EMPIRICALLY RELEVANT LONG-RUN RESTRICTIONS*

Martín Almuzara†

This version: March 2020
(Download here the latest version)

Abstract

I propose a new approach to the formulation of long-run restrictions that consists of constraining the local-to-zero impulse-response function of a multivariate time series—a transformation of the impulse-response function that focuses on low-frequency variation. I argue that researchers should restrict a range of low enough frequencies to capture an empirically motivated notion of the long run. I establish conditions under which that approach delivers identification of impulse-response functions and by embedding it into a dynamic model (such as a vector autoregression or a dynamic factor model) I show how to conduct inference. An application to factor-neutral technology shocks is discussed.

Keywords: Long-run restrictions, impulse-response functions, technology shocks.

JEL Classification: C32, C51.

*I have greatly benefited from comments and discussion with Manuel Arellano and Enrique Sentana. I also thank seminar participants at EEA-ESEM 2019 (Manchester) and SAEe 2019 (Alicante). Especial gratitude goes to Omar Rachedi for his comments on an (embarrassingly) preliminary draft. All remaining errors are my own responsibility. I acknowledge funding from the Santander Research Chair at CEMFI. The views expressed in this paper are those of the author and do not necessarily reflect the position of the Federal Reserve Bank of New York or the Federal Reserve System.

†Federal Reserve Bank of New York: martin.almuzara@ny.frb.org

1 Introduction

A long-run restriction is one that imposes that the cumulative effect of certain shock on certain variable is "eventually" zero. Economic research features many instances of such restrictions: assuming that factor-neutral technical change is the single driver of long-term movements in labor productivity is standard in business cycle analyses (see, e.g., Galí (1999) and Christiano et al. (2004)); the long-run neutrality of monetary policy, demand, and exchange rate shocks is frequent as well (see, e.g., Peersman and Smets (2003)). With an econometric perspective, the question is to what extent can long-run restrictions help identify, estimate, and test economic models given that, in the datasets often available to researchers, information about the long run is admittedly scarce.

In effect, interpreting "eventually" to mean "at an arbitrarily far horizon" deprives long-run restrictions of empirical content. In the data, one never sees the impulse and the response separated by eternity—and one never wishes so. Instead, the application usually indicates a more appealing empirically-grounded notion of long run. For example, in studying firm decisions, it is common to interpret the long run as a period long enough to allow the adjustment of all productive factors (and this may vary across industries); in business cycle analysis the long run represents variation taking place beyond the typical periodicity of the business cycle (beyond, say, an 8-year period to evoke the criterion adopted by Müller and Watson (2008)).

In this paper, I propose a novel approach to the formulation of long-run restrictions which enables researchers to clearly express what the long run means for their empirical analysis. I introduce what I call empirically relevant long-run restrictions. These are restrictions on the local-to-zero impulse-response function (IRF) of a multivariate time series—a transformation of the dynamic responses of variables to shocks which focuses on low-frequency components. Put simply, each variable is the sum of components that vary with different periodicities and an empirically relevant long-run restriction limits the contribution of a shock to the variance of the components the periodicities of which are too long—how long depends on the notion of long run that suits the application. The approach is introduced in detail in section 2.

Early methods for long-run restrictions can be traced back to Shapiro and Watson (1988), Blanchard and Quah (1989) and King et al. (1991). In these papers, a long-run restriction sets to zero the sum of all dynamic responses of a variable to a shock. This is equivalent to setting to zero the contribution of the shock to the long-run variance of the variable. Since the long-run variance may be regarded as the variance of a component with an infinitely long period, my approach generalizes the early ones. Doing so is important because long but not infinitely long periods do unfold their full cyclical patterns in realistic datasets (e.g., 70 years of GDP data do contain information about 8-year cycles). Consequently, one expects to be able to learn about such components by looking at statistics that capture genuine long-run variation. What is more, inference based on appropriately chosen statistics is presumably more transparent about the sources of sampling uncertainty.

In this context, my paper makes three contributions. The *first* is to study the identification content of empirically relevant long-run restrictions. The fundamental identification problem of macroeconometrics is to recover the IRF—the dynamic multipliers measuring the response of variables to shocks—from the autocorrelation structure of the data. Since such a recovery is not possible without restrictions, a large array of identification strategies has been developed.[1] An important class considers restrictions on a strictly invertible time series, i.e., a time series for which shocks are linear combinations of the past and the present of the series. The theory of identification I develop belongs in that class.[2] Within the class, the problem usually reduces to the identification of an $n \times n$ orthogonal matrix (n the number of shocks and variables) that transforms one-step ahead prediction errors into shocks. A key insight of my analysis is that, because the object to be identified is finite-dimensional but the empirically relevant long-run restrictions form a continuum, overidentification will generally occur. Owing to that, my long-run restrictions are generally testable. One may take two separate subsets of restrictions and, in principle, identify two seperate values for the IRF. Differences between the two values are evidence against the long-run restrictions. This is the subject of section 3.

The *second* contribution of my paper is to provide a framework that exploits empirically relevant long-run restrictions to conduct inference about the IRF. I focus on the construction of an approximate likelihood function for the parameters of interest because such a function serves as input of both frequentist and Bayesian calculations. The construction is as follows. I take a time series of length T and form two sets of statistics: on the one hand, a set of q projection coefficients of the time series onto cosine functions with long periods—the long-run statistics; on the other, a set of $T - q$ projection residuals—the short-run statistics. Periods of the cosine functions are selected to generate a grid of cyclical components in the domain of the empirically relevant notion of the long run. Long-run statistics then capture the trend behavior of the series while short-run statistics describe variation beyond trend. Just as in the traditional approach, both types of statistics are needed to learn about an IRF identified by

[1] Stock and Watson (2016, sec. 4) contains a good summary.

[2] See Plagborg-Møller (2019) and Plagborg-Møller and Wolf (2019) for work on noninvertible processes.

long-run restrictions.[3] Moreover, with T large and q fixed, long-run and short-run statistics are approximately independent and normally distributed (in a suitable sense).

I propose to model the distribution of long-run statistics in the spirit of the low-frequency econometrics literature.[4] In particular, I restrict the distribution to a convenient parametric class that accommodates more general dependence patterns than the standard I(0) and I(1) dynamics. For short-run statistics, I consider a vector autoregression (VAR) with normally distributed errors thinking of it as an approximation to a more general model. Once the likelihood function has been constructed, different inference approaches become available. Using EM ideas, furthermore, it is not hard to extend the framework to latent variables models (such as dynamic factor models). Section 4 develops the details.

The *third* contribution of my paper is empirical. I revisit a classical application of long-run restrictions to factor-neutral technology shocks. Two relevant quantities for the assessment of models of business cycle fluctuations are the response of hours to technology shocks and the contribution of technology shocks to the variance of labor productivity. Early empirical work focused on bivariate models of labor productivity and hours worked for which the effect of non-technology shocks on labor productivity sums to zero in the long run. Using my approach and US data, I impose a long-run restriction that excludes non-technology shocks from labor productivity, understanding long run as variation occurring beyond an 8-year period. Doing so leads to a negative (and significant) response of hours to technology shocks and a modest (but not negligible) fraction of the variation in productivity accounted for by technology shocks.

However, estimates are sensitive to the notion of long run: if restrictions are imposed only on components with periodicities well above 8 years, the response is estimated positive and the fraction high. This is a clear sign of the overidentifying restrictions being at fault, what I confirm by constructing and implementing a test. Part of the empirical contribution of my paper is, then, to give a reevaluation of the controversy in the evidence reported by Galí (1999) and by Christiano et al. (2004).[5] The empirical analysis is fully developed in section 5.

My paper relates to various different literatures. First, Rubio-Ramírez et al. (2010) gave a theory of global identification for finite-order VAR models. In contrast, my identification theory considers more general dynamic models. The motivation is that finite-order VARs have too strong implications for low-frequency variation and are, therefore, not very appealing in

[3] Long-run statistics generalize the long-run variance matrix.

[4] Müller and Watson (2017) provide an excellent survey.

[5] Evidence from bivariate models of the sort that is discussed here are not common in the recent literature. See Ramey (2016, sec. 5) for an up-to-date summary.

the study of long-run restrictions. Second, Francis et al. (2014) proposed an approach to long-run restrictions based on maximizing the contribution of a shock to the conditional variance of a variable at a finite horizon. My approach, per contra, uses the local-to-zero IRF. I argue that the local-to-zero IRF is a more convenient device and avoids some caveats that come with extrapolation of VAR dynamics (see the simulation experiment in section 4). Third, Müller and Watson (2008) pioneered the use of a small set of low-frequency statistics to address time series questions about long-run objects. My paper combines long-run and short-run statistics in a unified framework. Finally, Canova et al. (2010) highlighted that the treatment of trends in hours was key to the estimates of the response of hours to technology shocks. Relative to them, my paper provides a flexible approach to modeling trends and useful diagnostics for the adequacy of long-run restrictions. A discussion of this point is made in section 5 and section 6 concludes the paper.

Notation. For integers t_0, t_1 with $t_0 \leq t_1$, I use $\omega_{t_0:t_1}$ to denote the sequence $\{\omega_t\}_{t=t_0}^{t_1}$. When each ω_t is an array of dimension $d_1 \times d_2$, and if no confusion is possible, I also use $\omega_{t_0:t_1}$ to denote the $d_1 \times d_2(j_1 - j_0 + 1)$ array obtained by horizontal concatenation of the terms of $\{\omega_t\}_{t=t_0}^{t_1}$. I write $\operatorname{diag} \omega_{t_0:t_1}$ for the block diagonal matrix with blocks $\omega_{t_0}, \ldots, \omega_{t_1}$. I write $\mathbb{E}_T[\omega_t] := T^{-1} \sum_{t=1}^{T} \omega_t$ for the average of $\omega_{1:T}$, "$\sim$" for equality in distribution, "$\xrightarrow{p}$" for convergence in probability, and "$\Longrightarrow$" for weak convergence. If M is a matrix, $\|M\|$ is its Hilbert-Schmidt norm and, if M is symmetric positive definite, $\mathrm{Ch}(M)$ is its lower triangular Cholesky factor. I denote by $\mathbb{C}$ the complex plane and by $\mathbb{T} := \{z \in \mathbb{C} : |z| = 1\}$ the unit circle. I equip $\mathbb{T}$ with the Hausdorff-1 measure μ restricted to its Borel sets and I let $L^p = L^p(\mathbb{T}, d\mu)$ be the space of L^p-integrable functions with respect to μ. Likewise, I denote by $\mathbb{Z}$ the set of integers and I let $\ell^p = L^p(\mathbb{Z}, d\mu_c)$ be the space of ℓ^p-summable sequences, equivalently, L^p-integrable with respect to counting measure μ_c. I write $\|\cdot\|_{L^p}$ (the essential supremum with respect to μ when $p = \infty$) for the norm in L^p and $\|\cdot\|_{\ell^p}$ for the norm in ℓ^p. When a sequence is defined over a subset of $\mathbb{Z}$ its norm is computed by extending the original sequence to all of $\mathbb{Z}$ setting the image outside the original domain to zero.

2 Model

Consider the time series

$$
\begin{aligned}
y_t &= \sum_{l=0}^{\infty} \Theta_l \varepsilon_{t-l}, \\
\varepsilon_t &\overset{iid}{\sim} N\left(0_{n\times 1}, \Sigma_\varepsilon\right),
\end{aligned} \tag{1}
$$

where both y_t and ε_t are n-dimensional. The vector y_t collects the variables of interest (e.g., indicators of economic activity, inflation, interest rates, productivity, and employment). At this stage, it is not important whether y_t is directly observed or latent. The vector ε_t, in turn, contains the shocks driving y_t (e.g., changes in demand, monetary policy, and technology). As is standard, it is assumed that $\Sigma_\varepsilon = \operatorname{diag} \sigma_1^2, \ldots, \sigma_n^2$ is a diagonal matrix. This way, no shock can be predicted from past, present or future realizations of other shocks.[6] Normality of shocks can be dispensed with but, since it is convenient and occurs frequently in macroeconometrics, I maintain it.

The coefficient matrix Θ_l measures the response of y_{t+l} to the impulse ε_t and, owing to it, the sequence $\Theta_{0:\infty}$ receives the name of impulse-response function (IRF) for the system (1). Knowledge of $\{\Theta_{0:\infty}, \Sigma_\varepsilon\}$ completely determines the probability distribution of the time series y_t and, in particular, its dependence structure. Usual objects of interest are terms of the IRF and forecast error variance decompositions (FEVD), which quantify the contribution of each shocks to the conditional variance of the series at different horizons. These are functionals of the parameter $\{\Theta_{0:\infty}, \Sigma_\varepsilon\}$. Some examples are described below.

As a minimum requirement, the infinite sum in (1) should make sense. The assumption that follows takes care of it:

Assumption 1. *Let $\Theta(z) := \sum_{l=0}^{\infty} \Theta_l z^l$ be defined for $z \in \mathbb{T}$. Then,*

(i) $\Theta_{0:\infty} \in \ell^1$, *i.e.,* $\sum_{l=0}^{\infty} \|\Theta_l\| < \infty$, *and*

(ii) $\det\{\Theta(z)\} = 0$ *implies* $|z| > 1$.

To set some notation, y_t has autocovariance function $\Gamma_l := \mathbb{E}[y_t y_{t+l}]$ ($l \in \mathbb{Z}$) and spectrum $f(z) := \Theta(z)\Sigma_\varepsilon\Theta'(z^{-1}) = \sum_{l=-\infty}^{\infty} \Gamma_l z^l$ ($z \in \mathbb{T}$).[7] Assumption 1(i) imposes limited dependence:

[6] See Ramey (2016) for a discussion of this assumption.

[7] I call spectrum what is sometimes known as autocovariance generating function. The former is usually reserved to f viewed as a function of $\arg(z) \in (-\pi, \pi]$, the principal value of the argument of z.

it asks autocovariances to decay sufficiently fast. By 1(i), $\sum_{j=-\infty}^{\infty} \|\Gamma_j\| < \infty$ and, therefore, $\Gamma_{-\infty:\infty} \in \ell^1$ and $f \in L^\infty$. In particular, y_t is stationary and has short memory. Assumption 1(ii) makes f positive definite over all of $\mathbb{T}$. In particular, no linear combination of y_t has zero long-run variance and y_t admits a VAR(∞) representation. It rules out overdifferencing (as in Almuzara and Marcet (2019)) and noninvertibility (as in Plagborg-Møller (2019) and Plagborg-Møller and Wolf (2019)).

2.1 Empiricaly relevant long-run restrictions

It is time to introduce the main concept of the paper.

Definition 1. *Shock ε_{kt} satisfies an empirically relevant long-run restriction with respect to variable y_{jt} if the condition*

$$[\Theta(z)]_{jk} = \sum_{l=0}^{\infty} \Theta_{l,jk} z^l = 0 \tag{2}$$

holds for all $z = \exp(i\lambda)$ with $i := \sqrt{-1}$ and $0 \leq \lambda \leq \lambda_{\mathrm{LR}}$.

To interpret definition 1 consider first the case when $\lambda = 0$ ($z = 1$). Thus,

$$[\Theta(1)]_{jk} = \sum_{l=0}^{\infty} \Theta_{l,jk} = 0,$$

or, in other words, the response of y_{jt} to a unit change in ε_{kt} accumulates to zero at an infinite horizon. This is the basis for the classical long-run restriction approach (see, e.g., Shapiro and Watson (1988), Blanchard and Quah (1989), and King et al. (1991)).

One implication is that ε_{kt} does not contribute to the long-run variance of y_{jt}. The long-run variance of y_{jt} is the limit of the variance of its (suitably scaled) sample mean,[8]

$$\mathbb{V}_0\left(y_j\right) := \lim_{T\to\infty} T\mathrm{Var}\left(\frac{1}{T}\sum_{t=1}^{T} y_{jt}\right) = \sum_{k=1}^{n} [\Theta(1)]_{jk}^2 \sigma_k^2,$$

Because the different shocks are uncorrelated to each other the long-run variance of y_{jt}

[8]The calculation is as follows:

$$\mathrm{Var}\left(\mathbb{E}_T\left[y_{jt}\right]\right) = \sum_{k=1}^{n}\left[\sum_{t=-\infty}^{0}\mathrm{Var}\left(\frac{1}{T}\left(\sum_{l=1}^{T}\Theta_{l-t,jk}\right)\varepsilon_{kt}\right) + \sum_{t=1}^{T}\mathrm{Var}\left(\frac{1}{T}\left(\sum_{l=0}^{T-t}\Theta_{l,jk}\right)\varepsilon_{kt}\right)\right] = \left(\sum_{k=1}^{n}\left(\sum_{l=0}^{\infty}\Theta_{l,jk}\right)^2\frac{\sigma_k^2}{T}\right)(1+o(1)),$$

where the second equality follows from dominated convergence applied to counting measure.

decomposes into the sum of the contribution of each shock. Setting $[\Theta(1)]_{jk} = 0$ corresponds to setting the contribution of the k-th shock to the long-run variance of y_{jt} to zero.

Heuristically, one may think of the long-run variance of y_{jt} as the variance of a "cycle" of frequency $\lambda = 0$ (or infinitely long period) and of an equally weighted average as the natural way to extract it from y_{jt}. A basic insight from the spectral analysis of time series (see, e.g., Granger and Watson (1984)) is that, to extract a cycle of frequency λ, one should form an average with trigonometric weights. The variance of such an average is

$$\mathbb{V}_\lambda\left(y_j\right) := \lim_{T\to\infty} T\text{Var}\left(\frac{1}{T}\sum_{t=1}^{T} y_{jt}\exp\left(\frac{t-1}{T}i\lambda\right)\right) = \sum_{k=1}^{n}\left[\Theta(z)\right]_{jk}^{2}\sigma_k^2,$$

where $z = \exp(i\lambda)$. In light of that, definition 1 indicates that ε_{kt} satisfies a long-run restriction relative to y_{jt} if it does not contribute to the variance of cycles in y_{jt} with frequency λ below the cutoff frequency λ_{LR}.

The threshold frequency λ_{LR}, I argue, ought to be selected to reflect the idea of long run that emerges from the application. So defined, a cycle of frequency λ repeats itself every $2\pi/\lambda$ periods. If, e.g., the time unit is a quarter and there is a desire to let the long-run represent variation below, say, an 8-year (i.e., 32-quarter) period one would simply take $\lambda_{\text{LR}} = 2\pi/32$.

Technology shocks. Let $y_t := (n_t, x_t)'$ with n_t hours worked and x_t an indicator of aggregate labor productivity. Also, let $\varepsilon_t := (\varepsilon_{\text{N}t}, \varepsilon_{\text{T}t})'$ where $\varepsilon_{\text{N}t}$ is a non-technology shock and $\varepsilon_{\text{T}t}$ is a factor-neutral technology shock. In line with (1), the dynamics of y_t are given by

$$\begin{aligned} y_t = \begin{bmatrix} n_t \\ x_t \end{bmatrix} &= \sum_{l=0}^{\infty}\begin{bmatrix}\Theta_{\ell,n\text{N}} & \Theta_{\ell,n\text{T}} \\ \Theta_{l,x\text{N}} & \Theta_{l,x\text{T}}\end{bmatrix}\begin{bmatrix}\varepsilon_{\text{N},t-l} \\ \varepsilon_{\text{T},t-l}\end{bmatrix} = \sum_{l=0}^{\infty}\Theta_l\varepsilon_{t-l}, \\ \varepsilon_t &\overset{iid}{\sim} N\left(0_{2\times1}, \Sigma_\varepsilon\right), \end{aligned}$$

and $\Sigma_\varepsilon = \text{diag}\,\sigma_{\text{N}}^2, \sigma_{\text{T}}^2$.

As discussed in the introduction (section 1), a key object for business cycles models is the IRF itself. The sign and magnitude of the dynamic effect $\Theta_{l,n\text{T}}$ of technology shocks on hours for low l (often $l = 0$) have received a lot of attention in empirical work.

Another object of interest is the FEVD. This is the answer to the question of what fraction of the forecast error variance of a given variable can be attributed to a certain shock. For example, the fraction of the forecast error variance of labor productivity at horizon h accounted for by technology shocks is defined as

$$\text{FEVD}_h(x, \varepsilon_\text{T}) := \frac{\sum_{l=0}^{h} \Theta_{l,x\text{T}}^2 \sigma_\text{T}^2}{\sum_{l=0}^{h} \left(\Theta_{l,x\text{T}}^2 \sigma_\text{T}^2 + \Theta_{l,x\text{N}}^2 \sigma_\text{N}^2 \right)}.$$

Clearly, $\text{FEVD}_h(x, \varepsilon_\text{T}) + \text{FEVD}_h(x, \varepsilon_\text{N}) = 1$ and $\text{FEVD}_h(n, \varepsilon_\text{T}) + \text{FEVD}_h(n, \varepsilon_\text{N}) = 1$ at each horizon h and, in consequence, the more important technology shocks are in explaining the variation of aggregate labor productivity and hours, the less important non-technology shocks are.

Galí (1999) and Christiano et al. (2004) were among the first to analyze the bivariate model of hours and productivity through long-run restrictions. They both imposed that

$$[\Theta(1)]_{21} = \sum_{l=0}^{\infty} \Theta_{l,x\text{N}} = 0$$

and employed US quarterly data to measure n_t and x_t. In this setup, it is natural to extend the long-run restriction to an empirically relevant range of below-business-cycle frequencies, that is, imposing

$$[\Theta(1)]_{21} = \sum_{l=0}^{\infty} \Theta_{l,x\text{N}} = 0$$

for $z = \exp(i\lambda)$, $0 \leq \lambda \leq \lambda_\text{LR}$, with a choice of λ_LR reflecting variation beyond the business cycle. I pursue the idea in section 5.

3 Identification

The identification problem is to recover $\{\Theta_{0:\infty}, \Sigma_\varepsilon\}$ from knowledge of $\Gamma_{-\infty:\infty}$ (or f). Without restrictions (beyond assumption 1), however, it is not possible to determine a unique value of $\{\Theta_{0:\infty}, \Sigma_\varepsilon\}$ from $\Gamma_{-\infty:\infty}$. To see this, note that under assumption 1, y_t admits the representation

(3)
$$\begin{aligned} y_t &= \sum_{l=0}^{\infty} \Omega_l u_{t-l}, \\ u_t &\overset{iid}{\sim} N\left(0_{n\times 1}, \Sigma_u\right), \end{aligned}$$

where $\Omega_0 = I_n$, $\Omega_l = \Theta_l \Theta_0^{-1}$, all $l \geq 1$, and $\Sigma_u = \Theta_0 \Sigma_\varepsilon \Theta_0'$.

There is a one-to-one correspondence between $\Gamma_{-\infty:\infty}$ and $\{\Omega_{0:\infty}, \Sigma_u\}$ and the latter can be obtained from limits of linear projections of y_t onto its past. Thus, the identification problem is, equivalently, to determine $\{\Theta_{0:\infty}, \Sigma_\varepsilon\}$ from $\{\Omega_{0:\infty}, \Sigma_u\}$ and the equation

$$\Theta(z)\Sigma_\varepsilon\Theta'(z^{-1}) = \left(\sum_{l=0}^{\infty}\Omega_l z^l\right)\Sigma_u\left(\sum_{l=0}^{\infty}\Omega_l' z^{-l}\right), \text{ all } z \in \mathbb{T}. \tag{4}$$

To simplify the exposition, I consider the normalization $\Sigma_\varepsilon = I_n$. Alternatively, one could restrict elements of Θ_0 and leave the diagonal of Σ_ε free. Let $\tilde{Q}$ be an $n \times n$ orthogonal matrix and let the sequence $\tilde{\Theta}_{0:\infty}$ be defined by $\tilde{\Theta}_l := \Theta_l\tilde{Q}$, all $l \geq 0$. Then,

$$\begin{aligned}\left(\sum_{l=0}^{\infty}\tilde{\Theta}_l z^l\right)\left(\sum_{l=0}^{\infty}\tilde{\Theta}_l' z^{-l}\right) &= \left(\sum_{l=0}^{\infty}\Theta_l z^l\right)\tilde{Q}\tilde{Q}'\left(\sum_{l=0}^{\infty}\Theta_l' z^{-l}\right)\\ &= \Theta(z)\Theta'(z^{-1}), \text{ all } z \in \mathbb{T},\end{aligned}$$

since $\tilde{Q}\tilde{Q}' = I_n$ by orthogonality. In consequence, if $\Theta_{0:\infty}$ solves (4), so does $\tilde{\Theta}_{0:\infty}$. The idea is to distinguish $\Theta_{0:\infty}$ from alternative $\tilde{\Theta}_{0:\infty}$ by demanding that $\Theta_{0:\infty}$ satisfies other restrictions, such as long-run restrictions of the form (2). When this can be achieved, one says that $\Theta_{0:\infty}$ is globally identified. In some situations, $\Theta_{0:\infty}$ is not globally identified but, for every $\tilde{\Theta}_{0:\infty}$ satisfying (4), $\tilde{\Theta}_{0:\infty}e_k = \Theta_{0:\infty}e_k$ being e_k the k-th column of I_n. In that case, one says the IRF of the k-th shock is globally identified.

A last remark prior to getting into the details of the identification results is that if $\Theta_{0:\infty}$ solves (4), one can find an $n \times n$ orthogonal matrix Q such that $\Theta_l = \Omega_j \mathrm{Ch}(\Sigma_u)Q$, all $l \geq 0$. The problem is, in sum, to identify the orthogonal matrix Q that transforms the reduced form $\{\Omega_{0:\infty}, \Sigma_u\}$ into $\Theta_{0:\infty}$. In fact, because of this, for any two observationally equivalent values of the parameter $\Theta_{0:\infty}$ and $\tilde{\Theta}_{0:\infty}$ (i.e. for any two values that fit (4)), there is an $n \times n$ orthogonal matrix $\tilde{Q}$ such that $\Theta_{0:\infty}\tilde{Q} = \tilde{\Theta}_{0:\infty}$.

3.1 Representation of identifying restrictions

My identification theory builds on that of Rubio-Ramírez et al. (2010). For any sequence $M_{l_0:l_1}$ of $n_1 \times n_2$ matrices and for any $n_2 \times n_3$ matrix M, let $M_{l_0:l_1}M$ denote the sequence of $n_1 \times n_3$ matrices with l-th term $M_l M$.

To state the main results, I need a rotation-equivariant transformation g mapping $\Theta_{0:\infty}$ to an $m \times n$ matrix on which restrictions will be placed. Let $\mathcal{P}$ denote the parameter space, i.e., the set of all possible values for $\Theta_{0:\infty}$. Let $g : \mathcal{P} \to \mathbb{R}^{m\times n}$ be a function such that

$$g\left(\tilde{\Theta}_{0:\infty}\tilde{Q}\right) = g\left(\tilde{\Theta}_{0:\infty}\right)\tilde{Q}$$

for all $\tilde{\Theta}_{0:\infty} \in \mathcal{P}$ and $n \times n$ orthogonal matrix $\tilde{Q}$.

The notation $\Theta_{0:\infty}$ is reserved for the true value of the parameter while $\tilde{\Theta}_{0:\infty}$ is a generic element of $\mathcal{P}$. Identifying restrictions will be represented by the $m \times m$ matrices $\{R_1, \ldots, R_n\}$ which select the linear combinations of $g\left(\tilde{\Theta}_{0:\infty}\right)$ equal to zero,

(5)
$R_k g\left(\tilde{\Theta}_{0:\infty}\right) e_k = 0_{m\times 1},$

The matrix R_k imposes restrictions on the k-th column of $g\left(\tilde{\Theta}_{0:\infty}\right)$. Let columns be ordered in such a way that $r_k := \text{rank}(R_k)$ decreases with k, i.e.,

$$r_1 \geq \cdots \geq r_n,$$

and let $r = r_1 + \cdots + r_n$ be the total number of restrictions.

To explain the role of g, let me select a grid of frequencies so that $0 \leq \lambda_1 \leq \cdots \leq \lambda_{n_{\text{fr}}} \leq \lambda_{\text{LR}}$. Define the function $g_{\text{LR}} : \mathcal{P} \to \mathbb{R}^{m\times n}$ with $m = n_{\text{fr}} n$ by

$$g_{\text{LR}}\left(\tilde{\Theta}_{0:\infty}\right) := \begin{bmatrix} \sum_{l=0}^{\infty} \tilde{\Theta}_l e^{i\lambda_1 l} \\ \vdots \\ \sum_{l=0}^{\infty} \tilde{\Theta}_l e^{i\lambda_{n_{\text{fr}}} l} \end{bmatrix}$$

Then, g_{LR} is rotation-equivariant since

$$g_{\text{LR}}\left(\tilde{\Theta}_{0:\infty}\tilde{Q}\right) = \begin{bmatrix} \sum_{l=0}^{\infty} \tilde{\Theta}_l \tilde{Q} e^{i\lambda_1 l} \\ \vdots \\ \sum_{l=0}^{\infty} \tilde{\Theta}_l \tilde{Q} e^{i\lambda_{n_{\text{fr}}} l} \end{bmatrix} = \begin{bmatrix} \sum_{l=0}^{\infty} \tilde{\Theta}_l e^{i\lambda_1 l} \\ \vdots \\ \sum_{l=0}^{\infty} \tilde{\Theta}_l e^{i\lambda_{n_{\text{fr}}} l} \end{bmatrix} \tilde{Q} = g_{\text{LR}}\left(\tilde{\Theta}_{0:\infty}\right)\tilde{Q}.$$

To represent the implications of an empirically relevant long-run restriction of shock k with respect to variable j, one would choose R_k to have ones in the entry $(i_{\text{fr}}, n(i_{\text{fr}} - 1) + j)$ for each $i_{\text{fr}} = 1, \ldots, n_{\text{fr}}$ and zeros elsewhere. Of course , if identification holds for the grid $\left\{\lambda_1, \ldots, \lambda_{n_{\text{fr}}}\right\}$ of frequencies, it also does for the continuum $0 \leq \lambda \leq \lambda_{\text{LR}}$.

In practice, g_{LR} will often be combined with other functions to produce identifying restrictions. To accommodate contemporaneous and long-run restrictions, e.g., one would use

$$g\left(\tilde{\Theta}_{0:\infty}\right) := \begin{bmatrix} \tilde{\Theta}_0 \\ g_{\text{LR}}\left(\tilde{\Theta}_{0:\infty}\right) \end{bmatrix}$$

which is easily seen to be rotation-equivariant.

Finally, as in Rubio-Ramírez et al. (2010), a normalization rule must be decided to settle the sign of each equation. A normalization rule in my setup is a subset $\mathcal{N} \subset \mathcal{P}$ such that, for each $\tilde{\Theta}_{0:\infty} \in \mathcal{P}$, there is a unique $n \times n$ diagonal matrix D with ones or minus ones on the main diagonal such that $\tilde{\Theta}_{0:\infty} D \in \mathcal{N}$.

3.2 Identification results

The main result is as follows:

Theorem 1. *Assume $\Theta_{0:\infty}$ satisfies (5) and the normalization $\Theta_{0:\infty} \in \mathcal{N}$. If,*

$$\operatorname{rank}\left(\begin{bmatrix} R_k g\left(\Theta_{0:\infty}\right) \\ \begin{bmatrix} I_k & 0_{k\times(n-k)} \end{bmatrix}\end{bmatrix}\right) = n, \text{ for each } k = 1, \ldots, K,$$

then, $\tilde{\Theta}_{0:\infty} e_K = \Theta_{0:\infty} e_K$ for all $\tilde{\Theta}_{0:\infty}$ that solves (4) and satisfies (5) and the normalization rule. In other words, the IRF of the K-th shock is globally identified. If moreover $K = n$, then $\Theta_{0:\infty}$ is globally identified.

The proof of theorem 1 can be found in appendix A. Despite the argument being similar to that of Rubio-Ramírez et al. (2010), theorem 1 applies to a large class of strictly invertible time series, with the restriction to a VAR model unnecessary. This is relevant for the study of long-run restrictions because VAR models are highly restrictive for the form low-frequency variability—a finite number of autocovariances determine the whole dependence structure of the time series.

With the construction of the g_{LR} function given in the previous section, theorem 1 covers empirically relevant long-run restrictions (possibly combined with other types of identifying restrictions). This I illustrate in the technology shocks example introduced in section 2.

Technology shocks. Consider the bivariate model of hours and productivity I described before. In order to simplify, let me construct the g_{LR} function with $n_{\text{fr}} = 1$, letting $\lambda_{n_{\text{fr}}} = \bar{\lambda}$,

$$g_{\text{LR}}\left(\tilde{\Theta}_{0:\infty}\right) := \sum_{l=0}^{\infty} \tilde{\Theta}_l e^{i\bar{\lambda} l}.$$

Thus, $m = n = 2$. Recall that, in my ordering of variables and shocks,

$$\Theta_l = \begin{bmatrix} \Theta_{\ell,n\mathrm{N}} & \Theta_{\ell,n\mathrm{T}} \\ \Theta_{l,x\mathrm{N}} & \Theta_{l,x\mathrm{T}} \end{bmatrix}$$

and, in consequence, the first column of g_{LR} refers to non-technology shocks. The empirically relevant long-run restriction implies

$$R_k g_{\mathrm{LR}}\left(\tilde{\Theta}_{0:\infty}\right) e_k = 0_{2\times 1},$$

with

$$R_1 := \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \text{ and } R_2 := 0_{2\times 2}.$$

Applying theorem 1, one concludes that $\Theta_{0:\infty}$ is globally identified if, for some $\bar{\lambda} \leq \lambda_{\mathrm{LR}}$,

$$\sum_{l=0}^{\infty} \Theta_{l,x\mathrm{T}} e^{i\bar{\lambda}l} \neq 0.$$

This, in turn, holds for all $\bar{\lambda} \leq \lambda_{\mathrm{LR}}$ by assumption 1(ii). Then, empirically relevant long-run restrictions which exclude non-technology shocks from low-frequency variation in productivity suffice for the global identification of the IRF.

3.3 Overidentification

To be completed

4 Inference

It will be apparent later that sufficiently general dynamics in y_t must be allowed for in order to strengthen the credibility of the identification approach I propose. Thus, I take Θ to be implicitly indexed by the sample size T and, consequently, one should formally think of y_t as originating in a triangular array $\{\{y_{Tt}\}_{t=1}^{T}\}_{T=1}^{\infty}$. However, in the interest of simplifying notation a sub-index T is omitted from Θ, y_t and other objects that may vary with T (e.g., the autocovariance function and spectrum of y_t).

However, to capture the scarcity of low-frequency information in the data I allow λ_{LR} to be implicitly indexed by T.

To explain the approach to inference I advocate under the new method some preliminary tools are needed. A central device from the low-frequency econometrics literature is the cosine transform (as before, see Müller and Watson (2017)). Given a time series $z_1, \ldots, z_T$ the j-th cosine transform is defined as

$$Z_{jT} := T^{-1} \sum_{t=1}^{T} \Psi_j \left(\frac{t - 1/2}{T} \right) z_t,$$

where $\Psi_j(s) := \sqrt{2} \cos(sj\pi)$ for each j.

We must begin by noting that $\Psi_1, \ldots, \Psi_q$ are orthonormal and may be regarded as a set of basis functions. In this sense, it would be possible to employ sines or Fourier functions to construct the basis but Müller and Watson (2008) give very compelling reasons why the cosine transforms are preferred over those. I do not elaborate on those reasons here.

One simple interpretation of the cosine transforms is that $Z_T \equiv [Z_{1T}, \ldots, Z_{qT}]'$ are projection coefficients onto $\Psi_1, \ldots, \Psi_q$. In addition to that, note that the function Ψ_j is periodic with period $2/j$. Therefore, it is valid (and very useful for extracting intuition) to think of Z_{jT} as a summary of the strength of a component with period $2T/j$ underlying the dynamics of z_t.

One can then collectively interpret the full set of cosine transforms of the data as a filter that isolates cycles representing long-run variation. A key insight from the low-frequency econometrics literature is that q must taken small to reflect the scarcity of information about such variation.

Technology shocks example. Let X_T and N_T be the cosine transforms of productivity and employment. Also, let $\hat{x}_T = \Psi X_T$ and $\hat{n}_T = \Psi N_T$ be the long-run projections.

For future reference, let $Y_T := [X_{1T}, N_{1T}, \ldots, X_{qT}, N_{qT}]'$.

Let Y_T be the cosine transforms of the time series y_t. Because Y_T consists of weighted averages, when T is large and q is fixed one must expect

$$Y_T \overset{\text{approx}}{\sim} \mathrm{N}\left(0_{(2q,1)}, \Omega\right).$$

Ω is determined by the local-to-zero spectrum $S(\lambda)$: This is the L_1-limit of

$$f_y(e^{i\lambda/T}) = \Theta(e^{i\lambda/T}) \Sigma_\varepsilon \Theta'(e^{-i\lambda/T})$$

.

Recall from the calculations of the section that introduces the empirically relevant long-run restrictions that we can think of $f_y(e^{i\lambda})$ as the variance of cycles with frequency λ in y_t. Recall, as well, that the IRF operator $\Theta(z) = \sum_{\ell=0}^{\infty} \Theta\ell z^{\ell}$ may vary with T (but dependence on the index is omitted for ease of notation). One of the reasons why I do this is because it allows me to accommodate scaling and, importantly, consider processes more general than I(0) and I(1) processes.

The bottomline from the preceding discussion, which I borrow from the low-frequency econometrics literature, is that one must only model S to obtain an approximate likelihood for Y_T. And, moreover, simple models of S are good enough to capture interesting features of low-frequency variability. The approach I propose for empirically relevant long-run restrictions builds on the following:

Lemma 1. *Under assumptions in Müller and Watson (2017, Thm. 1) each Wiener-Hopf factor of $f_y(e^{i\lambda/T})$ converges (in L_1) to a factor of* $\mathrm{S}(\lambda)$.

This rather technical lemma has a very simple interpretation in terms of the triangular factorization. There are many ways to decompose $f_y(e^{i\lambda}) = a(e^{i\lambda})a'(e^{-i\lambda})$ and each $a(e^{i\lambda/T})$ converges to some $a(\lambda)$ such that $S(\lambda) = a(\lambda)a'(-\lambda)$.

Joining this lemma to the reasoning above we obtain the corollary that modeling S and modeling a are equivalent tasks. Heuristically, we may regard a as capturing the part of the dynamics in y_t that dominate trending behavior. This makes perfect sense because this is the part of the IRF of y_t that are local to the long-run multipliers restricted under the traditional approach to long-run restrictions.

Therefore, lemma 1 must be taken as a low-frequency analogue to the Wold decomposition theorem which, in turn, is not any different from a triangular decomposition theorem for Toeplitz operators. The limit of $\Theta(e^{i\lambda/T})$ will be called the local-to-zero IRF.

Univariate example. Imagine that $y_t = \theta(\mathrm{L})\varepsilon_t$ where for simplicity we take $\mathrm{Var}(\varepsilon_t) = 1$ and $n = 1$. Consider the following two cases. First, if y_t is an I(0) process, it is well-known the local-to-zero spectrum S is flat at the level of the long-run variance of y_t. The local-to-zero IRF becomes flat as well. A simple calculation delivers $\theta(e^{i\lambda/T}) \to \theta(1)$. Second, if y_t is an I(1) process, the local-to-zero spectrum S becomes proportional to λ^{-2} in the vicinity of the zero-frequency. In turn, the local-to-zero IRF behaves as $\theta(e^{i\lambda/T}) = T^{-1}(1 - e^{i\lambda/T})^{-1}\tilde{\theta}(e^{i\lambda/T}) \to i\lambda^{-1}\tilde{\theta}(1)$ where $\tilde{\theta}$ is an IRF displaying I(0) behavior.

The key to the approach I propose is the following. The dynamics of y_t (after appropriate re-scaling by T) are modeled as

$$y_t = \Theta(\mathrm{L})\varepsilon_t = \alpha_T(\mathrm{L})\mathrm{H}^{-1}\eta_t,$$

with η_t white noise having $\mathrm{var}(\eta_t) = I_n$.

The identification assumptions are: (i) H^{-1} is a square root of Σ_ε,

$$\Sigma_\varepsilon = \mathrm{H}^{-1}(\mathrm{H}^{-1})'.$$

and (ii) a set of empirically relevant long-run restrictions hold, e.g.,

$$\alpha(\lambda)\mathrm{H}^{-1} = \lim_{T\to\infty} \alpha_T(e^{i\lambda/T})\mathrm{H}^{-1} \text{ is lower triangular.}$$

The intuition for the restriction is as discussed before but it is instructive to reconsider it in the context of the newly defined local-to-zero IRF and in the context of our technology shock example.

Technology shocks example. Recall that $n = 2$ and define $h(\lambda) := \alpha(\lambda)\mathrm{H}^{-1}$. The corresponding local-to-zero spectrum is $S(\lambda) = h(\lambda)h(-\lambda)$. The element $h_{11}(\lambda) := [h(\lambda)]_{11}$ refers to the multipliers of $\eta_{\mathrm{T}t}$ on x_t while the element $S_{11}(\lambda) := [S(\lambda)]_{11}$ is the variance of a cycle of frequency λ/T in x_t.

The assumption that $h(\lambda)$ is lower triangular means that

$$\mathrm{S}_{11}(\lambda) = h_{11}(\lambda)h_{11}(-\lambda),$$

or, put in words, that only technology shocks contribute variance to cycles of frequency λ/T in average productivity. Empirical relevance originates from restricting a range of frequencies that represent below business cycle variation.

4.1 Parametric specification

Low-frequency econometrics reduces inference to a small-sample problem. An in a small-sample problem there is no option but to proceed in a parametric way. That is why I postulate a family of parametric representations for the local-to-zero IRF. Concretely, I provide the following simple representation of the local-to-zero IRF, called the (A, c) model.

In the bivariate case, the (A, c) model is

$$\alpha(\lambda) = \mathrm{A}\begin{bmatrix} (c_1 + i\lambda)^{-1} & 0 \\ 0 & (c_2 + i\lambda)^{-1} \end{bmatrix} \mathrm{A}^{-1} \cdot \tilde{\alpha},$$

where $\tilde{\alpha} = \tilde{\alpha}(1)$ is the long-run multiplier of some I(0) process.

One interpretation of the (A, c) model is as a low-frequency version of a VAR(1) model. To see the connection consider the process

$$\tilde{y}_t = \left\{ A e^{-T^{-1}\mathrm{diag}(c_1, c_2)} A^{-1} \right\} \tilde{y}_{t-1} + T^{-1}\tilde{u}_t,$$

with $\tilde{u}_t = \tilde{\alpha}(\mathrm{L})\tilde{v}_t$ an I(0) process.

I will now construct a low-frequency likelihood corresponding to the (A, c) model. Recall that the low-frequency likelihood is simply the approximate likelihood of the fixed set of cosine transforms under an asymptotic approximation in which $T \to \infty$ with q (the number of such transforms) fixed.

The local-to-zero spectrum for the (A, c) model is given by

$$\mathrm{S}(\lambda) = \mathrm{AC}(\lambda)\mathrm{A}^{-1}\tilde{\alpha}\Sigma_v\tilde{\alpha}'(\mathrm{A}^{-1})'\mathrm{C}'(-\lambda)\mathrm{A}',$$

where $\mathrm{C}(\lambda) = \mathrm{diag}\{(c_1 + i\lambda)^{-1}, (c_2 + i\lambda)^{-1}\}$.

The implied Ω in the asymptotic jointly normal distribution of Y_T is nearly block diagonal. Each block of Ω is a weighted average of S, namely:

$$\begin{aligned} \Omega_{jk} &= \mathrm{Cov}\Big(Y_{jT}, Y_{kT}\Big) \\ &\to \int_{\mathbb{R}} \left(\int_{[0,1]} \Psi_j(s) e^{i\lambda s} \right) \mathrm{S}(\lambda) \left(\int_{[0,1]} \Psi_k(s) e^{-i\lambda s} \right) \mathrm{d}\lambda. \end{aligned}$$

As a practical matter, Ω can be approximated by $\Psi_T' \mathrm{Var}(\tilde{y}_{1:T})\, \Psi_T$ for large T.

Now, Σ_ε remains the MSE matrix of the infinite-past prediction of y_t. The presence of this matrix means that it is not possible to construct inference about the structural parameters identified by long-run restrictions using low-frequency variation information alone. We need an additional result.

Lemma 2. *There is a root-T uniformly consistent estimator of* Σ_ε.

Let $\hat{\Sigma}_\varepsilon$ be this estimator. Uniform consistency refers to the set of all possible values that $(\mathrm{A}, c, \tilde{\alpha})$ may take. The usefulness of the preceding lemma is that it justifies replacing Σ_ε by $\hat{\Sigma}_\varepsilon$ into the low-frequency likelihood. The large-T approximate low-frequency likelihood is therefore

$$\mathrm{p}(Y_T|\mathrm{A}, c, \tilde{\alpha}) \propto \prod_{j=1}^{q} \{\det(\Omega_{jj})\}^{-\frac{1}{2}} \exp\left\{ -\frac{1}{2} Y_{jT}' \Omega_{jj}^{-1} Y_{jT} \right\},$$

with $\Omega_{jj} = \Omega_{jj}(\mathrm{A}, c, \tilde{\alpha})$ computed from $\mathrm{S}(\lambda) = \mathrm{S}(\lambda|\mathrm{A}, c, \tilde{\alpha})$.

Technology shocks example. Any square-root of Σ_ε can be written as $\mathrm{Ch}(\Sigma_\varepsilon)\mathrm{P}(\theta)$ where $\mathrm{Ch}(\Sigma_v)$ is the lower triangular Cholesky factor of Σ_ε and $\mathrm{P}(\theta)$ is an orthogonal matrix or rotation. In the technology shock example we just need to consider two-dimensional rotations and these form a one-parameter family,

$$\mathrm{P}(\theta) = \begin{bmatrix} \cos(\pi\theta) & -\sin(\pi\theta) \\ \sin(\pi\theta) & \cos(\pi\theta) \end{bmatrix} \text{ for } \theta \in (-1, 1].$$

Inference about H reduces to inference about θ, the angle of the rotation relative to the Cholesky factor. This, as was anticipated, is a nonstandard estimation problem. The reason why this is a nonstandard estimation problem is that uncertainty about $\mathrm{Ch}(\Sigma_\varepsilon)$ is negligible for large T but uncertainty about $(\mathrm{A}, c, \tilde{\alpha})$ remains in large samples and translates into uncertainty about θ.

Here is an important observation. The normalized angle θ is determined by the restriction

$$h_{12}(\lambda|\theta, \mathrm{A}, c, \tilde{\alpha}) \equiv \Big[\mathrm{A}\mathrm{C}(\lambda)\mathrm{A}^{-1}\tilde{\alpha}\mathrm{Ch}(\Sigma_v)\mathrm{P}(\theta)\Big]_{12} = 0.$$

This amounts to a continuum of constraints. Trying to set them all may rule out plausible values of $(\mathrm{A}, c, \tilde{\alpha})$ and, therefore, the question is how should we proceed. The simplest possible option is to target just-identification: Enforce the constraint for a single $\lambda = \lambda_0$ or, even better, enforce a weighted average of the constraints. Another option is to allow for over-identification: Enforce a small number of (weighted averages of) the constraints. In the implementation appearing in the empirical application section I enforce a weighted average of the constraints, which is the conceptually simpler case.

Frequentist inference is relatively straightforward in this context but in the implementation

I show how to conduct Bayesian inference. These are the details: The low-frequency posterior distribution of $(\mathrm{A}, c, \tilde{\alpha})$ given Y_T follows from Bayes rule, namely:

$$\mathrm{p}(\mathrm{A}, c, \tilde{\alpha}|Y_T) \propto \mathrm{p}(Y_T|\mathrm{A}, c, \tilde{\alpha}) \cdot \mathrm{p}(\mathrm{A}, c, \tilde{\alpha}).$$

I use a flat prior on a bounded support of $(\mathrm{A}, c, \tilde{\alpha})$. The low-frequency likelihood is, of course, as described above. To proceed, we implicitly condition on Σ_ε. An algorithm to implement the procedure is to first simulate draws from $\mathrm{p}(\Sigma_v|y_{1:T})$, and then to simulate draws from $\mathrm{p}(\mathrm{A}, c, \tilde{\alpha}|Y_T)$ and compute θ. Draws of θ solve the equation

$$\int_{\mathbb{R}} h_{12}(\lambda|\theta, \mathrm{A}, c, \tilde{\alpha})\, \mathrm{dW}(\lambda) = 0.$$

A more flexible representation of $\alpha(\lambda)$: The (A, b, c, d) model is

$$\alpha(\lambda) = \mathrm{A} \begin{bmatrix} b_1 + (c_1 + i\lambda)^{-d_1} & 0 \\ 0 & b_2 + (c_2 + i\lambda)^{-d_2} \end{bmatrix} \mathrm{A}^{-1} \cdot \tilde{\alpha},$$

where $\tilde{\alpha} = \tilde{\alpha}(1)$ is the long run multiplier of some I(0) process.

In constructing set estimators one may want to derive robustified set estimators: Inference about θ with nuisance parameters $(\mathrm{A}, b, c, d, \tilde{\alpha})$ using the corresponding least-favorable-distribution implied set estimators or even derive bet-proof set estimators.

It must be noted that the difference in the rates at which information accumulates for the two blocks produces a nonstandard inference problem in which it is not generally possible to construct a consistent estimator of the structural IRF. Yet, meaningful inference is feasible and relatively simple to obtain and in the paper I examine both frequentist and Bayesian procedures with good statistical properties.

4.2 Comparison to other approaches

Blanchard & Quah (1989) determine H from the assumptions that

$$\Sigma_v = \mathrm{H}^{-1}(\mathrm{H}^{-1})' \text{ and } \tilde{\alpha}\mathrm{H}^{-1} \text{ is lower triangular,}$$

after suitable transformation of y_t into an I(0) process. Thus,

$$\mathrm{H}^{-1} = \tilde{\alpha}^{-1} \cdot \mathrm{Ch}(\tilde{\alpha}\Sigma_v\tilde{\alpha}')$$

Implementation is done by VAR models with lag length constraints.

Shapiro & Watson (1988) provide an IV representation of this approach. That is the classical long-run restriction approach.

The comparison with my approach is as follows. In the present context such transformation depends on (A, c). There is an implicit presumption of certainty about (A, c). Large-T consistent estimation of H follows from this certainty. Since y_t is I(0), S is flat and there is no information beyond $\lambda = 0$: I exploit information on frequencies in a T^{-1}-neighborhood of $\lambda = 0$.

In my approach $(A, c, \tilde{\alpha})$ are not consistently estimable: Uncertainty about those parameter does not vanish even as $T \to \infty$. Consistent estimation of H is not possible as a consequence. But perfectly valid and informative inference can be performed.

Recent studies focus on inference robust to near UR dynamics in n_t: Near non-invertibility of $\tilde{\alpha}$. Yet identification by a restriction in the limit as $t \to \infty$. Some references in this literature are Gosporodinov (2010), Chevillon et al. (2017).

An approach that is taking off in empirical macroeconomics: Identification by maximizing contribution to variance at long horizons. Econometrics of this method are not well understood: Probably subject to bias. Fragile when two shocks are nearly equally important. A reference is Neville et al. (2014).

5 Empirical analysis

In this aside the debate about the impact of factor-neutral technology shocks on aggregate employment and their importance for business cycle fluctuations in the U.S. economy is revisited.

A huge amount of work has been done in this area and a good part of it is not directly related to long-run restrictions. See Ramey (2016) for a recent and complete account. From the point of view of the proximity to the methodology of this paper I must mention Galí (1999) and Christiano et al. (2004).

The real business cycle theory put forward the idea that a simple model featuring a representative consumer optimally responding to shocks to the production technology of the economy is well able to reproduce the fluctuations from a large complex economy such us the U.S. Galí (1999) challenged this idea empirically. He modeled the joint dynamics of (the first differences of logs of) labor productivity and (the first differences of) hours worked as a VAR. Next, he separated technology from non-technology shocks by a long-run exclusion restriction: the responses of labor productivity to non-technology shocks accumulates to zero. As a result, he estimates a small contribution to business cycle fluctuations (perhaps, too small compared to the succeeding evidence) together with a negative contemporaneous response of employment to technology shocks. Such a negative response is not compatible with the basic real business cycle model (for reasonable parameter values).

Later, Christiano et al. (2004) report evidence against the conclusions of Galí (1999). They modeled the joint dynamics of (the first differences of logs of) labor productivity and hours worked (in levels) as a VAR. They also obtain identification of technology shocks by means of a long-run exclusion restriction that states that the responses of labor productivity to non-technology shocks sums to zero. They document that technology shocks matter significantly for business cycle fluctuations and that the response of employment to them is actually positive.

The main difference between the two papers lies in the assumption that hours are either I(0) or I(1). This is a secondary assumption for the identification of technology shocks and it is somewhat unpleasant that results hinge so crucially on a secondary assumption.

In this context, I estimate a dynamic system that allows hours to display low-frequency variation compatible with a range of models that nests the I(0) and I(1) alternatives. Long-run and short-run statistics are displayed in the figures that follow.

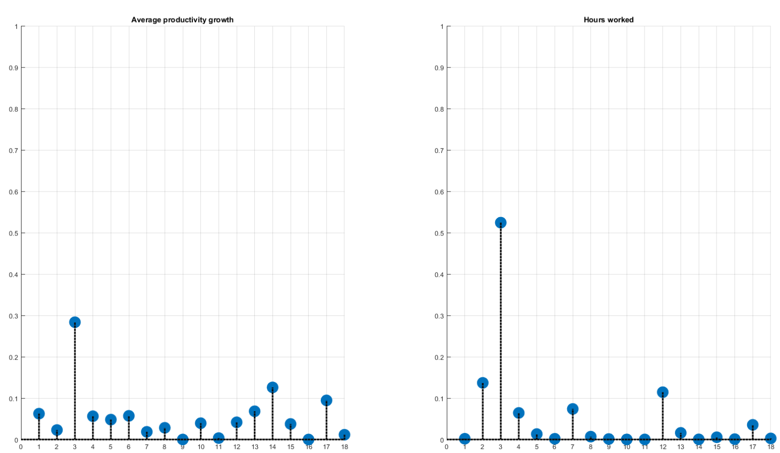

FIGURE 1. Normalized square cosine transforms.

Next, I impose the long-run restriction that excludes non-technology shocks from labor productivity understanding the long-run as variation taking place below a 8-year period. When I

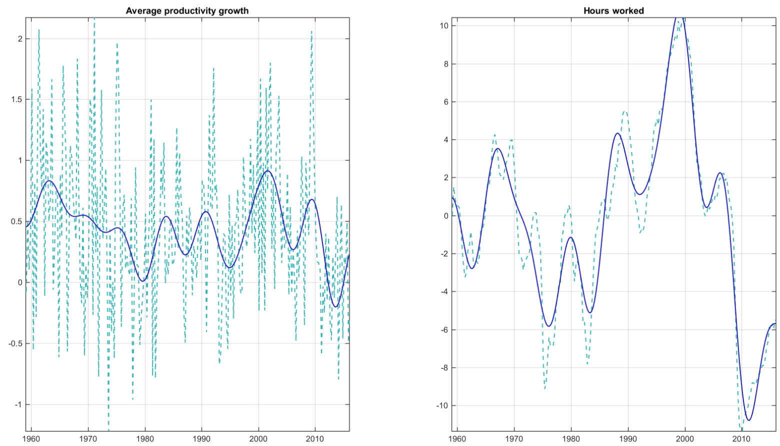

FIGURE 2. Series (green/dashed) and long-run projections (blue/solid).

do this I obtain a negative (and significantly different from zero) response of hours to technology shocks and a modest (but not negligible) fraction of the variation in productivity accounted for by technology shocks. A 90%-credible set for the response of a positive technology shock (normalized to fit the units of labor productivity) is $[-0.58, -0.19]$. Regarding the importance of technology shocks, a 90%-credible set for forecast error variance decomposition at horizon $h = 0$ (i.e., at impact) is $[0.02, 0.19]$.

For comparability, some results are replicated when the long-run is understood as variation below a 30-year period instead of our shorter notion. Only in that case do I get a positive response of hours and a larger forecast error variance decomposition at horizon $h = 0$. See figures 3 and 4 in appendix A.

That the results differ so much is an indication that overidentifying restrictions may be at fault. This is addressed in the following subsection.

6 Conclusion

This paper introduces an empirically relevant approach to long-run restrictions and argues that these are natural and appealing in applied practice. The context typically provides a precise and well-motivated notion of the long-run, and it is conceptually straightforward to impose

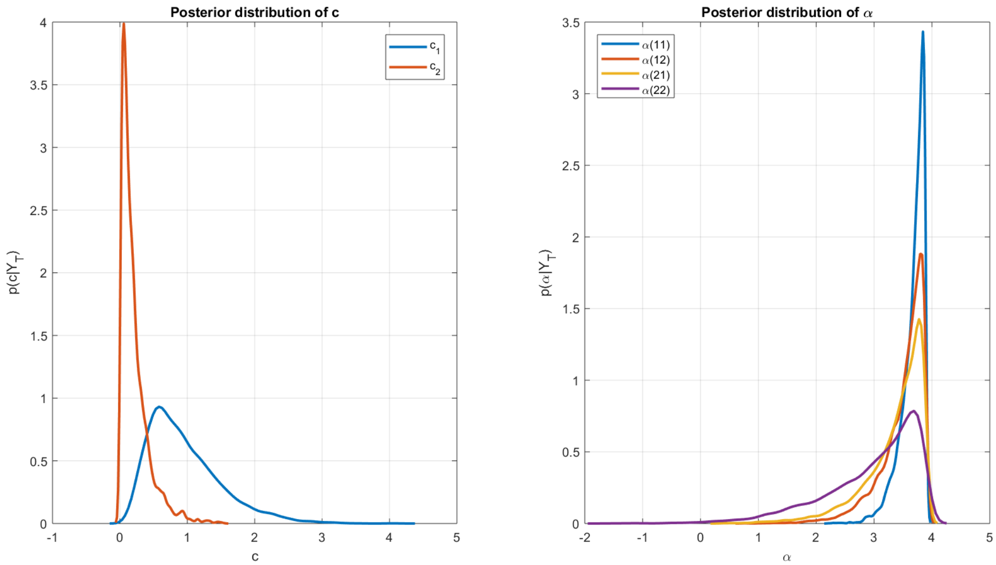

FIGURE 3. Posterior densities from simulation draws.

and interpret restrictions under such a notion by means of the local-to-zero IRF. The identification content of the empirically relevant approach to long-run restrictions is thoroughly explored by developing a general theory of identification.

Moreover, a convenient way to conduct inference about structural parameters and dynamic causal effects is obtained by expressing the data in terms of two approximately independent blocks: one that captures low-frequency variability and another that represents high-frequency variability. Approximate likelihoods can be derived for each block (justified under a natural asymptotic framework) from where both frequentist and Bayesian inferences follow rather easily. Because in this framework information about low-frequency variability accumulates slowly, the inference problem is nonstandard with the implication that it is not generally possible to produce a consistent point estimator of the parameters of interest. However, meaningful set estimators can be constructed and the two applications developed in the paper indicate that empirically meaningful inference is possible.

The research I present in this paper speaks to a broader project of developing methods to combine more realistic models of low-frequency variability with standard descriptions of short-run dynamics. The spirit is eloquently expressed in the commentary made by Sims (2005) to Galí (2005). I believe the likelihood separation principle I introduce in the paper is the natural vehicle to perform such a combination and its possibilities should be explored in other contexts.

On a more concrete ground, I would like to highlight two additional directions for further research. The first one is to investigate the possibility of formulating long-run sign restrictions

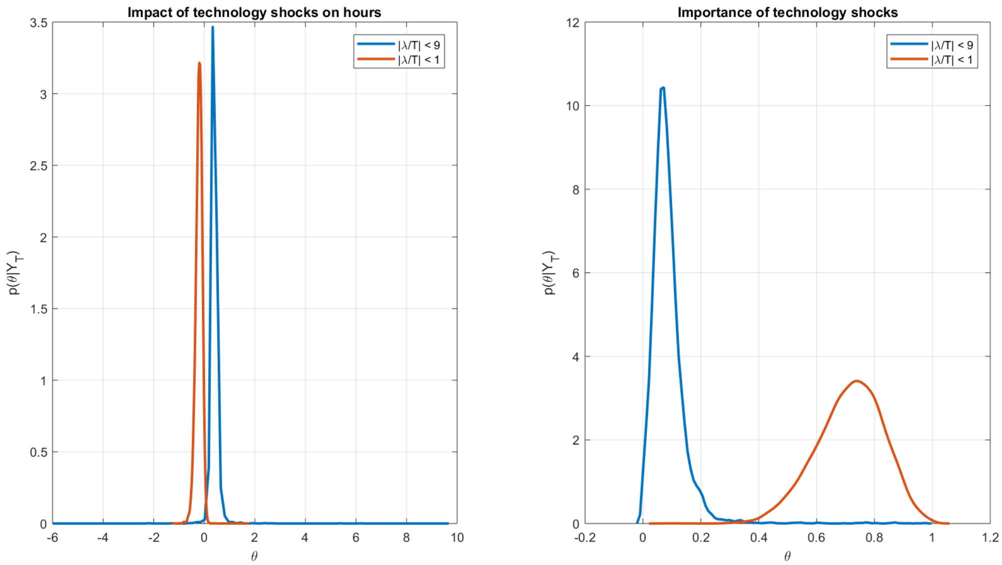

FIGURE 4. Posterior densities from simulation draws.

or, more generally, sign restrictions on certain range of frequencies. The second one follows more directly from the method explained in the paper. One may wish to use a long-run restriction in such a way that reflects a shock having a predominant (albeit not total) role in accounting for the variance of components of a certain variable with sufficiently low frequencies. More specifically, one could say that ε_{kt} accounts for a fraction within $1-\delta$ of the totality of the variance of a cycle of frequency λ in y_{jt}, i.e., $(1-\delta)[\Theta(z)]^2_{(j,k)} \leq \mathbb{V}_\lambda\left(y_j\right) \leq [\Theta(z)]^2_{(j,k)}$.

A Proofs

A.1 Theorem 1

Proof of theorem 1. The argument parallels that of Rubio-Ramírez et al. (2010, th. 1).

Suppose there is $\tilde{\Theta}_{0:\infty}$ that satisfies the identification condition (4) and the restrictions (5) together with the normalization but for which $\tilde{\Theta}_{0:\infty} e_K \neq \Theta_{0:\infty} e_K$. There is, then, an orthogonal matrix $\tilde{Q}$ with at least one non-zero off-diagonal element such that $\Theta_{0:\infty}\tilde{Q} = \tilde{\Theta}_{0:\infty}$. Let k be the index of the first column of $\tilde{Q}$ at which a non-zero off-diagonal element occurs.

For $\tilde{\Theta}_{0:\infty} e_K \neq \Theta_{0:\infty} e_K$ to be true it must be that $k \leq K$. Form the vector

$$v_k := \tilde{Q} e_k - \tilde{Q}_{kk} e_k,$$

where e_k is the k-th column of I_n and $\tilde{Q}_{kk}$ is the (k,k) entry of $\tilde{Q}$. The choice of k guarantees that $v_k \neq 0_{n\times 1}$. Now, note that

$$R_k g\left(\Theta_{0:\infty}\right) v_k = \left[R_k g\left(\tilde{\Theta}_{0:\infty}\right) - \tilde{Q}_{kk} R_k g\left(\Theta_{0:\infty}\right)\right] e_k = 0_{m\times 1}$$

since both $\tilde{\Theta}_{0:\infty}$ and $\Theta_{0:\infty}$ satisfy (5). Also, note that

$$\begin{bmatrix} I_k & 0_{k\times(n-k)} \end{bmatrix} v_k = 0_{k\times 1},$$

by choice of k and because the k-th entry of v_k is zero. Thus,

$$\begin{bmatrix} R_k g\left(\Theta_{0:\infty}\right) \\ \begin{bmatrix} I_k & 0_{k\times(n-k)} \end{bmatrix} \end{bmatrix} v_k = 0_{(m+k)\times 1}$$

for some nonzero vector v_k indicating that, for some $k \leq K$,

$$\operatorname{rank}\left(\begin{bmatrix} R_k g\left(\Theta_{0:\infty}\right) \\ \begin{bmatrix} I_k & 0_{k\times(n-k)} \end{bmatrix} \end{bmatrix}\right) < n$$

This establishes the first part of theorem 1. The second part follows from the first. ■

References

ALMUZARA, M. AND A. MARCET (2019): "To diff, or not to diff?" Working paper.

BLANCHARD, O. J. AND D. QUAH (1989): "Dynamic effects of aggregate demand and supply disturbances." *American Economic Review*, 79, 655–673.

CANOVA, F., D. LÓPEZ-SALIDO, AND C. MICHELACCI (2010): "The effects of technology shocks on hours and output: A robustness analysis." *Journal of Applied Econometrics*, 25, 755–773.

CHRISTIANO, L. J., M. EICHENBAUM, AND R. VIGFUSSON (2004): "The response of hours to a technology shock: Evidence based on direct measures of technology." *Journal of the European Economic Association*, 2, 381–395.

FRANCIS, N., M. T. OWYANG, J. E. ROUSH, AND R. DICECIO (2014): "A flexible finite-horizon alternative to long-run restrictions with an application to technology shocks." *Review of Economics and Statistics*, 96, 638–647.

GALÍ, J. (1999): "Technology, employment, and the business cycle: Do technology shocks explain aggregate fluctuations." *American Economic Review*, 89, 249–271.

——— (2005): "Trends in hours, balanced growth, and the role of technology in the business cycle." *Federal Reserve Bank of St. Louis Review*, 87, 459–486.

GRANGER, C. W. J. AND M. W. WATSON (1984): "Time series and spectral methods in econometrics." in *Handbook of Econometrics*, ed. by Z. Griliches and M. D. Intriligator, North-Holland, vol. 2, chap. 17.

KING, R. G., C. I. PLOSSER, J. H. STOCK, AND M. W. WATSON (1991): "Stochastic trends and economic fluctuations." *American Economic Review*, 81, 819–840.

MÜLLER, U. K. AND M. W. WATSON (2008): "Testing models of low-frequency variability." *Econometrica*, 76, 979–1016.

——— (2017): "Low-frequency econometrics." in *Advances in Economics and Econometrics: Eleventh World Congress of the Econometric Society*, ed. by B. Honoré and L. Samuelson, Cambridge University Press, vol. 2.

PEERSMAN, G. AND F. SMETS (2003): *The Monetary Transmission Mechanism in the Euro Area: Evidence from VAR Analysis*, Cambridge University Press.

PLAGBORG-MØLLER, M. (2019): "Bayesian inference on structural impulse response functions." *Quantitative Economics*, 10, 145–184.

PLAGBORG-MØLLER, M. AND C. K. WOLF (2019): "Instrumental variable identification of dynamic variance decompositions." Working paper.

RAMEY, V. (2016): "Macroeconomic shocks and their propagation." in *Handbook of Macroeconomics*, ed. by J. B. Taylor and H. Uhlig, North-Holland, vol. 2, chap. 2.

RUBIO-RAMÍREZ, J. F., D. F. WAGGONER, AND T. ZHA (2010): "Structural vector autoregressions: Theory of identification and algorithms for inference." *Review of Economic Studies*, 77, 665–696.

SHAPIRO, M. D. AND M. W. WATSON (1988): "Sources of business cycle fluctuations." *NBER Macroeconomics Annual*, 3, 111–156.

SIMS, C. A. (2005): "Comment on Galí (2005)." *Federal Reserve Bank of St. Louis Review*, 87, 489–492.

STOCK, J. H. AND M. W. WATSON (2016): "Dynamic factor models, factor-augmented vector autoregressions, and structural vector autoregressions in macroeconomics." in *Handbook of Macroeconomics*, ed. by J. B. Taylor and H. Uhlig, North-Holland, vol. 2, chap. 8.