y}_{t+1}, G_{t+1}, m_{t+1}, \vec{\varepsilon}^{\,'}_{G_{t+1}}\right)'\right] = A\left(\vec{Y}_t, G_t, m_t, \vec{\varepsilon}^{\,'}_{G_t}\right)' + \Theta\vec{\xi}_t,$$

where $\vec{Y}_t := \left(\widetilde{Y}_{1t}, ..., \widetilde{Y}_{Nt}\right)'$, $\vec{\xi}_t = \left(\xi_t, \varepsilon_{\xi^1_t}, ..., \varepsilon_{\xi^N_t}, \xi^g_t, \epsilon^g_{1t}, ..., \epsilon^g_{Nt}, \varepsilon_{m_{1t}}, ..., \varepsilon_{m_{Nt}}, \xi^4_t\right)'$, $\vec{\varepsilon}_{G_t} = \left(\varepsilon_{G_{1t}}, ..., \varepsilon_{G_{Nt}}\right)'$ *and:*

$$A = \begin{pmatrix} 1 & \cdots & 0 & b_1 & 0 & c_{11} & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & 0 & \vdots & \ddots & \vdots \\ 0 & \cdots & 1 & b_N & 0 & 0 & \cdots & c_{NN} \\ b_y\frac{1}{N} & \cdots & b_y\frac{1}{N} & 1 & b_m & 0 & 0 & 0 \\ d_y\frac{1}{N} & \cdots & d_y\frac{1}{N} & d_g & 1 & 0 & 0 & 0 \\ -\delta_Y + \delta_Y\frac{1}{N} & \cdots & \delta_Y\frac{1}{N} & 0 & 0 & 1 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \delta_Y\frac{1}{N} & \cdots & \delta_Y\frac{1}{N} - \delta_Y & 0 & 0 & 0 & \cdots & 1 \end{pmatrix},$$

$$\Theta = \begin{pmatrix} \theta_{11} & v_1 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \theta_{NN} & 0 & \cdots & v_N & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & \cdots & 0 & \theta_{gg} & 0 & \cdots & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 & \theta_{mm} \\ 0 & 0 & \cdots & 0 & \vartheta_{1g} & \vartheta^{\epsilon}_{11} & \cdots & 0 & \vartheta^m_{11} & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & \vartheta_{Ng} & 0 & \cdots & \vartheta^{\epsilon}_{NN} & 0 & \cdots & \vartheta^m_{NN} & 0 \end{pmatrix}.$$

Suppose as well that for all N:

1. *All eigenvalues of A lie outside the unit circle,*
2. $\vec{\xi}_t$ *is a vector martingale difference sequence.*

Then, there are restrictions on the matrices A and Θ *and on the processes in* $\vec{\xi}_t$ *such that the estimator in Proposition 16 is consistent.*

Thus, Proposition 17 shows that the methods we have already covered could be useful in these situations as well. Of course, a more general study of these is required to make more general claims. Furthermore, it would be interesting to see if and how the methods presented in this paper can help in situations in which part of the policy changes are anticipated by economic agents.[36]

[36] See Ramey (2016) for a review of these problems and the solutions adopted in different applications.

Another key condition in Assumption 1 says that the errors can be a composite of the usual unexplained shock plus terms that satisfy $\frac{1}{N}\sum_{n=1}^{N} u_{nt} = o_p(1)$. Similarly to the comment at the beginning of this subsection, the most natural way to think about this condition being violated points to adding an extra macro shock to F_t. In other words, for this condition to fail, we need enough cross-sectional correlation among the region's u_{nt}, even after differencing all the common region-invariant variables that might be driving that correlation. For example, in the models of Section 3, it is necessary that, even after controlling for $\widetilde{a}_t$, the aggregate TFP shock, and other potential macro shocks, the $\varepsilon_{\widetilde{a}_{nt}}$'s still display enough cross-sectional correlation for $\frac{1}{N}\sum_{n=1}^{N} u_{nt} \neq o_p(1)$. That is, and to give an example with only two shocks, even in the case that $\varepsilon_{\widetilde{a}_{nt}} = \delta_{1n}\widetilde{a}_t + \delta_{2n}\zeta_t + u_{nt}$, we would still need u_{nt} to be highly cross-sectionally correlated.

Finally, I give a big-picture overview of the methods discussed that will also help me relate them to the usual fixed-effects strategies. I focus the discussion on the third empirical strategy for the sake of concreteness, although what I say here holds for the other strategies as well.

As in all of them, the key of the strategy is to get estimates of $F_{1t}, F_{2t}, u_t^s, u_t^Z$.[37] There is no restriction on the processes of those unobserved macro shocks, so in this sense we can label the strategy as a "fixed effects" one. We can do so both because we treat the unobserved macro shocks as parameters to be estimated and because we obtain consistent estimates for the η_{MG}^n's without imposing any restriction on the distribution of those unobserved shocks conditional on the observables of the model ($s_t, \varepsilon_{s_{nt}}, \ldots$).

How is this possible? The superficial reason is obvious: we have managed to get our hands in an instrument, $\hat{u}_t^Z$, or, equivalently, on controls for F_{1t}, F_{2t}, u_t^s (the $\hat{F}_{1t}, \hat{F}_{2t}, \hat{u}_t^s$).[38] If we take the instrument view, note that the typical IV strategy consists of picking one observable as the candidate instrument and arguing for the exclusion restriction.[39] One way to think about this is that we are arguing for a particular configuration of the SEM. Conditional on this configuration, we are then able to recover the coefficients of interest. In (56)-(60) the logic is the same, with the important difference that the system implies that there are no observables that we could use as instruments for s_t. In this sense, the assumptions made are weaker than usual. How, then, could we get an instrument if there are no observables that serve such a purpose? Although (56)-(60) implies there are no observable instruments, it also implies there are functions of observable variables that can serve as instruments, or put differently, (56)-(60) implies that:

(a) there are unobservable variables that can serve as instruments,

(b) we can use the observables to estimate those unobservables.

The unobservables that serve as instruments are clear; for example, in this case, u_t^Z plays that role. What allows the use of observables to estimate those unobservables is the structure assumed for the unobserved heterogeneity of the form $\Omega_{nt} = \alpha_n \xi_t$, i.e., as the interaction of a purely cross-sectional

[37] Of course, when $\varepsilon_{s_{nt}}$ is correlated not only with the unobserved macro shocks, but also with ω_{nt}^Y, observing $\varepsilon_{Z_{nt}}$ is also crucial. But this is a well understood feature, so I will focus here exclusively on the logic behind getting at estimates of $F_{1t}, F_{2t}, u_t^s, u_t^Z$.

[38] To be precise, we are only getting at an estimate of the space spanned by F_{1t} and F_{2t}, but this is enough to control for their variation.

[39] Weak instruments are of course a concern, but one that comes, conceptually, later in the strategy.

heterogeneity term and a purely time series one. This structure is what allows the interactive-effects methods to construct estimates like $\hat{F}_{1t}$, i.e., to "extract" the unobservable controls or instruments from the observables of the model. The justification for the particular structure of Ω_{nt}, in turn, comes from macroeconomic theory: if we focus on models with heterogeneous agents, regions, etc., we are naturally led to structural equations in which part of the variation we see in the observable variables is driven by unobserved macro shocks (shocks to the Taylor rule, aggregate TFP shocks, etc) that have different effects in different agents, regions, etc. Moreover, as I explained in detail in Section 1, this is a particular case of a more general panel data model in which all parameters are allowed to vary. However, as I argued there, such a model is too general in the sense that is not estimable.

Does the unobservability of the instruments imply that this strategy is more robust than the usual IV strategies? No, because in both cases we rely on a particular configuration of the SEM or class of SEMs, and thus a model misspecification would break down the strategy in both cases. However, it does imply that even if no instrument is observed, we can in some cases recover an unobserved one.

The same comments apply when we think about using $\hat{F}_{1t}, \hat{F}_{2t}, \hat{u}_t^s$ as controls, i.e., we usually have to argue that these are all the "good" controls we need, after picking some observable candidates, whereas in (56)-(60) there are no observables that could serve this purpose. Hence, this view of the strategy is akin to the control function approach.

This discussion also means that part of the understanding that we get in the usual empirical strategies from having an observable instrument, for example due to its mechanics, here is washed away by the "anonymity" of the relevant variables. Probably the best that can be offered in this case is a hint of what these the unobserved macro shocks estimates could be representing.

Now, in a more subtle comparison, what is the relationship with the usual additive effects? Additive effects are a particular case of interactive effects. Since we are interested in recovering the coefficient on a region-invariant variable like s_t, the interesting comparison is to the presence of additive time fixed effects. Having an additive effect in (56)-(60) would imply that, for example, $\delta_{1n} = 1$. The assumptions in our empirical strategy allow this, so all of the analysis we have done applies to this situation as well. This means that we can still get consistent estimators $\hat{F}_{1t}, \hat{F}_{2t}, \hat{u}_t^s$ to recover η_{MG}^n. Alternatively, we can contrast this to the approach in Hausman and Taylor (1981): $\hat{u}_t^Z$ serves as an instrument for s_t, so if we were to estimate the first equation in (56)-(60) with time dummies, and thus get estimates of $\eta_{MG}^n - \frac{1}{N}\sum_{n=1}^N \eta_{MG}^n$, we could then use the residuals to run 2SLS using $\hat{u}_t^Z$ as an instrument for s_t to recover $\frac{1}{N}\sum_{n=1}^N \eta_{MG}^n$. With $\frac{1}{N}\sum_{n=1}^N \eta_{MG}^n$ and $\eta_{MG}^n - \frac{1}{N}\sum_{n=1}^N \eta_{MG}^n$ we can reconstruct the η_{MG}^n's by summing the former to the latter. Thus, the approach is similar to Hausman and Taylor (1981) in the sense that the instrument comes from within the system, but it differs in that it is not observable.

Inference Now, turning to inference, I briefly describe two alternative ways of thinking about what does it means to let $N \to \infty$ in these setups, because this will turn out to be important for understanding the discussion of inference.

The specification so far treats the regions in the sample as the entire population, because, for example, if we use data on the 50 US states, the most natural way to think about that sample is that one has the whole population of the states.[40] In particular, this means that the usual standard errors will

[40] This is not business as usual in econometrics, and recent papers such as Abadie, Athey, Imbens and Wooldridge (2014)

overestimate the true amount of noise. It also means that when we carry the thought experiment of $N \to \infty$, to prove the consistency of the estimators, we can think of it as making the grid finer; for example, using counties instead of states, etc. I will refer to this as the "refining grid metaphor." However, we can also think of N as being a sample from a very large population, which is the usual practice in econometrics. I will refer to this as the "infinite population metaphor." In this case, increasing N simply means increasing the sample size. For this second case, a future extension of these methods would be to apply the results in Mikusheva and Anatolyev (2018) to our specifications.

Nonetheless, each of these approaches requires its own study, so although they are important, I leave them to future work. Instead, in this paper I use the Bootstrap to get the standard errors. The bootstrap methods applied here in principle require a theoretical justification for some of the cases described in the previous subsections. Again, this is beyond the scope of this paper, so I adapt the methods in Gonçalves and Perron (2014) and Djogbenou, Gonçalves and Perron (2015). These are based on a general residual-based bootstrap for factor-augmented regression models, like those we have covered so far. The particular implementation shown in those papers, and applied here as well, is a multi-step wild bootstrap scheme. As Horowitz (2001) shows, in the context of a heteroskedastic regression model, the numerical performance of the wild bootstrap is much more accurate, especially in small samples, than that of a paired or nonparametric bootstrap. However, since the multi-step wild bootstrap is less comparable across cases for the methods proposed in this paper, I also provide results using a simple nonparametric bootstrap in which regions are resampled at random, with replacement. Because of this resampling, the nonparametric bootstrap has a better fit with the infinite population metaphor. In contrast, the wild bootstrap has a better fit with the refining grid metaphor, since it maintains fixed the panel dimension.

Choosing K The different specifications covered so far offer different possibilities for estimating the number of unobserved macro shocks. Nonetheless, the idea is always to apply the results in Bai and Ng (2002) to one of the reduced forms of the models. For example, for case 0 we can apply it to the reduced form of $\varepsilon_{s_{nt}}$ and use the estimated number of unobserved macro shocks to extract that amount from it. In case 3 we can use the reduced form of Y_{nt} to do this, etc.

6 Fiscal Multipliers in the US

In this section, I offer a detailed application of the procedure proposed in this paper to the case of fiscal multipliers in the US. I use a balanced panel of the 50 US states plus Washington, DC from 1971 to 2008. In Subsection 6.1 I detail the equations and variables to be used in the application. In Subsection 6.2, I describe the data I use and their sources. I then show my main results in Subsection 6.3 and offer a discussion comparing these to other numbers obtained in the literature. Finally, in Subsection 6.4, I show additional results and conduct robustness checks to increase the confidence in the estimates obtained.

deal with these situations.

6.1 Recovering Fiscal Multipliers from Regional GDP Regressions

All of the variables used are in real per capita terms. The main equation of interest, following the examples in the previous section, is:

$$Y_{nt} = \eta_{MG}^{n} G_t + \eta_{ML}^{n} \varepsilon_{G_{nt}} + F_t \lambda_n + \epsilon_{nt}^{Y}. \tag{74}$$

In equation (74), the dependent variable, Y_{nt}, is the real per capita GDP growth rate of state n in period t. G_t is the average of the growth rates of real per capita spending in every state, and $\varepsilon_{G_{nt}}$ is the deviation of state n from G_t. F_t is a vector of K unobserved macro shocks hitting state n, and ϵ_{nt} contains all the other shocks hitting state n.[41] I estimate variations of (74) including lags of both spending and GDP; I detail the exact equation in each case.

In the context of the menu of strategies discussed in the previous section, this application is usually thought to lie somewhere between cases 1 to 3. Hence, these are the cases for which I show results later in this section. The main obstacle to analyzing government spending is that this policy could be simultaneously determined with output, for example, increasing during recessions to ameliorate their depth. If we focus exclusively on this aspect, we should observe small (or negative) multipliers when doing naïve OLS regressions, because this property would generate a negative bias in the estimates. As I show later in this section, this is indeed what we get.

The main difference among these three cases lies between the third one, which allows regional spending to respond to regional GDP, and the first two which do not. Consequently, the third one is more robust. However, since it's much more involved than the first two, the cost in terms of power is higher. Moreover, to the extent that regional spending captures motivations that are further away from the typical stabilization motives of the federal government, this concern is smaller for regional spending.

Nonetheless, as the examples in Section 3 show, the estimators used are consistent for models with price rigidities, flexible prices, different market structures, etc. Being agnostic about these important issues is a key advantage of the results to follow.

In the cases that need a special numerical procedure, those with single-equation interactive-effects estimators, I use the strategy described in Bai (2009) and Ando and Bai (2015).

The aggregator of the models that justify (74) is used to aggregate the micro-global elasticities:

$$\hat{\eta}_{macro} = \sum_{n} \frac{1}{N} \frac{\overline{Y_n}}{Y} \hat{\eta}_{MG}^{n}. \tag{75}$$

See Section 3 for details.

To convert the macro elasticity into the typical fiscal multiplier used in the literature, I multiply (75) by the average ratio of GDP to government spending in the period 1971-2008, which is 3.16.

For inference I use a wild bootstrap and a nonparametric bootstrap. Since in every regression I control for the lag of the dependent variable, and because these variables are measured in growth rates, autocorrelation is not a concern. Therefore, for the wild bootstrap I apply the methods of Gonçalves and

[41] I refer to Y_{nt}, loosely, as "state n's GDP" and to G_t as "aggregate government spending," but please keep in mind that every regression uses real per capita variables.

Perron (2014). In contrast, heteroskedasticity is likely a concern, but the wild bootstrap is especially designed to handle it. For comparison purposes, I also report results using a simple nonparametric bootstrap. See Section 5 for details.

Finally, I briefly discuss the meaning of η^n_{MG} in (74). As Farhi and Werning (2016) clearly explain, what we are getting in (74) is a summary fiscal multiplier.[42] The intuition of why this is an interesting measure is that since all spending must be financed to get a given amount of response in output, we need to look at the response in output vis-à-vis the total spending required to achieve that change. Nonetheless, note that our strategies also allow the computation of the impulse responses themselves. For example, in case 3 we are getting an estimate of u^s_t that could be used for these purposes.

6.2 Data Description and Sources

To construct Y_{nt}, I started by collecting data on Gross State Product (GSP) by state from the BEA website.[43]

For G_t I collected data on federal, state and local spending. The federal spending for the period 1981-2008 was obtained from the CFFR[44], and for the period 1971-1980 from the ICPSR.[45] 2010 is the last year available for the CFFR, and to my knowledge there is still not a reliable source of data for federal spending past that point that is also completely consistent with the previous series. Thus, I choose to work with the sample from 1971 to 2008. I exclude 2009 and 2010 because of the crisis. For the state and local spending, as well as revenues, the period 1971-2008 was obtained from the Census Bureau.[46] Data prior to 1971 are available, but there is no match of the federal spending for those years. Thus, the main shortcoming in terms of limiting the sample size in this study comes from the limited availability of federal spending figures by state. The measure of total spending by state is obtained by adding spending on the federal, state and local levels and subtracting the revenues from the federal government of each state. Since a large number of state and local expenses, like education, are financed to some degree by federal programs, subtracting the money states and local governments get from the federal government helps to minimize the double-counting problem.

The population of every state comes from the Census Bureau.[47] The price index comes from the BEA and is the GDP chain-type price index.[48]

[42] The fiscal multiplier is usually defined in dollar units instead of elasticities as in (74), but this is inessential for this particular discussion. In my empirical results in the next subsections, I will report the typical fiscal multiplier instead of the elasticities because this is the common practice in the literature.

[43] U.S. Bureau of Economic Analysis, "ANNUAL GROSS DOMESTIC PRODUCT (GDP) BY STATE," https://www.bea.gov/itable/iTable.cfm?ReqID=70&step=1#reqid=70&step=1&isuri=1&7003=200&7001=1200&7002=1&7090=70.

[44] U.S. Census Bureau, "CONSOLIDATED FEDERAL FUNDS REPORT," https://www.census.gov/govs/cffr/.

[45] Anton, Thomas. FEDERAL BUDGET OUTLAYS, 1971-1980 [UNITED STATES]. Compiled by University of Michigan, Intergovernmental Fiscal Analysis Project. ICPSR ed. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [producer and distributor], 1984. http://doi.org/10.3886/ICPSR08199.v1

[46] U.S. Census Bureau, Annual Surveys of State and Local Government Finances (1992-2015) and U.S. Census Bureau, Annual Survey of State Government Finances and Census of Governments (1971-1991).

[47] I provide only the citation for Massachusetts since all the others are similar: "U.S. Bureau of the Census, Resident Population in Massachusetts [MAPOP], retrieved from FRED, Federal Reserve Bank of St. Louis; https://fred.stlouisfed.org/series/MAPOP."

[48] U.S. Bureau of Economic Analysis, Gross domestic product (chain-type price index) [A191RG3A086NBEA], retrieved from

For $\varepsilon_{Z_{nt}}$ I use declarations of natural disasters by the federal government: upon a natural disaster inside state n, the state has the option to apply for federal funds, arguing that its own funds are not enough to deal with the disaster. If the president declares it a natural disaster, federal funds are released to the state.[49] Since the declarations might be correlated with the actual disasters that could be hitting the states, I also control for the number of disasters in the regressions. However, because not all disasters are available, I control for the number of severe storms (tornado, hail and damaging wind). The data on declarations were obtained from FEMA[50] and the data on severe storms from the SPC.[51]

Finally, I exclude 4 states because they have yearly variations of above 20% in their GSP, which translate into unreliably large micro-global multipliers, and a very small population: North Dakota (0.23% of the US population), South Dakota (0.27%), Wyoming (0.18%) and Alaska (0.23%).

6.3 Main Results

Table 1 contains the main results obtained.[52] Column 1 is a simple time series OLS regression of Y_t on a constant, G_t, G_{t-1}, G_{t-2} and Y_{t-1}. This is just a benchmark to compare the different specifications. Note that if none of the threats to identification discussed earlier were present, we could just run an aggregate time series OLS regression, as Column 1 shows. Columns 2 – 4 estimate (74) for cases 1, 2 and 3, respectively. Since the model selection techniques of Bai and Ng (2002) applied to Y_{nt} and $\varepsilon_{G_{nt}}$ point to a number of unobserved macro shocks from 3 to 5, 5 is the number I use for most of the results; however, I also show what happens when we vary the number of estimated unobserved macro shocks.

A consistent pattern emerges from these numbers. The "naïve" column 1 shows a huge negative bias in the estimates, consistent with the main concern we have when looking at fiscal multipliers: if spending increases when the economy is doing poorly, a naïve OLS regression should pick this up this as small or even negative coefficients for government spending. The three strategies in columns 2 – 4 show a reversal to zero of the contemporaneous spending. They also show a very precisely estimated positive fiscal multiplier for lagged spending. Columns 2 and 4 show a small and non-significant estimate for the second lag of spending, whereas column 3 shows a negative coefficient significant at the 5% level (with a p-value of 0.048). This fact is not robust, though, as Table 2 shows. Varying the number of estimated unobserved macro shocks changes the point estimates for the second lag; in contrast, the coefficient on the first lag is very stable across all specifications. Moreover, Table 9 shows that only the first lag is significant when we instead apply the nonparametric bootstrap. To further investigate the effect of the first lag, Table 3 shows the results for case 2 under different specifications using only G_{t-1}. As the table shows, we now get estimates in the range 0.82 – 1.2, all significant at the 1% level. Hence, the results point towards a positive and precisely estimated fiscal multiplier of lagged spending. Of course, there could still be some bias in these estimates as well, and it might be the case

FRED, Federal Reserve Bank of St. Louis; https://fred.stlouisfed.org/series/A191RG3A086NBEA.

[49] See https://www.fema.gov.

[50] https://www.fema.gov. This product uses the Federal Emergency Management Agency's API, but is not endorsed by FEMA.

[51] https://www.spc.noaa.gov/.

[52] All of the tables and figures that are not present in the main text can be found in Appendix B.

	Dep. Variable: Real GDP Per Capita Growth			
	(1) OLS	(2) Model Selection	(3) Instrument	(4) SEM
G_t	-1.08***	-0.07	0.02	0.01
	(0.38)	(0.17)	(0.14)	(0.59)
G_{t-1}	0.28	0.65***	0.80***	0.69***
	(0.48)	(0.10)	(0.11)	(0.29)
G_{t-2}	0.38	0.08	-0.34**	0.03
	(0.39)	(0.08)	(0.17)	(0.24)
$State\ Time\ Trend$		✓	✓	✓
$Time\ Fixed\ Effect$				
$Interactive\ FE$		✓	✓	✓
$Number\ of\ IE$	0	5	5	5
Observations	37	1,739	1,739	1,739
Period	1971-2008	1971-2008	1971-2008	1971-2008
Number of Regions	-	47	47	47

Table 1: This table presents estimates of the fiscal multiplier. Column 1 is a simple time series OLS regression of Y_t on a constant, G_t, G_{t-1}, G_{t-2} and Y_{t-1}. Columns 2 to 4 estimate (74) for cases 1, 2 and 3, respectively. All of the regressions control for Y_{nt-1}. I exclude 4 states because they have yearly variations of above 20% in their GSP, which translate into unreliably large micro-global multipliers, and a very small population: North Dakota (0.23% of the US population), South Dakota (0.27%), Wyoming (0.18%) and Alaska (0.23%). Standard errors are in parentheses and are obtained with the wild bootstrap; see Section 5 for details. The coefficients with $***$ are significant at the 1% confidence level; with $**$ are significant at the 5% confidence level; and with $*$ are significant at the 10% confidence level.

that we are just approximating more closely the true effect. In the next subsection, I perform further robustness checks that reinforce these conclusions.

Thus, a first question of whether the fiscal multiplier is positive or negative is answered affirmatively, and with very precise estimates. However, these results do not allow a strong answer to the question of whether the fiscal multiplier is above or below 1. The estimates point to a multiplier somewhere between $0.7 - 1.2$. Therefore, it is not possible to discard the possibility that government spending crowds out/in private spending. However, the estimates on the lower end fall in the range of $0.7 - 0.9$, suggesting that if there is crowding out, it is not severe; and, similarly, the estimates on the higher end fall in the range of $1.1 - 1.2$, so if there is crowding in, it does not seem to be high.

Moreover, given the similarity of the estimates across the different cases, it does not seem possible to favor one strategy over the rest. In principle, as we argued in the first subsection, case 3 is the most robust of the three. This little difference could, in principle, be due to the fact that the simultaneous equation bias that regional spending might be suffering from in cases 1 and 2 is not strong enough for the estimates on aggregate spending to show large differences. Nonetheless, note that there is a

	Dep. Variable: Real GDP Per Capita Growth			
	(1)	(2)	(3)	(4)
	OLS	Instrument	Instrument	Instrument
G_t	-1.08***	0.34***	0.22**	0.02
	(0.38)	(0.10)	(0.11)	(0.14)
G_{t-1}	0.28	0.70***	0.72***	0.80***
	(0.48)	(0.07)	(0.07)	(0.11)
G_{t-2}	0.38	0.24**	0.28**	-0.34**
	(0.39)	(0.12)	(0.11)	(0.17)
$State\ Time\ Trend$		✓	✓	✓
$Time\ Fixed\ Effect$				
$Interactive\ FE$		✓	✓	✓
$Number\ of\ IE$	0	3	4	5
Observations	37	1,739	1,739	1,739
Period	1971-2008	1971-2008	1971-2008	1971-2008
Number of Regions	-	47	47	47

Table 2: This table presents estimates of the fiscal multiplier. Column 1 is a simple time series OLS regression of Y_t on a constant, G_t, G_{t-1}, G_{t-2} and Y_{t-1}. Columns 2 to 4 estimate (74) for case 2. All of the regressions control for Y_{nt-1}. I exclude 4 states because they have yearly variations of above 20% in their GSP, which translate into unreliably large micro-global multipliers, and a very small population: North Dakota (0.23% of the US population), South Dakota (0.27%), Wyoming (0.18%) and Alaska (0.23%). Standard errors are in parentheses and are obtained with the wild bootstrap; see Section 5 for details. The coefficients with $***$ are significant at the 1% confidence level; with $**$ are significant at the 5% confidence level; and with $*$ are significant at the 10% confidence level.

significant precision loss in case 3. Thus, the fact the there is not a significant difference could also be due to power loss, and if maybe we had a sample with a large enough N and T, we might we able to see sharper differences among them. This is the main disadvantage of case 3 in samples of this size.

Figure 4 shows on a heat map the micro-global multipliers that lie behind the fiscal multiplier of case 2 using only G_{t-1} and controlling for Y_{nt-1}. The fiscal multiplier in this case is 0.79 (0.09) and the 4 excluded states are fixed at that level. As the figure shows, the results seem to be coherent in the sense that states next to each other tend to have similar micro-global multipliers. Furthermore, the Far West, Southwest and Great Lakes are the regions with the highest multipliers, while most of the Mideast and New England is populated with states of low multipliers. It is worth noting that this pattern is also robust across specifications (unreported to save space): the particular values of the micro-global multipliers change, but states next to each other tend to have similar multipliers, and the geographical distribution of high and low multipliers remains the same.

Before turning to the next subsection, let me reflect on these results and compare them with previous results in the literature. Ramey (2016) provides a thorough review of the literature on fiscal multipliers.

	Dep. Variable: Real GDP Per Capita Growth				
	(1)	(2)	(3)	(4)	(5)
	OLS	Instrument	Instrument	Instrument	Instrument
G_{t-1}	-0.19	0.89***	0.82***	1.13***	1.20***
	(0.35)	(0.12)	(0.07)	(0.06)	(0.11)
State Time Trend		✓	✓		
Time Fixed Effect			✓	✓	
Interactive FE		✓	✓	✓	✓
Number of IE	0	5	5	5	5
Observations	37	1,739	1,739	1,739	1,739
Period	1971-2008	1971-2008	1971-2008	1971-2008	1971-2008
Number of Regions	-	47	47	47	47

Table 3: This table presents estimates of the fiscal multiplier. Column 1 is a simple time series OLS regression of Y_t on a constant, G_{t-1} and Y_{t-1}. Columns 2 to 5 estimate (74) for case 2 for the different specifications detailed in the checkboxes. All of the regressions control for Y_{nt-1}. I exclude 4 states because they have yearly variations of above 20% in their GSP, which translate into unreliably large micro-global multipliers, and a very small population: North Dakota (0.23% of the US population), South Dakota (0.27%), Wyoming (0.18%) and Alaska (0.23%). Standard errors are in parentheses and are obtained with the wild bootstrap; see Section 5 for details. The coefficients with $***$ are significant at the 1% confidence level; with $**$ are significant at the 5% confidence level; and with $*$ are significant at the 10% confidence level.

For our estimates, a common theme when carrying out the comparison is that it is hard to find studies that use similar time periods, and many use quarterly data instead of annual data. With this caveat in mind, there are some papers that get higher multipliers (for samples beginning in 1947 or 1960), but most of the studies find lower estimates. Maybe the easiest papers to compare against are those of Barro and Redlick (2011) and Hall (2009), because the equations they estimate are similar. In comparison to the estimates in these papers, here I get higher and more precise effects of government spending. One of the common features of those two papers is that the estimates of the spending variables are not significant for time periods close to those used here. Although the comparison is more difficult in the case of Blanchard and Perotti (2002), because the comparable summary fiscal multiplier is not provided, it is safe to say that the estimates here are larger. As Ramey (2016) points out, computing the summary fiscal multipliers from impulse responses in various papers usually gives estimates below 1, even when the peak output response against the initial government spending effect is above 1.

6.4 Additional Results and Robustness Checks

Robustness Checks I now present various robustness checks and present also other specifications that might be of interest to see how the strategies presented in this paper perform. The nonparametric bootstrap is in principle better for comparing across cases, given that the methodology is the same across

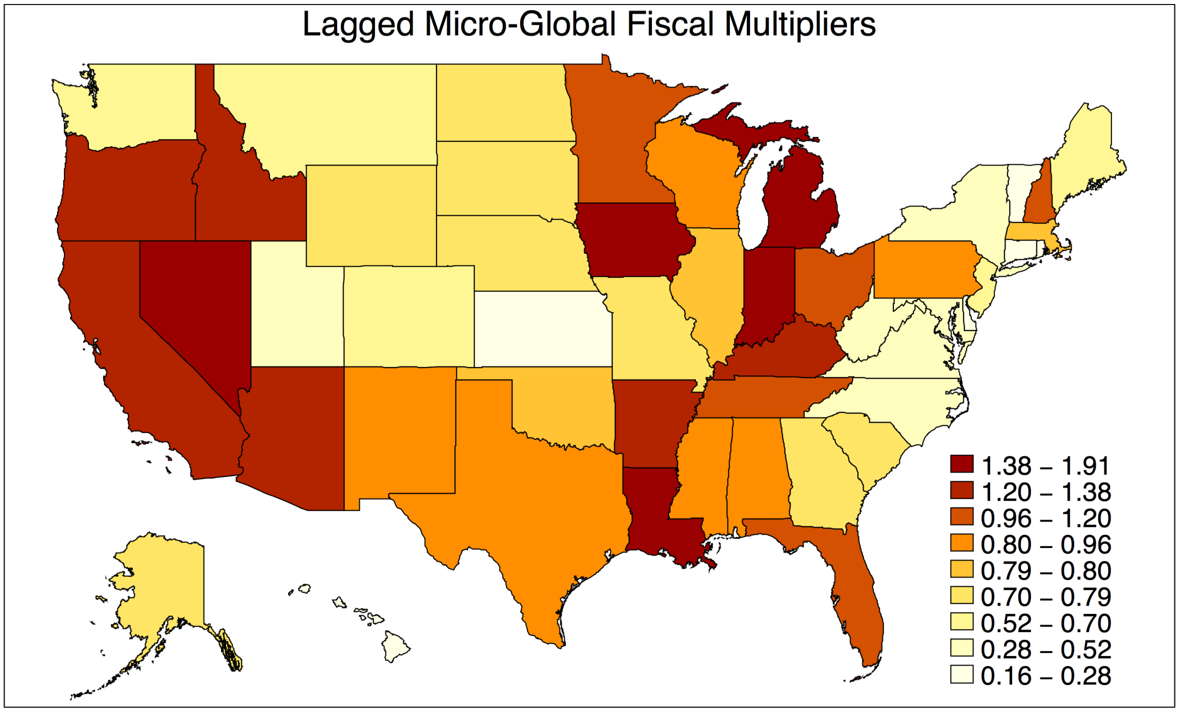

Figure 4: This figure shows a heat map of the micro-global multipliers that lie behind the fiscal multiplier of case 1.

the three cases. However, the wild bootstrap probably provides a better approximation for the sample size we have in this application; consequently, to clearly show the differences across cases, some of the results presented here rely on the nonparametric bootstrap, and the rest rely on the wild bootstrap. I indicate in each table which one is used. The general picture that emerges from these robustness checks in this respect is that the wild bootstrap tends to deliver lower standard errors, and that case 3 is the one with the least precise estimates, as expected. Nonetheless, the main message remains the same: the results point towards a positive effect of lagged fiscal stimulus, and we can neither confirm nor reject a multiplier above 1.

In Table 4, I show what happens when one explicitly accounts for time fixed effects. The results show very little difference from the results we obtained before. In Table 5, I leave the time fixed effects but now remove the state time trends. As the table shows, this increases the estimates by some margin, both of contemporaneous spending and of lagged spending. In Table 6, I remove both the time fixed effects and the state time trends, and there we see again slightly higher estimates but in the range we obtained before.

In Table 7, for case 2, I show what happens when we vary the amount of estimated unobserved macro shocks in columns 1 through 6. As the table shows, the estimates increase when we reach 4 and 5 unobserved macro shocks. Moreover, it shows that the difference between using 1 or 2 unobserved macro shocks and using 4 or 5 can increase the multiplier by 50%. This table also shows the results of including the lead of spending in column 7, which comes with a point estimate of 0.08 (0.16). However, as I discussed in Section 5, a detailed study of anticipation in the context of the methods presented in

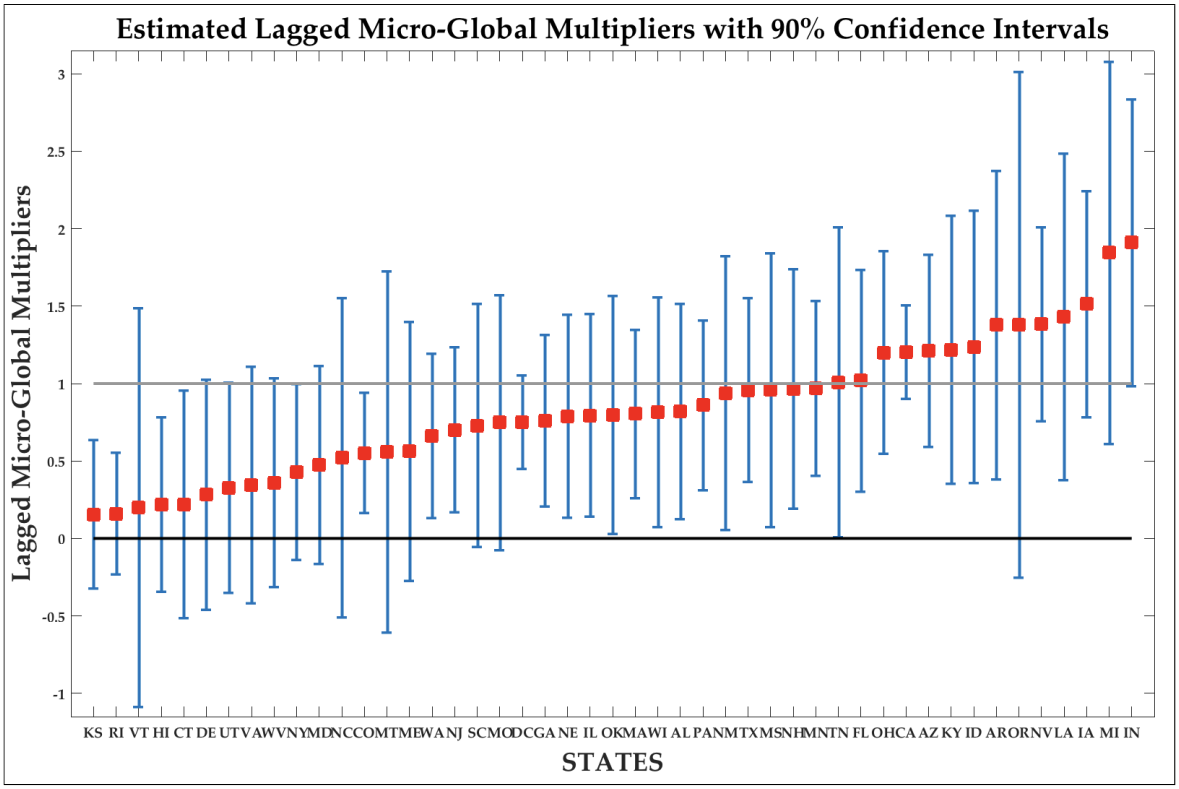

Figure 5: This figure shows the lagged micro-global multipliers of case 2 using only G_{t-1} with their 90% confidence intervals. All of the regressions control for Y_{nt-1}. I exclude 4 states because they have yearly variations of above 20% in their GSP, which translate into unreliably large micro-global multipliers, and a very small population: North Dakota (0.23% of the US population), South Dakota (0.27%), Wyoming (0.18%) and Alaska (0.23%).

this paper would be needed to completely discard foresight problems in the estimates. Figure 10 plots this last column to convey a better sense of the estimates.

In Table 8, I show what happens when we go from 1 to 5 unobserved macro shocks for case 3 and the three spending variables. Here again we observe a considerable difference between using, for example, 1 unobserved macro shock and using 5. Moreover, note here how much precision is lost in case 3 in comparison to the other cases, just as we observed in the main results, a consequence of being a much more involved estimator. This is one of the drawbacks of this case, especially in samples like this one in which neither N nor T is very large.

Finally, a comment on the strength of the instrument used. Case 1 can't be affected by a weak instrument problem because it does not rely on one, but cases 2 and 3 can. And, in principle, it is hard to know how to diagnose weak instruments in these setups given how non-standard they are. A thorough treatment of this issue would probably need its own study. However, if we look at the first stage F values for case 2, even forgetting about the fact that the unobserved macro shocks are generated regressors, there are many states for which this value is below 5. This suggests that strategies 2 and 3 may be suffering from a weak instruments problem.

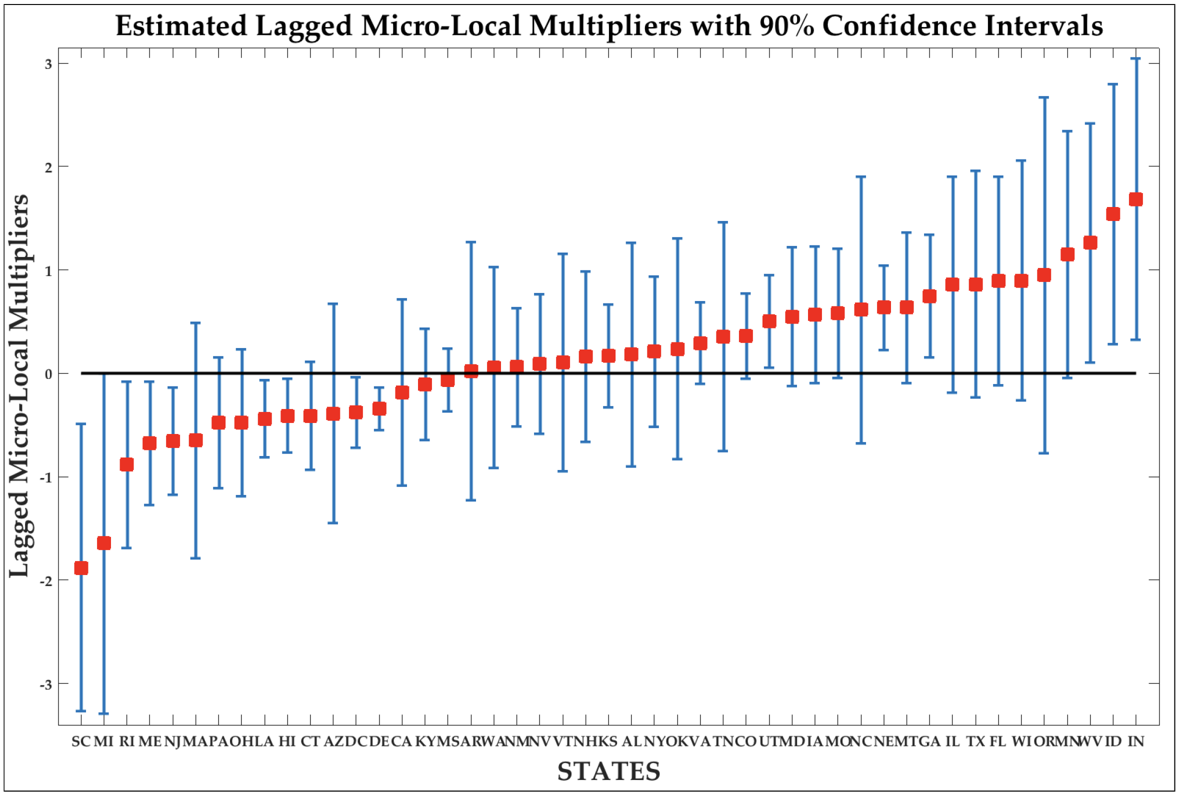

Figure 6: This figure shows the lagged micro-local multipliers of case 2 using only G_{t-1} with their 90% confidence intervals. All of the regressions control for Y_{nt-1}. I exclude 4 states because they have yearly variations of above 20% in their GSP, which translate into unreliably large micro-global multipliers, and a very small population: North Dakota (0.23% of the US population), South Dakota (0.27%), Wyoming (0.18%) and Alaska (0.23%).

This potential weak instruments problem, along with the relatively small N and T and the potential anticipation problems, are the three main concerns of this empirical application. Nonetheless, the main takeaways from these results are consistent across all of the specifications: contemporaneous aggregate spending does not seem to play an important role, and lagged spending appears to have a positive impact with a multiplier in the range $0.65 - 1.23$.

Additional Results Although Figure 4 shows the geographical distribution of the micro-global multipliers, it does not show the precision of the estimates. Moreover, it is interesting to compare these estimates to the micro-local multipliers to see if the theoretical differences we have pointed out actually show up in them. Figures 5 and 6 show the micro-global and micro-local multipliers of case 1, using only G_{t-1}, in ascending order. Figure 7 shows a scatter of these estimates against each other along with a 45-degree line. I use case 1 because of the potential weak instruments problem described in the previous subsection, and thus these results should be taken cautiously. As the figures show, the estimated micro-global multipliers are much higher than the micro-local ones for a large fraction of states. Furthermore, although there are many states with similar values for both, there are many that have micro-global multipliers of around 1, along with negative micro-local multipliers. For example,

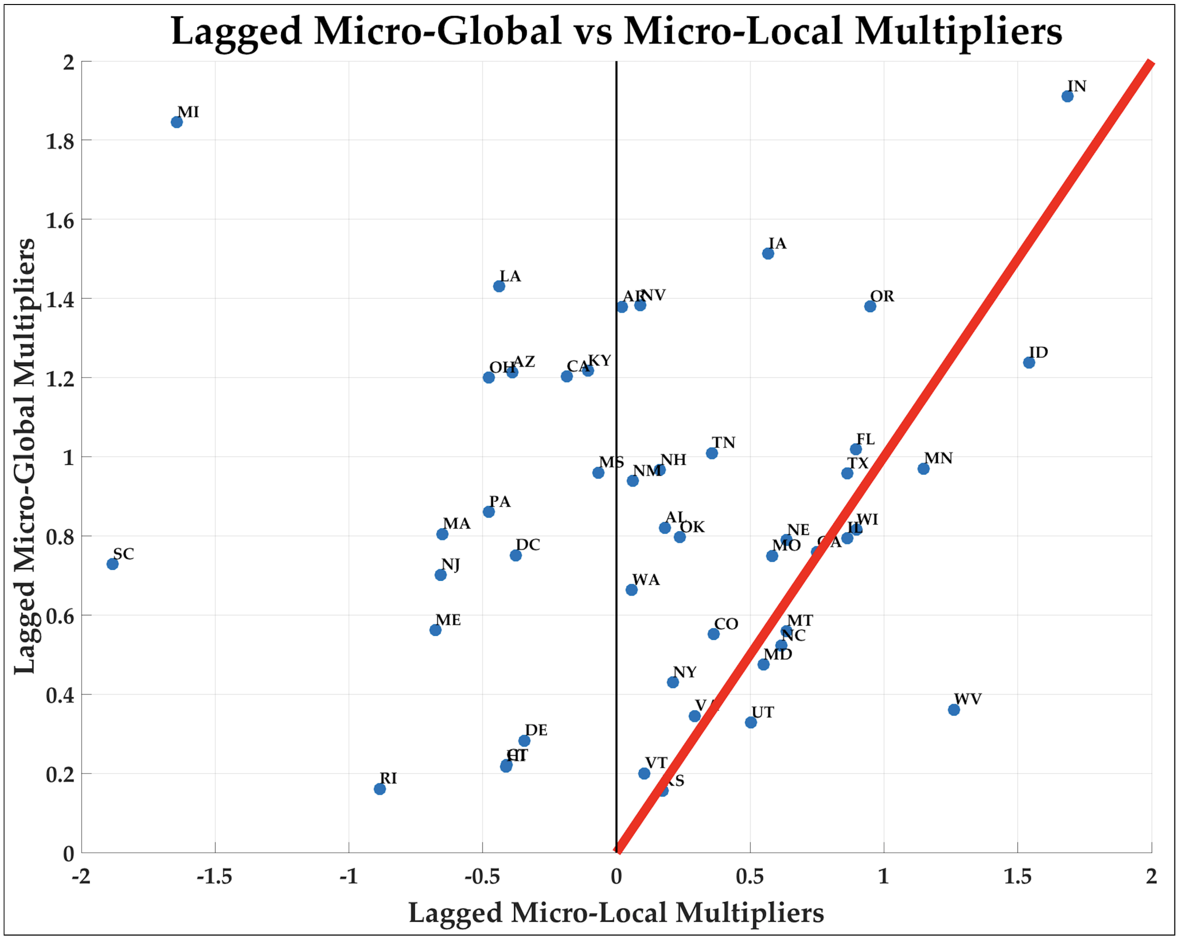

Figure 7: This figure plots the estimates in Figure 5 against those in Figure 6, along with a 45-degree line.

Massachusetts has a micro-global multiplier of 0.80 (0.35) and a micro-local multiplier of −0.65 (0.72).

Of course, bias could be a part of this disparity, but, more likely, there are other aspects that could help explain the difference. First, since the micro-global multipliers capture the effect of spillovers across regions, it is reasonable to expect differences in them. This could be the explanation for some of the states for which the disparity of both micro multipliers is rather small. For example, Minnesota has a micro-global multiplier of 0.97 (0.34) and a micro-local multiplier of 1.15 (0.71). Nevertheless, since the method in this paper does not allow us, without further assumptions, to attribute the difference to a particular source of spillovers, we cannot be entirely sure.

However, states like Massachusetts display such a large disparity that it is probably caused by more than spillover effects. To the extent that regional spending captures government spending that is more tightly connected to state and local spending, we should probably observe lower micro-local multipliers. The reason is that this spending is a closer substitute for private spending, since it is related particularly to, for example, education and transportation. Moreover, we should also probably observe lower local-multipliers for states with high debt to GSP ratios, since it is plausible to think that the private sector in these states would react more violently to cutting private spending, because of the anticipation of debt-related problems in the near future. In fact, Ilzetzki, Mendoza and Végh (2013) show evidence that this could also show up as negative multipliers for countries in a similar situation. This explanation could also account for the negative micro-local multipliers we observe for some of the states in our sample.

Finally, these figures also highlight the importance of relying on the micro-global elasticities for computing the macro elasticity: if we were to instead use the micro-local multipliers, we would get an estimate of 0.11 (statistically indistinguishable from zero), compared to the 0.79 (significant at the 1% level) we get when we use the micro-global multipliers.

7 Conclusion

This paper proposed a new estimation framework to recover macro elasticities, using regional data. The procedure is robust to many macro models that might be generating the data. Examples of the models considered are a regional new Keynesian model and a regional RBC model. Models with incomplete and complete markets and under different market structures are included. The estimation procedure delivers consistent estimates of the macro elasticities using as inputs the micro-global elasticities of every region.

An application to the case of fiscal multipliers in the US is provided. Regional regressions with state specific time-trends, time fixed effects and unobserved macro shocks are considered. The results point to higher effects of government spending than previous estimates have suggested, and are shown to hold under different specifications. Moreover, current spending does not seem to have an effect on current GDP growth, with lagged spending capturing the whole effect.

A Appendix: Proofs

Proposition 1

Proof. We have the following system of equations:

$$
\begin{aligned}
p_{it}^{d} &= -\rho_i^d y_{it} + \alpha_1 Y_t + \alpha_2 \xi_{1t} + u_{it}^{d} \\
p_{it}^{s} &= \rho_i^s y_{it} + \beta_1 s_t + \beta_1 \varepsilon_{s_{it}} + \beta_2 \xi_{2t} + u_{it}^{s} \\
p_{it}^{d} &= p_{it}^{s} = p_{it}.
\end{aligned}
$$

Just by replacing the first and second equations in the last one:

$$
\begin{aligned}
y_{it} = &\left(\frac{\alpha_1}{\rho_i^s + \rho_i^d}\right) Y_t - \left(\frac{\beta_1}{\rho_i^s + \rho_i^d}\right) s_t - \left(\frac{\beta_1}{\rho_i^s + \rho_i^d}\right) \varepsilon_{s_{it}} \\
&- \left(\frac{\beta_2}{\rho_i^s + \rho_i^d}\right) \xi_{2t} + \left(\frac{\alpha_2}{\rho_i^s + \rho_i^d}\right) \xi_{1t} + \frac{u_{it}^d - u_{it}^s}{\rho_i^s + \rho_i^d},
\end{aligned}
$$

and this this implies that:

$$
\begin{aligned}
Y_t = &-\frac{\int_0^1 \left(\frac{\beta_1}{\rho_i^s + \rho_i^d}\right) di}{\left[1 - \int_0^1 \left(\frac{\alpha_1}{\rho_i^s + \rho_i^d}\right) di\right]} s_t - \frac{\int_0^1 \left(\frac{\beta_1}{\rho_i^s + \rho_i^d}\right) \varepsilon_{s_{it}} di}{\left[1 - \int_0^1 \left(\frac{\alpha_1}{\rho_i^s + \rho_i^d}\right) di\right]} - \frac{\int_0^1 \left(\frac{\beta_2}{\rho_i^s + \rho_i^d}\right) di}{\left[1 - \int_0^1 \left(\frac{\alpha_1}{\rho_i^s + \rho_i^d}\right) di\right]} \xi_{2t} \\
&+ \frac{\int_0^1 \left(\frac{\alpha_2}{\rho_i^s + \rho_i^d}\right) di}{\left[1 - \int_0^1 \left(\frac{\alpha_1}{\rho_i^s + \rho_i^d}\right) di\right]} \xi_{1t} + \frac{\int_0^1 \frac{u_{it}^d - u_{it}^s}{\rho_i^s + \rho_i^d} di}{\left[1 - \int_0^1 \left(\frac{\alpha_1}{\rho_i^s + \rho_i^d}\right) di\right]}.
\end{aligned}
$$

And thus replacing in the original equation:

$$
\begin{aligned}
y_{it} = &-\left\{\left(\frac{\beta_1}{\rho_i^s + \rho_i^d}\right) + \left(\frac{\alpha_1}{\rho_i^s + \rho_i^d}\right) \frac{\int_0^1 \left(\frac{\beta_1}{\rho_i^s + \rho_i^d}\right) di}{\left[1 - \int_0^1 \left(\frac{\alpha_1}{\rho_i^s + \rho_i^d}\right) di\right]}\right\} s_t - \left(\frac{\beta_1}{\rho_i^s + \rho_i^d}\right) \varepsilon_{s_{it}} \\
&- \left(\frac{\beta_2}{\rho_i^s + \rho_i^d} + \left(\frac{\alpha_1}{\rho_i^s + \rho_i^d}\right) \frac{\int_0^1 \left(\frac{\beta_2}{\rho_i^s + \rho_i^d}\right) di}{\left[1 - \int_0^1 \left(\frac{\alpha_1}{\rho_i^s + \rho_i^d}\right) di\right]}\right) \xi_{2t} + \left(\frac{\alpha_2}{\rho_i^s + \rho_i^d} + \left(\frac{\alpha_1}{\rho_i^s + \rho_i^d}\right) \frac{\int_0^1 \left(\frac{\alpha_2}{\rho_i^s + \rho_i^d}\right) di}{\left[1 - \int_0^1 \left(\frac{\alpha_1}{\rho_i^s + \rho_i^d}\right) di\right]}\right) \xi_{1t} \\
&- \left(\frac{\alpha_1 \beta_1}{\rho_i^s + \rho_i^d}\right) \frac{\int_0^1 \left(\frac{\varepsilon_{s_{it}}}{\rho_i^s + \rho_i^d}\right) di}{\left[1 - \int_0^1 \left(\frac{\alpha_1}{\rho_i^s + \rho_i^d}\right) di\right]} + \frac{u_{it}^d - u_{it}^s}{\rho_i^s + \rho_i^d} + \left(\frac{\alpha_1}{\rho_i^s + \rho_i^d}\right) \frac{\int_0^1 \frac{u_{it}^d - u_{it}^s}{\rho_i^s + \rho_i^d} di}{\left[1 - \int_0^1 \left(\frac{\alpha_1}{\rho_i^s + \rho_i^d}\right) di\right]}.
\end{aligned}
$$

So let us define:

$$\eta^i_{MG} := -\left\{\left(\frac{\beta_1}{\rho^s_i+\rho^d_i}\right)+\left(\frac{\alpha_1}{\rho^s_i+\rho^d_i}\right)\frac{\int_0^1\left(\frac{\beta_1}{\rho^s_i+\rho^d_i}\right)di}{\left[1-\int_0^1\left(\frac{\alpha_1}{\rho^s_i+\rho^d_i}\right)di\right]}\right\}, \qquad \eta^i_{ML} := -\left(\frac{\beta_1}{\rho^s_i+\rho^d_i}\right)$$

$$\delta^i_1 := \left(\frac{\alpha_2}{\rho^s_i+\rho^d_i}+\left(\frac{\alpha_1}{\rho^s_i+\rho^d_i}\right)\frac{\int_0^1\left(\frac{\alpha_2}{\rho^s_i+\rho^d_i}\right)di}{\left[1-\int_0^1\left(\frac{\alpha_1}{\rho^s_i+\rho^d_i}\right)di\right]}\right), \qquad \delta^i_2 := -\left(\frac{\beta_2}{\rho^s_i+\rho^d_i}+\left(\frac{\alpha_1}{\rho^s_i+\rho^d_i}\right)\frac{\int_0^1\left(\frac{\beta_2}{\rho^s_i+\rho^d_i}\right)di}{\left[1-\int_0^1\left(\frac{\alpha_1}{\rho^s_i+\rho^d_i}\right)di\right]}\right)$$

$$\eta_{macro} := -\frac{\int_0^1\left(\frac{\beta_1}{\rho^s_i+\rho^d_i}\right)di}{\left[1-\int_0^1\left(\frac{\alpha_1}{\rho^s_i+\rho^d_i}\right)di\right]}, \qquad \epsilon_{it} := \frac{u^d_{it}-u^s_{it}}{\rho^s_i+\rho^d_i}.$$

Thus, note that by direct computation we have that $\int_0^1 \eta^i_{MG} di = \eta_{macro}$. Moreover, since we are assuming that $\rho^s_i+\rho^d_i \in \mathbb{R}_{\neq 0}$ and given our assumptions on $\varepsilon_{s_{it}}$, u^d_{it} and u^s_{it}, it follows that $\int_0^1\left(\frac{\varepsilon_{s_{it}}}{\rho^s_i+\rho^d_i}\right)di$, $\int_0^1 \frac{u^d_{it}}{\rho^s_i+\rho^d_i}di$ and $\int_0^1 \frac{u^s_{it}}{\rho^s_i+\rho^d_i}di$ converge in $L_2 - norm$ to zero.[53] Finally, ϵ_{it} inherits the properties of the supply and demand shocks. This, along with the definitions, completes the proof. ■

Proposition 2

Proof. We will indicate how to derive the whole log-linearized equilibrium. As always in these models, we have the two-step solution which gives:

$$c_{it} = C_t\left(\frac{p_{it}}{P_t}\right)^{-\varepsilon}$$
$$P_t := \left(\int_0^1 p_{it}^{-(\varepsilon-1)}di\right)^{-\frac{1}{\varepsilon-1}}$$
$$\int_0^1 p_{it}c_{it}di = P_tC_t,$$

given some desired consumption C_t in period t. And the F.O.C.s for the representative agent are:

$$C^\sigma_t L^\varphi_t = \frac{W_t}{P_t}$$
$$1 = \mathbb{E}_t\left\{\beta\left(\frac{C_{t+1}}{C_t}\right)^{-\sigma}\frac{P_t}{P_{t+1}}(1+i_t)\right\}.$$

Now, moving on to pricing decisions, in every region the firms that get to reset their prices solve:

$$\max_{p^*_{it}}\sum_{k=0}^{\infty}\theta^k\mathbb{E}_t\left\{Q_{t,t+k}\left[p^*_{it}y_{it+k|t}-\Psi_{nt+k}\left(y_{it+k|t}\right)\right]\right\}$$
$$s.t.: y_{it+k|t} = (C_{t+k}+G_{t+k})\left(\frac{p^*_{it}}{P_{t+k}}\right)^{-\varepsilon}$$

[53] See Acemoglu and Jensen (2012).

where $y_{it+k|t}$ is the output of a firm i in period $t+k$ that is stuck with p^*_{it} since period t and P_{t+k}, C_{t+k}, G_{t+k} and $Q_{t,t+k}$ are taken as given. Note that since the problem that each firm reseting its price in region n faces is the same, all such firms will set the same price, which I will call p^*_{nt} from now on. Thus:

$$\forall n \in \{1, ..., N\}\,,\ p^*_{it} = p^*_{nt}\ \forall i \in [\overline{\omega}_{n-1}, \overline{\omega}_n)\,.$$

Moreover, note that from $y_{it+k|t} = (C_{t+k} + G_{t+k}) \left(\frac{p^*_{it}}{P_{t+k}}\right)^{-\varepsilon}$ this also implies that:

$$n \in \{1, ..., N\}\,,\ y_{it+k|t} = y_{nt+k|t}\ \forall i \in [\overline{\omega}_{n-1}, \overline{\omega}_n)\,.$$

And it also implies that:

$$P_t = \left((1-\theta) \sum_{n=1}^{N} \frac{1}{N} \left\{\sum_{k=0}^{\infty} p^{*-(\varepsilon-1)}_{n,t-k} \theta^k\right\}\right)^{-\frac{1}{\varepsilon-1}}. \tag{76}$$

The problem can be rewritten as:

$$\max_{p^*_{nt}} \sum_{k=0}^{\infty} \theta^k \mathbb{E}_t \left\{Q_{t,t+k} \left[(C_{t+k} + G_{t+k}) \frac{p^{*1-\varepsilon}_{nt}}{P^{-\varepsilon}_{t+k}} - \Psi_{nt+k} \left((C_{t+k} + G_{t+k}) \left(\frac{p^*_{nt}}{P_{t+k}}\right)^{-\varepsilon}\right)\right]\right\}$$

and the F.O.C. for this problem gives:

$$p^*_{nt} = \frac{\varepsilon}{\varepsilon - 1} \frac{\sum_{k=0}^{\infty} \theta^k \mathbb{E}_t \left\{Q_{t,t+k} y_{nt+k|t} \psi_{nt+k}\left(y_{nt+k|t}\right)\right\}}{\sum_{k=0}^{\infty} \theta^k \mathbb{E}_t \left\{Q_{t,t+k} y_{nt+k|t}\right\}},$$

which is the usual result of resetting the price to a weighted average of future marginal costs.

This, in turn, implies that the price level dynamics are:

$$P_t = \left((1-\theta) P^{*-(\varepsilon-1)}_t + \theta P^{-(\varepsilon-1)}_{t-1}\right)^{-\frac{1}{\varepsilon-1}}$$

which follows from (76) and from defining:

$$P^*_t := \left(\sum_{n=1}^{N} \frac{1}{N} p^{*-(\varepsilon-1)}_{n,t}\right)^{-\frac{1}{\varepsilon-1}},$$

which is the price index that would arise in this economy if $\theta \to 0$, i.e., if prices were flexible.[54]

Define the gross inflation rate as $\Pi_t := \frac{P_t}{P_{t-1}}$, and the net inflation rate as $\pi_t := \ln \Pi_t = \ln P_t - \ln P_{t-1}$. Moreover, define:

$$P_{n,t} := \left(N \int_{\overline{\omega}_{n-1}}^{\overline{\omega}_n} p^{-(\varepsilon-1)}_{it} di\right)^{-\frac{1}{\varepsilon-1}} = \left((1-\theta) p^{*-(\varepsilon-1)}_{n,t} + \theta P^{-(\varepsilon-1)}_{n,t-1}\right)^{-\frac{1}{\varepsilon-1}}$$

[54] This can be seen from:

$$\lim_{\theta\to 0} \left((1-\theta) \sum_{n=1}^{N} \frac{1}{N} \left\{\sum_{k=0}^{\infty} p^{*-(\varepsilon-1)}_{n,t-k} \theta^k\right\}\right)^{-\frac{1}{\varepsilon-1}} = \left(\sum_{n=1}^{N} \frac{1}{N} p^{*-(\varepsilon-1)}_{n,t}\right)^{-\frac{1}{\varepsilon-1}} = P^*_t.$$

and thus note that with the same arguments as before, we see that $p^*_{n,t}$ is the price index that would arise in region n if prices were flexible. Hence, if we define:

$$P^*_{n,t} := \left(N\int_{\overline{\omega}_{n-1}}^{\overline{\omega}_n} p_{it}^{*-(\varepsilon-1)}di\right)^{-\frac{1}{\varepsilon-1}} = p^*_{nt},$$

then $P_{n,t} = \left((1-\theta)P_{n,t}^{*-(\varepsilon-1)} + \theta P_{n,t-1}^{-(\varepsilon-1)}\right)^{-\frac{1}{\varepsilon-1}}$.

Now, as we move forward, the following definitions will become useful. First define $Y_t^C := \left[\int_0^1 y_{it}^{\frac{\varepsilon-1}{\varepsilon}}di\right]^{\frac{\varepsilon}{\varepsilon-1}} = C_t + G_t$. And moreover, since $P_tC_t = \int_0^1 p_{it}c_{it}di$ and $\int p_{it}G_t\left(\frac{p_{it}}{P_t}\right)^{-\varepsilon_j}di = P_tG_t$, we have that $P_tY_t^C = \int_0^1 p_{it}y_{it}di$ and $y_{it} = c_{it} + g_{it} = Y_t^C\left(\frac{p_{it}}{P_t}\right)^{-\varepsilon}$. Next, define:

$$Y_{n,t} := \left[N\int_{\overline{\omega}_{n-1}}^{\overline{\omega}_n} y_{it}^{\frac{\varepsilon-1}{\varepsilon}}di\right]^{\frac{\varepsilon}{\varepsilon-1}} = Y_t^C\left(\frac{P_{n,t}}{P_t}\right)^{-\varepsilon}$$

$$G_{n,t} := \left[N\int_{\overline{\omega}_{n-1}}^{\overline{\omega}_n} g_{it}^{\frac{\varepsilon-1}{\varepsilon}}di\right]^{\frac{\varepsilon}{\varepsilon-1}} = G_t\left(\frac{P_{n,t}}{P_t}\right)^{-\varepsilon}.$$

And note that $Y_t^C = Y_{n,t}\left(\frac{P_{n,t}}{P_t}\right)^{\varepsilon}$ and thus $y_{it} = Y_{n,t}\left(\frac{p_{it}}{P_{n,t}}\right)^{-\varepsilon}$. We also have that $\left(\sum_{n=1}^N \frac{1}{N}P_{n,t}^{-(\varepsilon-1)}\right)^{\left(-\frac{1}{(\varepsilon-1)}\right)} = P_t$ and $\left(\sum_{n=1}^N \frac{1}{N}Y_{n,t}^{\frac{\varepsilon-1}{\varepsilon}}\right)^{\frac{\varepsilon}{\varepsilon-1}} = Y_t^C$. Finally, define $PR_{nt} := \frac{P_{n,t}}{P_t}$.

In terms of the aggregate productions function, from $y_{it} = a_{nt}l_{it}^{1-\alpha_n}$ we have that $l_{it} = \left(\frac{y_{it}}{a_{nt}}\right)^{\frac{1}{1-\alpha_n}}$ and thus using $y_{it} = Y_{n,t}\left(\frac{p_{it}}{P_{n,t}}\right)^{-\varepsilon}$:

$$L_{n,t} := \int_{\overline{\omega}_{n-1}}^{\overline{\omega}_n} l_{it}di = a_{nt}^{-\frac{1}{1-\alpha_n}}Y_{n,t}^{\frac{1}{1-\alpha_n}}\int_{\overline{\omega}_{n-1}}^{\overline{\omega}_n}\left(\frac{p_{it}}{P_{n,t}}\right)^{\frac{-\varepsilon}{1-\alpha_n}}di$$

and hence $Y_{n,t} = a_{nt}L_{n,t}^{1-\alpha_n}\left[\int_{\overline{\omega}_{n-1}}^{\overline{\omega}_n}\left(\frac{p_{it}}{P_{n,t}}\right)^{\frac{-\varepsilon}{1-\alpha_n}}di\right]^{-(1-\alpha_n)}$. And defining:

$$D_{n,t} := \left[\int_{\overline{\omega}_{n-1}}^{\overline{\omega}_n}\left(\left(\frac{p_{it}}{P_{n,t}}\right)^{-\varepsilon}\right)^{\frac{1}{1-\alpha_n}}di\right]^{(1-\alpha_n)}$$

we get $Y_{n,t} = a_{n,t}L_{n,t}^{1-\alpha_n}D_{n,t}^{-1}$, and thus $L_{n,t} = \left(D_{n,t}^{-1}a_{n,t}\right)^{-\frac{1}{1-\alpha_n}}Y_{n,t}^{\frac{1}{1-\alpha_n}}$. Which implies:

$$L_t = \sum_{n=1}^N L_{n,t} = \sum_{n=1}^N\left(D_{n,t}^{-1}a_{n,t}\right)^{-\frac{1}{1-\alpha_n}}\left(\frac{P_{n,t}}{P_t}\right)^{\frac{-\varepsilon}{1-\alpha_n}}\left(Y_t^C\right)^{\frac{1}{1-\alpha_n}}.$$

Now, on to the non-stochastic steady state. We look for a non-stochastic steady state with no inflation, and symmetry inside each region. Thus we have that:

$$\frac{P_t}{P_{t-1}} = \overline{\Pi} = \frac{P_{nt}}{P_{nt-1}} = \overline{\Pi}_n = 1 = \Pi_n^* = \frac{P^*_{nt}}{P_{nt}}, \overline{\pi}_n = 0, \frac{W_t}{P_{nt}} = \overline{\left(\frac{W}{P_n}\right)}$$
$$\forall n \in \{1,\ldots,N\},\ p_{it} = p_{jt} =: \overline{p_{nt}}\ \forall i,j \in [\overline{\omega}_{n-1},\overline{\omega}_n).$$

Note that what we asked for inflation implies that in the non-stochastic steady state:

$$\frac{\Pi_{nt}}{\Pi_t} = \frac{\overline{PR_{nt}}}{\overline{PR_{nt-1}}} = 1 \implies \overline{PR_{nt}} = \overline{PR_{nt-1}} = \overline{PR_n}.$$

I omit the whole derivation to save space because, having derived all the equilibrium equations, it is straightforward. After manipulating the equilibrium equations, we arrive at:

$$1 = \sum_{n=1}^{N} \frac{1}{N} \left(\overline{Y}^C\right)^{-\frac{\varepsilon-1}{\varepsilon}} \left[\frac{\left(\frac{\varepsilon}{\varepsilon-1}\frac{1}{1-\alpha_n}\overline{a}_n^{-\frac{1}{1-\alpha_n}}\right)^{-\varepsilon}}{\left(\frac{\varepsilon}{\varepsilon-1}\frac{1}{1-\alpha_k}\overline{a}_k^{-\frac{1}{1-\alpha_k}}\right)^{-\varepsilon}} \right]^{\frac{\frac{\varepsilon-1}{\varepsilon}}{1+\varepsilon\frac{\alpha_n}{1-\alpha_n}}} \left(\overline{Y}^{C\frac{1+\varepsilon\frac{\alpha_k}{1-\alpha_k}}{1+\varepsilon\frac{\alpha_n}{1-\alpha_n}}} \left(\frac{\overline{p_k}}{P}\right)^{(-\varepsilon)\frac{1+\varepsilon\frac{\alpha_k}{1-\alpha_k}}{1+\varepsilon\frac{\alpha_n}{1-\alpha_n}}} \right)^{\frac{(\varepsilon-1)}{\varepsilon}}$$

$$\overline{L} = \sum_{n=1}^{N} \left\{ \frac{1}{N}^{1-\alpha_n} \left(\begin{array}{c} \left(\overline{Y}^C\right)^{\frac{1}{\varepsilon}} \left[\frac{\left(\frac{\varepsilon}{\varepsilon-1}\frac{1}{1-\alpha_n}\overline{a}_n^{-\frac{1}{1-\alpha_n}}\right)^{-\varepsilon}}{\left(\frac{\varepsilon}{\varepsilon-1}\frac{1}{1-\alpha_k}\overline{a}_k^{-\frac{1}{1-\alpha_k}}\right)^{-\varepsilon}} \right]^{\frac{-\frac{1}{\varepsilon}}{1+\varepsilon\frac{\alpha_n}{1-\alpha_n}}} \\ * \left(\overline{Y}^{C\frac{1+\varepsilon\frac{\alpha_k}{1-\alpha_k}}{1+\varepsilon\frac{\alpha_n}{1-\alpha_n}}} \left(\frac{\overline{p_k}}{P}\right)^{(-\varepsilon)\frac{1+\varepsilon\frac{\alpha_k}{1-\alpha_k}}{1+\varepsilon\frac{\alpha_n}{1-\alpha_n}}} \right)^{-\frac{1}{\varepsilon}} \end{array} \right)^{-\varepsilon} \overline{a}_{n,t}^{-1}\overline{Y}^C \right\}^{\frac{1}{1-\alpha_n}}$$

$$\left(\overline{Y}^C - \overline{G}\right)^{\sigma(\varepsilon-1)} \overline{L}^{\varphi(\varepsilon-1)}$$

$$= \sum_{n=1}^{N} \frac{1}{N} \left(\frac{\varepsilon}{\varepsilon-1} \frac{\overline{a}_n^{-\frac{1}{1-\alpha_n}}}{1-\alpha_n} \left(\overline{Y}^C \left(\begin{array}{c} \left(\overline{Y}^C\right)^{\frac{1}{\varepsilon}} \left[\frac{\left(\frac{\varepsilon}{\varepsilon-1}\frac{1}{1-\alpha_n}\overline{a}_n^{-\frac{1}{1-\alpha_n}}\right)^{-\varepsilon}}{\left(\frac{\varepsilon}{\varepsilon-1}\frac{1}{1-\alpha_k}\overline{a}_k^{-\frac{1}{1-\alpha_k}}\right)^{-\varepsilon}} \right]^{\frac{-\frac{1}{\varepsilon}}{1+\varepsilon\frac{\alpha_n}{1-\alpha_n}}} \\ * \left(\overline{Y}^{C\frac{1+\varepsilon\frac{\alpha_k}{1-\alpha_k}}{1+\varepsilon\frac{\alpha_n}{1-\alpha_n}}} \left(\frac{\overline{p_k}}{P}\right)^{(-\varepsilon)\frac{1+\varepsilon\frac{\alpha_k}{1-\alpha_k}}{1+\varepsilon\frac{\alpha_n}{1-\alpha_n}}} \right)^{-\frac{1}{\varepsilon}} \end{array} \right)^{-\varepsilon} \right)^{\frac{\alpha_n}{1-\alpha_n}} \right)^{-(\varepsilon-1)}$$

which is a system of 3 equations in $\overline{Y}^C$, $\overline{L}$ and $\left(\frac{\overline{p_k}}{P}\right)$. We assume there are economies for which there is a solution to this system. After solving for those three variables, we can recover the rest.

Now, we want to log-linearize around the previous non-stochastic steady state. Define $\widetilde{\iota}_t := \iota_t - \overline{\iota} \approx i_t - \overline{i}$, where $\overline{\iota} := \ln\left(1 + \overline{i}\right)$. Define also $\pi_t := \ln(\Pi_t) = \ln(P_t) - \ln(P_{t-1})$. After some computations with the equilibrium equations, we get the system:

$$
\begin{aligned}
\widetilde{L}_t &= \sum_{n=1}^{N} \frac{\overline{L}_n}{\overline{L}_t}\left(\frac{1}{1-\alpha_n}\right)\left(\widetilde{Y}_{n,t} - \widetilde{a}_{n,t}\right) \\
0 &= \sum_{n=1}^{N} \frac{1}{N}\overline{PR_n}^{-(\varepsilon-1)}\widetilde{PR_{n,t}} \\
\widetilde{Y}_t^C &= \sum_{n=1}^{N} \frac{1}{N}\left(\frac{\overline{Y}_n}{\overline{Y}^C}\right)^{\frac{\varepsilon-1}{\varepsilon}} \widetilde{Y}_{n,t} \\
\pi_{nt} &= (1-\theta)\left\{\ln P_{nt}^* - \ln P_{nt-1}\right\} \\
\sigma\widetilde{C}_t + \varphi\widetilde{L}_t &= \widetilde{\left(\frac{W_t}{P_t}\right)} \\
\widetilde{C}_t &= \frac{1}{(1-\mathcal{G})}\widetilde{Y}_t^C - \frac{\mathcal{G}}{(1-\mathcal{G})}\widetilde{G}_t \\
\widetilde{\left(\frac{W_t}{P_t}\right)} &= \widetilde{\left(\frac{W_t}{P_{nt}}\right)} + \frac{1}{\varepsilon}\widetilde{Y}_t^C - \frac{1}{\varepsilon}\widetilde{Y}_{nt} \\
\widetilde{Y}_t^C &= \widetilde{Y}_{n,t} + \varepsilon\widetilde{PR}_{nt} \\
0 &= \mathbb{E}_t\left\{-\frac{\sigma}{(1-\mathcal{G})}\widetilde{Y}_{t+1}^C + \frac{\sigma\mathcal{G}}{(1-\mathcal{G})}\widetilde{G}_{t+1} + \frac{\sigma}{(1-\mathcal{G})}\widetilde{Y}_t^C - \frac{\sigma\mathcal{G}}{(1-\mathcal{G})}\widetilde{G}_t - \pi_{t+1} + \widetilde{\imath}_t\right\} \\
\left[ln\left(P_{nt}^*\right) - ln\left(P_{nt-1}\right)\right] &= (1-\theta\beta)\left(\frac{1-\alpha_n}{1-\alpha_n+\varepsilon\alpha_n}\right)MC_{n,t}^{Reg} + \pi_{nt} + \theta\beta\mathbb{E}_t\left\{ln\left(P_{nt+1}^*\right) - ln\left(P_{nt}\right)\right\} \\
MC_{n,t}^{Reg} &= \widetilde{\left(\frac{W_t}{P_{nt}}\right)} + \frac{\alpha_n}{1-\alpha_n}\widetilde{Y_{n,t}} - \frac{1}{1-\alpha_n}\widetilde{a_{nt}} \\
\widetilde{PR}_{nt} - \widetilde{PR}_{nt-1} &= \pi_{nt} - \pi_t.
\end{aligned}
$$

And we have to add to the system the Taylor rule $\widetilde{\imath}_t = v_t + \phi_\pi \pi_t + \phi_y \widetilde{Y}_t^C$. If we define $\lambda_n := \left(\frac{1-\theta}{\theta}\right)(1-\theta\beta)\left(\frac{1-\alpha_n}{1-\alpha_n+\varepsilon\alpha_n}\right)$ and manipulate this system, we arrive at the equations in the main text. ■

Corollary 1

Proof. The result is a particular case of Proposition 12 and, moreover, since it is based only on the first part of that proposition, it is just a direct application of Blanchard and Kahn (1980). Moreover, (12) follows directly from log-linearizing the aggregator. ■

Proposition 3

Proof. The F.O.C.s of the RA in this case are given by:

$$
\begin{aligned}
C_t^\sigma l_{it}^{\varphi_i} &= \frac{W_{it}}{P_t} \\
1 &= \mathbb{E}_t\left\{\beta\left(\frac{C_{t+1}}{C_t}\right)^{-\sigma}\frac{P_t}{P_{t+1}}(1+i_t)\right\}.
\end{aligned}
$$

In every region, the firm solves:

$$\max_{p_{it}} \left[p_{it} y_{it} - W_{it} y_{it}^{\frac{1}{1-\alpha_i}} a_{it}^{-\frac{1}{1-\alpha_i}} \right]$$
$$s.t.: \; y_{it} = (C_t + G_t) \left(\frac{p_{it}}{P_t} \right)^{-\varepsilon}$$

and the F.O.C. for this problem gives $p_{it} = \left(\frac{\varepsilon}{\varepsilon-1}\right) W_{it} \frac{1}{1-\alpha_i} y_{it}^{\frac{\alpha_i}{1-\alpha_i}} a_{it}^{-\frac{1}{1-\alpha_i}}$, which is the usual markup over marginal cost.

Let me define $Y_t^C := \left[\int_0^1 y_{it}^{\frac{\varepsilon-1}{\varepsilon}} di \right]^{\frac{\varepsilon}{\varepsilon-1}} = C_t + G_t$. And moreover, since $P_t C_t = \int_0^1 p_{it} c_{it} di$ and $\int p_{it} G_t \left(\frac{p_{it}}{P_t}\right)^{-\varepsilon j} di = P_t G_t$, we have that $P_t Y_t^C = \int_0^1 p_{it} y_{it} di$. Also, $y_{it} = Y_t^C \left(\frac{p_{it}}{P_t}\right)^{-\varepsilon}$.

Now, for the non-stochastic steady state, the same steps as in Proposition 2 take to a single equation in one unknown. I assume there are economies for which there is a solution to this equation. Moving forward, now we want to log-linearize the equilibrium equations around the non-stochastic steady state. After working with the equilibrium equations, we obtain the system:

$$\left(\frac{\alpha_i + \varphi_i}{1-\alpha_i} + \frac{1}{\varepsilon} \right) \widetilde{Y}_{it} = \left(\frac{1}{\varepsilon} - \frac{\sigma}{(1-\mathcal{G})} \right) \int_0^1 \left(\frac{\overline{Y}_j}{\overline{Y}^C} \right)^{\frac{\varepsilon-1}{\varepsilon}} \widetilde{Y}_{jt} dj + \frac{\sigma \mathcal{G}}{(1-\mathcal{G})} \widetilde{G}_t + \left(\frac{1+\varphi_i}{1-\alpha_i} \right) \widetilde{a}_{it}$$

and thus:

$$\int_0^1 \left(\frac{\overline{Y}_i}{\overline{Y}^C} \right)^{\frac{\varepsilon-1}{\varepsilon}} \widetilde{Y}_{it} di = \frac{\int_0^1 \left(\frac{\overline{Y}_i}{\overline{Y}^C} \right)^{\frac{\varepsilon-1}{\varepsilon}} \frac{(1-\alpha_i)\varepsilon\sigma\mathcal{G}}{(1-\mathcal{G})\left\{(\alpha_i+\varphi_i)\varepsilon+1-\alpha_i\right\}} di}{\left[1 - \int_0^1 \left(\frac{\overline{Y}_i}{\overline{Y}^C} \right)^{\frac{\varepsilon-1}{\varepsilon}} \left(\frac{(1-\alpha_i)\{(1-\mathcal{G})-\varepsilon\sigma\}}{\left\{(\alpha_i+\varphi_i)\varepsilon+1-\alpha_i\right\}(1-\mathcal{G})} \right) di \right]} \widetilde{G}_t$$
$$+ \int_0^1 \frac{\left(\frac{\overline{Y}_i}{\overline{Y}^C} \right)^{\frac{\varepsilon-1}{\varepsilon}} \left(\frac{\varepsilon(1+\varphi_i)}{(\alpha_i+\varphi_i)\varepsilon+1-\alpha_i} \right)}{\left[1 - \int_0^1 \left(\frac{\overline{Y}_i}{\overline{Y}^C} \right)^{\frac{\varepsilon-1}{\varepsilon}} \left(\frac{(1-\alpha_i)\{(1-\mathcal{G})-\varepsilon\sigma\}}{\left\{(\alpha_i+\varphi_i)\varepsilon+1-\alpha_i\right\}(1-\mathcal{G})} \right) di \right]} \widetilde{a}_{it} di$$

so we get:[55]

$$\widetilde{Y}_{it} = \left\{ \begin{array}{c} \frac{(1-\alpha_i)\varepsilon\sigma\mathcal{G}}{(1-\mathcal{G})\left\{\left(\alpha_i+\varphi_i\right)\varepsilon+1-\alpha_i\right\}} \\ +\left(\frac{(1-\alpha_i)\{(1-\mathcal{G})-\varepsilon\sigma\}}{\left\{\left(\alpha_i+\varphi_i\right)\varepsilon+1-\alpha_i\right\}(1-\mathcal{G})}\right) \frac{\int_0^1 \left(\frac{\overline{Y}_i}{\overline{Y}^C}\right)^{\frac{\varepsilon-1}{\varepsilon}} \frac{(1-\alpha_i)\varepsilon\sigma\mathcal{G}}{(1-\mathcal{G})\left\{\left(\alpha_i+\varphi_i\right)\varepsilon+1-\alpha_i\right\}} di}{\left[1-\int_0^1 \left(\frac{\overline{Y}_i}{\overline{Y}^C}\right)^{\frac{\varepsilon-1}{\varepsilon}} \left(\frac{(1-\alpha_i)\{(1-\mathcal{G})-\varepsilon\sigma\}}{\left\{\left(\alpha_i+\varphi_i\right)\varepsilon+1-\alpha_i\right\}(1-\mathcal{G})}\right) di\right]} \end{array} \right\} \widetilde{G}_t$$

$$+\left(\frac{\varepsilon\left(1+\varphi_i\right)}{\left(\alpha_i+\varphi_i\right)\varepsilon+1-\alpha_i}\right)\widetilde{a}_{it}$$

$$+\left(\frac{(1-\alpha_i)\{(1-\mathcal{G})-\varepsilon\sigma\}}{\left\{\left(\alpha_i+\varphi_i\right)\varepsilon+1-\alpha_i\right\}(1-\mathcal{G})}\right)\int_0^1 \frac{\left(\frac{\overline{Y}_i}{\overline{Y}^C}\right)^{\frac{\varepsilon-1}{\varepsilon}} \left(\frac{\varepsilon\left(1+\varphi_i\right)}{\left(\alpha_i+\varphi_i\right)\varepsilon+1-\alpha_i}\right)}{\left[1-\int_0^1 \left(\frac{\overline{Y}_i}{\overline{Y}^C}\right)^{\frac{\varepsilon-1}{\varepsilon}} \left(\frac{(1-\alpha_i)\{(1-\mathcal{G})-\varepsilon\sigma\}}{\left\{\left(\alpha_i+\varphi_i\right)\varepsilon+1-\alpha_i\right\}(1-\mathcal{G})}\right) di\right]}\widetilde{a}_{it}di.$$

Finally, using that $\widetilde{a}_{it} = \widetilde{a}_t + \varepsilon_{\widetilde{a}_{it}}$ and given the assumptions on $\varepsilon_{\widetilde{a}_{it}}$ (as in Proposition 1), we have that:

$$\int_0^1 \frac{\left(\frac{\overline{Y}_i}{\overline{Y}^C}\right)^{\frac{\varepsilon-1}{\varepsilon}} \left(\frac{\varepsilon\left(1+\varphi_i\right)}{\left(\alpha_i+\varphi_i\right)\varepsilon+1-\alpha_i}\right)}{\left[1-\int_0^1 \left(\frac{\overline{Y}_i}{\overline{Y}^C}\right)^{\frac{\varepsilon-1}{\varepsilon}} \left(\frac{(1-\alpha_i)\{(1-\mathcal{G})-\varepsilon\sigma\}}{\left\{\left(\alpha_i+\varphi_i\right)\varepsilon+1-\alpha_i\right\}(1-\mathcal{G})}\right) di\right]}\widetilde{a}_{it}di = \int_0^1 \frac{\left(\frac{\overline{Y}_i}{\overline{Y}^C}\right)^{\frac{\varepsilon-1}{\varepsilon}} \left(\frac{\varepsilon\left(1+\varphi_i\right)}{\left(\alpha_i+\varphi_i\right)\varepsilon+1-\alpha_i}\right)}{\left[1-\int_0^1 \left(\frac{\overline{Y}_i}{\overline{Y}^C}\right)^{\frac{\varepsilon-1}{\varepsilon}} \left(\frac{(1-\alpha_i)\{(1-\mathcal{G})-\varepsilon\sigma\}}{\left\{\left(\alpha_i+\varphi_i\right)\varepsilon+1-\alpha_i\right\}(1-\mathcal{G})}\right) di\right]}di\widetilde{a}_t.$$

Hence, defining:

$$\eta^i_{MG} := \frac{(1-\alpha_i)\varepsilon\sigma\mathcal{G}}{(1-\mathcal{G})\left\{\left(\alpha_i+\varphi_i\right)\varepsilon+1-\alpha_i\right\}}$$
$$+\left(\frac{(1-\alpha_i)\{(1-\mathcal{G})-\varepsilon\sigma\}}{\left\{\left(\alpha_i+\varphi_i\right)\varepsilon+1-\alpha_i\right\}(1-\mathcal{G})}\right)\left(\frac{\int_0^1 \left(\frac{\overline{Y}_i}{\overline{Y}^C}\right)^{\frac{\varepsilon-1}{\varepsilon}} \frac{(1-\alpha_i)\varepsilon\sigma\mathcal{G}}{(1-\mathcal{G})\left\{\left(\alpha_i+\varphi_i\right)\varepsilon+1-\alpha_i\right\}} di}{\left[1-\int_0^1 \left(\frac{\overline{Y}_i}{\overline{Y}^C}\right)^{\frac{\varepsilon-1}{\varepsilon}} \left(\frac{(1-\alpha_i)\{(1-\mathcal{G})-\varepsilon\sigma\}}{\left\{\left(\alpha_i+\varphi_i\right)\varepsilon+1-\alpha_i\right\}(1-\mathcal{G})}\right) di\right]}\right)$$
$$\lambda_i := \left(\frac{\varepsilon\left(1+\varphi_i\right)}{\left(\alpha_i+\varphi_i\right)\varepsilon+1-\alpha_i}\right)$$
$$+\left(\frac{(1-\alpha_i)\{(1-\mathcal{G})-\varepsilon\sigma\}}{\left\{\left(\alpha_i+\varphi_i\right)\varepsilon+1-\alpha_i\right\}(1-\mathcal{G})}\right)\left(\frac{\int_0^1 \left(\frac{\overline{Y}_i}{\overline{Y}^C}\right)^{\frac{\varepsilon-1}{\varepsilon}} \left(\frac{\varepsilon\left(1+\varphi_i\right)}{\left(\alpha_i+\varphi_i\right)\varepsilon+1-\alpha_i}\right) di}{\left[1-\int_0^1 \left(\frac{\overline{Y}_i}{\overline{Y}^C}\right)^{\frac{\varepsilon-1}{\varepsilon}} \left(\frac{(1-\alpha_i)\{(1-\mathcal{G})-\varepsilon\sigma\}}{\left\{\left(\alpha_i+\varphi_i\right)\varepsilon+1-\alpha_i\right\}(1-\mathcal{G})}\right) di\right]}\right)$$
$$\lambda_i^{\varepsilon_a} := \left(\frac{\varepsilon\left(1+\varphi_i\right)}{\left(\alpha_i+\varphi_i\right)\varepsilon+1-\alpha_i}\right),$$

[55] Of course, I assumed $\left[1-\int_0^1 \left(\frac{\overline{Y}_i}{\overline{Y}^C}\right)^{\frac{\varepsilon-1}{\varepsilon}} \left(\frac{(1-\alpha_i)\{(1-\mathcal{G})-\varepsilon\sigma\}}{\left\{\left(\alpha_i+\varphi_i\right)\varepsilon+1-\alpha_i\right\}(1-\mathcal{G})}\right) di\right] \neq 0$.

where note that by direct computation:

$$\eta_{macro} := \frac{\partial \widetilde{Y}_t}{\partial \widetilde{G}_t} = \int_0^1 \omega_i \eta^i_{MG} di$$

with $\omega_i := \left(\frac{\overline{Y}_i}{\overline{Y}}\right)$ and $\int_0^1 \omega_i di = 1$. Thus, we have:

$$\widetilde{Y}_{it} = \eta^i_{MG}\widetilde{G}_t + \lambda_i \widetilde{a}_t + \lambda_i^{\varepsilon_a} \varepsilon_{\widetilde{a}_{it}}$$

which completes the proof. ■

Proposition 4

Proof. This result is just a special case of Proposition 5 with $\rho_i = 0$. ■

Proposition 5

Proof. Under perfect competition and CRS:

$$\max_{p_{it}} \left[p_{it} y_{it} - W_{it} y_{it} a_{it}^{-1} G_{it}^{-\rho_i}\right]$$

implies $p_{it} = \frac{W_{it}}{a_{it} G_{it}^{\rho_i}}$. Now, the non-stochastic steady state leads to a system of equations, as in Propositions 2 and 3, and I also assume here that there are economies for which there is a solution. I omit this here to save space. Moving forward, after we log-linearize the equilibrium equations around the non-stochastic steady state, we get:

$$\begin{aligned}\widetilde{y}_{it} = &\left[\frac{\frac{1}{\varepsilon}\left(\frac{\mathcal{G}_i}{1-\mathcal{G}_i}\right) + \rho_i + \varphi_i \rho_i}{\varphi_i + \frac{1}{\varepsilon}\left(\frac{1}{1-\mathcal{G}_i}\right)}\right] \widetilde{G}_{it} + \left(\frac{1+\varphi_i}{\varphi_i + \frac{1}{\varepsilon}\left(\frac{1}{1-\mathcal{G}_i}\right)}\right)\widetilde{a_{it}} - \left[\frac{\sigma - \frac{1}{\varepsilon}}{\varphi_i + \frac{1}{\varepsilon}\left(\frac{1}{1-\mathcal{G}_i}\right)}\right] \int_0^1 \left(\frac{\overline{c}_j}{\overline{\overline{C}}}\right)^{\frac{\varepsilon-1}{\varepsilon}} \left(\frac{1}{1-\mathcal{G}_j}\right) \widetilde{y}_{jt} dj \\ &+ \left[\frac{\sigma - \frac{1}{\varepsilon}}{\varphi_i + \frac{1}{\varepsilon}\left(\frac{1}{1-\mathcal{G}_i}\right)}\right] \int_0^1 \left(\frac{\overline{c}_j}{\overline{\overline{C}}}\right)^{\frac{\varepsilon-1}{\varepsilon}} \left(\frac{\mathcal{G}_j}{1-\mathcal{G}_j}\right) \widetilde{G}_{jt} dj.\end{aligned}$$

And, hence, with the same arguments that we made at the end of Proposition 3, we obtain equations (17) and (18). ■

Proposition 6

Proof. Instead of repeating the arguments in the proof of Proposition 2 for this case, we detail the steps in which the proof changes. Using that proof, it is easy to see that given the subsidy, we now have:

$$MC^{Reg}_{n,t} = \widetilde{\left(\frac{W_t}{P_{nt}}\right)} + \frac{\alpha_n}{1-\alpha_n} \widetilde{Y_{n,t}} - \frac{1}{1-\alpha_n} \widetilde{a_{nt}} + \widetilde{s}_{n,t}. \tag{77}$$

Then, it is easy to see that the pricing problem for firms resetting their prices does not change, and in the derivation of the recursive structure for the reset price nothing changes, so we still have the relationship:

$$\pi_{nt} = \left(\frac{1-\theta}{\theta}\right)(1-\theta\beta)\left(\frac{1-\alpha_n}{1-\alpha_n + \varepsilon\alpha_n}\right) MC^{Reg}_{n,t} + \beta \mathbb{E}_t \left\{\pi_{nt+1}\right\}.$$

By replacing the new expression (77), we get the NKPC displayed in the text. Moreover, by setting $\mathcal{G} = 0$ in the Euler equation, we get the one displayed in the text. I still assume that there are economies for which the equations of the non-stochastic steady state have a solution. ■

Proposition 7

Proof. Instead of repeating the arguments in the proof of Proposition 3 for this case, we detail the steps in which it needs to be adjusted. Given the subsidy, we now have:

$$\widetilde{\frac{W_{it}}{P_t}} = \widetilde{\frac{p_{it}}{P_t}} - \widetilde{s}_{it} + \frac{1}{1-\alpha_i}\widetilde{a}_{it} - \frac{\alpha_i}{1-\alpha_i}\widetilde{Y}_{it}.$$

And this implies now that:

$$\sigma \widetilde{Y}_t^C + \varphi_i \widetilde{l}_{it} = \widetilde{\frac{p_{it}}{P_t}} - \widetilde{s}_{it} + \frac{1}{1-\alpha_i}\widetilde{a}_{it} - \frac{\alpha_i}{1-\alpha_i}\widetilde{Y}_{it}$$

because there is no government spending. By following the same steps as before, we get the result in the text. I still assume that there are economies for which the equation of the non-stochastic steady state has a solution. ■

Proposition 8

Proof. From the firm's problem in every region, we get $W_{nt} = a_{nt}$. From the F.O.C. of representative agent n, we get that:

$$a_{nt} L_{nt} - T_{nt} = a_{nt}^{\frac{1}{\sigma_n^c}} L_{nt}^{-\frac{\varphi_n}{\sigma_n^c}}. \tag{78}$$

Thus, (78) defines $\overline{L}_n$ in the non-stochastic steady state, which then allows us to compute the rest of the variables. Then, after log-linearizing:

$$\widetilde{L_{nt}} = \left(\frac{\overline{Y_n}}{\frac{1}{N}\overline{G}} + \frac{\overline{a}_n^{\frac{1}{\sigma_n^c}} \overline{L}_{nt}^{-\frac{\varphi_n}{\sigma_n^c}}}{\frac{1}{N}\overline{G}} \frac{\varphi_n}{\sigma_n^c} \right)^{-1} \widetilde{G_t} - \frac{\left(\frac{\overline{Y_n}}{\frac{1}{N}\overline{G}} - \frac{\overline{a}_n^{\frac{1}{\sigma_n^c}} \overline{L}_{nt}^{-\frac{\varphi_n}{\sigma_n^c}}}{\frac{1}{N}\overline{G}} \frac{1}{\sigma_n^c} \right)}{\left(\frac{\overline{Y_n}}{\frac{1}{N}\overline{G}} + \frac{\overline{a}_n^{\frac{1}{\sigma_n^c}} \overline{L}_{nt}^{-\frac{\varphi_n}{\sigma_n^c}}}{\frac{1}{N}\overline{G}} \frac{\varphi_n}{\sigma_n^c} \right)} \widetilde{a_{nt}}$$

and thus:

$$\widetilde{Y}_{nt} = \left(\frac{\overline{Y_n}}{\frac{1}{N}\overline{G}} + \frac{\overline{a}_n^{\frac{1}{\sigma_n^c}} \overline{L}_{nt}^{-\frac{\varphi_n}{\sigma_n^c}}}{\frac{1}{N}\overline{G}} \frac{\varphi_n}{\sigma_n^c} \right)^{-1} \widetilde{G_t} + \widetilde{a_t} \left[1 - \frac{\left(\frac{\overline{Y_n}}{\frac{1}{N}\overline{G}} - \frac{\overline{a}_n^{\frac{1}{\sigma_n^c}} \overline{L}_{nt}^{-\frac{\varphi_n}{\sigma_n^c}}}{\frac{1}{N}\overline{G}} \frac{1}{\sigma_n^c} \right)}{\left(\frac{\overline{Y_n}}{\frac{1}{N}\overline{G}} + \frac{\overline{a}_n^{\frac{1}{\sigma_n^c}} \overline{L}_{nt}^{-\frac{\varphi_n}{\sigma_n^c}}}{\frac{1}{N}\overline{G}} \frac{\varphi_n}{\sigma_n^c} \right)} \right] + \varepsilon_{\widetilde{a}_{nt}} \left[1 - \frac{\left(\frac{\overline{Y_n}}{\frac{1}{N}\overline{G}} - \frac{\overline{a}_n^{\frac{1}{\sigma_n^c}} \overline{L}_{nt}^{-\frac{\varphi_n}{\sigma_n^c}}}{\frac{1}{N}\overline{G}} \frac{1}{\sigma_n^c} \right)}{\left(\frac{\overline{Y_n}}{\frac{1}{N}\overline{G}} + \frac{\overline{a}_n^{\frac{1}{\sigma_n^c}} \overline{L}_{nt}^{-\frac{\varphi_n}{\sigma_n^c}}}{\frac{1}{N}\overline{G}} \frac{\varphi_n}{\sigma_n^c} \right)} \right].$$

Note that since $\widetilde{G}_t = \widetilde{\frac{1}{N}G_t}$, this also gives the response to a change in per capita aggregate spending. Then, defining:

$$\eta^n_{MG} := \left(\frac{\overline{Y_n}}{\frac{1}{N}\overline{G}} + \frac{\overline{a}_n^{\frac{1}{\sigma_n^c}}\overline{L}_{nt}^{-\frac{\varphi_n}{\sigma_n^c}}}{\frac{1}{N}\overline{G}}\frac{\varphi_n}{\sigma_n^c}\right)^{-1}, \lambda_n := \left[1 - \frac{\left(\frac{\overline{Y_n}}{\frac{1}{N}\overline{G}} - \frac{\overline{a}_n^{\frac{1}{\sigma_n^c}}\overline{L}_{nt}^{-\frac{\varphi_n}{\sigma_n^c}}}{\frac{1}{N}\overline{G}}\frac{1}{\sigma_n^c}\right)}{\left(\frac{\overline{Y_n}}{\frac{1}{N}\overline{G}} + \frac{\overline{a}_n^{\frac{1}{\sigma_n^c}}\overline{L}_{nt}^{-\frac{\varphi_n}{\sigma_n^c}}}{\frac{1}{N}\overline{G}}\frac{\varphi_n}{\sigma_n^c}\right)}\right], \epsilon_{nt} := \varepsilon_{\widetilde{a}_{nt}}\left[1 - \frac{\left(\frac{\overline{Y_n}}{\frac{1}{N}\overline{G}} - \frac{\overline{a}_n^{\frac{1}{\sigma_n^c}}\overline{L}_{nt}^{-\frac{\varphi_n}{\sigma_n^c}}}{\frac{1}{N}\overline{G}}\frac{1}{\sigma_n^c}\right)}{\left(\frac{\overline{Y_n}}{\frac{1}{N}\overline{G}} + \frac{\overline{a}_n^{\frac{1}{\sigma_n^c}}\overline{L}_{nt}^{-\frac{\varphi_n}{\sigma_n^c}}}{\frac{1}{N}\overline{G}}\frac{\varphi_n}{\sigma_n^c}\right)}\right],$$

shows the first part of the proposition.

Finally, real per capita GDP is given by $Y_t = \sum_n \frac{1}{N} Y_{nt}$, and thus $\widetilde{Y}_t = \sum_n \frac{1}{N}\frac{\overline{Y_{nt}}}{\overline{Y_t}}\widetilde{Y}_{nt}$, where $\sum_n \frac{1}{N}\frac{\overline{Y_{nt}}}{\overline{Y_t}} = 1$. Defining $\omega_n := \frac{1}{N}\frac{\overline{Y_{nt}}}{\overline{Y_t}}$ completes the proof. ∎

Proposition 9

Proof. From the proof of Proposition 3, we know that in equilibrium:

$$\widetilde{Y}_{it} = \left\{ \begin{array}{c} \frac{(1-\alpha_i)\varepsilon\sigma\mathcal{G}}{(1-\mathcal{G})\left\{(\alpha_i+\varphi_i)\varepsilon+1-\alpha_i\right\}} \\ + \left(\frac{(1-\alpha_i)\{(1-\mathcal{G})-\varepsilon\sigma\}}{\left\{(\alpha_i+\varphi_i)\varepsilon+1-\alpha_i\right\}(1-\mathcal{G})}\right)\frac{\frac{1}{N}\sum_{i=1}^N\left(\frac{\overline{Y}_i}{\overline{Y}^C}\right)^{\frac{\varepsilon-1}{\varepsilon}}\frac{(1-\alpha_i)\varepsilon\sigma\mathcal{G}}{(1-\mathcal{G})\left\{(\alpha_i+\varphi_i)\varepsilon+1-\alpha_i\right\}}}{\left[1-\frac{1}{N}\sum_{i=1}^N\left(\frac{\overline{Y}_i}{\overline{Y}^C}\right)^{\frac{\varepsilon-1}{\varepsilon}}\left(\frac{(1-\alpha_i)\{(1-\mathcal{G})-\varepsilon\sigma\}}{\left\{(\alpha_i+\varphi_i)\varepsilon+1-\alpha_i\right\}(1-\mathcal{G})}\right)\right]} \end{array} \right\}\widetilde{G}_t$$

$$+ \left(\frac{\varepsilon\left(1+\varphi_i\right)}{\left(\alpha_i+\varphi_i\right)\varepsilon+1-\alpha_i}\right)\widetilde{a}_{it}$$

$$+ \left(\frac{(1-\alpha_i)\{(1-\mathcal{G})-\varepsilon\sigma\}}{\left\{\left(\alpha_i+\varphi_i\right)\varepsilon+1-\alpha_i\right\}(1-\mathcal{G})}\right)\frac{1}{N}\sum_{i=1}^N\frac{\left(\frac{\overline{Y}_i}{\overline{Y}^C}\right)^{\frac{\varepsilon-1}{\varepsilon}}\left(\frac{\varepsilon\left(1+\varphi_i\right)}{\left(\alpha_i+\varphi_i\right)\varepsilon+1-\alpha_i}\right)}{\left[1-\frac{1}{N}\sum_{i=1}^N\left(\frac{\overline{Y}_i}{\overline{Y}^C}\right)^{\frac{\varepsilon-1}{\varepsilon}}\left(\frac{(1-\alpha_i)\{(1-\mathcal{G})-\varepsilon\sigma\}}{\left\{\left(\alpha_i+\varphi_i\right)\varepsilon+1-\alpha_i\right\}(1-\mathcal{G})}\right)\right]}\widetilde{a}_{it}.$$

And thus replacing the process for technology:

$$\widetilde{Y}_{it} = \left\{ \begin{array}{c} \frac{(1-\alpha_i)\varepsilon\sigma\mathcal{G}}{(1-\mathcal{G})\left\{\left(\alpha_i+\varphi_i\right)\varepsilon+1-\alpha_i\right\}} \\ +\left(\frac{(1-\alpha_i)\{(1-\mathcal{G})-\varepsilon\sigma\}}{\left\{\left(\alpha_i+\varphi_i\right)\varepsilon+1-\alpha_i\right\}(1-\mathcal{G})}\right) \frac{\frac{1}{N}\sum_{i=1}^{N}\left(\frac{\overline{Y}_i}{\overline{Y}^C}\right)^{\frac{\varepsilon-1}{\varepsilon}} \frac{(1-\alpha_i)\varepsilon\sigma\mathcal{G}}{(1-\mathcal{G})\left\{\left(\alpha_i+\varphi_i\right)\varepsilon+1-\alpha_i\right\}}}{\left[1-\frac{1}{N}\sum_{i=1}^{N}\left(\frac{\overline{Y}_i}{\overline{Y}^C}\right)^{\frac{\varepsilon-1}{\varepsilon}}\left(\frac{(1-\alpha_i)\{(1-\mathcal{G})-\varepsilon\sigma\}}{\left\{\left(\alpha_i+\varphi_i\right)\varepsilon+1-\alpha_i\right\}(1-\mathcal{G})}\right)\right]} \end{array} \right\} \widetilde{G}_t$$

$$+\left(\begin{array}{c} \frac{\varepsilon\left(1+\varphi_i\right)}{\left(\alpha_i+\varphi_i\right)\varepsilon+1-\alpha_i} \\ +\left(\frac{(1-\alpha_i)\{(1-\mathcal{G})-\varepsilon\sigma\}}{\left\{\left(\alpha_i+\varphi_i\right)\varepsilon+1-\alpha_i\right\}(1-\mathcal{G})}\right)\frac{1}{N}\sum_{i=1}^{N}\frac{\left(\frac{\overline{Y}_i}{\overline{Y}^C}\right)^{\frac{\varepsilon-1}{\varepsilon}}\left(\frac{\varepsilon\left(1+\varphi_i\right)}{\left(\alpha_i+\varphi_i\right)\varepsilon+1-\alpha_i}\right)}{\left[1-\frac{1}{N}\sum_{i=1}^{N}\left(\frac{\overline{Y}_i}{\overline{Y}^C}\right)^{\frac{\varepsilon-1}{\varepsilon}}\left(\frac{(1-\alpha_i)\{(1-\mathcal{G})-\varepsilon\sigma\}}{\left\{\left(\alpha_i+\varphi_i\right)\varepsilon+1-\alpha_i\right\}(1-\mathcal{G})}\right)\right]} \end{array} \right)\widetilde{a}_t$$

$$+\left(\frac{\varepsilon\left(1+\varphi_i\right)}{\left(\alpha_i+\varphi_i\right)\varepsilon+1-\alpha_i}\right)\varepsilon_{\widetilde{a}_{it}}$$

$$+\left(\frac{(1-\alpha_i)\{(1-\mathcal{G})-\varepsilon\sigma\}}{\left\{\left(\alpha_i+\varphi_i\right)\varepsilon+1-\alpha_i\right\}(1-\mathcal{G})}\right)\frac{1}{N}\sum_{i=1}^{N}\frac{\left(\frac{\overline{Y}_i}{\overline{Y}^C}\right)^{\frac{\varepsilon-1}{\varepsilon}}\left(\frac{\varepsilon\left(1+\varphi_i\right)}{\left(\alpha_i+\varphi_i\right)\varepsilon+1-\alpha_i}\right)}{\left[1-\frac{1}{N}\sum_{i=1}^{N}\left(\frac{\overline{Y}_i}{\overline{Y}^C}\right)^{\frac{\varepsilon-1}{\varepsilon}}\left(\frac{(1-\alpha_i)\{(1-\mathcal{G})-\varepsilon\sigma\}}{\left\{\left(\alpha_i+\varphi_i\right)\varepsilon+1-\alpha_i\right\}(1-\mathcal{G})}\right)\right]}\varepsilon_{\widetilde{a}_{it}}.$$

For any variable x_t let us denote:

$$\Delta\widetilde{x}_t = \Delta\log\left(x_t\right).$$

Thus:

$$\Delta\log\left(Y_{it}\right) = \left\{ \begin{array}{c} \frac{(1-\alpha_i)\varepsilon\sigma\mathcal{G}}{(1-\mathcal{G})\left\{\left(\alpha_i+\varphi_i\right)\varepsilon+1-\alpha_i\right\}} \\ +\left(\frac{(1-\alpha_i)\{(1-\mathcal{G})-\varepsilon\sigma\}}{\left\{\left(\alpha_i+\varphi_i\right)\varepsilon+1-\alpha_i\right\}(1-\mathcal{G})}\right) \frac{\frac{1}{N}\sum_{i=1}^{N}\left(\frac{\overline{Y}_i}{\overline{Y}^C}\right)^{\frac{\varepsilon-1}{\varepsilon}} \frac{(1-\alpha_i)\varepsilon\sigma\mathcal{G}}{(1-\mathcal{G})\left\{\left(\alpha_i+\varphi_i\right)\varepsilon+1-\alpha_i\right\}}}{\left[1-\frac{1}{N}\sum_{i=1}^{N}\left(\frac{\overline{Y}_i}{\overline{Y}^C}\right)^{\frac{\varepsilon-1}{\varepsilon}}\left(\frac{(1-\alpha_i)\{(1-\mathcal{G})-\varepsilon\sigma\}}{\left\{\left(\alpha_i+\varphi_i\right)\varepsilon+1-\alpha_i\right\}(1-\mathcal{G})}\right)\right]} \end{array} \right\} \Delta\log\left(G_t\right)$$

$$+\left(\begin{array}{c} \frac{\varepsilon\left(1+\varphi_i\right)}{\left(\alpha_i+\varphi_i\right)\varepsilon+1-\alpha_i} \\ +\left(\frac{(1-\alpha_i)\{(1-\mathcal{G})-\varepsilon\sigma\}}{\left\{\left(\alpha_i+\varphi_i\right)\varepsilon+1-\alpha_i\right\}(1-\mathcal{G})}\right)\frac{1}{N}\sum_{i=1}^{N}\frac{\left(\frac{\overline{Y}_i}{\overline{Y}^C}\right)^{\frac{\varepsilon-1}{\varepsilon}}\left(\frac{\varepsilon\left(1+\varphi_i\right)}{\left(\alpha_i+\varphi_i\right)\varepsilon+1-\alpha_i}\right)}{\left[1-\frac{1}{N}\sum_{i=1}^{N}\left(\frac{\overline{Y}_i}{\overline{Y}^C}\right)^{\frac{\varepsilon-1}{\varepsilon}}\left(\frac{(1-\alpha_i)\{(1-\mathcal{G})-\varepsilon\sigma\}}{\left\{\left(\alpha_i+\varphi_i\right)\varepsilon+1-\alpha_i\right\}(1-\mathcal{G})}\right)\right]} \end{array} \right)\Delta\log\left(a_t\right)$$

$$+\left(\frac{\varepsilon\left(1+\varphi_i\right)}{\left(\alpha_i+\varphi_i\right)\varepsilon+1-\alpha_i}\right)\Delta\log\left(\varepsilon_{a_{it}}\right)$$

$$+\left(\frac{(1-\alpha_i)\{(1-\mathcal{G})-\varepsilon\sigma\}}{\left\{\left(\alpha_i+\varphi_i\right)\varepsilon+1-\alpha_i\right\}(1-\mathcal{G})}\right)\frac{1}{N}\sum_{i=1}^{N}\frac{\left(\frac{\overline{Y}_i}{\overline{Y}^C}\right)^{\frac{\varepsilon-1}{\varepsilon}}\left(\frac{\varepsilon\left(1+\varphi_i\right)}{\left(\alpha_i+\varphi_i\right)\varepsilon+1-\alpha_i}\right)}{\left[1-\frac{1}{N}\sum_{i=1}^{N}\left(\frac{\overline{Y}_i}{\overline{Y}^C}\right)^{\frac{\varepsilon-1}{\varepsilon}}\left(\frac{(1-\alpha_i)\{(1-\mathcal{G})-\varepsilon\sigma\}}{\left\{\left(\alpha_i+\varphi_i\right)\varepsilon+1-\alpha_i\right\}(1-\mathcal{G})}\right)\right]}\Delta\log\left(\varepsilon_{a_{it}}\right).$$

Moreover, note that:

$$Var\left(\frac{\left(\frac{\overline{Y}_i}{\overline{Y}^C}\right)^{\frac{\varepsilon-1}{\varepsilon}}\left(\frac{\varepsilon\left(1+\varphi_i\right)}{\left(\alpha_i+\varphi_i\right)\varepsilon+1-\alpha_i}\right)}{\left[1-\frac{1}{N}\sum_{i=1}^{N}\left(\frac{\overline{Y}_i}{\overline{Y}^C}\right)^{\frac{\varepsilon-1}{\varepsilon}}\left(\frac{(1-\alpha_i)\{(1-\mathcal{G})-\varepsilon\sigma\}}{\left\{\left(\alpha_i+\varphi_i\right)\varepsilon+1-\alpha_i\right\}(1-\mathcal{G})}\right)\right]}\Delta\log\left(\varepsilon_{a_{it}}\right)\right)$$

$$=\left(\frac{\left(\frac{\overline{Y}_i}{\overline{Y}^C}\right)^{\frac{\varepsilon-1}{\varepsilon}}\left(\frac{\varepsilon\left(1+\varphi_i\right)}{\left(\alpha_i+\varphi_i\right)\varepsilon+1-\alpha_i}\right)}{\left[1-\frac{1}{N}\sum_{i=1}^{N}\left(\frac{\overline{Y}_i}{\overline{Y}^C}\right)^{\frac{\varepsilon-1}{\varepsilon}}\left(\frac{(1-\alpha_i)\{(1-\mathcal{G})-\varepsilon\sigma\}}{\left\{\left(\alpha_i+\varphi_i\right)\varepsilon+1-\alpha_i\right\}(1-\mathcal{G})}\right)\right]}\right)^2\sigma^2$$

and:

$$\frac{1}{N^2}\sum_{i=1}^{N}Var\left(\frac{\left(\frac{\overline{Y}_i}{\overline{Y}^C}\right)^{\frac{\varepsilon-1}{\varepsilon}}\left(\frac{\varepsilon\left(1+\varphi_i\right)}{\left(\alpha_i+\varphi_i\right)\varepsilon+1-\alpha_i}\right)}{\left[1-\frac{1}{N}\sum_{i=1}^{N}\left(\frac{\overline{Y}_i}{\overline{Y}^C}\right)^{\frac{\varepsilon-1}{\varepsilon}}\left(\frac{(1-\alpha_i)\{(1-\mathcal{G})-\varepsilon\sigma\}}{\left\{\left(\alpha_i+\varphi_i\right)\varepsilon+1-\alpha_i\right\}(1-\mathcal{G})}\right)\right]}\Delta\log\left(\varepsilon_{a_{it}}\right)\right)$$

$$=\sigma^2\frac{1}{N}\frac{\frac{1}{N}\sum_{i=1}^{N}\left(\frac{\overline{Y}_i}{\overline{Y}^C}\right)^{2\left(\frac{\varepsilon-1}{\varepsilon}\right)}\left(\frac{\varepsilon\left(1+\varphi_i\right)}{\left(\alpha_i+\varphi_i\right)\varepsilon+1-\alpha_i}\right)^2}{\left[1-\frac{1}{N}\sum_{i=1}^{N}\left(\frac{\overline{Y}_i}{\overline{Y}^C}\right)^{\frac{\varepsilon-1}{\varepsilon}}\left(\frac{(1-\alpha_i)\{(1-\mathcal{G})-\varepsilon\sigma\}}{\left\{\left(\alpha_i+\varphi_i\right)\varepsilon+1-\alpha_i\right\}(1-\mathcal{G})}\right)\right]^2}=o\left(1\right)$$

where the last steps follow from the boundness and convergence assumptions made on the parameters and the sequences, respectively. And thus given our assumption that $\Delta\log\left(\varepsilon_{a_{it}}\right)$ is $i.i.d.$ by Chebychev's WLLN, we have:

$$\frac{1}{N}\sum_{i=1}^{N}\frac{\left(\frac{\overline{Y}_i}{\overline{Y}^C}\right)^{\frac{\varepsilon-1}{\varepsilon}}\left(\frac{\varepsilon\left(1+\varphi_i\right)}{\left(\alpha_i+\varphi_i\right)\varepsilon+1-\alpha_i}\right)}{\left[1-\frac{1}{N}\sum_{i=1}^{N}\left(\frac{\overline{Y}_i}{\overline{Y}^C}\right)^{\frac{\varepsilon-1}{\varepsilon}}\left(\frac{(1-\alpha_i)\{(1-\mathcal{G})-\varepsilon\sigma\}}{\left\{\left(\alpha_i+\varphi_i\right)\varepsilon+1-\alpha_i\right\}(1-\mathcal{G})}\right)\right]}\Delta\log\left(\varepsilon_{a_{it}}\right)=o_p\left(1\right).$$

Finally, note that (22) is trivially p.d. since in this case we have only one macro shock. Moreover, note that there are no state variables at all in this model. The rest of the assumptions directly imply that this model belongs to the CSVC, which completes the proof. ■

Proposition 10

Proof. Let us define:

$$v' := \left(\frac{1}{N},\ldots,\frac{1}{N}\right)$$
$$u' := \left(b_1,\ldots,b_N\right)$$
$$\Phi := A - uv'$$

then:

$$uv' = \begin{pmatrix} b_1 \\ \vdots \\ b_N \end{pmatrix} \left(\frac{1}{N}, \ldots, \frac{1}{N} \right) = \begin{pmatrix} b_1 \frac{1}{N} & \cdots & b_1 \frac{1}{N} \\ \vdots & \ddots & \vdots \\ b_N \frac{1}{N} & \cdots & b_N \frac{1}{N} \end{pmatrix}.$$

Now, using the Sherman-Morrison Formula, we get:

$$A^{-1} = \Phi^{-1} - \left(\frac{1}{1 + v'\Phi^{-1}u} \right) \Phi^{-1} uv' \Phi^{-1}$$

$$= \begin{pmatrix} a_{11}^{-1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & a_{NN}^{-1} \end{pmatrix} - \begin{pmatrix} \frac{a_{11}^{-2} b_1 \frac{1}{N}}{1 + \frac{1}{N} \sum_{n=1}^{N} a_{nn}^{-1} b_n} & \cdots & \frac{a_{NN}^{-1} a_{11}^{-1} b_1 \frac{1}{N}}{1 + \frac{1}{N} \sum_{n=1}^{N} a_{nn}^{-1} b_n} \\ \vdots & \ddots & \vdots \\ \frac{a_{11}^{-1} a_{NN}^{-1} b_N \frac{1}{N}}{1 + \frac{1}{N} \sum_{n=1}^{N} a_{nn}^{-1} b_n} & \cdots & \frac{a_{NN}^{-2} b_N \frac{1}{N}}{1 + \frac{1}{N} \sum_{n=1}^{N} a_{nn}^{-1} b_n} \end{pmatrix}$$

and thus:

$$A^{-1}\Theta = \begin{pmatrix} a_{11}^{-1}\theta_{11} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & a_{NN}^{-1}\theta_{NN} \end{pmatrix} + \frac{1}{N} \begin{pmatrix} a_{11}^{-1}\nu_1 & \cdots & a_{11}^{-1}\nu_1 \\ \vdots & \ddots & \vdots \\ a_{NN}^{-1}\nu_N & \cdots & a_{NN}^{-1}\nu_N \end{pmatrix}$$
$$- \frac{1}{N} \begin{pmatrix} a_{11}^{-1}\theta_{11} \frac{a_{11}^{-1} b_1}{1 + \frac{1}{N} \sum_{n=1}^{N} a_{nn}^{-1} b_n} + \left(\frac{a_{11}^{-1} b_1}{1 + \frac{1}{N} \sum_{n=1}^{N} a_{nn}^{-1} b_n} \right) \frac{1}{N} \sum_{n=1}^{N} a_{nn}^{-1} \nu_n & \cdots \\ \vdots & \ddots \\ a_{11}^{-1}\theta_{11} \frac{a_{NN}^{-1} b_N}{1 + \frac{1}{N} \sum_{n=1}^{N} a_{nn}^{-1} b_n} + \left(\frac{a_{NN}^{-1} b_N}{1 + \frac{1}{N} \sum_{n=1}^{N} a_{nn}^{-1} b_n} \right) \frac{1}{N} \sum_{n=1}^{N} a_{nn}^{-1} \nu_n & \cdots \end{pmatrix}.$$

Hence:

$$A^{-1}\Theta\vec{G}_t = \begin{pmatrix} a_{11}^{-1}\theta_{11}\widetilde{G}_{1t} \\ \vdots \\ a_{NN}^{-1}\theta_{NN}\widetilde{G}_{Nt} \end{pmatrix} + \begin{pmatrix} a_{11}^{-1}\nu_1 \left(\frac{1}{N} \sum_{n=1}^{N} \widetilde{G}_{nt} \right) \\ \vdots \\ a_{NN}^{-1}\nu_N \left(\frac{1}{N} \sum_{n=1}^{N} \widetilde{G}_{nt} \right) \end{pmatrix}$$
$$- \begin{pmatrix} \frac{1}{N} \sum_{n=1}^{N} \left\{ a_{nn}^{-1}\theta_{nn} \frac{a_{11}^{-1} b_1}{1 + \frac{1}{N} \sum_{h=1}^{N} a_{hh}^{-1} b_h} + \left(\frac{a_{11}^{-1} b_1}{1 + \frac{1}{N} \sum_{h=1}^{N} a_{hh}^{-1} b_h} \right) \left(\frac{1}{N} \sum_{h=1}^{N} a_{hh}^{-1} \nu_h \right) \right\} \widetilde{G}_{nt} \\ \vdots \\ \frac{1}{N} \sum_{n=1}^{N} \left\{ a_{nn}^{-1}\theta_{nn} \frac{a_{NN}^{-1} b_N}{1 + \frac{1}{N} \sum_{h=1}^{N} a_{hh}^{-1} b_h} + \left(\frac{a_{NN}^{-1} b_N}{1 + \frac{1}{N} \sum_{h=1}^{N} a_{hh}^{-1} b_h} \right) \left(\frac{1}{N} \sum_{h=1}^{N} a_{hh}^{-1} \nu_h \right) \right\} \widetilde{G}_{nt} \end{pmatrix}.$$

And thus because of Blanchard and Kahn (1980) and the fact that $\vec{G}_t$ is a martingale difference sequence of random vectors:

$$\widetilde{Y}_{nt} = -\left(a_{nn}^{-1}\theta_{nn} + a_{nn}^{-1}\nu_n - \frac{1}{N} \sum_{j=1}^{N} \left\{ a_{jj}^{-1}\theta_{jj} \frac{a_{nn}^{-1} b_n}{1 + \frac{1}{N} \sum_{h=1}^{N} a_{hh}^{-1} b_h} + \left(\frac{a_{nn}^{-1} b_n}{1 + \frac{1}{N} \sum_{h=1}^{N} a_{hh}^{-1} b_h} \right) \left(\frac{1}{N} \sum_{h=1}^{N} a_{hh}^{-1} \nu_h \right) \right\} \right) \widetilde{G}_t$$
$$- a_{nn}^{-1}\theta_{nn}\varepsilon_{\widetilde{G}_{Nt}} - \frac{1}{N} \sum_{j=1}^{N} \left[a_{nn}^{-1}\nu_n - a_{jj}^{-1}\theta_{jj} \frac{a_{nn}^{-1} b_n}{1 + \frac{1}{N} \sum_{h=1}^{N} a_{hh}^{-1} b_h} - \left(\frac{a_{nn}^{-1} b_n}{1 + \frac{1}{N} \sum_{h=1}^{N} a_{hh}^{-1} b_h} \right) \left(\frac{1}{N} \sum_{h=1}^{N} a_{hh}^{-1} \nu_h \right) \right] \varepsilon_{\widetilde{G}_{jt}}.$$

Now, note that:

$$
\begin{aligned}
&var\left\{\left[a_{nn}^{-1}\nu_n - a_{jj}^{-1}\theta_{jj}\frac{a_{nn}^{-1}b_n}{1+\frac{1}{N}\sum_{h=1}^{N}a_{hh}^{-1}b_h} - \left(\frac{a_{nn}^{-1}b_n}{1+\frac{1}{N}\sum_{h=1}^{N}a_{hh}^{-1}b_h}\right)\left(\frac{1}{N}\sum_{h=1}^{N}a_{hh}^{-1}\nu_h\right)\right]\varepsilon_{\widetilde{G}_{jt}}\right\} \\
&= \left[a_{nn}^{-1}\nu_n - a_{jj}^{-1}\theta_{jj}\frac{a_{nn}^{-1}b_n}{1+\frac{1}{N}\sum_{h=1}^{N}a_{hh}^{-1}b_h} - \left(\frac{a_{nn}^{-1}b_n}{1+\frac{1}{N}\sum_{h=1}^{N}a_{hh}^{-1}b_h}\right)\left(\frac{1}{N}\sum_{h=1}^{N}a_{hh}^{-1}\nu_h\right)\right]^2\sigma^2.
\end{aligned}
$$

And thus:

$$
\begin{aligned}
&\frac{1}{N^2}\sum_{i=1}^{N}\left[a_{nn}^{-1}\nu_n - a_{jj}^{-1}\theta_{jj}\frac{a_{nn}^{-1}b_n}{1+\frac{1}{N}\sum_{h=1}^{N}a_{hh}^{-1}b_h} - \left(\frac{a_{nn}^{-1}b_n}{1+\frac{1}{N}\sum_{h=1}^{N}a_{hh}^{-1}b_h}\right)\left(\frac{1}{N}\sum_{h=1}^{N}a_{hh}^{-1}\nu_h\right)\right]^2\sigma^2 \\
&= \frac{1}{N}\sigma^2\frac{1}{N}\sum_{i=1}^{N}\left[a_{nn}^{-1}\nu_n - a_{jj}^{-1}\theta_{jj}\frac{a_{nn}^{-1}b_n}{1+\frac{1}{N}\sum_{h=1}^{N}a_{hh}^{-1}b_h} - \left(\frac{a_{nn}^{-1}b_n}{1+\frac{1}{N}\sum_{h=1}^{N}a_{hh}^{-1}b_h}\right)\left(\frac{1}{N}\sum_{h=1}^{N}a_{hh}^{-1}\nu_h\right)\right]^2 = o\left(1\right),
\end{aligned}
$$

where the last steps follow from the boundness and convergence assumptions made about the parameters and the sequences, respectively. And thus given our assumption that $\varepsilon_{\widetilde{G}_{jt}}$ is $i.i.d.$ by Chebychev's WLLN, we have:

$$
\frac{1}{N}\sum_{j=1}^{N}\left[\begin{array}{c} a_{nn}^{-1}\nu_n - a_{jj}^{-1}\theta_{jj}\frac{a_{nn}^{-1}b_n}{1+\frac{1}{N}\sum_{h=1}^{N}a_{hh}^{-1}b_h} \\ -\left(\frac{a_{nn}^{-1}b_n}{1+\frac{1}{N}\sum_{h=1}^{N}a_{hh}^{-1}b_h}\right)\left(\frac{1}{N}\sum_{h=1}^{N}a_{hh}^{-1}\nu_h\right)\end{array}\right]\varepsilon_{\widetilde{G}_{jt}} = o_p\left(1\right).
$$

Moreover, note that there are no state variables at all in this model. The rest of the assumptions directly imply that this model belongs to the CSVC, which completes the proof. ■

Corollary 2

Proof. From the proof of Proposition 10, defining:

$$
\begin{aligned}
\gamma_n &:= -\left(a_{nn}^{-1}\theta_{nn} + a_{nn}^{-1}\nu_n - \frac{1}{N}\sum_{j=1}^{N}\left\{a_{jj}^{-1}\theta_{jj}\frac{a_{nn}^{-1}b_n}{1+\frac{1}{N}\sum_{h=1}^{N}a_{hh}^{-1}b_h} + \left(\frac{a_{nn}^{-1}b_n}{1+\frac{1}{N}\sum_{h=1}^{N}a_{hh}^{-1}b_h}\right)\left(\frac{1}{N}\sum_{h=1}^{N}a_{hh}^{-1}\nu_h\right)\right\}\right) \\
\beta_n &:= -a_{nn}^{-1}\theta_{nn}
\end{aligned}
$$

completes the first part of the proof. The fact that $\eta_{macro} = \frac{1}{N}\sum_{n=1}^{N}\gamma_n$ follows from the definition of a simple aggregator. ■

Proposition 11

Proof. The first part of the proof is just a direct application of Blanchard and Kahn (1980), so I refer the reader to that paper for more details.

For the second part, note that conditions 1 and 2 imply that:

$$\left(c_n^{(N)\prime}, d_n^{(N)\prime}, e_n^{(N)\prime}\right)\mathbf{1}\left\{\begin{array}{c} \sum_{j\neq n}^{N}\left(\frac{c_{nj}^{(N)}}{\left(c_n^{(N)\prime}, d_n^{(N)\prime}, e_n^{(N)\prime}\right)\mathbf{1}}\right)\varepsilon_{s_{jt}} \\ +\sum_{j\neq n}^{N}\left(\frac{d_{nj}^{(N)}}{\left(c_n^{(N)\prime}, d_n^{(N)\prime}, e_n^{(N)\prime}\right)\mathbf{1}}\right)\varepsilon_{s_{jt-1}} \\ +\sum_{j\neq n}^{N}\left(\frac{e_{nj}^{(N)}}{\left(c_n^{(N)\prime}, d_n^{(N)\prime}, e_n^{(N)\prime}\right)\mathbf{1}}\right)\varepsilon_{\xi_{1t}^{j}} \end{array}\right\} = o_p(1).$$

The form of the solution plus condition 1 imply that Assumption 1 is satisfied. Finally, since the state variables are common to all regions, the rest of the conditions directly imply that model M belongs to the CSVC. ■

Proposition 12

Proof. The proof is very similar to the that of Proposition 11. The first part is just a direct application of Blanchard and Kahn (1980), so I refer the reader to that paper for more details. For the second part, note that conditions 1 and 2 imply that:

$$\left(c_n^{(N)\prime}, d_n^{(N)\prime}, e_n^{(N)\prime}\right)\mathbf{1}\left\{\begin{array}{c} \sum_{j\neq n}^{N}\left(\frac{c_{nj}^{(N)}}{\left(c_n^{(N)\prime}, d_n^{(N)\prime}, e_n^{(N)\prime}\right)\mathbf{1}}\right)\varepsilon_{s_{jt}} \\ +\sum_{j\neq n}^{N}\left(\frac{d_{nj}^{(N)}}{\left(c_n^{(N)\prime}, d_n^{(N)\prime}, e_n^{(N)\prime}\right)\mathbf{1}}\right)\varepsilon_{s_{jt-1}} \\ +\sum_{j\neq n}^{N}\left(\frac{e_{nj}^{(N)}}{\left(c_n^{(N)\prime}, d_n^{(N)\prime}, e_n^{(N)\prime}\right)\mathbf{1}}\right)\varepsilon_{\xi_{1t}^{j}} \end{array}\right\} = o_p(1).$$

The form of the solution plus condition 1 imply that Assumption 1 is satisfied. Finally, in this case we have N regional state variables, so the rest of the conditions directly imply that model M belongs to the RSVC. ■

Lemma 1

Proof. Given the assumptions, we have that:

$$\begin{aligned}\varphi_{nt}^{a} &= a_t + u_{nt}^{a} \sim N\left(\mu_{a,n}, \sigma_{a,n}^2 + \sigma_{u^a,n}^2\right) \\ \varphi_{nt}^{v} &= v_t + u_{nt}^{v} \sim N\left(\mu_{v,n}, \sigma_{v,n}^2 + \sigma_{u^v,n}^2\right)\end{aligned}$$

and:

$$\begin{pmatrix} a_t \\ \varphi_{nt}^{a} \\ \varphi_{nt}^{v} \end{pmatrix} \sim N\left(\begin{bmatrix} \mu_{a,n} \\ \mu_{a,n} \\ \mu_{v,n} \end{bmatrix}, \begin{bmatrix} \sigma_{a,n}^2 & \sigma_{a,n}^2 & 0 \\ \sigma_{a,n}^2 & \sigma_{a,n}^2 + \sigma_{u^a,n}^2 & 0 \\ 0 & 0 & \sigma_{v,n}^2 + \sigma_{u^v,n}^2 \end{bmatrix}\right).$$

Thus:

$$a_t \left| \begin{pmatrix} \varphi_{nt}^{a} \\ \varphi_{nt}^{v} \end{pmatrix} \right. \sim N(\phi, \Omega)$$

where, applying the formulas for the Kalman Filter, we get $\phi = \mu_{a,n} + \frac{\tau_{u^a,n}}{\tau_{u^a,n}+\tau_{a,n}}\left(\varphi^a_{nt} - \mu_{a,n}\right)$. So, if $\mu_{a,n} = \mu_{v,n} = 0$:

$$\mathbb{E}\left[a_t \middle| \varphi^a_{nt}, \varphi^v_{nt}\right] = \frac{\tau_{u^a,n}}{\tau_{u^a,n} + \tau_{a,n}}\varphi^a_{nt}$$
$$\mathbb{E}\left[v_t \middle| \varphi^a_{nt}, \varphi^v_{nt}\right] = \frac{\tau_{u^v,n}}{\tau_{u^v,n} + \tau_{v,n}}\varphi^v_{nt}.$$

So we get:

$$\mathbb{E}\left[\varepsilon_{s_{nt}} a_t + \varepsilon_{s_{nt}} v_t \middle| \varphi^a_{nt}, \varphi^v_{nt}\right] - \varepsilon^2_{s_{nt}} = \varepsilon_{s_{nt}} \frac{\tau_{u^a,n}}{\tau_{u^a,n} + \tau_{a,n}}\varphi^a_{nt} + \varepsilon_{s_{nt}} \frac{\tau_{u^v,n}}{\tau_{u^v,n} + \tau_{v,n}}\varphi^v_{nt} - \frac{1}{2}\varepsilon^2_{s_{nt}}$$

which has F.O.C.:

$$\frac{\tau_{u^a,n}}{\tau_{u^a,n} + \tau_{a,n}}\varphi^a_{nt} + \frac{\tau_{u^v,n}}{\tau_{u^v,n} + \tau_{v,n}}\varphi^v_{nt} - \varepsilon_{s_{nt}} = 0.$$

And thus:

$$\begin{aligned}\varepsilon_{s_{nt}} &= \frac{\tau_{u^a,n}}{\tau_{u^a,n} + \tau_{a,n}}\varphi^a_{nt} + \frac{\tau_{u^v,n}}{\tau_{u^v,n} + \tau_{v,n}}\varphi^v_{nt} \\ &= \frac{\tau_{u^a,n}}{\tau_{u^a,n} + \tau_{a,n}}a_t + \frac{\tau_{u^v,n}}{\tau_{u^v,n} + \tau_{v,n}}v_t + \frac{\tau_{u^a,n}}{\tau_{u^a,n} + \tau_{a,n}}u^a_{nt} + \frac{\tau_{u^v,n}}{\tau_{u^v,n} + \tau_{v,n}}u^v_{nt}\end{aligned}$$

and thus defining:

$$\epsilon^s_{nt} := \frac{\tau_{u^a,n}}{\tau_{u^a,n} + \tau_{a,n}}u^a_{nt} + \frac{\tau_{u^v,n}}{\tau_{u^v,n} + \tau_{v,n}}u^v_{nt},$$

we complete the proof. ∎

Corollary 3

Proof. The proof is exactly the same as that of Lemma 1 with the obvious parameters' replacements, so we do not duplicate it here. ∎

Now, in some of the proofs that follow, the following two Lemmas will be useful. For an arbitrary matrix A, $\|A\|$ denotes its Frobenius norm.

Lemma 2. *Suppose we have a process x_n with:*

$$\frac{1}{NT}\sum_{n=1}^{N}\sum_{t=1}^{T} x^2_{nt} = O_p(1)$$

and a process θ_n with:

$$\theta_{nt} = \frac{1}{N}\sum_{n=1}^{N}\epsilon_{nt},$$

with ϵ_{nt} independent across n and t with $\mathbb{E}[\epsilon_{nt}] = 0$, $\mathbb{E}\left[|\epsilon_{nt}|^8\right] \leq M < \infty$ and identically distributed across t.

Then:

$$\frac{1}{NT}\sum_{n=1}^{N} x'_n\theta_n = o_p(1).$$

Proof. I keep the index n on θ first to work in a more general setup, and then I particularize to the specific case of the statement. First note that:

$$\left|\frac{1}{NT}\sum_{n=1}^{N}x_n^{'}\theta_n\right| \leq \frac{1}{NT}\sum_{n=1}^{N}\left|x_n^{'}\theta_n\right| \leq \sqrt{\frac{1}{NT}\sum_{n=1}^{N}\sum_{t=1}^{T}x_{nt}^2}\sqrt{\frac{1}{NT}\sum_{n=1}^{N}\sum_{t=1}^{T}\theta_{nt}^2}.$$

Moreover:

$$\begin{aligned}
\mathbb{V}\left[\frac{1}{NT}\sum_{n=1}^{N}\sum_{t=1}^{T}\theta_{nt}^2\right] &= \mathbb{E}\left[\frac{1}{NT}\sum_{n=1}^{N}\sum_{t=1}^{T}\theta_{nt}^2\right]^2 - \left(\mathbb{E}\left[\frac{1}{NT}\sum_{n=1}^{N}\sum_{t=1}^{T}\theta_{nt}^2\right]\right)^2 \\
&\leq \left[\frac{1}{N^2T^2}\sum_{n=1}^{N}\sum_{t=1}^{T}\sum_{j=1}^{N}\sum_{s=1}^{T}\sqrt{\mathbb{E}\theta_{nt}^4}\sqrt{\mathbb{E}\theta_{js}^4}\right] - \left(\mathbb{E}\left[\frac{1}{NT}\sum_{n=1}^{N}\sum_{t=1}^{T}\theta_{nt}^2\right]\right)^2 \\
&= \mathbb{E}\theta_{nt}^4 - \left(\mathbb{E}\left[\frac{1}{NT}\sum_{n=1}^{N}\sum_{t=1}^{T}\theta_{nt}^2\right]\right)^2
\end{aligned}$$

Thus, given that $\theta_{nt} = o_p(1)$, we know $\theta_{nt}^4 = o_p(1)$ and we also have uniform integrability:

$$\mathbb{E}\left[\left|\theta_{nt}^4\right|^2\right] = \mathbb{E}\left[\theta_{nt}^{4\,2}\right] \leq \frac{1}{N^8}\sum_{n,j,k,i,u,p,l,r=1}^{N}\sqrt{\sqrt{\sqrt{\mathbb{E}\epsilon_{nt}^8\mathbb{E}\epsilon_{jt}^8}\,\mathbb{E}\left(\epsilon_{kt}\epsilon_{it}\right)^4}\,\mathbb{E}\left(\epsilon_{ut}\epsilon_{pt}\epsilon_{lt}\epsilon_{rt}\right)^2}$$

and thus, given that we are assuming bounded eighth moments: $\mathbb{E}\epsilon_{nt}^8 \leq M < \infty$, then $\mathbb{E}\left[\left|\theta_{nt}^4\right|^2\right] < \infty$. Hence, θ_{nt}^4 converges to zero in probability, and being uniformly integrable this means it converges in L^1 and thus the mean converges to zero as well. Moreover:

$$\mathbb{E}\left[\frac{1}{NT}\sum_{n=1}^{N}\sum_{t=1}^{T}\theta_{nt}^2\right] = \left[\frac{1}{NT}\sum_{n=1}^{N}\sum_{t=1}^{T}\mathbb{E}\left(\theta_{nt}^2\right)\right] = \mathbb{E}\left(\theta_{nt}^2\right).$$

And thus with the same arguments, we have that:

$$\mathbb{E}\left[\frac{1}{NT}\sum_{n=1}^{N}\sum_{t=1}^{T}\theta_{nt}^2\right] \underset{N,T\to\infty}{\longrightarrow} 0.$$

So combining the previous results:

$$\mathbb{V}\left[\frac{1}{NT}\sum_{n=1}^{N}\sum_{t=1}^{T}\theta_{nt}^2\right] = \underbrace{\mathbb{E}\theta_{nt}^4}_{\underset{N,T\to\infty}{\longrightarrow}0} - \Bigg(\underbrace{\mathbb{E}\left[\frac{1}{NT}\sum_{n=1}^{N}\sum_{t=1}^{T}\theta_{nt}^2\right]}_{\underset{N,T\to\infty}{\longrightarrow}0}\Bigg)^2 \underset{N,T\to\infty}{\longrightarrow} 0.$$

Which implies:

$$\frac{1}{NT}\sum_{n=1}^{N}\sum_{t=1}^{T}\theta_{nt}^2 = o_p(1).$$

And thus we conclude:

$$\left|\frac{1}{NT}\sum_{n=1}^{N}x_n^{'}\theta_n\right| \leq \sqrt{\frac{1}{NT}\sum_{n=1}^{N}\sum_{t=1}^{T}x_{nt}^2}\sqrt{\frac{1}{NT}\sum_{n=1}^{N}\sum_{t=1}^{T}\theta_{nt}^2} = O_p(1)\,o_p(1) = o_p(1).$$

■

Lemma 3. *Suppose*

$$\theta_t = \frac{1}{N}\sum_{n=1}^{N}\epsilon_{nt},$$

with ϵ_{nt} independent across n and t with $\mathbb{E}[\epsilon_{nt}] = 0, \mathbb{E}\left[|\epsilon_{nt}|^8\right] \leq M < \infty$ and identically distributed across t, $\mathbb{E}\left[x_{nt}^8\right] \leq M$, $\mathbb{E}\left[\|Z_{nt}\|^{16}\right] \leq M$, $\mathbb{E}\left\{\left(\sum_{u=1}^{P}\left[\iota_{nu}^T\right]^{-2}\right)^2\right\} \leq M$, $\forall T, n$ where ι_{nu}^T are the eigenvalues of $\frac{Z_n^{'}Z_n}{T}$. Then:

$$\frac{1}{NT}\sum_{n=1}^{N}x_n^{'}Z_n\left(Z_n^{'}Z_n\right)^{-1}Z_n^{'}\theta = o_p(1). \tag{79}$$

Proof. To prove this result, first note that:

(80)

$$\left|\frac{1}{NT}\sum_{n=1}^{N}x_n^{'}Z_n\left(Z_n^{'}Z_n\right)^{-1}Z_n^{'}\theta\right| \leq \left(\frac{1}{N}\sum_{n=1}^{N}\left\|\frac{x_n}{\sqrt{T}}\right\|^8\right)^{\frac{1}{8}}\left(\frac{1}{N}\sum_{n=1}^{N}\left\|\frac{Z_n}{\sqrt{T}}\right\|^{16}\right)^{\frac{1}{8}}\left(\frac{1}{N}\sum_{n=1}^{N}\left\|\left(\frac{Z_n^{'}Z_n}{T}\right)^{-1}\right\|^4\right)^{\frac{1}{4}}\left\|\frac{\theta}{\sqrt{T}}\right\|.$$

Now, note that:

$$\left\|\frac{\theta}{\sqrt{T}}\right\|^2 = \frac{1}{T}\sum_{t=1}^{T}\theta_t^2 = o_p(1) \tag{81}$$

by Lemma 2. We also have:

$$\mathbb{E}\left|\frac{1}{NT^4}\sum_{n=1}^{N}\|x_n\|^8\right| \leq \frac{1}{NT^4}\sum_{n=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{u=1}^{T}\sum_{v=1}^{T}\left(\mathbb{E}\left[x_{nt}^8\right]\mathbb{E}\left[x_{ns}^8\right]\mathbb{E}\left[x_{nu}^8\right]\mathbb{E}\left[x_{nv}^8\right]\right)^{\frac{1}{4}}$$

so $\mathbb{E}\left[x_{nt}^8\right] \leq M$ implies $\mathbb{E}\left|\frac{1}{NT^4}\sum_{n=1}^{N}\|x_n\|^8\right| \leq \vec{M} < \infty$ and thus:

$$\left(\frac{1}{NT^4}\sum_{n=1}^{N}\|x_n\|^8\right)^{\frac{1}{8}} = O_p(1). \tag{82}$$

Also:

$$\mathbb{E}\left|\frac{1}{NT^8}\sum_{n=1}^{N}\|Z_n\|^{16}\right| \leq \frac{1}{NT^8}\sum_{n=1}^{N}\sum_{t,s,u,v,w,f,g,h=1}^{T}\left\{\begin{array}{c}\mathbb{E}\left[\|Z_{nt}\|^{16}\right]\mathbb{E}\left[\|Z_{ns}\|^{16}\right]\mathbb{E}\left[\|Z_{nu}\|^{16}\right]\mathbb{E}\left[\|Z_{nv}\|^{16}\right]\\ \mathbb{E}\left[\|Z_{nw}\|^{16}\right]\mathbb{E}\left[\|Z_{nvf}\|^{16}\right]\mathbb{E}\left[\|Z_{ng}\|^{16}\right]\mathbb{E}\left[\|Z_{nh}\|^{16}\right]\end{array}\right\}^{\frac{1}{8}},$$

so $\mathbb{E}\left[\|Z_{nt}\|^{16}\right] \leq M < \infty$ implies $\mathbb{E}\left|\frac{1}{NT^8}\sum_{n=1}^{N}\|Z_n\|^{16}\right| \leq \vec{M} < \infty$ and thus:

$$\left(\frac{1}{NT^8}\sum_{n=1}^{N}\|Z_n\|^{16}\right)^{\frac{1}{8}} = O_p(1). \tag{83}$$

Moreover:

$$\mathbb{E}\left|\frac{1}{N}\sum_{n=1}^{N}\left\|\left(\frac{Z_n'Z_n}{T}\right)^{-1}\right\|^4\right| = \frac{1}{N}\sum_{n=1}^{N}\mathbb{E}\left\|\left(\frac{Z_n'Z_n}{T}\right)^{-1}\right\|^4 = \frac{1}{N}\sum_{n=1}^{N}\mathbb{E}\left\{\left[tr\left(D^{-2}\right)\right]^2\right\}.$$

So $\mathbb{E}\left\{\left[tr\left(D^{-2}\right)\right]^2\right\} = \mathbb{E}\left\{\left(\sum_{u=1}^{P}\left[\iota_{nu}^{T}\right]^{-2}\right)^2\right\} \leq M < \infty$, $\forall T, n$ where ι_{nu}^{T} are the eigenvalues of $\frac{Z_n'Z_n}{T}$, implies:

$$\left(\sum_{n=1}^{N}\frac{1}{N}\left\|\left(\frac{Z_n'Z_n}{T}\right)^{-1}\right\|^4\right)^{\frac{1}{4}} = O_p(1). \tag{84}$$

Thus, (81), (82), (83), (84) and (80) imply:

$$\left|\frac{1}{NT}\sum_{n=1}^{N}x_n'Z_n\left(Z_n'Z_n\right)^{-1}Z_n'\theta\right| \leq o_p(1)$$

and thus (79) holds. ■

Proposition 13

Proof. I show the second claim; the first one follows as a simple particular case. I sometimes choose not to make explicit the index N in $\hat{\beta}_n^{(N)}$ (or $\hat{\gamma}_n^{(N)}$, etc) and just write $\hat{\beta}_n$ to make the reading easier, but of course, it should always be kept in mind that these, alongside Z_n, are indexed by N, because L_n and γ_n change their dimensionality and also because δ_n is allowed to converge to a vector as N increases. I denote as F^0 the true unobserved macro shocks.

Let me also elaborate on the paths under which I am allowing $N, T \to \infty$. I denote by Q the total number of regressors in $\mathbb{X}_n$ and p the number of regressors in X_n. Clearly, $Q = p + N + K$. Clearly, for (21) to be p.d., we need $T \geq Q$. But it is also likely that for this to be true in the limit when $N, T \to \infty$, we also need $\frac{Q}{T} \underset{N,T\to\infty}{\longrightarrow} \rho \in (0,1)$. Indeed, Bai and Yin (1993) prove that if $\mathbb{X}_n$ has i.i.d. entries with mean zero, variance 1 and finite fourth moment, then $\frac{Q}{T} \underset{N,T\to\infty}{\longrightarrow} 1$ implies $\lambda_{min} \overset{a.s.}{\to} 0$, where λ_{min} is the smallest eigenvalue of $(1/T)\,\mathbb{X}_n\mathbb{X}_n'$. Additionally, they show $\frac{Q}{T} \underset{N,T\to\infty}{\longrightarrow} \rho \in (0,1)$ implies $\lambda_{min} \overset{a.s.}{\to} \left(1-\sqrt{\rho}\right)^2 > 0$. Similar results can be found in Silverstein (1985) for Wishart matrices. The setup in this paper, however, does not match these assumptions, in particular the independence one. Nonetheless, we will assume $\frac{Q}{T} \underset{N,T\to\infty}{\longrightarrow} \rho \in (0,1)$ when the regressors $L_n^{(N)}$ are present. Note that when they are not, Q is fixed with respect to T, N, and we do not have to worry about this. Finally, note that $\frac{Q}{T} \underset{N,T\to\infty}{\longrightarrow} \rho \in (0,1)$ rules out the case with $T = Q + J$ for any fixed integer J, but we could have $T = (1+v)\,Q$ for a small $v > 0$. Thus, given that we already had the restriction $T \geq Q$, assuming $\frac{Q}{T} \underset{N,T\to\infty}{\longrightarrow} \rho \in (0,1)$ does not seem like a big compromise. When regressors $L_n^{(N)}$ are absent, we can take the simultaneous limit without restrictions.

I will derive the results assuming that the only thing we observe is $s_{nt} = s_t + \varepsilon_{s_{nt}}$, so we will estimate $\hat{s}_t = \frac{1}{N}\sum_{j=1}^N s_{jt}$ and $\hat{\varepsilon}_{s_{nt}} = s_{nt} - \hat{s}_t$. If $\varepsilon_{s_{nt}}$ and s_t are observed, the related arguments can be put aside and the proof is simplified. Under our assumptions:

$$\hat{\varepsilon}_{s_{nt}} = s_{nt} - \hat{s}_t = \theta_n^{\varepsilon\prime} F_t^0 + \epsilon_{nt}^s - \frac{1}{N}\sum_{j=1}^N \epsilon_{jt}^s = \varepsilon_{s_{nt}} - \frac{1}{N}\sum_{j=1}^N \epsilon_{jt}^s$$

$$\hat{s}_t = \theta^{s\prime} F_t^0 + u_t^s + \frac{1}{N}\sum_{j=1}^N \epsilon_{jt}^s = s_t + \frac{1}{N}\sum_{j=1}^N \epsilon_{jt}^s.$$

Now, for the estimation of the unobserved macro shocks, we have:

$$\min_{F:\frac{F'F}{T}=I_K} \frac{1}{NT} S_{NT}(F) := \frac{1}{NT}\sum_{n=1}^N \hat{\varepsilon}_{s_n}' M_F \hat{\varepsilon}_{s_n} - \frac{1}{NT}\sum_{n=1}^N \epsilon_n^{s\prime} M_{F^0} \epsilon_n^s.$$

Now, note that:

$$\begin{aligned}
\frac{1}{NT} S_{NT}(F) &= \frac{1}{NT}\sum_{n=1}^N \left(F^0\theta_n^\varepsilon + \epsilon_n^s - \frac{1}{N}\sum_{j=1}^N \epsilon_j^s\right)' M_F \left(F^0\theta_n^\varepsilon + \epsilon_n^s - \frac{1}{N}\sum_{j=1}^N \epsilon_j^s\right) - \frac{1}{NT}\sum_{n=1}^N \epsilon_n^{s\prime} M_{F^0}\epsilon_n^s \\
&= \frac{1}{NT}\sum_{n=1}^N \theta_n^{\varepsilon\prime} F^{0\prime} M_F F^0 \theta_n^\varepsilon + \frac{2}{NT}\sum_{n=1}^N \theta_n^{\varepsilon\prime} F^{0\prime} M_F \epsilon_n^s + \frac{1}{NT}\sum_{n=1}^N \epsilon_n^{s\prime}\left(M_F - M_{F^0}\right)\epsilon_n^s \\
&\quad - \frac{2}{NT}\sum_{n=1}^N \epsilon_n^{s\prime} M_F \frac{1}{N}\sum_{j=1}^N \epsilon_j^s + \frac{1}{T}\frac{1}{N}\sum_{j=1}^N \epsilon_j^{s\prime} M_F \frac{1}{N}\sum_{j=1}^N \epsilon_j^s - \frac{2}{NT}\sum_{n=1}^N \theta_n^{\varepsilon\prime} F^{0\prime} M_F \frac{1}{N}\sum_{j=1}^N \epsilon_j^s.
\end{aligned}$$

Now, Lemma 3 and our assumptions imply:

$$\sup_F -\frac{2}{NT}\sum_{n=1}^N \epsilon_n^{s\prime} M_F \frac{1}{N}\sum_{j=1}^N \epsilon_j^s + \sup_F \frac{1}{T}\frac{1}{N}\sum_{j=1}^N \epsilon_j^{s\prime} M_F \frac{1}{N}\sum_{j=1}^N \epsilon_j^s - \sup_F \frac{2}{NT}\sum_{n=1}^N \theta_n^{\varepsilon\prime} F^{0\prime} M_F \frac{1}{N}\sum_{j=1}^N \epsilon_j^s = o_p(1).$$

Thus, this means we have:

$$\frac{1}{NT} S_{NT}(F) = \frac{1}{NT}\sum_{n=1}^N \theta_n^{\varepsilon\prime} F^{0\prime} M_F F^0 \theta_n^\varepsilon + \frac{2}{NT}\sum_{n=1}^N \theta_n^{\varepsilon\prime} F^{0\prime} M_F \epsilon_n^s + \frac{1}{NT}\sum_{n=1}^N \epsilon_n^{s\prime}\left(M_F - M_{F^0}\right)\epsilon_n^s + o_p(1).$$

Now, following the arguments in, for example, Stock and Watson (2002), Bai and Ng (2002) and Bai (2009), we have that for (29):

$$\left\|M_{\hat{F}} - M_{F^0}\right\|^2 = \left\|P_{F^0} - P_{\hat{F}}\right\|^2 = o_p(1).$$

Next, note that:

$$\left|\frac{1}{T}\left(Y_n - X_n\beta_n - L_n\gamma_n\right)'\left(M_{F^0} - M_{\hat{F}}\right)\left(Y_n - X_n\beta_n - L_n\gamma_n\right)\right| = \frac{1}{T}\left\|Y_n - X_n\beta_n - L_n\gamma_n\right\|^2 \left\|\left(M_{F^0} - M_{\hat{F}}\right)\right\|.$$

Moreover, $\frac{1}{T}\left\|Y_n - X_n\beta_n - L_n\gamma_n\right\|^2 = O_p(1)$ over bounded β_n and γ_n. Thus:

$$\frac{1}{T}\left\|Y_n - X_n\beta_n - L_n\gamma_n\right\|^2 \left\|\left(M_{F^0} - M_{\hat{F}}\right)\right\| = O_p(1)\, o_p(1) = o_p(1).$$

This implies, then:

$$\left|\frac{1}{T}\left(Y_n - X_n\beta_n - L_n\gamma_n\right)'\left(M_{F^0} - M_{\hat{F}}\right)\left(Y_n - X_n\beta_n - L_n\gamma_n\right)\right| \leq o_p(1). \tag{85}$$

Let me define:

$$\widetilde{\beta}_n, \widetilde{\gamma}_n \coloneqq \underset{\beta_n, \gamma_n}{\arg\min} \frac{1}{T}\left(Y_n - X_n\beta_n - L_n\gamma_n\right)' M_{F^0}\left(Y_n - X_n\beta_n - L_n\gamma_n\right),$$

then (85) implies:

$$\left\|\begin{pmatrix} \hat{\beta}_n \\ \hat{\gamma}_n \end{pmatrix} - \begin{pmatrix} \widetilde{\beta}_n \\ \widetilde{\gamma}_n \end{pmatrix}\right\| = o_p(1). \tag{86}$$

Furthermore, note that:

$$\begin{aligned}
\frac{1}{T} RS(\beta, \gamma) &\coloneqq \frac{1}{T}\left(Y_n - X_n\beta_n - L_n\gamma_n\right)' M_{F^0}\left(Y_n - X_n\beta_n - L_n\gamma_n\right) - \frac{1}{T}\epsilon_n' M_{F^0}\epsilon_n \\
&= \frac{1}{T}\left[X_n\Delta\beta_n + L_n\Delta\gamma_n\right]' M_{F^0}\left[X_n\Delta\beta_n + L_n\Delta\gamma_n\right] + \frac{2}{T}\left[X_n\Delta\beta_n + L_n\Delta\gamma_n\right]' M_{F^0}\widetilde{\epsilon}_n \\
&+ \frac{1}{T}\widetilde{\epsilon}_n' M_{F^0}\widetilde{\epsilon}_n - \frac{1}{T}\epsilon_n' M_{F^0}\epsilon_n.
\end{aligned}$$

where $\widetilde{\epsilon}_n = \epsilon_n + \left(\beta_{2n} - \beta_{1n}\right)\left(\frac{1}{N}\sum_{j=1}^N \epsilon_j^s\right)$.

Note that:

$$\begin{aligned}
\sup_\gamma \frac{2}{T}\left|\Delta\gamma_n^{(N)'} L_n^{(N)'} F^0\left(F^{0\prime}F^0\right)^{-1} F^{0\prime} u_n\right| &= \sup_\gamma 2\left|\Delta\gamma_n^{(N)'}\frac{L_n^{(N)'}F^0}{T}\left(\frac{F^{0\prime}F^0}{T}\right)^{-1}\frac{F^{0\prime}u_n}{T}\right| \\
&\leq \sup_\gamma 2\frac{\sqrt{N}\left\|\Delta\gamma_n^{(N)}\right\|}{\sqrt{N}}\left\|\frac{L_n^{(N)'}F^0}{T}\right\|\left\|\left(\frac{F^{0\prime}F^0}{T}\right)^{-1}\right\|\sqrt{\sum_{j=1}^K\left(\frac{\sum_{t=1}^T F_{jt}^0 u_{nt}}{T}\right)^2} \\
&= \sup_\gamma 2\sqrt{N}\left\|\Delta\gamma_n^{(N)}\right\|\sqrt{\sum_{k=1}^K\frac{\sum_{j=1}^N\left(\frac{\sum_{t=1}^T L_{njt}F_{kt}^0}{T}\right)^2}{N}}\, O_p(1)\, o_p(1) = o_p(1),
\end{aligned}$$

where:

$$\mathbb{E}\left|\frac{\sum_{j=1}^N\left(\frac{\sum_{t=1}^T L_{njt}F_{kt}^0}{T}\right)^2}{N}\right| \leq \frac{1}{NT^2}\sum_{j=1}^N\sum_{t=1}^T\sum_{s=1}^T\left(\mathbb{E}\left[L_{njt}^4\right]\mathbb{E}\left[\left(F_{kt}^0\right)^4\right]\mathbb{E}\left[L_{njs}^4\right]\mathbb{E}\left[\left(F_{ks}^0\right)^4\right]\right)^{\frac{1}{4}},$$

so the bounded moments condition in Assumption 4 implies $\mathbb{E}\left|\frac{\sum_{j=1}^{N}\left(\frac{\sum_{t=1}^{T}L_{njt}F_{kt}^{0}}{T}\right)^2}{N}\right| \leq M < \infty$ and thus $\frac{\sum_{j=1}^{N}\left(\frac{\sum_{t=1}^{T}L_{njt}F_{kt}^{0}}{T}\right)^2}{N} = O_p(1)$.

Also:

$$\begin{aligned}
\sup_{\gamma}\frac{2}{T}\left|\Delta\gamma_n^{(N)\prime}L_n^{(N)\prime}u_n\right| &\leq \sup_{\gamma}\frac{2}{T}\left\|\Delta\gamma_n^{(N)}\right\|\left\|L_n^{(N)\prime}u_n\right\| \\
&= \sup_{\gamma}2\sqrt{N}\left\|\Delta\gamma_n^{(N)}\right\|\sqrt{\frac{\sum_{j=1}^{N}\left(\frac{\sum_{t=1}^{T}L_{njt}u_{nt}}{T}\right)^2}{N}} \\
&= \sup_{\gamma}2\sqrt{N}\left\|\Delta\gamma_n^{(N)}\right\|o_p(1) = o_p(1),
\end{aligned}$$

where:

$$\mathbb{E}\left|\frac{1}{N}\sum_{j=1}^{N}\frac{1}{T}\left(\sum_{t=1}^{T}L_{njt}u_{nt}\right)^2\right| \leq \frac{1}{N}\frac{1}{T}\sum_{j=1}^{N}\sum_{t=1}^{T}\sqrt{\mathbb{E}\left\{L_{njt}^4\right\}\mathbb{E}\left\{u_{nt}^4\right\}}$$

and thus the bounded moments conditions on Assumptions 1 and 4 imply $\mathbb{E}\left|\frac{1}{N}\sum_{j=1}^{N}\frac{1}{T}\left(\sum_{t=1}^{T}L_{njt}u_{nt}\right)^2\right| \leq M < \infty$ and thus $\frac{1}{N}\sum_{j=1}^{N}\frac{1}{T}\left(\sum_{t=1}^{T}L_{njt}u_{nt}\right)^2 = O_p(1)$, which then implies:

$$\frac{\sum_{j=1}^{N}\left(\frac{\sum_{t=1}^{T}L_{njt}u_{nt}}{T}\right)^2}{N} = O_p(1)\frac{1}{T} = o_p(1).$$

With the same arguments, we can also show:

$$\frac{2}{T}\Delta\beta_n' X_n' M_{F^0} u_n = o_p(1). \tag{87}$$

The same computations show that the same is true if we replace u_{nt} with the rest of the terms in $\widetilde{\epsilon}_n$ that are $o_p(1)$ as in Lemma 3. The result for the terms involving β_{1n} and β_{2n} is obtained for bounded values of these, as usual. Thus:

$$\begin{aligned}
\frac{1}{T}RS(\beta,\gamma) &= \frac{1}{T}\left[X_n\Delta\beta_n + L_n\Delta\gamma_n\right]' M_{F^0}\left[X_n\Delta\beta_n + L_n\Delta\gamma_n\right] + o_p(1) \\
&= \frac{1}{T}\Delta\delta_n' Z_n' M_{F^0} Z_n \Delta\delta_n + o_p(1) \\
&= \Delta\delta_n' \frac{Z_n' M_{F^0} Z_n}{T}\Delta\delta_n + o_p(1)
\end{aligned}$$

and by Assumption 4, $\frac{Z_n' M_{F^0} Z_n}{T}$ is p.d. and thus:

$$\Delta\delta_n' \frac{Z_n' M_{F^0} Z_n}{T}\Delta\delta_n \geq 0 \tag{88}$$

for all $\Delta\delta_n$. Moreover, note that:

$$\frac{1}{T} RS\left(\beta^0, \gamma^0\right) = 0 \tag{89}$$

and also:

$$\Delta\widetilde{\delta}_n' \frac{Z_n' M_{F^0} Z_n}{T} \Delta\widetilde{\delta}_n + o_p(1) \leq \frac{1}{T} RS\left(\beta^0, \gamma^0\right) \tag{90}$$

and thus (88), (89) and (90) imply:

$$\Delta\widetilde{\delta}_n' \frac{Z_n' M_{F^0} Z_n}{T} \Delta\widetilde{\delta}_n = o_p(1).$$

Thus, together with Assumption 4, this implies:

$$\Delta\widetilde{\delta}_n = o_p(1). \tag{91}$$

Finally, (86) and (91) imply:

$$\left\| \begin{pmatrix} \hat{\beta}_n \\ \hat{\gamma}_n \end{pmatrix} - \begin{pmatrix} \beta_n^0 \\ \gamma_n^0 \end{pmatrix} \right\| = o_p(1),$$

which completes the proof. ■

Proposition 14

Proof. As in Proposition 13, I assume that $\varepsilon_{s_{nt}}$ and s_t are not directly observed; if they are the proof is simplified in that dimension. With the same arguments as in the proof of Proposition 13, (29) gives:

$$\left\| M_{\hat{F}} - M_{F^S} \right\|^2 = \left\| P_{F^S} - P_{\hat{F}} \right\|^2 = o_p(1).$$

Moreover, we know (see, for example, Bai and Ng (2002) and Bai (2009)) that:

$$\frac{1}{\sqrt{T}} \left\| M_{\hat{F}} F^S \right\| = o_p(1)$$

and, furthermore, there is an invertible matrix H^* such that:

$$\frac{1}{\sqrt{T}} \left\| \hat{F} - F^S H^* \right\| = o_p(1).$$

Thus, for each subset (k), we can define, to ease notation, a selection matrix H_{Selec}, with $H^{(k)} := H^* H_{Selec}$, such that:

$$\frac{1}{\sqrt{T}} \left\| \hat{F}^{(k)} - F^S H^{(k)} \right\| = o_p(1).$$

Now, note that:

$$
\begin{aligned}
&\frac{1}{NT}\sum_{n=1}^{N}\left(Y_n-\hat{X}_n\beta_n-F\lambda_n\right)'\left(Y_n-\hat{X}_n\beta_n-F\lambda_n\right)\\
&=\frac{1}{NT}\sum_{n=1}^{N}\left(Y_n-X_n\beta_n-F\lambda_n+\left(\beta_{2n}-\beta_{1n}\right)\varphi\right)'\left(Y_n-X_n\beta_n-F\lambda_n+\left(\beta_{2n}-\beta_{1n}\right)\varphi\right)\\
&=\frac{1}{NT}\sum_{n=1}^{N}\left(Y_n-X_n\beta_n-F\lambda_n\right)'\left(Y_n-X_n\beta_n-F\lambda_n\right)+\frac{2}{NT}\sum_{n=1}^{N}\left(Y_n-X_n\beta_n-F\lambda_n\right)'\left(\beta_{2n}-\beta_{1n}\right)\varphi\\
&+\frac{\varphi'\varphi}{T}\frac{1}{N}\sum_{n=1}^{N}\left(\beta_{2n}-\beta_{1n}\right)^2
\end{aligned}
$$

where φ are the $o_p(1)$ terms of the measurement error, i.e., the terms by which s_t differs from $\hat{s}_t$, which by the assumptions in this proposition on the errors of (39) are $o_p(1)$. Moreover, by Lemma 3, we have that:

$$
2\left(\beta_{2n}-\beta_{1n}\right)\frac{\varphi'\left[\frac{1}{N}\sum_{n=1}^{N}\left(Y_n-X_n\beta_n-F\lambda_n\right)\right]}{T}+\frac{\varphi'\varphi}{T}\frac{1}{N}\sum_{n=1}^{N}\left(\beta_{2n}-\beta_{1n}\right)^2=o_p(1),
$$

uniformly for bounded β_n and λ_n. And thus:

$$
\frac{1}{NT}\sum_{n=1}^{N}\left(Y_n-\hat{X}_n\beta_n-F\lambda_n\right)'\left(Y_n-\hat{X}_n\beta_n-F\lambda_n\right)=\frac{1}{NT}\sum_{n=1}^{N}\left(Y_n-X_n\beta_n-F\lambda_n\right)'\left(Y_n-X_n\beta_n-F\lambda_n\right)+o_p(1).
$$

Now define:

$$
\begin{aligned}
NSSR(F)&:=\frac{1}{NT}\sum_{n=1}^{N}\left(Y_n-X_n\beta_n(F)-F\lambda_n(F)\right)'\left(Y_n-X_n\beta_n(F)-F\lambda_n(F)\right)\\
&=\frac{1}{NT}\sum_{n=1}^{N}\lambda_n'F^{0\prime}M_{[X_n:F]}F^0\lambda_n+\frac{1}{NT}\sum_{n=1}^{N}\epsilon_n'M_{[X_n:F]}\epsilon_n+\frac{2}{NT}\sum_{n=1}^{N}\lambda_n'F^{0\prime}M_{[X_n:F]}\epsilon_n,
\end{aligned}
$$

where $\beta_n(F), \lambda_n(F)$ are the OLS estimators when regressing Y_n on $[X_n : F]$, and $M_{[X_n:F]}$ is the usual "residual maker":

$$
M_{[X_n:F]}:=I_T-[X_n:F]\left([X_n:F]'[X_n:F]\right)^{-1}[X_n:F]'.
$$

We will also denote:

$$
P_{[X_n:F]}:=[X_n:F]\left([X_n:F]'[X_n:F]\right)^{-1}[X_n:F]'
$$

the corresponding projection matrix.

Now, the first step is to prove:

$$\frac{1}{NT}\sum_{n=1}^{N}\epsilon_{n}^{'}M_{\left[X_n:\hat{F}^{(k)}\right]}\epsilon_n \underset{N,T\to\infty}{\overset{P}{\longrightarrow}} \frac{1}{NT}\sum_{n=1}^{N}\epsilon_{n}^{'}M_{\left[X_n:F^SH^{(k)}\right]}\epsilon_n$$

$$\frac{2}{NT}\sum_{n=1}^{N}\lambda_{n}^{'}F^{0'}M_{\left[X_n:\hat{F}^{(k)}\right]}\epsilon_n \underset{N,T\to\infty}{\overset{P}{\longrightarrow}} \frac{2}{NT}\sum_{n=1}^{N}\lambda_{n}^{'}F^{0'}M_{\left[X_n:F^SH^{(k)}\right]}\epsilon_n$$

$$\frac{1}{NT}\sum_{n=1}^{N}\lambda_{n}^{'}F^{0'}M_{\left[X_n:\hat{F}^{(k)}\right]}F^0\lambda_n \underset{N,T\to\infty}{\overset{P}{\longrightarrow}} \frac{1}{NT}\sum_{n=1}^{N}\lambda_{n}^{'}F^{0'}M_{\left[X_n:F^SH^{(k)}\right]}F^0\lambda_n.$$

So for the last term we have that:

(92)

$$\left|\frac{1}{NT}\sum_{n=1}^{N}\lambda_{n}^{'}F^{0'}\left(M_{\left[X_n:\hat{F}^{(k)}\right]}-M_{\left[X_n:F^SH^{(k)}\right]}\right)F^0\lambda_n\right| \leq \vec{\lambda}^2\frac{1}{T}\left(\sum_{t=1}^{T}\left\|F_t^0\right\|^2\right)\left(\frac{1}{N}\sum_{n=1}^{N}\left\|M_{\left[X_n:\hat{F}^{(k)}\right]}-M_{\left[X_n:F^SH^{(k)}\right]}\right\|^2\right)^{\frac{1}{2}}$$

and thus:

$$\mathbb{E}\left|\frac{1}{T}\sum_{t=1}^{T}\left\|F_t^0\right\|^2\right| = \frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\left[\left\|F_t^0\right\|^2\right] < \infty$$

where the inequality follows because our assumptions on moments, in Subsection (4.1), imply $\mathbb{E}\left[\left\|F_t^0\right\|^2\right] \leq M < \infty$. Thus $\frac{1}{T}\sum_{t=1}^{T}\left\|F_t^0\right\|^2 = O_p(1)$. Moreover, by direct computation, note that:

$$\begin{aligned}\frac{1}{N}\sum_{n=1}^{N}\left\|M_{\left[X_n:\hat{F}^{(k)}\right]}-M_{\left[X_n:F^SH^{(k)}\right]}\right\|^2 &= \frac{1}{N}\sum_{n=1}^{N}\left(2tr\left(I_{K-2}\right)-2tr\left(Z_n\left(\frac{Z_n'Z_n}{T}\right)^{-1}\frac{Z_n'\hat{Z}_n}{T}\left(\frac{\hat{Z}_n'\hat{Z}_n}{T}\right)^{-1}\frac{\hat{Z}_n'}{T}\right)\right)\\ &= 2\left[P+K-2-\frac{1}{N}\sum_{n=1}^{N}tr\left(\left(\frac{Z_n'Z_n}{T}\right)^{-1}\frac{Z_n'\hat{Z}_n}{T}\left(\frac{\hat{Z}_n'\hat{Z}_n}{T}\right)^{-1}\frac{\hat{Z}_n'Z_n}{T}\right)\right]\end{aligned}$$

where to ease notation, we are denoting $Z_n = Z_n\left(F^SH^{(k)}\right)$ and $\hat{Z}_n = Z_n\left(\hat{F}^{(k)}\right)$. Now, after some computations, we get:

$$\begin{aligned}&tr\left(\left(\frac{Z_n'Z_n}{T}\right)^{-1}\frac{Z_n'\hat{Z}_n}{T}\left(\frac{\hat{Z}_n'\hat{Z}_n}{T}\right)^{-1}\frac{\hat{Z}_n'Z_n}{T}\right)\\ &= tr\left(I_P\right)+tr\left(b_{FS}b_{\hat{F}S}\right)-tr\left(b_{FS}\left(\frac{\hat{F}^{(k)'}\hat{F}^{(k)}}{T}\right)^{-1}\frac{\hat{F}^{(k)'}}{T}X_n\vec{b}_{2n}\right)\\ &-tr\left[b_{\hat{F}S}\left(\frac{H^{(k)'}F^{S'}F^SH^{(k)}}{T}\right)^{-1}\frac{H^{(k)'}F^{S'}}{T}X_nb_{2n}\right]\\ &+tr\left[\left(\frac{H^{(k)'}F^{S'}F^SH^{(k)}}{T}\right)^{-1}\frac{H^{(k)'}F^{S'}}{T}X_nb_{2n}\left(\frac{\hat{F}^{(k)'}\hat{F}^{(k)}}{T}\right)^{-1}\frac{\hat{F}^{(k)'}}{T}X_n\vec{b}_{2n}\right]\end{aligned}$$

with:

$$\begin{aligned}
b_{FS} &= \left(\frac{H^{(k)\prime}F^{S\prime}F^{S}H^{(k)}}{T}\right)^{-1}\frac{H^{(k)\prime}F^{S\prime}}{T}\hat{F}^{(k)} \\
b_{\hat{F}S} &= \left(\frac{\hat{F}^{(k)\prime}\hat{F}^{(k)}}{T}\right)^{-1}\frac{\hat{F}^{(k)\prime}}{T}F^{S}H^{(k)} \\
b_{2n} &= \left(X_n^{\prime}M_{F^S H^{(k)}}X_n\right)^{-1}X_n^{\prime}M_{F^S H^{(k)}}\hat{F}^{(k)} \\
\vec{b}_{2n} &= \left(X_n^{\prime}M_{\hat{F}^{(k)}}X_n\right)^{-1}X_n^{\prime}M_{\hat{F}^{(k)}}F^{S}H^{(k)}.
\end{aligned}$$

Now, note that:

$$\begin{aligned}
tr\left(b_{FS}b_{\hat{F}S}\right) &= tr\left[\left(\frac{H^{(k)\prime}F^{S\prime}F^{S}H^{(k)}}{T}\right)^{-1}\frac{H^{(k)\prime}F^{S\prime}\hat{F}^{(k)}}{T}\left(\frac{\hat{F}^{(k)\prime}\hat{F}^{(k)}}{T}\right)^{-1}\frac{\hat{F}^{(k)\prime}F^{S}H^{(k)}}{T}\right] \\
&= tr\left[\frac{H^{(k)\prime}F^{S\prime}\hat{F}^{(k)}}{T}+\frac{H^{(k)\prime}F^{S\prime}}{\sqrt{T}}\left\{\frac{\hat{F}^{(k)}}{\sqrt{T}}-\frac{F^{S}H^{(k)}}{\sqrt{T}}\right\}\frac{\hat{F}^{(k)\prime}F^{S}H^{(k)}}{T}\left(\frac{H^{(k)\prime}F^{S\prime}F^{S}H^{(k)}}{T}\right)^{-1}\right]
\end{aligned}$$

And from:

$$\frac{1}{T}\left\|\hat{F}^{(k)}-F^{S}H^{(k)}\right\|^{2} = tr\left(I_{K-2}\right)-tr\left(\frac{\hat{F}^{(k)\prime}F^{S}H^{(k)}}{T}\right)$$

we get:

$$tr\left(b_{FS}b_{\hat{F}S}\right) = \underbrace{tr\left(\frac{H^{(k)\prime}F^{S\prime}\hat{F}^{(k)}}{T}\right)}_{\xrightarrow[N,T\to\infty]{P} tr(I_{K-2})=K-2.}+tr\left\{\left\{\frac{\hat{F}^{(k)}}{\sqrt{T}}-\frac{F^{S}H^{(k)}}{\sqrt{T}}\right\}\frac{\hat{F}^{(k)\prime}F^{S}H^{(k)}}{T}\left(\frac{H^{(k)\prime}F^{S\prime}F^{S}H^{(k)}}{T}\right)^{-1}\frac{H^{(k)\prime}F^{S\prime}}{\sqrt{T}}\right\}.$$

And note that:

$$\begin{aligned}
&\left|tr\left\{\left\{\frac{\hat{F}^{(k)}}{\sqrt{T}}-\frac{F^{S}H^{(k)}}{\sqrt{T}}\right\}\frac{\hat{F}^{(k)\prime}F^{S}H^{(k)}}{T}\left(\frac{H^{(k)\prime}F^{S\prime}F^{S}H^{(k)}}{T}\right)^{-1}\frac{H^{(k)\prime}F^{S\prime}}{\sqrt{T}}\right\}\right| \\
&\leq \left\|\left\{\frac{\hat{F}^{(k)}}{\sqrt{T}}-\frac{F^{S}H^{(k)}}{\sqrt{T}}\right\}\right\|\left\|\frac{\hat{F}^{(k)\prime}F^{S}H^{(k)}}{T}\right\|\left\|\left(\frac{H^{(k)\prime}F^{S\prime}F^{S}H^{(k)}}{T}\right)^{-1}\right\|\left\|\frac{H^{(k)\prime}F^{S\prime}}{\sqrt{T}}\right\| \\
&= o_p(1)\,O_p(1)\,O_p(1)\,O_p(1) = o_p(1).
\end{aligned}$$

Hence:

$$tr\left(b_{FS}b_{\hat{F}S}\right) = \underbrace{tr\left(\frac{H^{(k)\prime}F^{S\prime}\hat{F}^{(k)}}{T}\right)}_{\xrightarrow[N,T\to\infty]{P} tr(I_{K-2})=K-2}+\underbrace{tr\left\{\left\{\frac{\hat{F}^{(k)}}{\sqrt{T}}-\frac{F^{S}H^{(k)}}{\sqrt{T}}\right\}\frac{\hat{F}^{(k)\prime}F^{S}H^{(k)}}{T}\left(\frac{H^{(k)\prime}F^{S\prime}F^{S}H^{(k)}}{T}\right)^{-1}\frac{H^{(k)\prime}F^{S\prime}}{\sqrt{T}}\right\}}_{\xrightarrow[N,T\to\infty]{P} 0}\xrightarrow[N,T\to\infty]{P} K-2.$$

Moreover:

$$\left|\frac{1}{N}\sum_{n=1}^{N} tr\left(b_{FS}\left(\frac{\hat{F}^{(k)\prime}\hat{F}^{(k)}}{T}\right)^{-1}\frac{\hat{F}^{(k)\prime}}{T}X_n\vec{b}_{2n}\right)\right|$$
$$\leq \left\|\frac{1}{\sqrt{T}}M_{\hat{F}^{(k)}}F^S H^{(k)}\right\|\left\|b_{FS}\left(\frac{\hat{F}^{(k)\prime}\hat{F}^{(k)}}{T}\right)^{-1}\frac{\hat{F}^{(k)\prime}}{\sqrt{T}}\right\|\frac{1}{N}\sum_{n=1}^{N}\left\|\frac{1}{\sqrt{T}}X_n\left(\frac{X_n' M_{\hat{F}^{(k)}}X_n}{T}\right)^{-1}\frac{1}{\sqrt{T}}X_n'\right\| = o_p(1).$$

And:

$$\left|\frac{1}{N}\sum_{n=1}^{N} tr\left(b_{\hat{F}S}\left(\frac{H^{(k)\prime}F^{S\prime}F^S H^{(k)}}{T}\right)^{-1}\frac{H^{(k)\prime}F^{S\prime}}{T}X_n b_{2n}\right)\right|$$
$$\leq \left\|b_{\hat{F}S}\left(\frac{H^{(k)\prime}F^{S\prime}F^S H^{(k)}}{T}\right)^{-1}\frac{H^{(k)\prime}F^{S\prime}}{\sqrt{T}}\right\|\frac{1}{N}\sum_{n=1}^{N}\left\|\frac{1}{\sqrt{T}}X_n b_{2n}\right\| = o_p(1).$$

And also:

$$\left|\frac{1}{N}\sum_{n=1}^{N} tr\left(\left(\frac{H^{(k)\prime}F^{S\prime}F^S H^{(k)}}{T}\right)^{-1}\frac{H^{(k)\prime}F^{S\prime}}{T}X_n b_{2n}\left(\frac{\hat{F}^{(k)\prime}\hat{F}^{(k)}}{T}\right)^{-1}\frac{\hat{F}^{(k)\prime}}{T}X_n\vec{b}_{2n}\right)\right|$$
$$\leq \left(\left\|\left(\frac{H^{(k)\prime}F^{S\prime}F^S H^{(k)}}{T}\right)^{-1}\frac{H^{(k)\prime}F^{S\prime}}{\sqrt{T}}\right\|^2\frac{1}{N}\sum_{n=1}^{N}\left\|\frac{1}{\sqrt{T}}X_n b_{2n}\right\|^2\right)^{\frac{1}{2}}\left(\left\|\left(\frac{\hat{F}^{(k)\prime}\hat{F}^{(k)}}{T}\right)^{-1}\frac{\hat{F}^{(k)\prime}}{\sqrt{T}}\right\|^2\frac{1}{N}\sum_{n=1}^{N}\left\|\frac{1}{\sqrt{T}}X_n\vec{b}_{2n}\right\|^2\right)^{\frac{1}{2}}$$
$$= \left[o_p(1)\right]^{\frac{1}{2}}\left[o_p(1)\right]^{\frac{1}{2}} = o_p(1).$$

Thus:

$$\frac{1}{N}\sum_{n=1}^{N}\left\|M_{\left[X_n:\hat{F}^{(k)}\right]} - M_{\left[X_n:F^S H^{(k)}\right]}\right\|^2 = o_p(1), \tag{93}$$

and then (92) implies:

$$\frac{1}{NT}\sum_{n=1}^{N}\lambda_n' F^{0\prime}M_{\left[X_n:\hat{F}^{(k)}\right]}F^0\lambda_n \underset{N,T\to\infty}{\overset{P}{\longrightarrow}} \frac{1}{NT}\sum_{n=1}^{N}\lambda_n' F^{0\prime}M_{\left[X_n:F^S H^{(k)}\right]}F^0\lambda_n.$$

I omit the proof for the other two terms to save space, because it mimics the one I just showed. Thus, we have that:

$$NSSR\left(\hat{F}^{(k)}\right) = \frac{1}{NT}\sum_{n=1}^{N}\lambda_n' F^{0\prime}M_{\left[X_n:\hat{F}^{(k)}\right]}F^0\lambda_n + \frac{1}{NT}\sum_{n=1}^{N}\epsilon_n' M_{\left[X_n:\hat{F}^{(k)}\right]}\epsilon_n + \frac{2}{NT}\sum_{n=1}^{N}\lambda_n' F^{0\prime}M_{\left[X_n:\hat{F}^{(k)}\right]}\epsilon_n$$
$$\underset{N,T\to\infty}{\overset{P}{\longrightarrow}} \frac{1}{NT}\sum_{n=1}^{N}\lambda_n' F^{0\prime}M_{\left[X_n:F^S H^{(k)}\right]}F^0\lambda_n + \frac{1}{NT}\sum_{n=1}^{N}\epsilon_n' M_{\left[X_n:F^S H^{(k)}\right]}\epsilon_n + \frac{2}{NT}\sum_{n=1}^{N}\lambda_n' F^{0\prime}M_{\left[X_n:F^S H^{(k)}\right]}\epsilon_n$$
$$= NSSR\left(F^S H^{(k)}\right).$$

Now, note that the second term in $NSSR\left(F^S H^{(k)}\right)$ evaluated at $\hat{F}^{(k)}$ equals:

$$\frac{1}{NT}\sum_{n=1}^{N}\epsilon_n^{'}M_{\left[X_n:F^S H^{(k)}\right]}\epsilon_n = \frac{1}{NT}\sum_{n=1}^{N}\epsilon_n^{'}\epsilon_n - \frac{1}{NT}\sum_{n=1}^{N}\epsilon_n^{'}P_{\left[X_n:F^S H^{(k)}\right]}\epsilon_n.$$

Then, note that:

$$\mathbb{E}\left[\frac{1}{NT}\sum_{n=1}^{N}\epsilon_n^{'}\epsilon_n\right] = \left[\frac{1}{NT}\sum_{n=1}^{N}\sum_{t=1}^{T}\mathbb{E}\left(\epsilon_{nt}^2\right)\right] = \frac{1}{NT}\sum_{n=1}^{N}\sum_{t=1}^{T}\mathbb{E}\left(\epsilon_{nt}^2\right) = \mathbb{E}\left(\epsilon_{nt}^2\right)$$

and:

$$\mathbb{V}\left[\frac{1}{NT}\sum_{n=1}^{N}\epsilon_n^{'}\epsilon_n\right] = \frac{1}{N^2T^2}\sum_{n=1}^{N}\sum_{t=1}^{T}\mathbb{V}\left(\epsilon_{nt}^2\right) = \frac{1}{NT}\mathbb{V}\left(\epsilon_{nt}^2\right)\underset{N,T\to\infty}{\longrightarrow} 0, \tag{94}$$

where the limit in (94) is an ordinary limit. Thus:

$$\frac{1}{NT}\sum_{n=1}^{N}\epsilon_n^{'}\epsilon_n \underset{N,T\to\infty}{\overset{P}{\longrightarrow}} \mathbb{E}\left(\epsilon_{nt}^2\right). \tag{95}$$

Now, for $\frac{1}{NT}\sum_{n=1}^{N}\epsilon_n^{'}P_{\left[X_n:F^S H^{(k)}\right]}\epsilon_n$ note that:[56]

$$\begin{aligned}&\left|\frac{1}{NT}\sum_{n=1}^{N}\epsilon_n^{'}Z_n\left(Z_n^{'}Z_n\right)^{-1}Z_n^{'}\epsilon_n\right|\\ &\leq\left(\frac{1}{NT^4}\sum_{n=1}^{N}\|Z_n\|^8\right)^{\frac{1}{8}}\left(\frac{1}{NT^4}\sum_{n=1}^{N}\|\epsilon_n\|^8\right)^{\frac{1}{8}}\left(\sum_{n=1}^{N}\frac{1}{N}\left\|\left(\frac{Z_n^{'}Z_n}{T}\right)^{-1}\right\|^4\right)^{\frac{1}{4}}\sqrt{\frac{1}{N}\sum_{n=1}^{N}\left\|\frac{Z_n^{'}\epsilon_n}{T}\right\|^2}.\end{aligned}$$

So, the second term:

$$\mathbb{E}\left|\frac{1}{NT^4}\sum_{n=1}^{N}\|\epsilon_n\|^8\right| \leq \frac{1}{NT^4}\sum_{n=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{u=1}^{T}\sum_{v=1}^{T}\left(\mathbb{E}\left[\epsilon_{nt}^8\right]\mathbb{E}\left[\epsilon_{ns}^8\right]\mathbb{E}\left[\epsilon_{nu}^8\right]\mathbb{E}\left[\epsilon_{nv}^8\right]\right)^{\frac{1}{4}}$$

so $\mathbb{E}\left[\epsilon_{nt}^8\right] \leq M$ implies $\mathbb{E}\left|\frac{1}{NT^4}\sum_{n=1}^{N}\|\epsilon_n\|^8\right| \leq \vec{M} < \infty$ and thus:

$$\left(\frac{1}{NT^4}\sum_{n=1}^{N}\|\epsilon_n\|^8\right)^{\frac{1}{8}} = O_p(1).$$

Moving now to the first term:

$$\mathbb{E}\left|\frac{1}{NT^4}\sum_{n=1}^{N}\|Z_n\|^8\right| \leq \frac{1}{NT^4}\sum_{n=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{u=1}^{T}\sum_{v=1}^{T}\left\{\mathbb{E}\left[\|Z_{nt}\|^8\right]\mathbb{E}\left[\|Z_{ns}\|^8\right]\mathbb{E}\left[\|Z_{nu}\|^8\right]\mathbb{E}\left[\|Z_{nv}\|^8\right]\right\}^{\frac{1}{4}}$$

[56] Again, let us denote $Z_n = Z_n\left(F^S H^{(k)}\right)$.

so $\mathbb{E}\left[\|Z_{nt}\|^8\right] \leq M < \infty$ implies $\mathbb{E}\left|\frac{1}{NT^4}\sum_{n=1}^{N}\|Z_n\|^8\right| \leq \vec{M} < \infty$ and thus $\left(\frac{1}{NT^4}\sum_{n=1}^{N}\|Z_n\|^8\right)^{\frac{1}{8}} = O_p(1)$.
For the third term:

$$\mathbb{E}\left|\frac{1}{N}\sum_{n=1}^{N}\left\|\left(\frac{Z_n'Z_n}{T}\right)^{-1}\right\|^4\right| = \frac{1}{N}\sum_{n=1}^{N}\mathbb{E}\left\|\left(\frac{Z_n'Z_n}{T}\right)^{-1}\right\|^4 = \frac{1}{N}\sum_{n=1}^{N}\mathbb{E}\left\{\left[tr\left(D^{-2}\right)\right]^2\right\}$$

So $\mathbb{E}\left\{\left[tr\left(D^{-2}\right)\right]^2\right\} = \mathbb{E}\left\{\left(\sum_{u=1}^{P}\left[\iota_{n,u}^T\right]^{-2}\right)^2\right\} \leq M < \infty,\ \forall T, n$ where ι_{nu}^T are the eigenvalues of $\frac{Z_n'Z_n}{T}$ implies:

$$\left(\sum_{n=1}^{N}\frac{1}{N}\left\|\left(\frac{Z_n'Z_n}{T}\right)^{-1}\right\|^4\right)^{\frac{1}{4}} = O_p(1).$$

Finally:

$$\frac{1}{N}\sum_{n=1}^{N}\left\|\frac{Z_n'\epsilon_n}{T}\right\|^2 = \frac{1}{N}\frac{1}{T^2}\sum_{n=1}^{N}\left\|Z_n'\epsilon_n\right\|^2 = \frac{1}{N}\frac{1}{T^2}\sum_{n=1}^{N}\left(\sqrt{\sum_{p=1}^{P}\left(\sum_{t=1}^{T}Z_{ntp}\epsilon_{nt}\right)^2}\right)^2$$
$$= \sum_{p=1}^{P}\frac{1}{N}\frac{1}{T^2}\sum_{n=1}^{N}\left\{\sum_{t=1}^{T}Z_{ntp}^2\epsilon_{nt}^2 + \sum_{t=1}^{T}\sum_{s\neq t}^{T}Z_{ntp}Z_{nsp}\epsilon_{nt}\epsilon_{ns}\right\}.$$

Thus:

$$\mathbb{E}\left|\frac{1}{N}\sum_{n=1}^{N}\left\|\frac{Z_n'\epsilon_n}{\sqrt{T}}\right\|^2\right| = \mathbb{E}\left\{\sum_{p=1}^{P}\frac{1}{N}\frac{1}{T}\sum_{n=1}^{N}\left\{\sum_{t=1}^{T}Z_{ntp}^2\epsilon_{nt}^2 + \sum_{t=1}^{T}\sum_{s\neq t}^{T}Z_{ntp}Z_{nsp}\epsilon_{nt}\epsilon_{ns}\right\}\right\}$$
$$\leq \sum_{p=1}^{P}\frac{1}{N}\frac{1}{T}\sum_{n=1}^{N}\sum_{t=1}^{T}\left(\mathbb{E}\left[Z_{ntp}^4\right]\right)^{\frac{1}{2}}\left(\mathbb{E}\left[\epsilon_{nt}^4\right]\right)^{\frac{1}{2}}$$

and, hence, since $\mathbb{E}\left[Z_{ntp}^4\right], \mathbb{E}\left[\epsilon_{nt}^4\right] \leq M < \infty$ because of our assumption of bounded eighth moments:

$$\frac{1}{N}\sum_{n=1}^{N}\left\|\frac{Z_n'\epsilon_n}{\sqrt{T}}\right\|^2 = O_p(1).$$

And this implies:

$$\frac{1}{\sqrt{T}}\frac{1}{N}\sum_{n=1}^{N}\left\|\frac{Z_n'\epsilon_n}{\sqrt{T}}\right\|^2 = \frac{1}{N}\sum_{n=1}^{N}\left\|\frac{Z_n'\epsilon_n}{T}\right\|^2 = o_p(1).$$

Putting all the pieces back together:

$$\left|\frac{1}{NT}\sum_{n=1}^{N}\epsilon_n'Z_n\left(Z_n'Z_n\right)^{-1}Z_n'\epsilon_n\right| \leq O_p(1)\,O_p(1)\,O_p(1)\,o_p(1) = o_p(1) \tag{96}$$

and thus:

$$\frac{1}{NT}\sum_{n=1}^{N}\epsilon_n'P_{\left[X_n:F^SH^{(k)}\right]}\epsilon_n \xrightarrow[N,T\to\infty]{P} 0. \tag{97}$$

Thus, (95) and (97) imply:

$$\frac{1}{NT}\sum_{n=1}^{N}\epsilon_n' M_{[X_n:F^S H^{(k)}]}\epsilon_n \underset{N,T\to\infty}{\overset{P}{\longrightarrow}} \mathbb{E}\left(\epsilon_{nt}^2\right). \tag{98}$$

The previous computations assume there are no $o_p(1)$ terms in ϵ_n, but Lemma (3) implies that these terms vanish as well, and thus (98) still holds. Moreover, if the reduced forms come explicitly from the system (45) and (45) and in ϵ_n one of the $o_p(1)$ terms is, for example, a weighted average of $\varepsilon_{s_{nt}}$, the regressors will have similar $o_p(1)$ terms that will be correlated with ϵ_n, but again applying Lemma (3), these terms vanish as well. In that case, we would have orthogonal regressors asymptotically, and, although the last two conditions in Assumption 1 would not hold, the results would still go through.

Now for the third term in $NSSR(.)$, we have that:

$$\frac{1}{NT}\sum_{n=1}^{N}\lambda_n' F^{0\prime} M_{[X_n:F^S H^{(k)}]}\epsilon_n = \frac{1}{NT}\sum_{n=1}^{N}\lambda_n' F^{0\prime}\epsilon_n - \frac{1}{NT}\sum_{n=1}^{N}\lambda_n' F^{0\prime} P_{[X_n:F^S H^{(k)}]}\epsilon_n$$

and note that:

$$\left|\frac{1}{NT}\sum_{n=1}^{N}\lambda_n' F^{0\prime}\epsilon_n\right| \leq \vec{\lambda}\frac{1}{N}\sum_{n=1}^{N}\left\|\frac{F^{0\prime}\epsilon_n}{T}\right\|$$

and:

$$\begin{aligned}\frac{1}{N}\sum_{n=1}^{N}\left\|\frac{F^{0\prime}\epsilon_n}{T}\right\| &= \frac{1}{NT}\sum_{n=1}^{N}\sqrt{\sum_{k=1}^{K-2}\left(\sum_{t=1}^{T}F_{kt}^0\epsilon_{nt}\right)^2}\\ &= \frac{1}{NT}\sum_{n=1}^{N}\sqrt{\left\{\sum_{k=1}^{K-2}\sum_{t=1}^{T}\left(F_{kt}^0\right)^2\epsilon_{nt}^2 + \sum_{k=1}^{K-2}\sum_{t=1}^{T}\sum_{s\neq t}^{T}F_{kt}^0\epsilon_{nt}F_{st}^0\epsilon_{st}\right\}}\end{aligned}$$

and note that:

$$\mathbb{E}\left|\frac{1}{N}\sum_{n=1}^{N}\left\|\frac{F^{0\prime}\epsilon_n}{\sqrt{T}}\right\|\right| \leq \frac{1}{N\sqrt{T}}\sum_{n=1}^{N}\sqrt{\sum_{k=1}^{K-2}\sum_{t=1}^{T}\sqrt{\mathbb{E}\left[\left(F_{kt}^0\right)^4\right]\mathbb{E}\left[\epsilon_{nt}^4\right]}}$$

and thus since $\mathbb{E}\left[\left(F_{kt}^0\right)^4\right], \mathbb{E}\left[\epsilon_{nt}^4\right] \leq M < \infty$ because of our assumption of bounded moments $\frac{1}{N}\sum_{n=1}^{N}\left\|\frac{F^{0\prime}\epsilon_n}{\sqrt{T}}\right\| = O_p(1)$ and thus it follows that $\frac{1}{N}\sum_{n=1}^{N}\left\|\frac{F^{0\prime}\epsilon_n}{T}\right\| = o_p(1)$ and then $\left|\frac{1}{NT}\sum_{n=1}^{N}\lambda_n' F^{0\prime}\epsilon_n\right| \leq o_p(1)$, so:

$$\frac{1}{NT}\sum_{n=1}^{N}\lambda_n' F^{0\prime}\epsilon_n \underset{N,T\to\infty}{\overset{P}{\longrightarrow}} 0.$$

Moreover, we have that:[57]

$$\left|\frac{1}{NT}\sum_{n=1}^{N}\lambda_n' F^{0\prime} P_{\left[X_n:F^S H^{(k)}\right]}\epsilon_n\right|$$
$$\leq \left(\frac{1}{NT^4}\sum_{n=1}^{N}\|Z_n\|^8\right)^{\frac{1}{8}}\left(\frac{\vec{\lambda}}{NT^4}\sum_{n=1}^{N}\left\|F^0\right\|^8\right)^{\frac{1}{8}}\left(\sum_{n=1}^{N}\frac{1}{N}\left\|\left(\frac{Z_n' Z_n}{T}\right)^{-1}\right\|^4\right)^{\frac{1}{4}}\sqrt{\frac{1}{N}\sum_{n=1}^{N}\left\|\frac{Z_n'\epsilon_n}{T}\right\|^2}$$

so the only new term here is $\frac{\vec{\lambda}}{NT^4}\sum_{n=1}^{N}\left\|F^0\right\|^8$, and with the assumption on bounded moments in Subsection (4.1), and the same logic as before, we know that this term is $O_p(1)$, and thus using (96):

$$\frac{1}{NT}\sum_{n=1}^{N}\lambda_n' F^{0\prime} P_{\left[X_n:F^S H^{(k)}\right]}\epsilon_n \xrightarrow[N,T\to\infty]{P} 0.$$

Then, we have shown that:

$$\frac{1}{NT}\sum_{n=1}^{N}\lambda_n' F^{0\prime} M_{\left[X_n:F^S H^{(k)}\right]}\epsilon_n \xrightarrow[N,T\to\infty]{P} 0.$$

And again Lemma (3) implies that if we incorporate the $o_p(1)$ terms in ϵ_n or the regressors, the same computations go through.

Now, for the first term of $NSRR(.)$, we next show that:

$$\frac{1}{NT}\sum_{n=1}^{N}\lambda_n' F^{0\prime} M_{\left[X_n:F^S H^{(k)}\right]}F^0\lambda_n = o_p(1) \iff C\left(F^S H^{(k)}\right)\ni F^0. \tag{99}$$

We start by first showing that $\frac{1}{NT}\sum_{n=1}^{N}\lambda_n' F^{0\prime} M_{\left[X_n:F^S H^{(k)}\right]}F^0\lambda_n$ converges in probability. To that end, first note that:

$$\frac{1}{T}u_n^{(k)\prime} M_{\left[X_n:F^S H^{(k)}\right]}u_n^{(k)}$$

is such that:

$$\mathbb{E}\left(\left|\frac{1}{T}\lambda_n' F^{0\prime} F^0\lambda_n\right|^2\right) < \infty$$
$$\mathbb{E}\left(\left|\frac{1}{T}\lambda_n' F^{0\prime} P_{\left[X_n:F^S H^{(k)}\right]}F^0\lambda_n\right|^2\right) < \infty$$

implying that $\frac{1}{T}u_n^{(k)\prime} M_{\left[X_n:F^S H^{(k)}\right]}u_n^{(k)}$ is uniform integrable. For the second term, note that:

$$\mathbb{E}\left(\left|\frac{1}{T}\lambda_n' F^{0\prime} P_{\left[X_n:F^S H^{(k)}\right]}F^0\lambda_n\right|^2\right) \leq (P+K-2)\,\vec{\lambda}^4\frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\sqrt{\mathbb{E}\left[\left\|F_t^0\right\|^4\right]\mathbb{E}\left[\left\|F_s^0\right\|^4\right]}$$

so the assumption on bounded moments in Subsection (4.1) implies:

$$\mathbb{E}\left(\left|\frac{1}{T}\lambda_n' F^{0\prime} P_{\left[X_n:F^S H^{(k)}\right]}F^0\lambda_n\right|^2\right) \leq \vec{M} < \infty$$

[57] Again, let us denote $Z_n = Z_n\left(F^S H^{(k)}\right)$.

and thus $\frac{1}{T}\lambda_n' F^{0\prime} P_{\left[X_n : F^S H^{(k)}\right]} F^0 \lambda_n$ is uniform integrable. Now, note that:

$$\mathbb{E}\left(\left|\frac{1}{T}\lambda_n' F^{0\prime} F^0 \lambda_n\right|^2\right) \leq \mathbb{E}\left(\left\|\frac{\lambda_n' F^{0\prime}}{\sqrt{T}}\right\|^4\right) \tag{100}$$

which by the previous argument is also uniform integrable. So given that $\frac{1}{T} u_n^{(k)\prime} M_{\left[X_n : F^S H^{(k)}\right]} u_n^{(k)}$ is uniform integrable, now note that:

$$\frac{1}{T}\lambda_n' F^{0\prime} F^0 \lambda_n \xrightarrow[N,T\to\infty]{P} \mathbb{E}\left[\left(F_t^{0\prime}\lambda_n\right)^2\right]$$

$$\frac{1}{T}\lambda_n' F^{0\prime} Z_n \left(Z_n' Z_n\right)^{-1} Z_n' F^0 \lambda_n \xrightarrow[N,T\to\infty]{P} Q_{ZF}' \left(Q_{ZZ}\right)^{-1} Q_{ZF}$$

and thus:

$$\frac{1}{T}\mathbb{E}\left[\lambda_n' F^{0\prime} F^0 \lambda_n\right] \xrightarrow[N,T\to\infty]{} \mathbb{E}\left[\left(F_t^{0\prime}\lambda_n\right)^2\right]$$

$$\frac{1}{T}\mathbb{E}\left[\frac{1}{T}\lambda_n' F^{0\prime} Z_n \left(Z_n' Z_n\right)^{-1} Z_n' F^0 \lambda_n\right] \xrightarrow[N,T\to\infty]{} \mathbb{E}\left[Q_{ZF}' \left(Q_{ZZ}\right)^{-1} Q_{ZF}\right]$$

so:

$$\mathbb{E}\left[\frac{1}{T}\lambda_n' F^{0\prime} P_{\left[X_n : F^S H^{(k)}\right]} F^0 \lambda_n\right] \xrightarrow[N,T\to\infty]{} \mathbb{E}\left[\left(F_t^{0\prime}\lambda_n\right)^2\right] - \mathbb{E}\left[Q_{ZF}' \left(Q_{ZZ}\right)^{-1} Q_{ZF}\right].$$

Now, we show that:

$$\mathbb{V}\left(\frac{1}{NT}\sum_{n=1}^{N} u_n^{(k)\prime} M_{\left[X_n : F^S H^{(k)}\right]} u_n^{(k)}\right) \xrightarrow[N,T\to\infty]{} 0.$$

Note that:

$$\begin{aligned}
&\mathbb{V}\left(\frac{1}{NT}\sum_{n=1}^{N} u_n^{(k)\prime} M_{\left[X_n : F^S H^{(k)}\right]} u_n^{(k)}\right) \\
&= \frac{1}{N^2T^2}\sum_{n=1}^{N}\mathbb{V}\left\{\lambda_n' F^{0\prime} M_{\left[X_n : F^S H^{(k)}\right]} F^0 \lambda_n\right\} + \frac{N(N-1)}{N^2T^2} cov\left\{\lambda_n' F^{0\prime} M_{\left[X_n : F^S H^{(k)}\right]} F^0 \lambda_n, \lambda_j' F^{0\prime} M_{\left[X_j : F^S H^{(k)}\right]} F^0 \lambda_j\right\}
\end{aligned}$$

and:

$$\begin{aligned}
&\left|\frac{N(N-1)}{N^2T^2} cov\left\{\lambda_n' F^{0\prime} M_{\left[X_n : F^S H^{(k)}\right]} F^0 \lambda_n, \lambda_j' F^{0\prime} M_{\left[X_j : F^S H^{(k)}\right]} F^0 \lambda_j\right\}\right| \\
&\leq \frac{N(N-1)}{N^2}\sqrt{\mathbb{E}\left[\left(\frac{\lambda_n' F^{0\prime} M_{\left[X_n : F^S H^{(k)}\right]} F^0 \lambda_n}{T} - \mathbb{E}\left[\frac{\lambda_n' F^{0\prime} M_{\left[X_n : F^S H^{(k)}\right]} F^0 \lambda_n}{T}\right]\right)^2\right]\mathbb{E}\left[\ldots\right]}
\end{aligned}$$

and since $\hat{u}_{nt}^{(k)2}$ is stationary with absolutely summable autocovariances:

$$\mathbb{E}\left[\left(\frac{\lambda_n' F^{0\prime} M_{\left[X_n : F^S H^{(k)}\right]} F^0 \lambda_n}{T} - \mathbb{E}\left[\frac{\lambda_n' F^{0\prime} M_{\left[X_n : F^S H^{(k)}\right]} F^0 \lambda_n}{T}\right]\right)^2\right] \xrightarrow[N,T\to\infty]{} 0$$

and thus:

$$\frac{N(N-1)}{N^2T^2} cov\left\{\lambda_n' F^{0\prime} M_{\left[X_n:F^S H^{(k)}\right]} F^0 \lambda_n, \lambda_j' F^{0\prime} M_{\left[X_j:F^S H^{(k)}\right]} F^0 \lambda_j\right\} \underset{N,T\to\infty}{\longrightarrow} 0. \tag{101}$$

Moreover, note that:

$$\begin{aligned}\mathbb{V}\left\{\frac{1}{T}\lambda_n' F^{0\prime} M_{\left[X_n:F^S H^{(k)}\right]} F^0 \lambda_n\right\} &= \mathbb{E}\left(\left[\frac{1}{T}\lambda_n' F^{0\prime} M_{\left[X_n:F^S H^{(k)}\right]} F^0 \lambda_n\right]^2\right) - \left(\mathbb{E}\left[\frac{1}{T}\lambda_n' F^{0\prime} M_{\left[X_n:F^S H^{(k)}\right]} F^0 \lambda_n\right]\right)^2 \\ &\leq \mathbb{E}\left(\left[\frac{1}{T}\lambda_n' F^{0\prime} F^0 \lambda_n\right]^2\right) \leq \vec{M} < \infty \end{aligned} \tag{102}$$

where the inequality follows from (100). To recap we have shown:

$$\begin{aligned}&\mathbb{V}\left(\frac{1}{NT}\sum_{n=1}^N u_n^{(k)\prime} M_{\left[X_n:F^S H^{(k)}\right]} u_n^{(k)}\right) \\ &\leq \underbrace{\frac{1}{N}\vec{M}}_{\underset{N,T\to\infty}{\longrightarrow} 0} + \underbrace{\frac{N(N-1)}{N^2}}_{\underset{N,T\to\infty}{\longrightarrow} 1} \underbrace{cov\left\{\frac{\lambda_n' F^{0\prime} M_{\left[X_n:F^S H^{(k)}\right]} F^0 \lambda_n}{T}, \frac{\lambda_j' F^{0\prime} M_{\left[X_j:F^S H^{(k)}\right]} F^0 \lambda_j}{T}\right\}}_{\underset{N,T\to\infty}{\longrightarrow} 0} \underset{N,T\to\infty}{\longrightarrow} 0.\end{aligned}$$

Now that we know that $\frac{1}{NT}\sum_{n=1}^N \lambda_n' F^{0\prime} M_{\left[X_n:F^S H^{(k)}\right]} F^0 \lambda_n$ converges in probability, we show that this limit cannot be zero. First, note that:

$$\begin{aligned}&\mathbb{E}\left(\left|\frac{1}{NT}\sum_{n=1}^N \lambda_n' F^{0\prime} F^0 \lambda_n\right|^2\right) \\ &\leq \frac{1}{N^2T^2}\sum_{n=1}^N\sum_{j=1}^N \sqrt{\left[\sum_{t=1}^T\sum_{s=1}^T \sqrt{\left\{\sum_{u=1}^{K-2}\sum_{v=1}^{K-2}\sum_{x=1}^{K-2}\sum_{w=1}^{K-2} \left(\vec{\lambda}\right)^4 \sqrt{\left(\sqrt{\mathbb{E}\left[\left(F_{kt}^0\right)^4\right]\mathbb{E}\left[\left(F_{ut}^0\right)^4\right]}\right)(\ldots)}\right\}\ldots}\right]\ldots}\end{aligned}$$

which implies that $\frac{1}{NT}\sum_{n=1}^N \lambda_n' F^{0\prime} F^0 \lambda_n$ is uniformly integrable and thus since:

$$\left|\frac{1}{NT}\sum_{n=1}^N \lambda_n' F^{0\prime} M_{\left[X_n:F^S H^{(k)}\right]} F^0 \lambda_n\right| \leq \left|\frac{1}{NT}\sum_{n=1}^N \lambda_n' F^{0\prime} F^0 \lambda_n\right|$$

we have that $\frac{1}{NT}\sum_{n=1}^N \lambda_n' F^{0\prime} M_{\left[X_n:F^S H^{(k)}\right]} F^0 \lambda_n$ is uniformly integrable as well. But this implies that since $\frac{1}{NT}\sum_{n=1}^N \lambda_n' F^{0\prime} M_{\left[X_n:F^S H^{(k)}\right]} F^0 \lambda_n$ converges in probability, then it also converges in L^1, and thus it implies $\mathbb{E}\left[\frac{1}{NT}\sum_{n=1}^N \lambda_n' F^{0\prime} M_{\left[X_n:F^S H^{(k)}\right]} F^0 \lambda_n\right]$ converges to the expected limit. We now show that this cannot be zero:[58]

$$\mathbb{E}\left[\frac{1}{NT}\sum_{n=1}^{N}\lambda_n' F^{0\prime} M_{\left[X_n:F^S H^{(k)}\right]} F^0 \lambda_n\right] = \frac{1}{NT}\sum_{n=1}^{N}\mathbb{E}\left[\lambda_n' F^{0\prime} M_{\left[X_n:F^S H^{(k)}\right]} F^0 \lambda_n\right]$$
$$= \frac{1}{NT}\sum_{n=1}^{N}\mathbb{E}\left[u_n^{(k)\prime} M_{\left[X_n:F^S H^{(k)}\right]} u_n^{(k)}\right]$$
$$> \frac{1}{NT}\sum_{n=1}^{N} O(T)$$
$$= \frac{O(T)}{T}$$
$$\underset{N,T\to\infty}{\longrightarrow} V^{(k)} > 0.$$

Thus we know not only $\frac{1}{N}\sum_{n=1}^{N}\lambda_n' \frac{F^{0\prime} M_{\left[X_n:F^{S(k)}\right]} F^0}{T}\lambda_n \geq 0$ but also $\frac{1}{N}\sum_{n=1}^{N}\lambda_n' \frac{F^{0\prime} M_{\left[X_n:F^{S(k)}\right]} F^0}{T}\lambda_n \overset{P}{\nrightarrow} 0$.

Then, putting all the pieces together, we have that:

$$NSSR\left(\hat{F}^{(k)}\right) \underset{N,T\to\infty}{\overset{P}{\longrightarrow}} \begin{cases} \mathbb{E}\left[\epsilon_{nt}^2\right] & if \quad (k) = (k)^* \\ > \mathbb{E}\left[\epsilon_{nt}^2\right] + V^{(k)} & if \quad (k) \neq (k)^* \end{cases}. \tag{103}$$

This result means that:

$$\left\| M_{F^0} - M_{\hat{F}^*} \right\| = o_p(1). \tag{104}$$

Now, using (104), with the same argument as those at the end of Proposition 13, we have that:

$$\left\|\hat{\beta}_n^{(N)}\left(\hat{F}^*\right) - \beta_n^{(N)}\right\| = o_p(1),$$

which completes the proof. ∎

Corollary 4

Proof. The proof follows the exact same steps as Lemma 1, so I omit them here. ∎

Proposition 15

Proof. The first part of the proof is exactly the same as the one in Ando and Bai (2015). The only modification needed is the use of Lemma (3) to show that the $o_p(1)$ terms that come from the indirect observability of s_t and $\varepsilon_{s_{nt}}$ vanish. This follows the same steps as the beginning of Proposition 13, which uses Lemma (3), and thus we do not repeat it here. With a consistent estimator of the space spanned by F^0, the rest of the proof is identical to that of Proposition 13. ∎

Proposition 16

[58] The $O(T)$ term should be $O^{(k)}(T)$ to make clear that $O^{(k)}(T)$ is an $O(T)$ term that depends on the subset (k), but we omit the (k) in the superscript for a cleaner notation.

Proof. The proof will follow the steps described in Subsection 5.4 in order to move forward in an organized manner. I will allow for a term like $\varphi_n \frac{1}{N}\sum_{j=1}^{N}\vec{\beta}_{2j}\varepsilon_{s_{jt}}$ in (56), for some arbitrary weights $\vec{\beta}_{2j}$, to mimic some of the models in Section 3 more closely. Thus, the error term in (56) after the unobserved macro shocks becomes $\omega_{nt}^{Y}+\varphi_n \frac{1}{N}\sum_{j=1}^{N}\vec{\beta}_{2j}\varepsilon_{s_{jt}}$. This is just to make the proof a bit more general. Also, I will assume $\sum_{n=1}^{N}\theta_{Zn}\frac{\vec{\beta}_{2n}}{1-\delta_Y\beta_{2n}}=\sum_{n=1}^{N}\theta_{u^s n}\frac{\vec{\beta}_{2n}}{1-\delta_Y\beta_{2n}}x_n=0, x_n=\kappa_{0n},\kappa_{1n},\kappa_{u^s,n},\kappa_{u^Z,n}$. Let me note that this assumption, like the equivalent one for β_{2n} in point 5 of the statement in this proposition, could be weakened by requiring θ_{Zn} and $\theta_{u^s n}$ to have mean zero and be independent across n; this is one place where I choose to make a stronger assumption to make notation easier at no conceptual cost in terms of the strategy. Moreover, I will also assume we only observe $s_{nt}:=s_t+\varepsilon_{s_{nt}}$, for a slightly more general proof as well. This then implies $\hat{\varepsilon}_{s_{nt}}=\varepsilon_{s_{nt}}-\frac{1}{N}\sum_{n=1}^{N}\theta_{Zn}\omega_{nt}^{Z}-\frac{1}{N}\sum_{n=1}^{N}\omega_{nt}^{s}$.

With the same arguments as in Propositions 13, 14 and 15, the first step in (66) gives:

$$\left\|M_{\vec{F}}-M_{\hat{F}}\right\|=o_p(1) \tag{105}$$

$$\left|\Omega_{nt}-\hat{F}_t'\hat{\phi}_n^{y}\right|=o_p(1). \tag{106}$$

Now, in step 2 we can rewrite:

$$\frac{1}{NT}\sum_{n=1}^{N}\sum_{t=1}^{T}\left(\hat{\varepsilon}_{s_{nt}}-\phi_n^{\varepsilon}u_t^{s}-\phi_{Zn}^{\varepsilon}\varepsilon_{Z_{nt}}-\phi_{\Omega n}^{\varepsilon}\hat{F}_t'\hat{\phi}_n^{y}\right)^2=$$

$$\frac{1}{NT}\sum_{n=1}^{N}\sum_{t=1}^{T}\left(\varepsilon_{s_{nt}}-\frac{1}{N}\sum_{n=1}^{N}\theta_{Zn}\omega_{nt}^{Z}-\frac{1}{N}\sum_{n=1}^{N}\omega_{nt}^{s}-\phi_n^{\varepsilon}u_t^{s}-\phi_{Zn}^{\varepsilon}\varepsilon_{Z_{nt}}-\phi_{\Omega n}^{\varepsilon}\left[\hat{F}_t'\hat{\phi}_n^{y}-\Omega_{nt}+\Omega_{nt}\right]\right)^2=$$

$$\frac{1}{NT}\sum_{n=1}^{N}\sum_{t=1}^{T}\left(\varepsilon_{s_{nt}}-\phi_n^{\varepsilon}u_t^{s}-\phi_{Zn}^{\varepsilon}\varepsilon_{Z_{nt}}-\phi_{\Omega n}^{\varepsilon}\Omega_{nt}-\phi_{\Omega n}^{\varepsilon}\left[\hat{F}_t'\hat{\phi}_n^{y}-\Omega_{nt}\right]-\frac{1}{N}\sum_{n=1}^{N}\theta_{Zn}\omega_{nt}^{Z}-\frac{1}{N}\sum_{n=1}^{N}\omega_{nt}^{s}\right)^2.$$

Note that by (106) and our assumptions $\phi_{\Omega n}^{\varepsilon}\left[\hat{F}_t'\hat{\phi}_n^{y}-\Omega_{nt}\right]-\frac{1}{N}\sum_{n=1}^{N}\theta_{Zn}\omega_{nt}^{Z}-\frac{1}{N}\sum_{n=1}^{N}\omega_{nt}^{s}=o_p(1)$ for bounded $\phi_{\Omega n}^{\varepsilon}$. Using the same arguments as in Lemma 3 and Proposition 14, we get from (67):

$$\frac{1}{\sqrt{T}}\left\|\hat{u}^{s}-u^{s}h^{*}\right\|=o_p(1), \tag{107}$$

for some non-zero constant h^{*}.

Moving forward, using (107) and standard arguments for the consistency of IV estimators (see, for example, White (2001)), together with Lemma 3 to take into account the non-observability of s_t, steps 3 and 4 give:

$$\left|\hat{u}_t^{Z}-u_t^{Z}\right|=o_p(1).$$

Applying the same machinery to step 5, and using Assumption 11, we have that:

$$\left\|M_{\hat{F}^{RC}}-M_{F^0}\right\|=o_p(1) \tag{108}$$

where $\hat{F}^{RC}=\left[\hat{F}_1^{RC},\hat{F}_2^{RC},\hat{F}_3^{RC}\right]$ and $F^0=\left[F_1,F_2,u_t^{s}\right]$. And using (108) with the arguments of Proposition 13, we get (69). ■

Proposition 17

Proof. The first part of the results follows simply from direct computation. I do not give all the details to save space, but note for example that we can compute the inverse of A as follows. Let us define:

$$v' = \left(\frac{1}{N}, \dots, \frac{1}{N}, 0, 0, 0, \dots, 0\right)$$
$$u' = (0, \dots, 0, b_y, d_y, \delta_Y, \dots, \delta_Y)$$

then:

$$uv' = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ b_y \\ d_y \\ \delta_Y \\ \vdots \\ \delta_Y \end{pmatrix} \left(\frac{1}{N}, \dots, \frac{1}{N}, 0, 0, 0, \dots, 0\right) = \begin{pmatrix} 0 & \cdots & 0 & 0 & 0 & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & 0 & 0 & 0 & \cdots & 0 \\ b_y \frac{1}{N} & \cdots & b_y \frac{1}{N} & 0 & 0 & 0 & \cdots & 0 \\ d_y \frac{1}{N} & \cdots & d_y \frac{1}{N} & 0 & 0 & 0 & \cdots & 0 \\ \delta_Y \frac{1}{N} & \cdots & \delta_Y \frac{1}{N} & 0 & 0 & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \delta_Y \frac{1}{N} & \cdots & \delta_Y \frac{1}{N} & 0 & 0 & 0 & \cdots & 0 \end{pmatrix}.$$

Also:

$$\begin{aligned} \Phi &= A - uv' \\ &= \begin{pmatrix} 1 & \cdots & 0 & b_1 & 0 & c_{11} & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 1 & b_N & 0 & 0 & \cdots & c_{NN} \\ 0 & \cdots & 0 & 1 & b_m & 0 & \cdots & 0 \\ 0 & \cdots & 0 & d_g & 1 & 0 & \cdots & 0 \\ -\delta_Y & \cdots & 0 & 0 & 0 & 1 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & -\delta_Y & 0 & 0 & 0 & \cdots & 1 \end{pmatrix}. \end{aligned}$$

After some computations, we get:

$$\Phi^{-1} = \begin{pmatrix} 1 - \frac{\delta_Y c_{11}}{1+\delta_Y c_{11}} & \cdots & 0 & J_1 & \hat{J}_1 & -\frac{c_{11}}{1+\delta_Y c_{11}} & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 1 - \frac{\delta_Y c_{NN}}{1+\delta_Y c_{NN}} & J_N & \hat{J}_N & 0 & \cdots & -\frac{c_{NN}}{1+\delta_Y c_{NN}} \\ 0 & \cdots & 0 & \frac{1}{1-b_m d_g} & -\frac{b_m}{1-b_m d_g} & 0 & \cdots & 0 \\ 0 & \cdots & 0 & -\frac{d_g}{1-b_m d_g} & \frac{1}{1-b_m d_g} & 0 & \cdots & 0 \\ \delta_Y \frac{1}{1+\delta_Y c_{11}} & \cdots & 0 & \varphi_1 & \hat{\varphi}_1 & \frac{1}{1+\delta_Y c_{11}} & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & \delta_Y \frac{1}{1+\delta_Y c_{NN}} & \varphi_N & \hat{\varphi}_N & 0 & \cdots & \frac{1}{1+\delta_Y c_{NN}} \end{pmatrix},$$

where:

$$
\begin{aligned}
J_n &= -b_n \frac{1}{1-b_m d_g} + c_{nn}\delta_Y b_n \frac{1}{1-b_m d_g}\frac{1}{1+\delta_Y c_{nn}} \\
\hat{J}_n &= b_n \frac{b_m}{1-b_m d_g} - c_{nn}\frac{b_m}{1-b_m d_g}\delta_Y b_n \frac{1}{1+\delta_Y c_{nn}} \\
\varphi_n &= -\delta_Y b_n \frac{1}{1-b_m d_g}\frac{1}{1+\delta_Y c_{nn}} \\
\hat{\varphi}_n &= \frac{b_m}{1-b_m d_g}\delta_Y b_n \frac{1}{1+\delta_Y c_{nn}}.
\end{aligned}
$$

Now, using the Sherman-Morrison Formula, we get A^{-1} (I omit the computations here to save space). With this matrix, we can now use Blanchard and Kahn (1980) and compute the solution with the aid of $A^{-1}\Theta$ plus the assumptions in this proposition. After this, we can eliminate from the system the equation for G_t and repeat these steps. Using the solution to these two systems, it is easy to see that we can impose restrictions on A and Θ and in the processes $\vec{\xi}_t$, for example requiring that $\varepsilon_{\xi_t^n}$ be independent for all n and t and identically distributed across t with $\mathbb{E}\left[\varepsilon_{\xi_t^1}\right] = 0$ and $\mathbb{E}\left[\left|\varepsilon_{\xi_t^1}\right|^{16}\right] < C$, for (68) to be consistent. ■

B Appendix: Simulations and Empirical Results

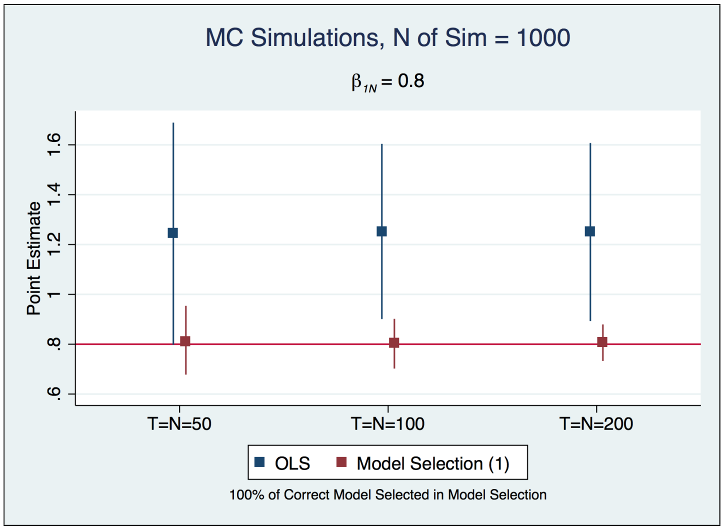

Figure 8: This figure shows the Monte Carlo simulations for the micro-global elasticity in Case 1.

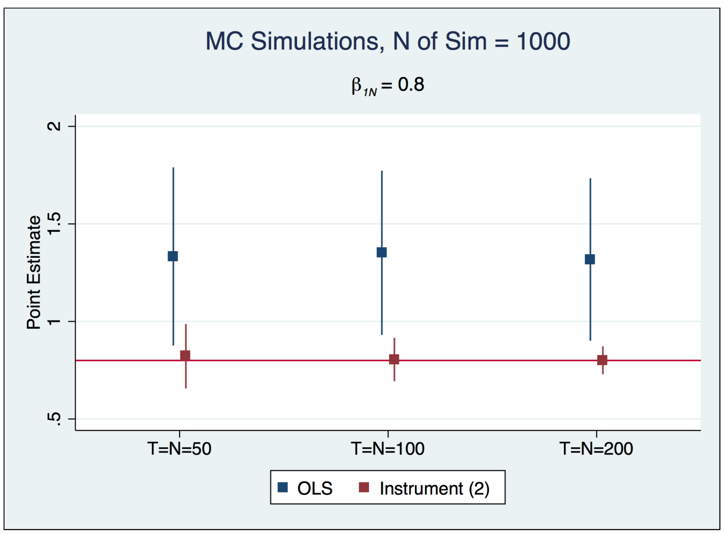

Figure 9: This figure shows the Monte Carlo simulations for the micro-global elasticity in Case 2.

	Dep. Variable: Real GDP Per Capita Growth			
	(1)	(2)	(3)	(4)
	OLS	Model Selection	Instrument	SEM
G_t	-1.08***	-0.08	-0.08	0.07
	(0.38)	(0.34)	(0.319)	(0.82)
G_{t-1}	0.28	0.69***	0.77***	0.69***
	(0.48)	(0.18)	(0.18)	(0.28)
G_{t-2}	0.38	-0.254	-0.23	0.04
	(0.39)	(0.344)	(0.34)	(0.30)
$State\ Time\ Trend$		✓	✓	✓
$Time\ Fixed\ Effect$		✓	✓	✓
$Interactive\ FE$		✓	✓	✓
$Number\ of\ IE$	0	5	5	5
Observations	37	1,739	1,739	1,739
Period	1971-2008	1971-2008	1971-2008	1971-2008
Number of Regions	-	47	47	47

Table 4: This table presents estimates of the fiscal multiplier. Column 1 is a simple time series OLS regression of Y_t on a constant, G_t, G_{t-1}, G_{t-2} and Y_{t-1}. Columns 2 to 4 estimate (74) for cases 1, 2 and 3, respectively. All of the regressions control for Y_{nt-1}. I exclude 4 states because they have yearly variations of above 20% in their GSP, which translate into unreliably large micro-global multipliers, and a very small population: North Dakota (0.23% of the US population), South Dakota (0.27%), Wyoming (0.18%) and Alaska (0.23%). Standard errors are in parentheses and are obtained with the nonparametric bootstrap; see Section 5 for details. The coefficients with $***$ are significant at the 1% confidence level; with $**$ are significant at the 5% confidence level; and with $*$ are significant at the 10% confidence level.

	Dep. Variable: Real GDP Per Capita Growth			
	(1)	(2)	(3)	(4)
	OLS	Model Selection	Instrument	SEM
G_t	-0.85**	0.13*	0.63**	0.23
	(0.40)	(0.07)	(0.33)	(0.35)
G_{t-1}	0.65	0.80***	0.86***	1.01***
	(0.48)	(0.19)	(0.22)	(0.34)
G_{t-2}	0.58	0.01	-0.23	0.33
	(0.40)	(0.40)	(0.60)	(0.34)
$State\ Time\ Trend$				
$Time\ Fixed\ Effect$		✓	✓	✓
$Interactive\ FE$		✓	✓	✓
$Number\ of\ IE$	0	5	5	5
Observations	37	1,739	1,739	1,739
Period	1971-2008	1971-2008	1971-2008	1971-2008
Number of Regions	-	47	47	47

Table 5: This table presents estimates of the fiscal multiplier. Column 1 is a simple time series OLS regression of Y_t on G_t, G_{t-1}, G_{t-2} and Y_{t-1}. Columns 2 to 4 estimate (74) for cases 1, 2 and 3, respectively. All of the regressions control for Y_{nt-1}. I exclude 4 states because they have yearly variations of above 20% in their GSP, which translate into unreliably large micro-global multipliers, and a very small population: North Dakota (0.23% of the US population), South Dakota (0.27%), Wyoming (0.18%) and Alaska (0.23%). Standard errors are in parentheses and are obtained with the nonparametric bootstrap; see Section 5 for details. The coefficients with $***$ are significant at the 1% confidence level; with $**$ are significant at the 5% confidence level; and with $*$ are significant at the 10% confidence level.

	Dep. Variable: Real GDP Per Capita Growth			
	(1) OLS	(2) Model Selection	(3) Instrument	(4) SEM
G_t	-0.85**	0.14**	0.42	-0.10
	(0.40)	(0.07)	(0.27)	(0.48)
G_{t-1}	0.65	0.76***	0.86***	0.94***
	(0.48)	(0.17)	(0.20)	(0.38)
G_{t-2}	0.58	0.14	0.04	0.38
	(0.40)	(0.33)	(0.37)	(0.31)
$State\ Time\ Trend$				
$Time\ Fixed\ Effect$				
$Interactive\ FE$		✓	✓	✓
$Number\ of\ IE$	0	5	5	5
Observations	37	1,739	1,739	1,739
Period	1971-2008	1971-2008	1971-2008	1971-2008
Number of Regions	-	47	47	47

Table 6: This table presents estimates of the fiscal multiplier. Column 1 is a simple time series OLS regression of Y_t on G_t, G_{t-1}, G_{t-2} and Y_{t-1}. Columns 2 to 4 estimate (74) for cases 1, 2 and 3, respectively. All of the regressions control for Y_{nt-1}. I exclude 4 states because they have yearly variations of above 20% in their GSP, which translate into unreliably large micro-global multipliers, and a very small population: North Dakota (0.23% of the US population), South Dakota (0.27%), Wyoming (0.18%) and Alaska (0.23%). Standard errors are in parentheses and are obtained with the nonparametric bootstrap; see Section 5 for details. The coefficients with $***$ are significant at the 1% confidence level; with $**$ are significant at the 5% confidence level; and with $*$ are significant at the 10% confidence level.

	(1) Inst.	(2) Inst.	(3) Inst.	(4) Inst.	(5) Inst.	(6) Inst.	(7) Inst.
G_{t+1}	-	-	-	-	-	-	0.08
							(0.15)
G_t	-	-	-	-	-	-	-0.01
							(0.17)
G_{t-1}	0.93***	0.88***	0.83***	0.75***	1.23***	1.20***	0.85***
	(0.07)	(0.07)	(0.07)	(0.08)	(0.11)	(0.11)	(0.14)
G_{t-2}	-	-	-	-	-	-	-0.35
							(0.21)
STT							✓
TFE							
IE		✓	✓	✓	✓	✓	✓
$N° \ of\ IE$	0	1	2	3	4	5	5
Obs.	1,739	1,739	1,739	1,739	1,739	1,739	1,739
Period	1971-2008	1971-2008	1971-2008	1971-2008	1971-2008	1971-2008	1971-2008
Regions	47	47	47	47	47	47	47

Table 7: This table presents estimates of the fiscal multiplier. Columns 1 through 6 show what happens in case 2 when we go from 0 to 5 IE, when using only G_{t-1}, and without using state time trends. Column 7 shows what happens in case 2 when we include a lead. All of the regressions control for Y_{nt-1}. I exclude 4 states because they have yearly variations of above 20% in their GSP, which translate into unreliably large micro-global multipliers, and a very small population: North Dakota (0.23% of the US population), South Dakota (0.27%), Wyoming (0.18%) and Alaska (0.23%). Standard errors are in parentheses and are obtained with the wild bootstrap; see Section 5 for details. The coefficients with $***$ are significant at the 1% confidence level; with $**$ are significant at the 5% confidence level; and with $*$ are significant at the 10% confidence level.

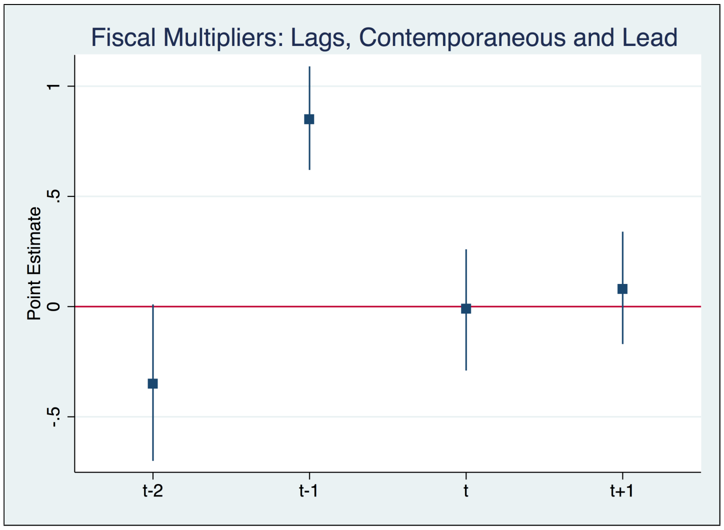

Figure 10: This figure shows the point estimates from column 7 in Table 7, along with their 90% confidence intervals.

	(1) SEM	(2) SEM	(3) SEM	(4) SEM	(5) SEM
G_t	-0.62***	0.14	-0.08	0.18	0.01
	(0.23)	(0.61)	(0.44)	(0.72)	(0.59)
G_{t-1}	0.25**	0.24	0.35**	0.29	0.69***
	(0.13)	(0.16)	(0.16)	(0.24)	(0.29)
G_{t-2}	-0.05	0.28	0.28**	0.36**	0.03
	(0.14)	(0.20)	(0.14)	(0.20)	(0.24)
$State\ Time\ Trend$	✓	✓	✓	✓	✓
$Time\ Fixed\ Effect$					
$Interactive\ FE$	✓	✓	✓	✓	✓
$Number\ of\ IE$	1	2	3	4	5
Observations	1,739	1,739	1,739	1,739	1,739
Period	1971-2008	1971-2008	1971-2008	1971-2008	1971-2008
Number of Regions	47	47	47	47	47

Table 8: This table presents estimates of the fiscal multiplier. Columns 1 through 5 show what happens in case 3 when we go from 1 to 5 IE, when using G_t, G_{t-1}, G_{t-2}. All of the regressions control for Y_{nt-1}. I exclude 4 states because they have yearly variations of above 20% in their GSP, which translate into unreliably large micro-global multipliers, and a very small population: North Dakota (0.23% of the US population), South Dakota (0.27%), Wyoming (0.18%) and Alaska (0.23%). Standard errors are in parentheses and are obtained with the wild bootstrap; see Section 5 for details. The coefficients with $***$ are significant at the 1% confidence level; with $**$ are significant at the 5% confidence level; and with $*$ are significant at the 10% confidence level.

	Dep. Variable: Real GDP Per Capita Growth			
	(1)	(2)	(3)	(4)
	OLS	Model Selection	Instrument	SEM
G_t	-1.08***	-0.07	0.02	0.01
	(0.38)	(0.28)	(0.32)	(0.71)
G_{t-1}	0.28	0.65***	0.80***	0.69***
	(0.48)	(0.15)	(0.25)	(0.32)
G_{t-2}	0.38	0.08	-0.34	0.03
	(0.39)	(0.11)	(0.51)	(0.35)
$State\ Time\ Trend$		✓	✓	✓
$Time\ Fixed\ Effect$				
$Interactive\ FE$		✓	✓	✓
$Number\ of\ IE$	0	5	5	5
Observations	37	1,739	1,739	1,739
Period	1971-2008	1971-2008	1971-2008	1971-2008
Number of Regions	-	47	47	47

Table 9: This table presents estimates of the fiscal multiplier. Column 1 is a simple time series OLS regression of Y_t on a constant, G_t, G_{t-1}, G_{t-2} and Y_{t-1}. Columns 2 to 4 estimate (74) for cases 1, 2 and 3, respectively. All of the regressions control for Y_{nt-1}. I exclude 4 states because they have yearly variations of above 20% in their GSP, which translate into unreliably large micro-global multipliers, and a very small population: North Dakota (0.23% of the US population), South Dakota (0.27%), Wyoming (0.18%) and Alaska (0.23%). Standard errors are in parentheses and are obtained with the nonparametric bootstrap; see Section 5 for details. The coefficients with $***$ are significant at the 1% confidence level; with $**$ are significant at the 5% confidence level; and with $*$ are significant at the 10% confidence level.

References

Abadie, Alberto, Susan Athey, Guido W. Imbens, and Jeffrey M. Wooldridge, "Finite Population Causal Standard Errors," NBER Working Papers 20325, National Bureau of Economic Research, Inc July 2014.

Acemoglu, Daron and Martin Kaae Jensen, "Robust Comparative Statics in Large Dynamic Economies," NBER Working Papers 18178, National Bureau of Economic Research, Inc June 2012.

_ **and Pascual Restrepo**, "Robots and Jobs: Evidence from US Labor Markets," NBER Working Papers 23285, National Bureau of Economic Research, Inc March 2017.

Ahn, Seung C., Young H. Lee, and Peter Schmidt, "Panel data models with multiple time-varying individual effects," *Journal of Econometrics*, 2013, *174* (1), 1–14.

Ahn, Seung Chan, Young Hoon Lee, and Peter Schmidt, "GMM estimation of linear panel data models with time-varying individual effects," *Journal of Econometrics*, April 2001, *101* (2), 219–255.

Al-Najjar, Nabil Ibraheem, "Decomposition and Characterization of Risk with a Continuum of Random Variables," *Econometrica*, September 1995, *63* (5), 1195–1224.

Ando, Tomohiro and Jushan Bai, "Asset Pricing with a General Multifactor Structure," *Journal of Financial Econometrics*, 2015, *13* (3), 556–604.

Andrews, Donald W. K., "Cross-Section Regression with Common Shocks," *Econometrica*, 09 2005, *73* (5), 1551–1585.

Bai, Jushan, "Panel Data Models With Interactive Fixed Effects," *Econometrica*, 07 2009, *77* (4), 1229–1279.

_ **and Serena Ng**, "Determining the Number of Factors in Approximate Factor Models," *Econometrica*, January 2002, *70* (1), 191–221.

Bai, Z. D. and Y. Q. Yin, "Limit of the Smallest Eigenvalue of a Large Dimensional Sample Covariance Matrix," *Ann. Probab.*, 07 1993, *21* (3), 1275–1294.

Barro, Robert J. and Charles J. Redlick, "Macroeconomic Effects From Government Purchases and Taxes," *The Quarterly Journal of Economics*, 2011, *126* (1), 51–102.

Beraja, Martin, "Counterfactual Equivalence in Macroeconomics," Technical Report 2017.

_ **, Andreas Fuster, Erik Hurst, and Joseph Vavra**, "Regional Heterogeneity and the Refinancing Channel of Monetary Policy*," *The Quarterly Journal of Economics*, 2018, p. qjy021.

_ **, Erik Hurst, and Juan Ospina**, "The Aggregate Implications of Regional Business Cycles," NBER Working Papers 21956, National Bureau of Economic Research, Inc February 2016.

Blanchard, Olivier and Roberto Perotti, "An Empirical Characterization of the Dynamic Effects of Changes in Government Spending and Taxes on Output," *The Quarterly Journal of Economics*, 2002, *117* (4), 1329–1368.

Blanchard, Olivier Jean and Charles M Kahn, "The Solution of Linear Difference Models under Rational Expectations," *Econometrica*, July 1980, *48* (5), 1305–11.

Calvo, Guillermo A., "Staggered prices in a utility-maximizing framework," *Journal of Monetary Economics*, September 1983, *12* (3), 383–398.

Chodorow-Reich, Gabriel, "Geographic Cross-Sectional Fiscal Spending Multipliers: What Have We Learned?," 2017 2017.

_ , Laura Feiveson, Zachary Liscow, and William Gui Woolston, "Does State Fiscal Relief during Recessions Increase Employment? Evidence from the American Recovery and Reinvestment Act," *American Economic Journal: Economic Policy*, August 2012, *4* (3), 118–145.

Connor, Gregory and Robert A. Korajczyk, "Performance measurement with the arbitrage pricing theory : A new framework for analysis," *Journal of Financial Economics*, March 1986, *15* (3), 373–394.

Djogbenou, Antoine, Sílvia Gonçalves, and Benoit Perron, "Bootstrap Inference in Regressions with Estimated Factors and Serial Correlation," *Journal of Time Series Analysis*, 2015, *36* (3), 481–502.

Farhi, E. and I. Werning, "Chapter 31 - Fiscal Multipliers: Liquidity Traps and Currency Unions," in John B. Taylor and Harald Uhlig, eds., *John B. Taylor and Harald Uhlig, eds.*, Vol. 2 of *Handbook of Macroeconomics*, Elsevier, 2016, pp. 2417 – 2492.

Gonçalves, Sílvia and Benoit Perron, "Bootstrapping factor-augmented regression models," *Journal of Econometrics*, 2014, *182* (1), 156–173.

Hagedorn, Marcus, Iourii Manovskii, and Kurt Mitman, "The Impact of Unemployment Benefit Extensions on Employment: The 2014 Employment Miracle?," CEPR Discussion Papers 11060, C.E.P.R. Discussion Papers January 2016.

Hall, Robert E., "By How Much Does GDP Rise If the Government Buys More Output?," *Brookings Papers on Economic Activity*, 2009, *40* (2 (Fall)), 183–249.

Hausman, Jerry A and William E Taylor, "Panel Data and Unobservable Individual Effects," *Econometrica*, November 1981, *49* (6), 1377–1398.

Holtz-Eakin, Douglas, Whitney Newey, and Harvey S Rosen, "Estimating Vector Autoregressions with Panel Data," *Econometrica*, November 1988, *56* (6), 1371–95.

Horowitz, Joel L., "Chapter 52 - The Bootstrap," in James J. Heckman and Edward Leamer, eds., *James J. Heckman and Edward Leamer, eds.*, Vol. 5 of *Handbook of Econometrics*, Elsevier, 2001, pp. 3159 – 3228.

Ilzetzki, Ethan, Enrique G. Mendoza, and Carlos A. Végh, "How big (small?) are fiscal multipliers?," *Journal of Monetary Economics*, 2013, *60* (2), 239–254.

Leeb, Hannes and Benedikt M. Pötscher, "Can One Estimate The Unconditional Distribution Of Post-Model-Selection Estimators?," *Econometric Theory*, April 2008, *24* (02), 338–376.

Mian, Atif and Amir Sufi, "House Prices, Home Equity-Based Borrowing, and the US Household Leverage Crisis," *American Economic Review*, August 2011, *101* (5), 2132–2156.

_ **and** _ , "What Explains the 2007-2009 Drop in Employment?," *Econometrica*, 2014, *82* (6), 2197–2223.

_ **, Kamalesh Rao, and Amir Sufi**, "Household Balance Sheets, Consumption, and the Economic Slump," *The Quarterly Journal of Economics*, 2013, *128* (4), 1687–1726.

Mikusheva, Anna and Stanislav Anatolyev, "Limit Theorems for Factor Models," *Unpublished manuscript, MIT*, 2018.

Nakamura, Emi and John Steinsson, "Fiscal Stimulus in a Monetary Union: Evidence from US Regions," *American Economic Review*, March 2014, *104* (3), 753–92.

Parker, Jonathan A., "On Measuring the Effects of Fiscal Policy in Recessions," *Journal of Economic Literature*, September 2011, *49* (3), 703–718.

Pesaran, M. Hashem, "Estimation and Inference in Large Heterogeneous Panels with a Multifactor Error Structure," *Econometrica*, 07 2006, *74* (4), 967–1012.

Ramey, V.A., "Macroeconomic Shocks and Their Propagation," *Handbook of Macroeconomics*, 2016, *2*, 71 – 162.

Ramey, Valerie A., "Identifying Government Spending Shocks: It's all in the Timing," *The Quarterly Journal of Economics*, 2011, *126* (1), 1–50.

_ **and Matthew D. Shapiro**, "Costly capital reallocation and the effects of government spending," *Carnegie-Rochester Conference Series on Public Policy*, June 1998, *48* (1), 145–194.

Serrato, Juan Carlos SuÃ¡rez and Philippe Wingender, "Estimating Local Fiscal Multipliers," NBER Working Papers 22425, National Bureau of Economic Research, Inc July 2016.

Silverstein, Jack W., "The Smallest Eigenvalue of a Large Dimensional Wishart Matrix," *Ann. Probab.*, 11 1985, *13* (4), 1364–1368.

Smets, Frank and Rafael Wouters, "Shocks and Frictions in US Business Cycles: A Bayesian DSGE Approach," *American Economic Review*, June 2007, *97* (3), 586–606.

Stock, J. H. and M. W. Watson, "Forecasting Using Principal Components From a Large Number of Predictors," *Journal of the American Statistical Association*, December 2002, *97*, 1167–1179.

Stock, James H. and Mark W. Watson, "Implications of Dynamic Factor Models for VAR Analysis," NBER Working Papers 11467, National Bureau of Economic Research, Inc July 2005.

Uhlig, Harald, "A law of large numbers for large economies (*)," *Economic Theory*, 1996, *8* (1), 41–50.

_ , "Some Fiscal Calculus," *American Economic Review*, May 2010, *100* (2), 30–34.

Varah, J.M., "A lower bound for the smallest singular value of a matrix," *Linear Algebra and its Applications*, 1975, *11* (1), 3 – 5.

White, H., *Asymptotic Theory for Econometricians* Economic theory, econometrics, and mathematical economics, Academic Press, 2001.

Wilson, Daniel J., "Fiscal Spending Jobs Multipliers: Evidence from the 2009 American Recovery and Reinvestment Act," *American Economic Journal: Economic Policy*, August 2012, *4* (3), 251–82.

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