University of Glasgow

Adam Smith Business School

WORKING PAPER SERIES

Frequency-band estimation of the number of factors detecting the main business cycle shocks.

Marco Avarucci, Maddalena Cavicchioli, Mario Forni, Paolo Zaffaroni

Paper no. 2022-13
November 2022

FREQUENCY-BAND ESTIMATION OF THE NUMBER OF FACTORS DETECTING THE MAIN BUSINESS CYCLE SHOCKS

MARCO AVARUCCI
Adam Smith Business School, University of Glasgow

MADDALENA CAVICCHIOLI
Department of Economics "Marco Biagi", University of Modena and Reggio Emilia

MARIO FORNI
Department of Economics "Marco Biagi", University of Modena and Reggio Emilia

PAOLO ZAFFARONI
Business School, Imperial College London

November 3, 2022

We introduce consistent estimators for the number of shocks driving large-dimensional dynamic factor models. Our estimators are functions of the eigenvalues of the spectral density matrix of the observables and can be applied to single frequencies as well as to specific frequency bands, making them suitable for disentangling shocks affecting dynamic macroeconomic models with a factor model representation. Their small-sample performance in simulations is excellent, even in estimating the number of shocks that drive medium-sized DSGE models. We apply our estimators to the FRED-QD dataset, finding that the U.S. macroeconomy is driven by two shocks: an inflationary demand shock and a deflationary supply shock.

JEL CLASSIFICATION: E00, C01, C55

KEYWORDS: Frequency Bands, Dynamic Eigenvalue Ratio, Generalized Dynamic Factor Models, Business Cycle, Permanent Component, DSGE.

1. INTRODUCTION

Real Business Cycle (RBC) models postulate the existence of a single supply shock that drives variables related to real economic activity (Kydland and Prescott, 1982). In contrast, modern Dynamic Stochastic General Equilibrium (DSGE) models feature several shocks (Smets and Wouters, 2007). The task of choosing between these rival models raises two fundamental empirical questions: How many shocks are there in the economy overall? And, even more importantly, how many shocks drive the business cycle, and how many, instead, do so over the long run? Recently, Angeletos et al. (2020) (henceforth ACD) show empirically that just one shock (a noninflationary demand shock) explains the bulk of the business cycle fluctuations of real macroeconomic variables, and they uncover a second shock that drives the long

Marco Avarucci: marco.avarucci@glasgow.ac.uk
Maddalena Cavicchioli: maddalena.cavicchioli@unimore.it
Mario Forni: mario.forni@unimore.it
Paolo Zaffaroni: p.zaffaroni@imperial.ac.uk

We thank Seung (Min) Ahn, Juhee Bae, Luca Gambetti, Marc Hallin, Paymon Khorrami, Marco Lippi, Roman Liška, Emi Nakamura, Alexei Onatski, Luca Sala, and Stefano Soccorsi for sharing codes and providing useful suggestions. A special thanks goes to Georges-Marios Angeletos, whose suggestions were very helpful to improve the empirical part of the paper; to Fabrice Collard who shared his codes and patiently explained to us how to use them; and to Alex Horenstein, who provided useful comments about the motivation of our preferred estimator. Forni acknowledges financial support by the PRIN-MIUR Grant 20174ESLX7.

run. These two shocks appear orthogonal to one another, even though ACD's methodology does not impose orthogonality from the outset. As ACD elicit, addressing these questions sheds light on which approach to macroeconomic modelling appears more appropriate. More generally, it can help to clarify issues at the heart of the business cycle debate, namely whether the main macroeconomic shocks are permanent or transitory and whether they are mainly demand or supply shocks.

Structural Vector Autoregression (SVAR) models—the most popular tool in business cycle analysis— are not suitable to identify the number of shocks, given that in SVARs the number of shocks is necessarily equal to the number of variables. In this paper we depart from SVAR and frame the problem within the realm of Generalized Dynamic Factor Models (GDFM), naturally designed to investigate the effect of macroeconomic shocks at any desired frequency band, such as cyclical and long-run frequencies, because GDFM are naturally wired into the frequency domain. Noticeably, in GDFM the number of shocks is typically much smaller than the number of variables described by the model, which is an important feature also shared by several DSGE models. Indeed, strong analogies exists between GDFM and DSGE models because both admit a *singular finite-order* VAR representation (i.e., when the observables satisfy a vector autoregressive model with a finite number of lags together with a singular covariance matrix of the shocks). For DSGE models, this representation is derived by log-linearizing the models around their steady-state, whereas for GDFM it holds under mild conditions on the impulse response functions (IRFs).

This paper proposes new consistent estimators for the number of common shocks within the realm of GDFMs. Our main estimator is the Dynamic Difference Ratio (DDR) estimator but we also study two additional criteria: the Dynamic Eigenvalue Ratio (DER) and the Dynamic Growth Ratio (DGR) criteria. We can apply our main estimator DDR, DER, and DGR, to single frequencies as well as to specific frequency bands of interest. Our estimators perform extremely well in a wide range of simulated environments against existing methodologies.

To illustrate the powerfulness of our method, we develop an extensive empirical application, applying our criteria to the FRED-QD dataset. We demonstrate that two dominant shocks drive the U.S. economy, an inflationary demand shock and a deflationary supply shock, corroborating ACD findings. Noticeably, the two shocks appear largely disconnected, along the lines of ACD, in the sense that they affect the economy at different frequency bands, with modest overlap. The demand shock explains most of the cyclical fluctuations in the main macroeconomic aggregates; the supply shock affects mainly the long-run. However, unlike ACD, we find a close connection between our demand shock and inflation: the shock is inflationary at business-cycle frequencies and explains inflation in the long-run.[1]

Estimators of the Number of Common Shocks. Going more in detail on the core methodological contribution, let $\mathbf{x}_{nt}$ be an n-dimensional stochastic process and $\widehat{\mathbf{\Sigma}}_n(\omega_\ell)$ be a suitable estimator of the spectral density matrix of $\mathbf{x}_{nt}$ at the Fourier frequencies $\omega_\ell = 2\pi\ell/T$, $\ell = 0, 1, \ldots, T-1$, with T being the available number of time-series observations of $\mathbf{x}_{nt}$. Moreover, let $\hat{\mu}_{nk}(\omega_\ell)$ be the $k-$th eigenvalue of $\widehat{\mathbf{\Sigma}}_n(\omega_\ell)$ in decreasing order of magnitude, with $1 \leq k \leq n$, here denominated as the *dynamic eigenvalues*, to distinguish them from the *static eigenvalues* (i.e., the eigenvalues of the variance-covariance matrix of $\mathbf{x}_{nt}$). Finally, let $\hat{\mu}_{nk}$ be the average of these eigenvalues across frequencies within the frequency band $(\omega_{\underline{\ell}}, \omega_{\overline{\ell}})$:

[1] The reasons for this divergence is spelled out in Section B.2 of the Online Appendix.

$$\hat{\mu}_{nk} \equiv \sum_{\ell=\underline{\ell}}^{\overline{\ell}} \hat{\mu}_{nk}(\omega_\ell)/(\overline{\ell} - \underline{\ell} + 1).$$

Considering the ratio $\mathrm{DDR}(k) = (\hat{\mu}_{nk} - \hat{\mu}_{n,k+1})/(\hat{\mu}_{n,k+1} - \hat{\mu}_{n,k+2})$ for every $1 \leq k \leq q_{\max} < n$, our estimator of the number of shocks, $\hat{k}_{\mathrm{DDR}}$, is given by the value of k maximizing DDR, or $\hat{k}_{\mathrm{DDR}} \equiv \arg\max_{1\leq k\leq q_{\max}} (\hat{\mu}_{nk} - \hat{\mu}_{n,k+1})/(\hat{\mu}_{n,k+1} - \hat{\mu}_{n,k+2})$. We prove that $\hat{k}_{\mathrm{DDR}}$ converges in probability to the true number q of shocks, as n and T go to infinity at the same rate.

To provide an intuition for the DDR estimator, let us recall an important result in GDFM theory: a q-dimensional dynamic factor structure is characterized by the behaviour of the eigenvalues of the spectral density matrix $\mathbf{\Sigma}_n(\omega)$ of the $\mathbf{x}_{nt}$, as $n \to \infty$.[2] The q largest eigenvalues diverge, whereas the others are bounded.[3] We show here that, under suitable assumptions, similar properties hold for the sample analogue of such eigenvalues, that is $\hat{\mu}_{nk}(\omega)$. Now consider

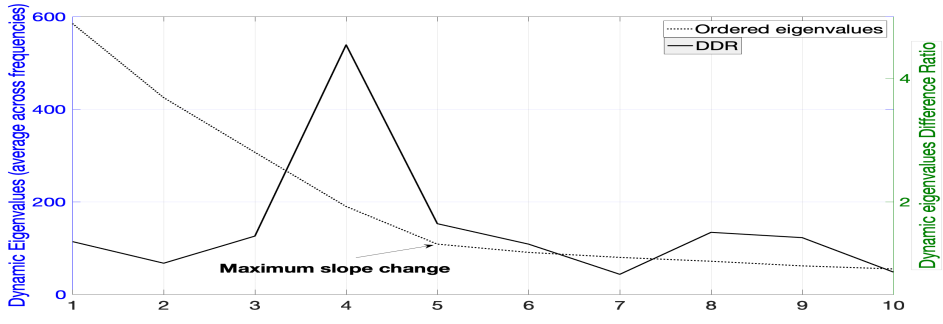

FIGURE 1.—The dotted line is the plot of $\hat{\mu}_{nk}$ (left y axis) as a function of k (x axis). The solid line is $\mathrm{DDR}(k)$ (right y axis). Estimates are obtained from data generated with the ARMA specification of the third experiment in Subsection B.6 of the Online Appendix, $(n, T) = (100, 200)$, $q = 4$. $\mathrm{DDR}(k)$ reaches its maximum in k^* when the maximum slope change of the dotted line is in $k^* + 1$.

the average sample eigenvalue $\hat{\mu}_{nk}$ as a function of k and the piecewise linear curve, often indicated as a polyline, linking the points $(k, \hat{\mu}_{nk})$ (dotted line, Figure 1). DDR is the ratio of the slopes of two adjacent segments; it measures the curvature of the polyline at $k + 1$. The larger DDR is, the smaller the angle above the polyline is at $k + 1$. By maximizing DDR we identify the point $k + 1$ where the steep descent ends and the slight slope begins. The basic idea is that when $k = q$, the descent must be steep, because $\hat{\mu}_{nq}$ is large whereas $\hat{\mu}_{nq+1}$ is small. On the other hand, at $k = q + 1$ the slope becomes small, because it is the difference between small eigenvalues.

Although DDR is the main methodology proposed here, it is not the only estimator studied in this paper. We analyze two additional criteria that represent the dynamic equivalents of two criteria Ahn and Horenstein (2013) propose to estimate the number of static factors within

[2] See Forni et al. (2000), Forni and Lippi (2001).

[3] Similarly, the existence of a *static* factor representation is linked to the behaviour of the eigenvalues of the *variance-covariance matrix* of the first n variables (Chamberlain and Rothschild, 1983), and not of their spectral density matrix.

a static factor model (i.e., the Eigenvalue Ratio and the Growth Ratio criteria). In analogy with these denominations, we call these new estimators the Dynamic Eigenvalue Ratio (DER) and the Dynamic Growth Ratio (DGR) estimators. We prove that these estimators converge in probability to q, as n and T go to infinity at the same rate. A noticeable feature of DDR, DER, and DGR is that they do not require discretionary preliminary choices of tuning parameters from the researcher, apart from the size of the spectral window. This is an advantage over existing criteria, which, besides the window size, require making additional choices, that may affect the results.

Small sample performances. We evaluate the finite-sample performance of DDR, DER, and DGR and compare it with two existing criteria, namely the ones Hallin and Liska (2007) (HL hereafter) and Onatski (2009) (O hereafter) propose, by means of Monte Carlo simulations. Particularly, we run four experiments. For the first three, we evaluate DDR, along with the competing estimators, on the whole interval $[0, \pi]$. We find that DDR dominates DER and DGR in the first three experiments and performs comparably with DGR in the fourth one. Moreover, DDR performs better than, or comparably with, HL and O for all experiments. We conclude that DDR dominates DER and DGR, and is an excellent alternative to existing criteria as far as finite-sample performance is concerned. In experiment four, the spectral density matrix of the variables have reduced rank $q - 1$ at frequency zero. The economic interpretation is that we have only real activity variables and there is a shock, such as a demand shock, having only transitory effects on all variables. We show that DDR is able to detect the rank reduction.

Applications. Our first analysis is based on a quarterly U.S. macroeconomic dataset: FRED-QD (McCracken and Ng, 2020). DDR provides a clear-cut result: two dominant shocks drive the U.S. macroeconomy, here dubbed *main business cycle shocks*. This result holds *both* on specific frequency bands and on the entire $[0, \pi]$ interval. Moreover, it holds *both* for the whole sample period and at several sub-periods. We find that two common shocks are sufficient to capture the bulk of the variance of the main macroeconomic aggregates, *both* at cyclical frequencies and in the long term.

In this paper, we do not proceed further with parametric estimation of the GDFM, but one could readily identify and estimate the shocks, along with the corresponding IRFs (see Forni et al. (2015, 2017) and Barigozzi et al. (2022)).

However, full estimation of the model is not necessary to obtain the decomposition of the spectral density produced by the dynamic eigenvalues and the corresponding eigenvectors. This decomposition corresponds to one of the many possible identification schemes, namely the one in which the first shock maximizes the sum of the explained variances of all variables in the panel (Brillinger, 1981). It turns out that this identification, although based on a statistical criterion, produces shocks that are *economically interpretable.* The first shock explains almost nothing of the long-run variance of GDP, consumption, investment, unemployment rate, and hours worked. Furthermore, it induces a positive covariance between GDP growth and inflation changes. Therefore, it has the salient features of a *demand* shock. This demand shock explains most of the cyclical fluctuations in GDP and other variables related to real economic activity. In addition, it explains most of the variance in inflation and the federal funds rate at the lower frequencies, including only the longer cyclical fluctuations (about four years and longer). The second shock has instead the typical traits of a *supply* shock, explaining the bulk of the long-term variance of the real activity variables and inducing negative covariance between GDP growth and inflation changes. The supply shock explains a minor, but not negligible, part of cyclical fluctuations of real activity variables, but it explains little about inflation and the federal funds rate at all frequencies.

These empirical results are incompatible with both the RBC model and the view that news shocks (i.e., shocks that best anticipate future productivity) explain the bulk of business cycle fluctuations (see Beaudry and Portier, 2006). The finding that the most important main business cycle shock has a transitory effect on output is in line with Angeletos et al. (2020). On the other hand, contrary to Angeletos et al. (2020), we find that it is not disconnected from inflation.

In our second analysis, we focus on DSGE models. We perform a further Monte Carlo exercise, illustrated in Section 4.2, tailored to assess the ability of our criteria to disentangle the number of structural shocks in medium-scale DSGE models, in view of the strong analogies between the GDFM and the reduced form of DSGE models, as they can both be represented as singular VARs of finite order. We first show how the DDR and DGR criteria are extremely reliable in correctly identifying the number of structural shocks driving two popular DSGE models, such as Justiniano et al. (2010) and Angeletos et al. (2018), even if this number is relatively large compared with the size of the cross-section. We conjecture that our methodology successfully applies to any DSGE model that admits a singular VAR representation (around steady-state).

We also demonstrate how to use our decomposition of the spectral density as a descriptive device to highlight important, but usually overlooked, frequency domain features of such medium-scale macroeconomic DSGE models. In particular, we evaluate how well the DSGE models in Justiniano et al. (2010) and Angeletos et al. (2018) reproduce the kind of empirical evidence described above.[4] This exercise, described in detail in Appendix B.3, provides further evidence on the usefulness of our methodology in a context even wider than GDFM, such as for DSGE models.

Related literature. This paper is related to the literature on factor model methods for estimating the number of latent factors, especially to the aforementioned Ahn and Horenstein (2013), Hallin and Liska (2007), and Onatski (2009), as these methodologies are all depending on the estimated eigenvalues of the sample covariance matrix and the sample spectral density matrix, respectively, with Ahn and Horenstein (2013) (hereafter AH) being the closest to our work. In particular, the DER and DGR are the dynamic counterparts (same estimator for the dynamic factor model instead of the static one) of the ER and GR estimators proposed by AH. This close analogy allowed us to adapt AH's proof to the dynamic setting, exploiting the static representation of the model when turned into the frequency domain and finding upper and lower bounds for the dynamic eigenvalues (as opposed to the static eigenvalues of AH) by means of random matrix theory results (see Bai and Yin (1993)).[5] On the other hand, Hallin and Liska (2007), and Onatski (2009) (hereafter HL, and O, respectively) are the only existing procedures to determine the number of factors in a dynamic factor model without assuming the static representation (with DDR being similar to the estimator used by O's test statistic) but whereas, like us, HL assume that n, T diverge simultaneously, O's test require that T diverges much faster than n. This aspect is relevant as in most macroeconomic and financial datasets T and n have similar sizes.

[4] We find that the model in Justiniano et al. (2010) is successful in reproducing the empirical features of the demand sources of variation in the data for most variables, but it is less able to reproduce the supply side of the economy. Recalling that the Angeletos et al. (2018) model does not model inflation, prompting us to look only at the first dynamic principal component, the model is akin to the data in capturing the explained variance of real variables, such as GDP, consumption, investment, and labor productivity.

[5] Like AH, we do not need to show the convergence of the smoothed periodogram to the population spectral density matrix and, like them, require the condition $n/T \to c$ for a positive bounded constant c (see our key result in Lemma A.13), but, differently from AH, we do not require independence between the latent factors and the idiosyncratic component.

Regarding the empirical results, our paper is related to a scant literature concerning the number of shocks driving the macroeconomy, with the closest paper being Angeletos et al. (2020), which proposes a frequency-domain method to identify business-cycle and long-run shocks. Sargent and Sims (1977), using a small-dimensional dynamic factor model, find that two shocks fit U.S. macroeconomic data reasonably well. Giannone et al. (2005) argue informally in favor of two shocks, based on the explained variances in the principal component series of a large factor model. Bai and Ng (2007) and Hallin and Liska (2007), using their proposed criteria for the number of dynamic factors, find four shocks. Amengual and Watson (2007) find seven shocks. Finally, Onatski (2009) finds that his proposed test cannot reject the null of two shocks versus the alternative, which assumes between three and seven shocks. Our finding of two main shocks is in line with Sargent and Sims (1977), Giannone et al. (2005) and Onatski (2009).

From the viewpoint of the business cycle literature, our paper is related to a wide array of literature about the role of demand and supply shocks. A few prominent papers are Kydland and Prescott (1982), Blanchard and Quah (1989), King et al. (1991), Galí (1999), Beaudry and Portier (2006), Bloom (2009), and in particular Angeletos et al. (2020). At first glance, the ACD approach may appear as in alternative to ours. Indeed, this is not the case. The purpose of our method is to estimate the number of common shocks, and thus it is best seen as complementary to the ACD method, which is about identifying and estimating the shocks themselves. We discuss this point in detail in the online Appendix, Section B.2.[6] Given the strong analogies between GDFM and DSGE models, our paper also contributes to this vast literature (see Smets and Wouters, 2007, Justiniano et al., 2010, and Angeletos et al., 2018, among many others), as it provides a sound way to estimate the number of shocks driving candidate DSGE models, thereby providing a taxonomy in terms of their frequency domain effect (i.e., their effect per frequency band). Finally, our paper is related to Forni et al. (2022), which focuses on VAR identification techniques in the frequency domain, extending ACD's method in various directions.

Exposition. The paper is organized as follows. Section 2 presents the factor model setup along with our estimators of the number of shocks and our consistency result. Section 3 presents a summary of the Monte Carlo exercises. Section 4 illustrates the practical powerfulness of our methodology: Section 4.1 presents the empirical application based on the large-dimensional dataset of the U.S. economy, and Section 4.2 demonstrates the usefulness of our methodology when applied to DSGE models. Section 5 concludes. Finally, the Appendix present the regularity assumptions and the formal proofs of our methodology. The Online Appendix presents a few auxiliary lemmas (Appendix B.1), an in-depth discussion about the relation with the method and results in Angeletos et al. (2020) (Appendix B.2), a frequency domain analysis of DSGE models (Appendix B.3), full details on the Monte Carlo experiments (Appendices B.4, B.5, B.6 and B.7) and additional material related to the empirical application (Appendices B.8 and B.9).

[6] Beaudry et al. (2020) is another interesting business cycle paper that uses frequency domain techniques; however, their goal differs from ours and so we abstain from illustrating further analogies and differences with our methodology.

2. METHODOLOGY: FREQUENCY-BAND ESTIMATION OF THE NUMBER OF SHOCKS

2.1. *The GDFM Setup*

The GDFM is a countably infinite set of observable stochastic processes x_{it} (Forni et al., 2000, Forni and Lippi, 2001) with the following decomposition:

$$x_{it} = \chi_{it} + e_{it} = \lambda_{i1}(L)f_{1t} + \lambda_{i2}(L)f_{2t} + \cdots + \lambda_{iq}(L)f_{qt} + e_{it}, \quad i \in \mathbb{N},\ t \in \mathbb{Z}, \tag{2.1}$$

where $\mathbf{f}_t = (f_{1t}\ f_{2t}\ \cdots\ f_{qt})'$ is a q-dimensional orthonormal unobservable white noise and the IRFs $\lambda_{ik}(L), i \in \mathbb{N},\ k = 1, \ldots, q$, are absolutely summable functions in the lag operator L.[7] Detailed assumptions on the *common components* χ_{it} and the *idiosyncratic components* e_{it} are given in Appendix A.1. Let us only recall here that the *common shocks* f_{kt}, $k = 1, \ldots, q$, often called *dynamic factors* in the literature and loaded through one-sided linear filters, that is the IRFs, $\lambda_{ij}(L)$, and the e_{it} are uncorrelated with the f_{jt} at any lead and lag. Moreover, the idiosyncratic components are weakly cross-correlated, uncorrelatedness (across units) being an extreme case. Weak cross-correlation essentially means that simple and weighted averages, such as $n^{-1}\sum_{i=1}^{n} e_{it}$, dissipate (in probability) when n becomes large.

A restriction that we do not impose in this paper, but one often assumed in the literature is that the common components are *contemporaneous* linear combinations of $r \geq q$ unobservable variables F_{ht}, $h = 1, \ldots, r$, often called "static factors" (see Stock and Watson, 2002a,b, Bai and Ng, 2002). In such a case, we say that the model admits a *static* factor representation; the dynamic nature of the model comes from the fact that the static factors have a dynamic representation in the common shocks.[8]

The shocks and the corresponding IRFs in (2.1) are not identified. This is best seen by writing the model in matrix form as

$$\mathbf{x}_{nt} = \boldsymbol{\chi}_{nt} + \mathbf{e}_{nt}, \qquad \boldsymbol{\chi}_{nt} = \boldsymbol{\Lambda}_n(L)\mathbf{f}_t \tag{2.2}$$

where $\mathbf{x}_{nt} = [x_{1t}\ \cdots\ x_{nt}]'$, $\boldsymbol{\chi}_{nt} = [\chi_{1t}\ \cdots\ \chi_{nt}]'$, $\mathbf{e}_{nt} = [e_{1t}\ \cdots\ e_{nt}]'$, and $\boldsymbol{\Lambda}_n(L)$ is the $n \times q$ matrix $[\lambda_{ik}(L) : i = 1, \ldots, n; k = 1, \ldots, q]$, and the prime ($'$) denotes transposition. It is easily seen that $\mathbf{x}_{nt}$ has the alternative representations

$$\mathbf{x}_{nt} = \boldsymbol{\Lambda}_n(L)\mathbf{Q}\mathbf{Q}'\mathbf{f}_t + \mathbf{e}_{nt} = \boldsymbol{\Lambda}_n^*(L)\mathbf{f}_t^* + \mathbf{e}_{nt}, \tag{2.3}$$

where $\boldsymbol{\Lambda}_n^*(L) = \boldsymbol{\Lambda}_n(L)\mathbf{Q}$ and $\mathbf{f}_t^* = \mathbf{Q}'\mathbf{f}_t$, $\mathbf{Q}$ being any orthogonal matrix (i.e., a matrix such that $\mathbf{Q}\mathbf{Q}' = \mathbf{I}_q$).

Forni et al. (2015, 2017) show that the GDFM is tightly connected to VAR models because, by assuming rational IRFs (see footnote 9), the common components and the observables in (2.2) have the VAR representations

$$\mathbf{A}_n(L)\boldsymbol{\chi}_{nt} = \mathbf{R}_n\mathbf{f}_t \qquad \mathbf{A}_n(L)\mathbf{x}_{nt} = \mathbf{R}_n\mathbf{f}_t + \boldsymbol{\eta}_{nt}, \tag{2.4}$$

[7] Orthonormality of the $\mathbf{f}_t$ means that $\mathbb{E}(\mathbf{f}_t) = \mathbf{0}_q, \mathbb{V}\mathrm{ar}(\mathbf{f}_t) = \boldsymbol{I}_q$, and $\mathbb{C}\mathrm{ov}(\mathbf{f}_t, \mathbf{f}_s') = \mathbf{0}_{q\times q}$ for every $t \neq s$. Absolute-summability of the filters $\lambda_{ik}(L)$ means that the sum of their coefficients, taken as absolute values, is finite.

[8] Large dynamic factor models have been applied successfully to the analysis of big panels of macroeconomic and financial time series. Early theoretical contributions are Forni and Reichlin (1998), Forni et al. (2000, 2005), Forni and Lippi (2001), Stock and Watson (2002b), Bai and Ng (2002, 2007), and Bai (2003). A partial list of early applications include forecasting in Stock and Watson (2002a,b), Boivin and Ng (2006), and D'Agostino and Giannone (2012), structural macroeconomic analysis in Bernanke and Boivin (2003), Bernanke et al. (2005), Favero et al. (2005), Eickmeier (2007), Forni et al. (2009), Forni and Gambetti (2010), as well as nowcasting and business cycle indicators in Forni and Lippi (2001), Cristadoro et al. (2005), Giannone et al. (2008), Altissimo et al. (2010)), and the analysis of financial markets in Ludvigson and Ng (2009).

where $\mathbf{A}_n(L)$ is a block-diagonal polynomial matrix of *finite* order, $\mathbf{R}_n$ is a $n \times q$ matrix, and the components $\boldsymbol{\eta}_t = \mathbf{A}_n(L)\mathbf{e}_{nt}$ are idiosyncratic. Clearly the autoregressive filter in (2.4) and the moving average filter in (2.2) are related by $\boldsymbol{\Lambda}_n(L) = \mathbf{A}_n^{-1}(L)\mathbf{R}_n$.[9] As this paper focuses on estimation of the number of common shocks, we rely on representation (2.2) and do not require rationality of the $\boldsymbol{\Lambda}_n(L)$, whereas, as indicated above, identification of the *structural* shocks and estimation of the corresponding IRFs can be carried out by estimating the VAR representation (2.4), along the lines of Forni et al. (2015, 2017) and Barigozzi et al. (2022).

2.2. *Estimators of the Number of Shocks in the Frequency-Domain*

Define the periodogram-smoothing estimator of the spectral density matrix of the $\mathbf{x}_{nt}$

$$\widehat{\boldsymbol{\Sigma}}_n(\omega_\ell) \equiv \frac{1}{2M_T+1} \sum_{j=-M_T}^{M_T} \widehat{\boldsymbol{\mathcal{I}}}_n\left(\omega_{\ell+j}\right) \tag{2.5}$$

at the Fourier frequency $-\pi < \omega_\ell \equiv 2\pi\ell/T \leq \pi$, assuming T even without loss of generality (see Brillinger (1981)[p.132]). Here $\{M_T\}$ is a sequence of positive integers and $\boldsymbol{\mathcal{I}}_n(\cdot)$ denotes the periodogram of the $\mathbf{x}_{nt}$,

$$\widehat{\boldsymbol{\mathcal{I}}}_n(\omega) \equiv \left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\mathbf{x}_{nt}e^{-\mathrm{i}\omega t}\right]\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\mathbf{x}'_{nt}e^{\mathrm{i}\omega t}\right], \tag{2.6}$$

where i in Roman font denotes the imaginary unit. Finally, recall that $\hat{\mu}_{nk}(\omega)$ denotes the $k-$th eigenvalue of $\widehat{\boldsymbol{\Sigma}}_n(\omega)$ in decreasing order of magnitude, with $1 \leq k \leq n$.

To construct our estimators, we need to evaluate $\widehat{\boldsymbol{\Sigma}}_n(\cdot)$ within any given frequency band $(\omega_{\underline{\ell}}, \omega_{\bar{\ell}})$, satisfying at minimum $-\pi \leq \omega_{\underline{\ell}} \leq \omega_{\bar{\ell}} < \pi$, yielding our DDR criterion[10]

$$\mathrm{DDR}_n^T(k) \equiv \frac{\sum_{\ell=\underline{\ell}}^{\bar{\ell}}\left(\hat{\mu}_{nk}(\omega_\ell) - \hat{\mu}_{n,k+1}(\omega_\ell)\right)}{\max\left(\sum_{\ell=\underline{\ell}}^{\bar{\ell}}\left(\hat{\mu}_{n,k+1}(\omega_\ell) - \hat{\mu}_{n,k+2}(\omega_\ell)\right), \sum_{\ell=\underline{\ell}}^{\bar{\ell}}\hat{\mu}_{n,m}(\omega_\ell)\right)}, \tag{2.7}$$

setting $m \equiv 2M_T + 1$, and the dynamic versions of Ahn and Horenstein (2013) two criteria,

$$\mathrm{DER}_n^T(k) \equiv \frac{\sum_{\ell=\underline{\ell}}^{\bar{\ell}}\hat{\mu}_{nk}(\omega_\ell)}{\sum_{\ell=\underline{\ell}}^{\bar{\ell}}\hat{\mu}_{n,k+1}(\omega_\ell)},$$

[9] The IRFs $\boldsymbol{\Lambda}_n(L)$ are rational functions when their $(i,j)$the entries satisfy $\lambda_{ij}(L) = \bar{\lambda}_{ij}(L)/\underline{\lambda}_{ij}(L)$, where $\bar{\lambda}_{ij}(L)$ and $\underline{\lambda}_{ij}(L)$ are finite-order polynomials in L. A simple example is $\bar{\lambda}_{ij}(L) = 1 + \bar{a}_{ij}L$ and $\underline{\lambda}_{ij}(L) = 1 - \underline{a}_{ij}L$ for some constants $\bar{a}_{ij}, \underline{a}_{ij}$.

[10] To avoid denominators very close to zero, we correct the denominator of the DDR criterion by taking, for each k, the maximum between $\sum_{\ell=\underline{\ell}}^{\bar{\ell}}\left(\hat{\mu}_{n,k+1}(\omega_\ell) - \hat{\mu}_{n,k+2}(\omega_\ell)\right)$ and $\sum_{\ell=\underline{\ell}}^{\bar{\ell}}\hat{\mu}_{n,m}(\omega_\ell)$. Notice that we allow the frequency band to reduce to a single frequency when $\omega_{\underline{\ell}} = \omega_{\bar{\ell}}$.

$$\mathrm{DGR}_n^T(k) \equiv \frac{\ln\left[V_n^T(k-1)/V_n^T(k)\right]}{\ln\left[V_n^T(k)/V_n^T(k+1)\right]} = \frac{\ln(1+\hat{\mu}_{nk}^*)}{\ln(1+\hat{\mu}_{n,k+1}^*)},$$

where $V_n^T(k) \equiv \sum_{\ell=k+1}^{n} \sum_{\ell=\underline{\ell}}^{\bar{\ell}} \hat{\mu}_{nl}(\omega_\ell), \hat{\mu}_{nk}^* \equiv \sum_{\ell=\underline{\ell}}^{\bar{\ell}} \hat{\mu}_{nk}(\omega_\ell)/V_n^T(k)$. Then, our estimators of the number k of common shocks are

$$\widehat{k}_{\mathrm{DDR}} \equiv \arg\max_{1\le k\le q_{\max}} \mathrm{DDR}_n^T(k), \tag{2.8}$$

$$\widehat{k}_{\mathrm{DER}} \equiv \arg\max_{1\le k\le q_{\max}} \mathrm{DER}_n^T(k), \tag{2.9}$$

$$\widehat{k}_{\mathrm{DGR}} \equiv \arg\max_{1\le k\le q_{\max}} \mathrm{DGR}_n^T(k). \tag{2.10}$$

where $q_{\max}$ is an upper bound for q chosen by the researcher.

THEOREM 1: *Suppose that Assumptions (1)- (5), listed in Appendix A.1, hold for some* $q \ge 1$. *Then,*

$$\lim_{n,T\to\infty} \Pr\left(\widehat{k}_{\mathrm{DDR}} = q\right) = 1, \tag{2.11}$$

$$\lim_{n,T\to\infty} \Pr\left(\widehat{k}_{\mathrm{DER}} = q\right) = 1, \tag{2.12}$$

$$\lim_{n,T\to\infty} \Pr\left(\widehat{k}_{\mathrm{DGR}} = q\right) = 1. \tag{2.13}$$

for any $q_{\max} \in (q, (2M_T+1) - q]$.

Proof. See Appendix A.3.

Notice that the computation of DDR, DER, and DGR does not require discretionary preliminary choices of tuning parameters by the researcher, apart from the size of the spectral window (the bandwidth) used in the estimation of the spectral density matrix (this cannot be avoided when using nonparametric estimators). This is an advantage over both HL and O, which, besides the window size, require making additional choices, that may influence the results.[11] Clearly, the DDR, DER, and DGR estimators, and the true number of common shocks q refer to the chosen band $(\omega_{\underline{\ell}}, \omega_{\bar{\ell}})$ but we do not write this explicitly to avoid cumbersome notation.

2.3. *Variance Decomposition in the Frequency Domain*

In the empirical section, besides estimating the number of macroeconomic shocks, we shall quantify their contribution to the variance of the observable variables at each frequency, making use of the empirical eigenvalues defined above, along with the corresponding empirical eigenvectors. Here we introduce the relevant estimators.

First, we want to estimate the amount of total variance explained by the common shocks for each variable and each frequency. Orthogonality of the common and the idiosyncratic components at all leads and lags imply that the spectral density matrix of the n observed variates

[11] More specifically, HL requires choosing the functional form for the penalty term and a grid n_j, T_j, $j = 1, \ldots, J$, which is needed for its calibration. O is a sequential-testing procedure and therefore requires choosing the significance level; furthermore, it requires choosing a grid of frequencies and a criterion for weighting the potentially different results obtained for different frequencies.

fulfils the decomposition

$$\mathbf{\Sigma}_n(\omega) = \mathbf{\Sigma}_n^{\chi}(\omega) + \mathbf{\Sigma}_n^{e}(\omega) \text{ for every } -\pi \leq \omega < \pi,$$

where $\mathbf{\Sigma}_n^{\chi}(\omega)$ and $\mathbf{\Sigma}_n^{e}(\omega)$ are the spectral density matrices of the common and idiosyncratic components, respectively. Now, under suitable conditions, the entries of $\mathbf{\Sigma}_n^{\chi}(\omega)$ can be estimated consistently by

$$\hat{\mathbf{\Sigma}}_n^{\chi}(\omega) = \sum_{k=1}^{q} \hat{\mathbf{v}}_{nk}(\omega)\hat{\mu}_{nk}(\omega)\hat{\mathbf{v}}'_{nk}(-\omega), \tag{2.14}$$

where $\hat{\mathbf{v}}_{nk}(\omega)$ is the eigenvector corresponding to the eigenvalue $\hat{\mu}_{nk}(\omega)$ (see Forni et al., 2017, Proposition 7). The average of (2.14) over all frequencies provides an estimate of the covariance matrix of the common components, whereas the average over the frequencies in the band $\omega_{\underline{\ell}}$ to $\omega_{\bar{\ell}}$ quantifies the portion of such covariance matrix associated with that band.

In order to quantify the variance explained by each shock for a given identification matrix, full estimation of the GDFM is needed. Going back to equations (2.2) and (2.3), we see that

$$\mathbf{\Sigma}_n^{\chi}(\omega) = \frac{1}{2\pi}\mathbf{\Lambda}(e^{-i\omega})\mathbf{\Lambda}'(e^{i\omega}) = \frac{1}{2\pi}\sum_{k=1}^{q}\mathbf{\Lambda}(e^{-i\omega})\mathbf{q}_k\mathbf{q}_k'\mathbf{\Lambda}'(e^{i\omega}), \tag{2.15}$$

where $\mathbf{q}_k$ denotes the k-th column of the matrix $\mathbf{Q}$. The above equation shows that the summation on the right-hand side is independent of $\mathbf{Q}$, whereas the single terms of the summation depend on $\mathbf{Q}$.

Notice however that $\mathbf{\Sigma}_n^{\chi}(\omega)$ satisfies the so-called spectral decomposition

$$\mathbf{\Sigma}_n^{\chi}(\omega) = \sum_{k=1}^{q}\mathbf{v}_{nk}^{\chi}(\omega)\mu_{nk}^{\chi}(\omega)\mathbf{v}_{nk}^{\chi\prime}(-\omega), \tag{2.16}$$

where $\mu_{nk}^{\chi}(\omega)$ and $\mathbf{v}_{nk}^{\chi}(\omega)$ denote, respectively, the population eigenvalues and eigevectors of $\mathbf{\Sigma}_n^{\chi}(\omega)$. Comparing (2.15) and (2.16) shows that the spectral decomposition provides the variance decomposition for one of the many possible identifications of the common shocks. With the spectral decomposition identification, the first identified shock maximizes the sum of the explained variances of all variables (Brillinger, 1981). The decomposition in (2.16) can be estimated consistently by formula (2.14) without estimating the VAR (2.4) and thus without imposing explicit identification restrictions (see Forni et al. (2017), Lemmas 2 and 4).

3. MONTE CARLO SIMULATIONS

In this section we provide a summary of the thoughtful Monte Carlo experiments described in Section B.5 of the Online Appendix in detail, to which we refer, including tables reporting the numerical results.

As anticipated, we evaluate the finite-sample performance of our estimators, DDR, DER, and DGR, and we compare them with that of HL and O. We do not consider the methods proposed by Bai and Ng (2007) and Amengual and Watson (2007) since they assume a factor model that can be written in the static form, which is a restriction that we do not impose here.

We consider several data generating processes (DGPs) to demonstrate the validity of our methodology, with the true q ranging from 2 to 6 depending on the cases. We run four experiments. In the first three, we evaluate the competing estimators on the whole interval $[0, \pi]$. The

fourth experiment is devoted to checking the ability of DDR to detect a rank reduction in the spectral density matrix at specific frequencies of interest (see below).

Experiments one and two are based on DGPs already proposed in the econometric literature, namely in Hallin and Liska (2007) and in Onatski (2009). Experiment three is our own DGP. In experiment four we focus on DDR. The DGP is such that the spectral density matrix of the variables has reduced rank at frequency zero. One of the shocks is a supply shock and has permanent effects on several variables. The other is a demand shocks having transitory effects on all variables, which are supposed to be real activity variables. DDR should detect two shocks when evaluated on large frequency bands and just one shock when evaluated at frequency zero.

To compute DDR, DER, and DGR, we use the periodogram smoothing estimator (2.5) with the bandwidth parameter $M_T = [0.75\sqrt{T}]$, and we take the average of the eigenvalues evaluated in the frequency grid $\omega_\ell = 2\pi\ell/T$, $\ell = 1, \dots, T-1$. With regard to HL, we use the log information criterion $IC^T_{2;n}$ with penalty $p_1(n,T)$ and the Bartlett lag window with parameter $M_T = [0.75\sqrt{T}]$, which yield the best performance in Hallin and Liska (2007) simulations. The method requires evaluation of the loss function over a grid n_j, T_j, $j = 1, \dots, J$; we stick to the one proposed by the authors (i.e., $n_j = n - 10j$, $T_j = T - 10j$, $j = 0, 1, 2, 3$). When dealing with O, we use the procedure described in Section 5.3 of Onatski (2009). We find that the results are sensitive to the choice of the parameter m. For the second experiment, we stick to Onatski's choice, which is very effective ($m = 30$ for $(n,T) = (70,70)$, $m = 40$ for $(n,T) = (100,120)$, $m = 65$ for $(n,T) = (150,500)$). For the first DGP, we use $m = 15$; for the third experiment, we use $m = 15, 20, 30$ for $T = 80, 240, 480$, respectively. These values produce better results than the larger ones suggested in Onatski (2009). For all the experiments and all estimators, we set $q_{\max} = 8$ and we generate 500 artificial datasets. We evaluate the results in terms of the percentage of times we find the correct number of shocks.

The results are the following. For experiment one, DDR and HL perform similarly and dominate the other estimators. For experiment two, with MA loadings, O ranks first and does slightly better than DDR; with AR loadings, DDR dominates all other estimators. For experiment three, in the version with small idiosyncratic components, DDR and DGR perform similarly and dominate the other estimators; in the version with large idiosyncratic components, DDR outperforms the other estimators. Finally, in experiment four, we show that with DDR, when evaluated at a single frequency (or small frequency bands around this frequency) in which the spectral density matrix of the variables has reduced rank $q-1$, the number of shocks at this frequency is $q-1$. Overall, we conclude that DDR is an excellent alternative to existing criteria in finite samples.

4. APPLICATIONS

We illustrate the substance of our methodology with two applications. First, an empirical analysis, based on a large-dimensional dataset, inferring the number of shocks driving the U.S. economy. Second, a simulation analysis demonstrating how our methodology continues to be valid even when applied to medium-scaled DSGE models.

4.1. *Dissecting the U.S. Economy*

How many shocks drive the economy? And, more specifically, how many shocks affect the macroeconomic variables in the long run, and how many, instead, are the ones driving the business cycle? Finally, do demand or supply shocks dominate the economy?

The scant available evidence on the number of shocks is mixed. Sargent and Sims (1977), using a small-dimensional dynamic factor model, find that two shocks fit U.S. macroeconomic

data reasonably well. Giannone et al. (2005) argue informally in favor of two shocks, based on the explained variances of the principal-component series of a large factor model. Bai and Ng (2007), using their information criteria, find seven static factors and four shocks. Amengual and Watson (2007) find seven static factors and seven shocks. Hallin and Liska (2007) do not find a clear-cut result: four shocks, but perhaps only one. Onatski (2009) finds that his proposed test cannot reject the null of two shocks versus the alternative, which assumes between three and seven shocks.

For our empirical analysis, we use the U.S. quarterly macroeconomic dataset recently developed by McCracken and Ng (2020). Of this dataset, we consider the $n = 216$ series starting from the first quarter of 1960. The final date of the sample is the first quarter of 2020. As for the transformations, we deviate from those McCracken and Ng (2020) suggest for the interest rates, which we take in levels rather than in differences; furthermore, we take prices and other nominal variables in log-differences, rather than in double differences of the logs. The reason is that we want to avoid a possible over differentiation, which enhances the high frequencies and is of little interest for our analysis. The complete list of variables and transformations is in Appendix B.9. After the transformations, the number of observations over time is $T = 240$.

We find a clear-cut result: *two* common shocks drive the U.S. economy. These two shocks explain the bulk of the variance in the main macroeconomic variables, both at business-cycle frequencies and in the long term. We also find that the first dynamic factor behaves like a demand shock and the second one like a supply shock.

Finding (i): The Number of Shocks Driving the U.S. Economy We consider the entire sample and nine sub-samples: the five 40-year sub-samples 1960Q2-2000Q1, 1965Q2-2005Q1, 1970Q2-2010Q1, 1975Q2-2015Q1, and 1980Q2-2020Q1, and the four 30-year sub-samples 1960Q2-1990Q1, 1970Q2-2000Q1 , 1980Q2-2010Q1, and 1990Q2-2020Q1. To estimate q, we compute DDR on three frequency bands: the entire $[0, \pi]$ band, the $[0, 2\pi/6]$ band, which excludes fluctuations of less than 18 months and is of little interest for macroeconomic analysis, and the $[2\pi/32,\ 2\pi/6]$ cyclical band, which includes waves ranging from 18 months to eight years. The DDR estimator for the $[0, 2\pi/6]$ band is named DDRa; the one relating to the cyclical band is named DDRbc. For comparison, we consider also DER, DGR, HL, and O (only on the whole $[0, \pi]$ band).[12] For DER, DGR, and DDR we set the bandwidth parameter $M_T = [0.75\sqrt{T}]$ for the whole sample and $M_T = [\sqrt{T}]$ for all subsamples, according to the results of the simulation exercise reported in Appendix B.8. For all estimators, we set $q_{\max} = 8$.

Results are reported in Table I. Our main estimator DDR selects *two* shocks; this finding is reasonably consistent across sub-samples. The same holds for DDRa, which excludes the short-run frequencies, and DDRbc. DER and DGR behave similarly, except for the sub-samples 1960Q2-2000Q1 and 1965Q2-2005Q1, for which they select one dynamic factor only. By contrast, HL selects five factors, whereas O is in favor of three factors. As for the sub-samples, HL oscillates between two and four factors, with a prevalence of four factors, and O varies between two and five, with a prevalence of two factors. Overall, DDR, DDRa, and DDRbc are more parsimonious than HL and O and more consistent across sub-samples.

Why do DDR and HL differ so much? One possible explanation is that there are two large factors and three smaller factors and that the latter are elusive to DDR but not to HL. This interpretation contrasts, however, with the simulation of Table B.I, reported in the Online Appendix B.5, where there are factors of different variance and DDR has a performance similar to that of HL. An alternative explanation is that HL overestimates the number of factors, as in the simulation of Table B.IV in the Onine Appendix B.5.

[12] HL is calculated as in the simulations, with $M_T = [0.75\sqrt{T}]$ (see the online Appendix B for details). For O, we use $m = 20$ for the whole sample and for the 40-year sub-samples, and we use $m = 15$ for the 30-year sub-samples.

TABLE I

Sample span	DDR	DDRa	DDRbc	DER	DGR	HL	O
1960Q2-2020Q1	2	2	2	2	2	5	3
1960Q2-2000Q1	2	2	2	1	1	4	2
1965Q2-2005Q1	2	2	2	1	1	4	2
1970Q2-2010Q1	2	2	2	2	2	4	2
1975Q2-2015Q1	2	2	2	2	2	4	5
1980Q2-2020Q1	2	2	2	2	2	4	2
1960Q2-1990Q1	1	2	2	1	1	3	5
1970Q2-2000Q1	1	1	1	1	1	4	4
1980Q2-2010Q1	2	3	3	2	2	2	2
1990Q2-2020Q1	2	2	2	2	2	3	3

Note: Number of factors detected by the competing criteria for the whole sample and nine sub-samples. DDR: Dynamic Difference Ratio Estimator. DDRa: Dynamic Difference Ratio Estimator evaluated on the $[0, 2\pi/6]$ frequency band. DDRbc: Dynamic Difference Ratio Estimator evaluated on the cyclical band $[2\pi/32, 2\pi/6]$. DER: Dynamic Eigenvalue Ratio estimator. DGR: Dynamic Growth Ratio estimator. HL: Hallin and Liška (2007) estimator. O: Onatski (2009) estimator.

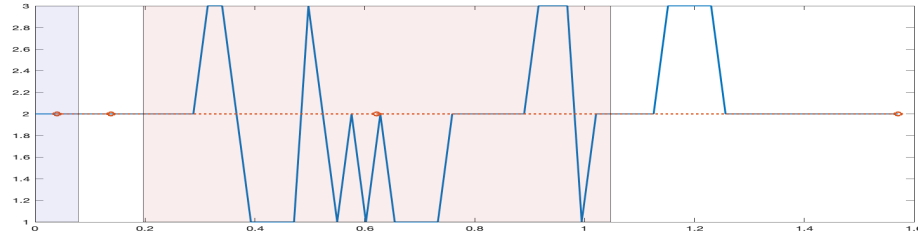

FIGURE 2.—Number of factors estimated by DDR when evaluated at each frequency (blue solid line) and on four frequency bands (red dotted line with circles).

For the subsequent empirical analysis, we focus solely on our main estimator, DDR. Figure 2 shows the number of estimated factors by DDR for each single Fourier frequency in the interval $[0, \pi/4]$ and each one of the frequency bands $[0, 2\pi/80]$ (long-run), $[2\pi/80, 2\pi/32]$ (long cycles), $[2\pi/32, 2\pi/6]$ (business cycle), and $[2\pi/6, \pi]$ (short-run). At frequency-zero DDR selects two factors; the estimator is somewhat unstable when evaluated at single frequencies, fluctuating between one and three factors. Its value, however, stabilizes at two when averaging the eigenvalues on the four bands above.

Our conclusion is that there are two major shocks driving the U.S. economy. This is true both in the long-run and at business cycle frequencies. The presence of other shocks cannot be ruled out, but these are not large or pervasive enough to be distinguished from idiosyncratic factors on the basis of existing data. The result of two shocks is in line with Sargent and Sims (1977), Giannone et al. (2005), and Onatski (2009), as cited above.

Finding (ii): Common and Idiosyncratic Variance In this section we use equation (2.14) to estimate, for seven key macroeconomic variables, the variance explained by the first and second dynamic factor along with the variance explained by the subsequent three factors (from the third to the fifth factor). The aim is to quantify the size of the explained variance for these variables when retaining only two factors and the size of the variance that is not accounted for with respect to the choice $q = 5$ suggested by HL. Table II shows the result. From the table one can see that on the bands of macroeconomic interest, namely the long run, the long

waves, and the business cycle, two factors capture about 85% of GDP growth fluctuations, 70% of consumption, 75%-85% of investment, 80%-90% of the unemployment rate variation, 85%-90% of hours worked, 85%-90% of inflation, and 80% of the federal funds rate. The variance not accounted for by selecting $q = 2$ instead of $q = 5$ is not negligible, particularly for consumption and the interest rate. Nevertheless, we conclude that two factors are enough to capture the bulk of the variance in the main macroeconomic aggregates on the frequency bands of main macroeconomic interest.

TABLE II

	PC series	All freq.	Long run > 20 years	Long cycles 8-20 years	Medium cycles 4-8 years	Short cycles 1.5-4 years	Bus. cycle 1.5-8 years
GDP	first 2	73.1	**84.6**	85.7	84.5	82.3	**83.1**
	next 3	16.4	**8.1**	5.9	7.1	9.3	**8.6**
Cons.	first 2	56.8	**70.9**	73.8	72.6	66.7	**68.7**
	next 3	16.6	**15.1**	13.3	14.9	11.1	**12.9**
Inv.	first 2	73.1	**74.7**	78.6	87.1	85.3	**85.7**
	next 3	8.8	**16.7**	11.2	4.6	3.0	**3.7**
U rate	ffirst 2	80.7	**80.3**	79.1	86.7	91.7	**89.7**
	next 3	8.5	**14.0**	15.4	8.9	2.3	**4.7**
Hours	first 2	77.3	**84.7**	85.3	91.4	91.2	**91.1**
	next 3	8.6	**10.0**	10.3	5.5	2.0	**3.4**
Inflation	first 2	84.5	**91.6**	91.1	91.3	63.2	**86.0**
	next 3	7.9	**5.7**	6.2	6.1	18.3	**8.5**
FFR	first 2	78.6	**78.9**	79.2	81.8	66.5	**79.7**
	next 3	16.1	**18.1**	17.2	13.6	20.1	**14.1**

Note: Percentage of variance explained by the first two dynamic factors and the following three for a few selected variables, by frequency band. All frequencies: $[0, \pi]$; Long run: $[0, 2\pi/80)$; Long cycles: $[2\pi/80, 2\pi/32)$; Medium cycles: $[2\pi/32, 2\pi/16)$; Short cycles: $[2\pi/16, 2\pi/6)$; Business cycle: $[2\pi/32, 2\pi/6)$.

Finding (iii): Cyclical Shock vs. Long-Run Shock We now investigate how much the first and second shock explain the variance in the above variables for each frequency band. To this end, we use again equation (2.14). As already observed at the end of Section 2.3, this decomposition is not obtained by imposing an economic identification criterion, as this would require estimation of (2.4) along with a choice for the identification matrix $\mathbf{Q}$. Despite this, it is possible to assign a precise economic meaning to the common shocks resulting from this statistical identification and to derive conclusions of sound economic interest.

The explained variances are shown in Table III. The first shock accounts for almost nothing of the variance in GDP in the long run and, more generally, over the long cycles band, which is instead explained by the second shock. The same result applies to all real activity variables. We therefore find ourselves, without having imposed it, in front of an identification *à la* Blanchard and Quah (1989): the first factor is a *transitory* shock, while the second one is a *permanent* shock. It is very much tempting to interpret the transitory shock as a demand shock and the long-run shock as a supply shock. To confirm this interpretation, we look at the covariances of GDP growth and inflation changes induced by the two shocks and find that, in fact, such covariance is *positive* for the transitory shock, which therefore has the features of a demand shock, and it is *negative* for the permanent shock, which can then be regarded as a supply shock (see the lower-right panel of Figure 3 below).

TABLE III

	Shocks	All freq.	Long run > 20 years	Long cycles 8-20 years	Medium cycles 4-8 years	Short cycles 1.5-4 years	Bus. cycle 1.5-8 years
GDP	1st	47.2	**1.4**	3.3	34.1	77.4	**62.2**
	2nd	25.9	**83.2**	82.5	50.3	4.9	**20.8**
Cons.	1st	29.5	**1.2**	1.9	24.8	57.0	**42.3**
	2nd	27.3	**69.7**	71.9	47.8	9.7	**26.4**
Inv.	1st	43.2	**4.0**	5.2	35.7	79.4	**63.1**
	2nd	29.9	**70.7**	73.3	51.4	5.9	**22.6**
U rate	1st	51.8	**9.9**	11.8	40.6	87.6	**70.8**
	2nd	28.9	**70.4**	67.3	46.1	4.1	**18.9**
Hours	1st	44.2	**0.8**	4.8	37.3	88.3	**68.5**
	2nd	33.1	**83.9**	80.4	54.0	3.0	**22.6**
Inflation	1st	78.3	**91.3**	89.7	81.8	30.9	**72.8**
	2nd	6.2	**0.3**	1.4	9.5	32.3	**13.2**
FFR	1st	75.2	**78.7**	78.6	76.4	52.7	**72.9**
	2nd	3.4	**0.2**	0.6	5.4	13.8	**6.8**

Note: Percentage of variance explained by the first and the second factor for a few selected variables, by frequency band. All frequencies: $[0, \pi]$; Long run: $[0, 2\pi/80)$; Long cycles: $[2\pi/80, 2\pi/32)$; Medium cycles: $[2\pi/32, 2\pi/16)$; Short cycles: $[2\pi/16, 2\pi/6)$; Business cycle: $[2\pi/32, 2\pi/6)$.

A caveat is necessary. Our shocks can in principle be, and probably are, mixtures of different underlying structural shocks, such as news technology shocks, uncertainty shocks, financial shocks, or policy shocks. Despite this, they may still be economically meaningful, if we can group the true structural shocks into the broader categories of demand and supply. This is possible if shocks belonging to the same group have similar effects on most macroeconomic variables.[13]

In the last column of the table we report the explained variances at business cycle frequencies. The demand shock is the most important cyclical shock for real activity variables. It accounts for 62% of GDP growth fluctuations, 42% of consumption, 63% of investment, 71% of unemployment, and 68% of hours worked. The contribution of the permanent, supply shock is still sizeable, particularly for consumption, but much smaller; for real activity variables, it varies between 19% (unemployment) and 26% (consumption).

Figure 3 illustrates the same points by showing the spectral density of the seven variables above, along with the spectra of the common components driven by the two shocks. The upper-left panel refers to GDP growth. The supply shock accounts for the long run and the long-medium cycles but explains almost nothing of short cycles and short-run frequencies as if it were cut by a low-pass filter canceling waves of periodicity shorter than five years. By contrast, the demand shock explains almost all cycles of four years or fewer and almost nothing of longer cycles. A similar result holds for the other variables related to real activity: consumption, investment, unemployment, and hours worked.

Demand shocks, both at business cycle frequencies and in the long run, almost exclusively explain the interest rate. The demand shock is therefore closely related to monetary policy, whereas the supply shocks is not. A possible explanation is that an expansionary supply shock reduces inflation, so that the Fed does not react to it.

[13]For instance, technology shocks can be interpreted as supply shocks, whereas financial and uncertainty shocks can be interpreted as demand shocks.

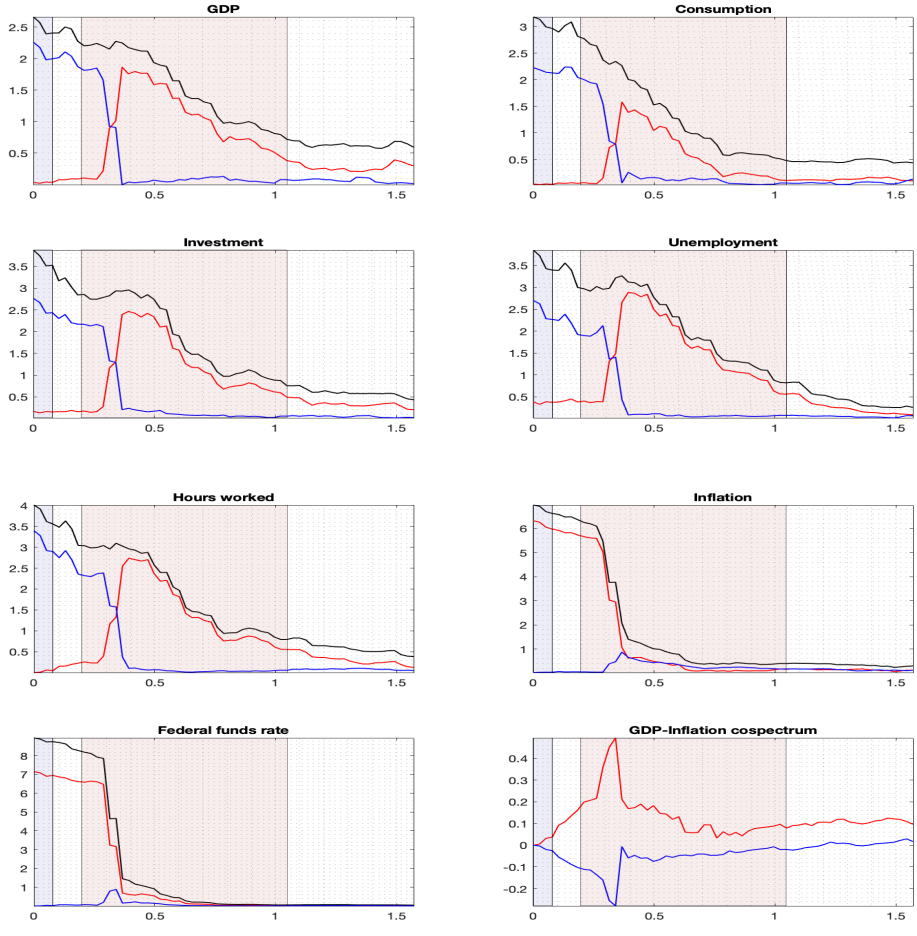

FIGURE 3.—The estimated spectral density functions of seven variables (black line) and the components are driven by the first factor (red line) and the second factor (blue line). The variables are: GDP growth, consumption growth, investment growth, unemployment rate changes, hours worked changes, GDP deflator inflation, and federal funds rate. Bottom-right panel: co-spectrum of GDP growth and inflation changes produced by the first factor (red line) and the second factor (blue line). Lilac shadowed area: long-run frequencies; pink shadowed area: business cycle frequencies.

Demand shocks explain most of inflation fluctuations, especially in the long run. This result, coupled with the previous one, explains why our criterion detects two shocks around frequency zero. If we want to explain long-run fluctuations of real activity variables, the supply shock is sufficient; but if we want to explain long-run fluctuations of all variables, including nominal variables and interest rates, we need two shocks.

The bottom-right panel of Figure 3 shows the co-spectra of GDP growth and inflation changes relative to the transitory shock (red line) and the permanent shock (blue line). As an-

ticipated, the transitory shock induces a positive covariance between GDP growth and inflation changes, whereas the opposite is true for the permanent shock.[14]

The above results are in line with Blanchard and Quah (1989) and King et al. (1991) and in sharp contrast with both the RBC model and the idea that news technology shocks (i.e., shocks anticipating future technology improvements) explain the bulk of business cycle fluctuations (Beaudry and Portier, 2006).[15] This confirms the ACD finding that long-run shock does not drive the bulk of business cycle fluctuations. On the other hand, the ACD hypothesis that there is just one non inflationary demand shock affecting real activity variables at business cycle frequencies is not in line with our results, mainly because the demand shock is closely related to inflation. The implication for macroeconomic modelling is that our analysis does not rule out standard demand shocks of the textbook type.

In Appendix B.2 we discuss in more detail the relation between our results and ACD's findings.

4.2. *Dynamic Principal Components for DSGE Models*

Strong analogies exist between certain DSGE models and the GDFM, particularly for the DSGE models of Angeletos et al. (2018) (ACD* henceforth) and of Justiniano et al. (2010) (JPT henceforth). In fact, these DSGE models and GDFMs admit a finite-order VAR representation with a singular covariance matrix of the error term, implied when the number of shocks is smaller than the number n of observables.[16] In view of this commonality, we develop a two-fold analysis, exploring how our methodology (i) is able to dissect (by frequency) the number of structural shocks driving DSGE models by means of a Monte Carlo experiment and (ii) provides a frequency domain variance decomposition of these DSGE structural shocks with the data used in Section 4.1. Part (i) is presented below and part (ii) in Section B.3 of the Online Appendix.

4.2.1. *Estimation of Number of Shocks in DSGE Models*

The aim of this experiment is twofold: test our estimators with a DGP from a macroeconomic model, and verify whether our criteria capture a relatively large value of q. In fact, the ACD* model has $q = 8$ shocks and $n_1 = 10$ variables; the JPT has $q = 7$ shocks and $n_1 = 11$ variables. This is developed in a Monte Carlo setting, with the DGPs for these variables reported in Section B.4 of the Online Appendix. We evaluate the results in terms of the percentage of times we find the correct number of shocks.

Table IV reports results of this experiment based on the DGP given by the JPT and ACD* models. HL performs poorly for both models and most q, n, T, s^2 configurations, with the noticeable exceptions of the ACD* model, large errors ($s^2 = 0.2$), case $q = 4$, $n = 60$, $T = 120$. [17] For almost all configurations DGR and DDR dominate DER, with no clear ranking between the DGR and DDR. When measurement errors are relatively small (1/11 of total variance, about

[14] We consider inflation changes in place of inflation because the latter exhibits a large negative co-spectrum with GDP growth for all of the first five dynamic factors, owing to the 1970s and the early 1980s, characterized by low growth and high inflation.

[15] Barsky and Sims (2011) and Forni et al. (2014) question this idea.

[16] For the DSGE models this is given by their reduced form, obtained by the standard practice of log-linearizing the models around their steady state (see (B.2)-(B.3) and (B.4)-(B.5) in Section B.4) of the Online Appendix whereas for GDFMs the VAR representation (see (2.4)) holds under suitable assumptions on the IRFs (see Forni et al. (2015)).

[17] One of the experiments of Appendix B.5, reported in Table D (variation of the third experiment), shows that HL performs very well for intermediate values of the idiosyncratic variance (within the HL calibration design) but less so for either very small or very large values of the idiosyncratic variance. This is the case in the current experiment.

TABLE IV

JPT model			$s^2 = 0.1$ (small errors)				$s^2 = 0.2$ (large errors)			
q	n	T	HL	DER	DGR	DDR	HL	DER	DGR	DDR
4	60	120	6.0	68.2	**93.8**	83.6	67.2	31.4	77.0	**84.2**
4	120	240	0.0	99.8	**100.0**	98.8	0.0	94.2	100.0	**98.6**
7	60	120	36.0	41.0	**91.2**	75.0	3.4	1.4	35.6	**50.8**
7	120	240	0.0	99.0	**100.0**	99.8	3.0	77.4	**99.4**	99.2
ACD model*			$s^2 = 0.1$ (small errors)				$s^2 = 0.2$ (large errors)			
q	n	T	HL	DER	DGR	DDR	HL	DER	DGR	DDR
4	60	120	9.0	78.0	**98.6**	93.0	91.8	26.8	76.4	**94.0**
4	120	240	0.0	**100.0**	**100.0**	99.2	0.0	98.4	**100.0**	**100.0**
8	60	120	47.0	13.4	79.4	**84.6**	3.0	0.0	8.8	**33.4**
8	120	240	0.0	99.8	**100.0**	**100.0**	36.0	38.8	98.4	**100.0**

Note: Experiment described in Section 4.2.1. Percentage of correct outcomes over 500 replications. HL: Hallin and Liška (2007) estimator, DER: Dynamic Eigenvalue Ratio estimator, DGR: Dynamic Growth Ratio estimator; DDR: Dynamic Difference Ratio estimator. Boldface numbers denote the estimator that performs best in each row.

9%) both criteria capture the correct number of shocks in most cases, even if there are seven or eight shocks. With $T = 240$, as is the case for the empirical exercise below, both criteria detect the correct number of shocks even when the errors are large (1/6 of total variance, about 17%). We omit results O because of the modest performance.

Summarizing, DDR and DGR detect correctly the number of shocks driving DSGE models even if the number of shocks is relatively large (seven and eight in the ACD* and JPT models, respectively).

5. SUMMARY AND CONCLUSIONS

Determining the number of different sources of fluctuations affecting the economy is a fundamental quest for macroeconomic modelling and the business cycle debate. However, existing methods to estimate the number of shocks are not entirely satisfactory, mainly because they do not focus on specific frequencies and cycles. In this paper, we study new criteria for single frequencies and selected frequency bands of interest, adapting Ahn and Horenstein (2013) criteria to the GDFM setting. Our estimators do not require pre-selecting penalty functions and nuisance parameters.

We establish the consistency of these estimators and evaluate their small-sample performance with Monte Carlo exercises. We find that one of them, the Dynamic Difference Ratio (DDR), dominates the other two in most simulations and performs better than, or comparably with, existing criteria. Noticeably, our estimators work even when applied to data stemming from DSGE models where the number of shocks is relatively large compared with the number of observables.

We apply DDR to the FRED-QD database and find a clear-cut result: two major shocks drive the U.S. economy; we dub them, following Angeletos et al. (2020), the main business cycle shocks. They account for the bulk of the variance in the main macroeconomic aggregates at both business-cycle and long-run frequencies. It turns out that the statistical identification

based on the eigenvalues leads a precise economic interpretation: the first factor appears to be a demand shock and the second factor a supply shock. The demand shock explains most of the business cycle fluctuations in real activity variables, as well as inflation and interest rate volatility.

REFERENCES

AHN, S. C. AND A. R. HORENSTEIN (2013): "Eigenvalue ratio test for the number of factors," *Econometrica*, 81, 1203–1227. [3, 5, 8, 18, 21]

ALTISSIMO, F., R. CRISTADORO, M. FORNI, AND G. LIPPI, M.AND VERONESE (2010): "New Eurocoin: Tracking Economic Growth in Real Time," *The Review of Economics and Statistics*, 92, 1024–1034. [7]

AMENGUAL, D. AND M. W. WATSON (2007): "Consistent estimation of the number of dynamic factors in a large N and T panel," *Journal of Business & Economic Statistics*, 25, 91–96. [6, 10, 12]

ANGELETOS, G., F. COLLARD, AND H. DELLAS (2020): "Business-Cycle Anatomy," *American Economic Review*, 110, 3030–3070. [1, 5, 6, 18, 2]

ANGELETOS, G. M., F. COLLARD, AND H. DELLAS (2018): "Quantifying confidence," *Econometrica*, 86, 1689–1726. [5, 6, 17, 7]

BAI, J. (2003): "Inferential theory for factor models of large dimensions," *Econometrica*, 71, 135–171. [7]

BAI, J. AND S. NG (2002): "Determining the number of factors in approximate factor models," *Econometrica*, 70, 191–221. [7]

——— (2007): "Determining the Number of Primitive Shocks in Factor Models," *Journal of Business & Economic Statistics*, 25, 52–60. [6, 7, 10, 12]

BAI, Z. AND J. W. SILVERSTEIN (2010): *Spectral Analysis of Large Dimensional Random Matrices*, vol. 20, Springer: New York, NY. [2]

BAI, Z. D. AND Y. Q. YIN (1993): "Limit of the smallest eigenvalue of a large dimensional sample covariance matrix," *The Annals of Probability*, 21, 1275–1294. [5, 2]

BARIGOZZI, M., M. HALLIN, M. LUCIANI, AND P. ZAFFARONI (2022): "Inferential Theory for Generalized Dynamic Factor Models," *Mimeo: Imperial College London*. [4, 8]

BARSKY, R. AND E. SIMS (2011): "News shocks and business cycles," *Journal of Monetary Economics*, 58, 273–289. [17]

BEAUDRY, P., D. GALIZIA, AND F. PORTIER (2020): "Putting the cycle back into business cycle analysis," *American Economic Review*, 110, 1–47. [6]

BEAUDRY, P. AND F. PORTIER (2006): "Stock prices, news, and economic fluctuations," *American Economic Review*, 96, 1293–1307. [5, 6, 17]

BERNANKE, B. AND J. BOIVIN (2003): "Monetary policy in a data-rich environment," *Journal of Monetary Economics*, 50, 525–546. [7]

BERNANKE, B., J. BOIVIN, AND P. ELIASZ (2005): "Measuring the effects of monetary policy: A factor-augmented vector autoregressive (FAVAR) approach," *The Quarterly Journal of Economics*, 120, 387–422. [7]

BERNSTEIN, D. S. (2009): *Matrix Mathematics: Theory, Facts, and Formulas*, Princeton University Press: Princeton (NJ). [1]

BLANCHARD, O. AND D. QUAH (1989): "The dynamic effects of aggregate demand and supply disturbances," *American Economic Review*, 79, 655–673. [6, 14, 17]

BLOOM, N. (2009): "The Impact of Uncertainty Shocks," *Econometrica*, 77, 623–685. [6]

BOIVIN, J. AND S. NG (2006): "Are more data always better for factor analysis?" *Journal of Econometrics*, 132, 169–194. [7]

BRILLINGER, D. (1981): *Time Series: Data Analysis and Theory*, SIAM: Philadelphia. [4, 8, 10, 3]

CHAMBERLAIN, G. AND M. ROTHSCHILD (1983): "Arbitrage, Factor Structure, and Mean-Variance Analysis on Large Asset Markets," *Econometrica*, 51, 1281–304. [3]

CRISTADORO, R., M. FORNI, L. REICHLIN, AND G. VERONESE (2005): "A core inflation indicator for the euro area," *Journal of Money, Credit and Banking*, 37, 539–60. [7]

D'AGOSTINO, A. AND D. GIANNONE (2012): "Comparing alternative predictors based on large-panel factor models," *Oxford Bulletin of Economics and Statistics*, 74, 306–326. [7]

EICKMEIER, S. (2007): "Business cycle transmission from the U.S. to Germany–A structural factor approach," *European Economic Review*, 51, 521–551. [7]

FAVERO, C. A., M. MARCELLINO, AND F. NEGLIA (2005): "Principal components at work: The empirical analysis of monetary policy with large data sets," *Journal of Applied Econometrics*, 20, 603–620. [7]

FORNI, M. AND L. GAMBETTI (2010): "The dynamic effects of monetary policy: A structural factor model approach," *Journal of Monetary Economics*, 57, 203–216. [7]

FORNI, M., L. GAMBETTI, A. GRANESE, L. SALA, AND S. SOCCORSI (2022): "Identifying Macroeconomic Shocks in the Frequency Domain," *Mimeo: U of Modena*. [6]
FORNI, M., L. GAMBETTI, AND L. SALA (2014): "No news in business cycles," *The Economic Journal*, 124, 1168–1191. [17]
FORNI, M., D. GIANNONE, M. LIPPI, AND L. REICHLIN (2009): "Opening the black box: Structural factor models with large cross-sections," *Econometric Theory*, 25, 1319–1347. [7]
FORNI, M., M. HALLIN, M. LIPPI, AND L. REICHLIN (2000): "The generalized dynamic-factor model: Identification and estimation," *The Review of Economics and Statistics*, 82, 540–554. [3, 7]
——— (2005): "The Generalized Dynamic Factor Model: One-Sided Estimation and Forecasting," *Journal of the American Statistical Association*, 100, 830–840. [7]
FORNI, M., M. HALLIN, M. LIPPI, AND P. ZAFFARONI (2015): "Dynamic factor models with infinite-dimensional factor spaces: One-sided representations," *Journal of Econometrics*, 185, 359–371. [4, 7, 8, 17]
——— (2017): "Dynamic factor models with infinite-dimensional factor space: Asymptotic analysis," *Journal of Econometrics*, 199, 74–92. [4, 7, 8, 10]
FORNI, M. AND M. LIPPI (2001): "The generalized dynamic factor model: Representation theory," *Econometric Theory*, 17, 1113–1141. [3, 7]
FORNI, M. AND L. REICHLIN (1998): "Let's Get Real: A Factor Analytical Approach to Disaggregated Business Cycle Dynamics," *Review of Economic Studies*, 65, 453–473. [7]
GALÍ, J. (1999): "Technology, Employment, and the Business Cycle: Do Technology Shocks Explain Aggregate Fluctuations?" *American Economic Review*, 89, 249–271. [6]
GIANNONE, D., L. REICHLIN, AND L. SALA (2005): "Monetary Policy in Real Time," in *NBER Macroeconomics Annual 2004, Volume 19*, National Bureau of Economic Research, Inc, 161–224. [6, 12, 13]
GIANNONE, D., L. REICHLIN, AND D. SMALL (2008): "Nowcasting: The real-time informational content of macroeconomic data," *Journal of Monetary Economics*, 55, 665–676. [7]
HALLIN, M. AND R. LISKA (2007): "Determining the number of factors in the general dynamic factor model," *Journal of the American Statistical Association*, 102, 603–617. [4, 5, 6, 11, 12, 13]
HORN, R. A. AND C. R. JOHNSON (1990): *Matrix Analysis*, Cambridge University Press: Cambridge (UK). [1]
——— (1991): *Topics in Matrix Analysis*, Cambridge University Press: Cambridge (UK). [1]
JUSTINIANO, A., G. PRIMICERI, AND A. TAMBALOTTI (2010): "Investment shocks and business cycles," *Journal of Monetary Economics*, 57, 132–145. [5, 6, 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. [6, 17]
KYDLAND, F. E. AND E. C. PRESCOTT (1982): "Time to Build and Aggregate Fluctuations," *Econometrica*, 50, 1345–1370. [1, 6]
LUDVIGSON, S. C. AND S. NG (2009): "Macro factors in bond risk premia," *The Review of Financial Studies*, 22, 5027–5067. [7]
LÜTKEPOHL, H. (1996): *Handbook of Matrices*, John Wiley & Sons Ltd: Chichester (UK). [1]
MCCRACKEN, M. W. AND S. NG (2020): "FRED-QD: A quarterly database for macroeconomic research," Working Papers 2020-005, Federal Reserve Bank of St. Louis. [4, 12]
ONATSKI, A. (2009): "Testing Hypotheses About the Number of Factors in Large Factor Models," *Econometrica*, 77, 1447–1479. [4, 5, 6, 11, 12, 13, 21, 24, 28, 1, 14, 15, 16]
SARGENT, T. AND C. SIMS (1977): "Business cycle modeling without pretending to have too much a priori economic theory," Working Papers 55, Federal Reserve Bank of Minneapolis. [6, 11, 13]
SMETS, F. AND R. WOUTERS (2007): "Shocks and frictions in U.S. business cycles: A bayesian DSGE approach," *American Economic Review*, 97, 586–606. [1, 6]
STOCK, J. AND M. WATSON (2002a): "Macroeconomic forecasting using diffusion indexes," *Journal of Business & Economic Statistics*, 20, 147–62. [7]
STOCK, J. H. AND M. W. WATSON (2002b): "Forecasting Using Principal Components from a Large Number of Predictors," *Journal of the American Statistical Association*, 97, 1167–1179. [7]

APPENDIX A: ASSUMPTIONS AND PROOFS OF THE MAIN RESULTS

A.1. *Assumptions and Notation*

Let $\mathbf{A}$ be a $(n \times m)$ matrix, with complex or real entries a_{ij}; in short $\mathbf{A} = [a_{ij} : 1 \leq i \leq n, 1 \leq j \leq m]$. For $m = n$, $\mu_1(\mathbf{A}) \geq \mu_2(\mathbf{A}) \geq \cdots \geq \mu_n(\mathbf{A})$ denote the eigenvalues of $\mathbf{A}$ in a decreasing order. The jth largest singular value of the matrix $\mathbf{A}$, $\sigma_j(\mathbf{A})$, is defined as

$\sqrt{\mu_j(\mathbf{A}'\mathbf{A})}$. The prime attached to the to a complex-valued matrix denotes the conjugate-complex transpose of the matrix, or the transpose when the matrix has real entries. The spectral norm $\|\mathbf{A}\| = \sqrt{\mu_1(\mathbf{A}'\mathbf{A})}$ is the largest singular value of $\mathbf{A}$, whereas $\|\boldsymbol{a}\|$ is the euclidean norm of the vector $\boldsymbol{a}$. The diagonal matrix $\mathbf{B} \equiv \text{diag}\{b_{11}, \dots b_{nn}\}$ has non-zero entries b_{ii}, for $1 \leq i \leq n$.

For the double-indexed process $\{y_{it} : 1 \leq i \leq n, 1 \leq t \leq T\}$, we write $\mathbf{y}_{nt} \equiv [y_{1t}, \cdots, y_{nt}]'$. The discrete Fourier transform (DFT) of $\{\mathbf{y}_{n1}, \dots, \mathbf{y}_{nT}\}$ at frequency ω_j is defined by $\widehat{\mathbf{y}}_{nj} \equiv T^{-1/2}\sum_{t=1}^{T}\mathbf{y}_t(n)e^{-\mathrm{i}\omega_j t}$ where $\omega_j = 2\pi j/T$ and T is assumed to be even, as discussed in Section 2.2 .

We now state the following Assumptions for the model introduced in Section 2.2

ASSUMPTION 1: *The process $\mathbf{f}_t$ is an orthonormal white noise.*

ASSUMPTION 2: *The coefficients of the filter $\mathbf{\Lambda}_n(L) \equiv \sum_{u=0}^{\infty}\mathbf{\Lambda}_n^{(u)}L^u$ are such that $\sum_{u=0}^{\infty}\|\mathbf{\Lambda}_n^{(u)}\|(1+u) = O(n^{1/2})$.*

ASSUMPTION 3: *Let ε_{jt}, $j \in \mathbb{N}$, $t \in \mathbb{Z}$ be independent and identically distributed (iid) zero-mean random variables with unit variance and $\mathbb{E}|\varepsilon_{jt}|^4 < \infty$. Further, let $\mathbf{u_{nt}} = [u_{1t}, \dots, u_{nt}]'$, with $u_{jt} \equiv C_j(L)\varepsilon_{jt} = \sum_{k=0}^{\infty}c_{jk}\varepsilon_{j,t-k}$ where $\sup_{j>0}\sum_{k=0}^{\infty}k|c_{jk}| < \infty$ and $\min_j|C_j(z)| > 0$ for $|z| \leq 1$, with $C(z) = \sum_{k=0}^{\infty}c_k z^k$ for $z \in \mathbb{C}$. We assume that the n-vector of the idiosyncratic components of the data is such that $\mathbf{e}_{nt} = \mathbf{A}_n\,\mathbf{u}_{nt}$,, where $\mathbf{A}_n$ is a non-singular lower triangular matrix with $\|\mathbf{A}_n\|$ uniformly bounded in n.*

ASSUMPTION 4: *The bandwidth parameter M_T is a function of T satisfying $1/M_T \to 0$ and $M_T^2/T = O(1)$ as $T \to \infty$. Moreover, as $n \to \infty$, $T \to \infty$, $n/T \to y \in (0,1) \cup (1,\infty)$.*

For a generic frequency ω_ℓ, $\underline{\ell} \leq \ell \leq \bar{\ell}$, we consider the index set

$$S(\ell) \equiv \{(\ell - M_T), (\ell - M_T + 1), \dots, \ell, (\ell+1), \dots (\ell + M_T)\}$$

The cardinality of $S(\ell)$ is denoted by $m \equiv 2M_T + 1$. Let s_ℓ be a generic element of $S(\ell)$, $\underline{s}_\ell \equiv \ell - M_T$ and $\bar{s}_\ell \equiv \ell + M_T$. To simplify the notation, in the following we will suppress the dependence on ℓ by setting $s \equiv s_\ell$ and $S \equiv S(\ell)$.

ASSUMPTION 5: *For any $\underline{\ell} \leq \ell \leq \bar{\ell}$, let $\hat{\nu}_{nk}^{\chi}(\omega_\ell) \equiv \mu_k\left[\widehat{\boldsymbol{\chi}}'\widehat{\boldsymbol{\chi}}/(mn)\right]$, for $k = 1, \dots, q$. Then, for each k, $\nu_k^{\chi}(\omega_\ell) \equiv \text{plim}_{n,T\to\infty}\hat{\nu}_{nk}^{\chi}(\omega_\ell)$ and assume that $0 < \nu_k^{\chi}(\omega_\ell) < \infty$ and $\nu_k^{\chi}(\omega_\ell) > \nu_{k+1}^{\chi}(\omega_\ell)$, for at least one ℓ in $[\underline{\ell}, \bar{\ell}]$, and $k = 1, \dots, q-1$.*

The first assumption is standard in the literature. Assumptions 2 and 3 are the same as Assumptions 2(i) and 2(ii)(a) in Onatski (2009). They implies regularity of the DFT's of $\boldsymbol{\chi}_{nt}$ and $\mathbf{e}_{nt}$ local to the frequency of interest ω_ℓ. Let $\widehat{\mathbf{F}} \equiv \left[\hat{\mathbf{f}}_{\underline{s}}, \hat{\mathbf{f}}_{\underline{s}+1}, \dots, \hat{\mathbf{f}}_{\bar{s}}\right]$, with $\hat{\mathbf{f}}_s \equiv \hat{\mathbf{f}}_{qs} = T^{-1/2}\sum_{t=1}^{T}\mathbf{f}_t e^{-i\omega_s t}$, and $\hat{\mathbf{\Lambda}}_n(s) \equiv \mathbf{\Lambda}_n\left(e^{-\mathrm{i}\omega_s}\right) = \sum_{u=1}^{\infty}\mathbf{\Lambda}_n^{(u)}e^{-\mathrm{i}\omega_s u}$, for any $s \in S$. Assumption 2 entails that, for n, T large, $\widehat{\boldsymbol{\chi}}$ is well approximated by $\hat{\mathbf{\Lambda}}_n(\ell)\widehat{\mathbf{F}}$ (see Lemma 1). Then, Assumption 5 could be formulated by imposing the same conditions on the eigenvalues of the matrix $[\hat{\mathbf{\Lambda}}_n(\ell)'\hat{\mathbf{\Lambda}}_n(\ell)/n](\widehat{\mathbf{F}}'\widehat{\mathbf{F}}/m)$, highlighting the similarities with Assumption A in Ahn and Horenstein (2013) (AH).

For any $s \in S$, $\hat{\mathbf{E}} \equiv [\hat{\mathbf{e}}_{n\underline{s}}, \hat{\mathbf{e}}_{n\underline{s}+1}, \dots, \hat{\mathbf{e}}_{n\bar{s}}]$, with $\hat{\mathbf{e}}_{ns} \equiv T^{-1/2}\sum_{t=1}^{T}\mathbf{f}_t e^{-i\omega_s t}$ and define $\tilde{\mathbf{E}} \equiv \mathbf{A}_n\,\hat{\mathbf{C}}_n(\ell)\hat{\boldsymbol{\mathcal{E}}}$, where $\hat{\mathbf{C}}_n(s) \equiv \text{diag}\left(\sum_{u=0}^{\infty}c_{1u}e^{-\mathrm{i}\omega_s u}, \dots, \sum_{u=0}^{\infty}c_{nu}e^{-\mathrm{i}\omega_s u}\right)$ is a $n \times n$ diagonal

matrix and $\hat{\boldsymbol{\mathcal{E}}} \equiv [\hat{\boldsymbol{\varepsilon}}_{n\underline{s}}, \hat{\boldsymbol{\varepsilon}}_{n,\underline{s}+1}, \ldots, \hat{\boldsymbol{\varepsilon}}_{n\bar{s}}]$ is a $n \times m$ matrix. Assumption 3, which can be seen as the dynamic counterpart of Assumption C in AH, guarantees that, for n, T large $\hat{\mathbf{E}}$ is well approximated by $\tilde{\mathbf{E}}$ (see Lemma 3). Lemma4 shows that the first q eigenvalues of $\tilde{\mathbf{E}}'\tilde{\mathbf{E}}/n$ are bounded and bounded away from zero. The latter results are obtained by using results from the large random matrix theory (see Lemma 17), requiring that $n/T \to y$, as postulated in Assumption 4 (see also Assumption C(i) in AH).

We conclude this section by introducing further notation used in Section A.2. $\tilde{\mathbf{X}} \equiv \hat{\mathbf{\Lambda}}_n(\ell)\hat{\mathbf{F}} + \tilde{\mathbf{E}}$. and denote by $\tilde{\nu}_{nk}(\omega_\ell) \equiv \mu_k\left(\tilde{\mathbf{X}}\tilde{\mathbf{X}}'/(nm)\right)$ its kth largest eigenvalue, for $k = 1, \ldots, m$. We recall that $\hat{\mathbf{X}} = \hat{\boldsymbol{\chi}} + \hat{\mathbf{E}}$ and $\hat{\nu}_{nk}(\omega_\ell) = \mu_k\left(\hat{\mathbf{X}}\hat{\mathbf{X}}'/(nm)\right)$.

Note that $\hat{\nu}_{nk}(\omega_\ell) = \hat{\mu}_{nk}(\omega_\ell)/n$, with $\hat{\mu}_{nk}(\omega_\ell) = \mu_k\left(\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{n}}(\omega_\ell)\right)$ and $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{n}}(\omega_\ell) = \hat{\mathbf{X}}\hat{\mathbf{X}}'/m$. We find it convenient using $\hat{\nu}_{nk}(\omega_\ell)$, $k = 1, \ldots, m$, in our derivations because they are bounded in probability.

A.2. *Main Lemmas*

This section introduces intermediate results necessary to prove Theorem 1 in Section A.3, referring to the Online Appendix Section A.4 for formal proofs.

LEMMA 1: *Under Assumptions 1 and 2,* $\|\hat{\boldsymbol{\chi}} - \hat{\mathbf{\Lambda}}_n(\ell)\hat{\mathbf{F}}\|^2 = O_p\left(mn/T\right)$

LEMMA 2: *Under Assumptions 1, 2 and 5, for* $k = 1, \ldots, q$

$$\sigma_k^2\left(\frac{\hat{\mathbf{\Lambda}}_n(\ell)\hat{\mathbf{F}}}{\sqrt{mn}}\right) - \sigma_k^2\left(\frac{\hat{\boldsymbol{\chi}}}{\sqrt{mn}}\right) = O_p\left(\frac{1}{\sqrt{T}}\right), \tag{A.1}$$

and

$$\tilde{\nu}_{nk}^{\chi}(\omega_\ell) \equiv \nu_k\left(\frac{\hat{\mathbf{\Lambda}}_n(\ell)\hat{\mathbf{F}}\hat{\mathbf{F}}'\hat{\mathbf{\Lambda}}_n'(\ell)}{mn}\right) = \mu_k^{\chi}(\omega_\ell) + o_p(1) \tag{A.2}$$

with $\nu_k^{\chi}(\omega_\ell)$ *defined in Assumption 5.*

LEMMA 3: *Under Assumption 3,* $\|\hat{\mathbf{E}} - \tilde{\mathbf{E}}\|^2 = O_p\left(mn/T\right)$,

LEMMA 4: *Let* $n, T \to \infty$. *Then, under Assumptions 3 and 4,* $\underline{l} \leq \mu_k\left(\tilde{\mathbf{E}}\tilde{\mathbf{E}}'/n\right) \leq \bar{l}$ *almost surely as, for* $1 \leq k \leq m$, *where* $\underline{l} \equiv \delta_k\left(1 - 1/\sqrt{y}\right)^2$, $\bar{l} \equiv \delta_k\left(1 + 1/\sqrt{y}\right)^2$ *and* δ_k *is a positive bounded real number satisfying*

$$\mu_n(\mathbf{A}_n\,\hat{\mathbf{C}}_n(\ell)\hat{\mathbf{C}}_n'(\ell)\,\mathbf{A}_n') \leq \delta_k \leq \mu_1(\mathbf{A}_n\,\hat{\mathbf{C}}_n(\ell)\hat{\mathbf{C}}_n'(\ell)\,\mathbf{A}_n'). \tag{A.3}$$

LEMMA 5: *Under Assumptions 1-3, for* m *sufficiently large and* $j = 1, \ldots, m - 2q$

$$\mu_{2q+j}\left(\frac{\tilde{\mathbf{E}}\tilde{\mathbf{E}}'}{mn}\right) \leq \tilde{\nu}_{n,q+j}(\omega_\ell) \leq \mu_1\left(\frac{\tilde{\mathbf{E}}\tilde{\mathbf{E}}'}{mn}\right), \tag{A.4}$$

LEMMA 6: *Let* $n, T \to \infty$. *Then, under Assumptions 1-5,* $\underline{l} \leq m\,\tilde{\nu}_{nk}(\omega_\ell) \leq \bar{l}$ *almost surely, for* $k = 1, \ldots, m - q$.

LEMMA 7: *Under Assumptions 1-5, for* $k = 1, \ldots, q$, $|\tilde{\nu}_{nk}(\omega_\ell) - \tilde{\nu}^{\chi}_{nk}(\omega_\ell)| = o_p(m^{-1/2})$, *where* $\tilde{\nu}^{\chi}_{nk}(\omega_\ell)$ *has been defined in display (A.2).*

LEMMA 8: *Under Assumptions 1-5,* $|\hat{\nu}_{nk}(\omega_\ell) - \tilde{\nu}_{nk}(\omega_\ell)| = O_p(T^{-1/2})$, *for* $k = 1, \ldots, q$, *and* $|\hat{\nu}_{nk}(\omega_\ell) - \tilde{\nu}_{nk}(\omega_\ell)| = O_p((mT)^{-1/2})$, *for* $k = q+1, \ldots, m$.

A.3. *Proof of Theorem 1*

In the following we write $\sum_\ell$ for $\sum_{\ell=\underline{\ell}}^{\bar{\ell}}$, and set $L \equiv \bar{\ell} - \underline{\ell} + 1$. Recall that the results derived in section A.2 hold for any $-\pi < \omega_\ell \leq \pi$ as n, T diverge. Note that our estimators are invariant to the scaling of matrix $\widehat{\mathbf{\Sigma}}_n(\omega_\ell)$. As explained at the end of section A.1, we find with convenient working with the eigenvalues of the matrix $\widehat{\mathbf{\Sigma}}_n(\omega_\ell)/n$, denoted by $\hat{\nu}_{nk}(\omega_\ell)$.

Lemmas 7,8, and 2 and Assumption 5 imply that

$$\mathrm{DER}^T_n(k) = \frac{\sum_\ell \nu^{\chi}_k(\omega_\ell)}{\sum_\ell \nu^{\chi}_{k+1}(\omega_\ell)} + o_p(1) = O_p(1) \quad \text{for} \quad k = 1, \ldots, q-1.$$

From Lemmas 8 and 6 it follows that

$$\mathrm{DER}^T_n(q) \geq m \left[\frac{L^{-1} \sum_\ell \nu^{\chi}_q(\omega_\ell) + o_p(1)}{\bar{l} + o_p(1)} \right] \xrightarrow{p} \infty.$$

Lemma 6 also entails that $\underline{l} + o_p(1) \leq m\hat{\nu}_{nk}(\omega_\ell) \leq \bar{l} + o_p(1)$, uniformly in $q < k \leq m - q$. Hence,

$$\mathrm{DER}^T_n(k) \leq \frac{\bar{l} + o_p(1)}{\underline{l} + o_p(1)}, \quad \text{for} \quad k = q+1, \ldots, q_{\max}.$$

These results show (2.11).

We now show the consistency of the DGR estimator. Lemmas 3 and 4 entail that for $k = q+1, \ldots, q_{\max}$, $m^{-1}(\underline{l} + o_p(1)) \leq L^{-1} \sum_\ell \hat{\nu}_{nk}(\omega_\ell) \leq m^{-1}(\bar{l} + o_p(1))$. Hence, setting $W^T_n \equiv V_T/n$ we have

$$\left[\frac{m - (q+1)}{m} \right] (\underline{l} + o_p(1)) \leq L^{-1} W^T_n(q+1) \leq \left[\frac{m - (q+1)}{m} \right] (\bar{l} + o_p(1))) \tag{A.5}$$

Using the inequalities $c/(1+c) < \ln(1+c) < \delta$ for any $c \in (0, \infty)$ we have

$$\mathrm{DGR}^T_n(k) = \frac{\ln(1 + \hat{\nu}^*_{nk})}{\ln(1 + \hat{\nu}^*_{n,k+1})} < \frac{\hat{\nu}^*_{nk}}{\hat{\nu}^*_{n,k+1}/(1 + \hat{\nu}^*_{n,k+1})} = \mathrm{DER}^T_n(k)$$

for $k = 1, 2, \ldots, q-1, q+1, \ldots, q_{\max}$. For $k = q$

$$\mathrm{DGR}^T_n(q) = \frac{\ln(1 + \hat{\nu}^*_{nq})}{\ln(1 + \hat{\nu}^*_{n,q+1})} > \frac{\hat{\nu}^*_{n,q}/(1 + \hat{\nu}^*_{n,q})}{\hat{\nu}^*_{n,q+1}} = \mathrm{DER}^T_n(q) \cdot \frac{W^T_n(q+1)}{W^T_n(q-1)} \xrightarrow{p} \infty$$

because by display (A.5)

$$\frac{W_n^T(q+1)}{W_n^T(q-1)} = \frac{W_n^T(q+1)}{\sum_\ell [\hat{\nu}_{nq}(\omega_\ell) + \hat{\nu}_{n,q+1}(\omega_\ell)] + W_n^T(q+1)} = 1 + O_p(1).$$

We conclude the proof showing the consistency of the DDR estimator. For $k \leq q-2$, by Lemma 7 and Assumption 5

$$\text{DDR}_n^T(k) = \frac{L^{-1}\sum_\ell \left(\nu_k^\chi(\omega_\ell) - \nu_{k+1}^\chi(\omega_\ell)\right) + o_p(1)}{L^{-1}\sum_\ell \left(\nu_{k+1}^\chi(\omega_\ell) - \nu_{k+2}^\chi(\omega_\ell)\right) + o_p(1)} = O_p(1)$$

For $k = q-1$, by Lemma 7, Assumption 5 and Lemmas 6 and 8

$$\text{DDR}_n^T(q-1) = \frac{L^{-1}\sum_\ell \left(\nu_{q-1}^\chi(\omega_\ell) - \nu_q^\chi(\omega_\ell)\right) + o_p(1)}{L^{-1}\sum_\ell \nu_q^\chi(\omega_{\ell,j}) + o_p(1)} = O_p(1)$$

For $q < k \leq q_{\max}$, by Lemmas 6 and 8

$$\text{DDR}_n^T(k) = \frac{L^{-1}\sum_\ell \left(\tilde{\nu}_{nk}(\omega_\ell) - \tilde{\nu}_{n,k+1}(\omega_\ell) + o_p(1)\right)}{\max\left(L^{-1}\sum_\ell \left(\tilde{\nu}_{n,k+1}(\omega_\ell) - \tilde{\nu}_{n,k+2}(\omega_\ell)\right), L^{-1}\sum_\ell \tilde{\nu}_{n,m}(\omega_\ell)\right) + o_p(1)} = O_p(1)$$

Finally, by Lemma 7, Assumption 5 and Lemmas 6 and 8

$$\begin{aligned}\text{DDR}_n^T(q) &= \frac{L^{-1}\sum_\ell \left(\nu_q^\chi(\omega_\ell) - \tilde{\nu}_{n,q+1}(\omega_\ell)\right) + o_p(1)}{\max\left(L^{-1}\sum_\ell \left(\tilde{\nu}_{n,q+1}(\omega_\ell) - \tilde{\nu}_{n,q+2}(\omega_\ell)\right), L^{-1}\sum_\ell \tilde{\nu}_{n,m}(\omega_\ell)\right) + O_p(T^{1/2})} \\ &= \frac{m\left(L^{-1}\sum_\ell^M \nu_q^\chi(\omega_\ell) + o_p(1)\right)}{O_p(1)} \xrightarrow{p} \infty.\end{aligned}$$

Q.E.D.

A.4. *Proof of the Main Lemmas (Section A.2)*

PROOF OF LEMMA 1: The proof follows closely (Onatski, 2009) (proof of Lemma 4, p. 1471) and it is reported for completeness. Write $\hat{\boldsymbol{\chi}} - \hat{\boldsymbol{\Lambda}}_n(\ell)\hat{\mathbf{F}}$ as $\boldsymbol{P}_1 + \boldsymbol{R}_1$, which are $(n \times m)$ matrices with $s-$th columns $\boldsymbol{p}_{1s} = \hat{\boldsymbol{\chi}}_{ns} - \hat{\boldsymbol{\Lambda}}_n(s)\hat{\mathbf{f}}_s$ and $\boldsymbol{r}_{1s} = (\hat{\boldsymbol{\Lambda}}_n(s) - \hat{\boldsymbol{\Lambda}}_n(\ell))\hat{\mathbf{f}}_s$, $s \in S$. Recall that

$$\hat{\boldsymbol{\Lambda}}_n(s)\hat{\mathbf{f}}_s = \sum_{u=1}^{\infty} \boldsymbol{\Lambda}_n^{(u)} e^{-\mathrm{i}\omega_s u}\left[T^{-1/2}\sum_{\tau=1}^{T} \mathbf{f}_\tau e^{-\mathrm{i}\omega_s \tau}\right] \tag{A.6}$$

and

$$\hat{\boldsymbol{\chi}}_{ns} = T^{-1/2} \sum_{t=1}^{T} \sum_{u=0}^{\infty} \boldsymbol{\Lambda}_n^{(u)} \mathbf{f}_{t-u} e^{-\mathrm{i}\omega_s t} \tag{A.7}$$

Interchanging the order of summation in (A.7) and changing the summation index t to $\tau \equiv t-u$, we obtain the representation

$$\hat{\boldsymbol{\chi}}_{ns} = T^{-1/2} \sum_{u=0}^{\infty} \boldsymbol{\Lambda}_n^{(u)} \sum_{\tau=1-u}^{T-u} \mathbf{f}_{\tau} e^{-\mathrm{i}\omega_s(\tau+u)} \tag{A.8}$$

Subtracting (A.6) from (A.8) we obtain

$$\boldsymbol{p}_{1s} = \hat{\boldsymbol{\chi}}_{ns} - \hat{\boldsymbol{\Lambda}}_n(s)\hat{\mathbf{f}}_s = T^{-1/2} \sum_{u=0}^{\infty} \boldsymbol{\Lambda}_n^{(u)} e^{-\mathrm{i}\omega_s u} \boldsymbol{d}_{1u}(s)$$

where,

$$\boldsymbol{d}_{1u}(s) = \sum_{\tau=1-u}^{\min(T-u,0)} \mathbf{f}_{\tau} e^{-\mathrm{i}\omega_s \tau} - \sum_{\tau=\max(T-u,0)+1}^{T} \mathbf{f}_{\tau} e^{-\mathrm{i}\omega_s \tau}.$$

Then,

$$\begin{aligned}
\mathbb{E}\|\boldsymbol{p}_{1s}\|^2 &\leq T^{-1} \sum_{u,v=0}^{\infty} \mathbb{E}\left|\boldsymbol{d}_{1u}'(s)\boldsymbol{\Lambda}_n^{(u)\prime}\boldsymbol{\Lambda}_n^{(v)}\boldsymbol{d}_{1v}(s)\right| \\
&\leq T^{-1} \sum_{u,v=0}^{\infty} \left\|\boldsymbol{\Lambda}_n^{(u)}\right\| \left\|\boldsymbol{\Lambda}_n^{(u)}\right\| \mathbb{E}\left(\|\boldsymbol{d}_{1u}(s)\| \, \|\boldsymbol{d}_{1v}(s)\|\right) \\
&\leq T^{-1} \sum_{u,v=0}^{\infty} \left\|\boldsymbol{\Lambda}_n^{(u)}\right\| \left\|\boldsymbol{\Lambda}_n^{(v)}\right\| \left(\mathbb{E}\|\boldsymbol{d}_{1u}(s)\|^2 \, \mathbb{E}\|\boldsymbol{d}_{1v}(s)\|^2\right)^{1/2} \\
&= T^{-1} \left[\sum_{u=0}^{\infty} \left\|\boldsymbol{\Lambda}_n^{(u)}\right\| \left(\mathbb{E}\|\boldsymbol{d}_{1u}(s)\|^2\right)^{1/2}\right]^2.
\end{aligned}$$

But

$$\mathbb{E}\|\boldsymbol{d}_{1u}(s)\|^2 = \sum_{\tau=1-u}^{\min(T-u,0)} \mathbb{E}\|\mathbf{f}_{\tau}\|^2 + \sum_{\tau=\max(T-u,0)+1}^{T} \mathbb{E}\|\mathbf{f}_{\tau}\|^2 = 2q\min(u,T).$$

because $\mathbf{f}_t$ is a q-dimensional orthonormal white noise. Therefore, by Lemma 9

$$\begin{aligned}
\mathbb{E}\|\boldsymbol{P}_1\|^2 \leq \sum_s \mathbb{E}\|\boldsymbol{p}_{1s}\|^2 &\leq \frac{1}{T} \sum_s \left[\sum_{u=0}^{\infty} \left\|\boldsymbol{\Lambda}_n^{(u)}\right\| \left(\mathbb{E}\|\boldsymbol{d}_{1u}(s)\|^2\right)^{1/2}\right]^2 \\
&= \frac{1}{T} \sum_s \left[\sum_{u=0}^{T} \left\|\boldsymbol{\Lambda}_n^{(u)}\right\| (2qu)^{1/2} + \sum_{u=T+1}^{\infty} \left\|\boldsymbol{\Lambda}_n^{(u)}\right\| (2qT)^{1/2}\right]^2
\end{aligned}$$

$$\leq \frac{m}{T}\left[\sum_{u=0}^{\infty}\left\|\mathbf{\Lambda}_n^{(u)}\right\|(2qu)^{1/2}\right]^2 = O\left(\frac{mn}{T}\right) \tag{A.9}$$

where we used $\sum_s$ as the shorthand notation for $\sum_{s=\underline{s}}^{\bar{s}}$.

Next, we consider the matrix $\boldsymbol{R}_1$. Note that

$$\begin{aligned}\left\|\hat{\mathbf{\Lambda}}_n(s)-\hat{\mathbf{\Lambda}}_n(\ell)\right\| &\leq \sum_{u=0}^{\infty}\left\|\mathbf{\Lambda}_n^{(u)}\right\||e^{-iu\omega_s}-e^{-iu\omega_\ell}|\\ &\leq \sum_{u=0}^{\infty}\left\|\mathbf{\Lambda}_n^{(u)}\right\| u\frac{2\pi m}{T} = O\left(\frac{mn^{1/2}}{T}\right)\end{aligned}$$

uniformly in s. Further, since $\mathbf{f}_t$ is a $q-$ dimensional orthonormal white noise, $\sum_{s\in S}\mathbb{E}\left\|\hat{\mathbf{f}}_s\right\|^2 = O(m)$. Finally, by Lemma 9

$$\begin{aligned}\mathbb{E}\left\|\boldsymbol{R}_1\right\|^2 &\leq \sum_{s\in S}\mathbb{E}\left\|\boldsymbol{r}_{1s}\right\|^2 \leq \max_s\left\|\hat{\mathbf{\Lambda}}_n(s)-\hat{\mathbf{\Lambda}}_n(\ell)\right\|^2\sum_{s\in S}\mathbb{E}\left\|\hat{\mathbf{f}}_s\right\|^2\\ &= O\left(\frac{m^2n}{T^2}\right)O_p(m) = O\left(\frac{n}{T}\right)O(m) = O\left(\frac{mn}{T}\right)\end{aligned} \tag{A.10}$$

where we use that $m^2/T = O(1)$ by Assumption 4. The statement follows from (A.9) and (A.10) and the fact that, by Lemma 10 $\left\|\hat{\boldsymbol{\chi}}-\hat{\mathbf{\Lambda}}_n(\ell)\hat{\mathbf{F}}\right\|^2 \leq \left[\|\boldsymbol{P}_1\|+\|\boldsymbol{R}_1\|\right]^2 = O_p\left(mn/T\right)$. *Q.E.D.*

PROOF OF LEMMA 2: By Lemma 16, Lemma 1 and Assumption 5

$$\begin{aligned}&\sigma_k^2\left(\frac{\hat{\mathbf{\Lambda}}_n(\ell)\hat{\mathbf{F}}}{\sqrt{mn}}\right)-\sigma_k^2\left(\frac{\hat{\boldsymbol{\chi}}}{\sqrt{mn}}\right)\\ &\leq \sigma_1\left(\frac{\hat{\boldsymbol{\chi}}-\hat{\mathbf{\Lambda}}_n(\ell)\hat{\mathbf{F}}}{\sqrt{mn}}\right)\left[2\sigma_k\left(\frac{\hat{\boldsymbol{\chi}}}{\sqrt{mn}}\right)\sigma_1\left(\frac{\hat{\boldsymbol{\chi}}-\hat{\mathbf{\Lambda}}_n(\ell)\hat{\mathbf{F}}}{\sqrt{mn}}\right)\right] = O_p\left(\frac{1}{\sqrt{T}}\right)\end{aligned}$$

proving (A.1). The result in (A.2) follows from Assumption 5 and (A.1), by noting that $\sigma_k^2\left((mn)^{-1/2}\hat{\boldsymbol{\chi}}\right) = \hat{\mu}_{nk}(\omega_\ell)$. *Q.E.D.*

PROOF OF LEMMA 3: Following the proof of Lemma 1 we write $\hat{\mathbf{E}}-\tilde{\mathbf{E}}$ as the sum of the $(n\times m)$ matrices $\boldsymbol{P}_2$ and $\boldsymbol{R}_2$. The $s-$th column of $\boldsymbol{P}_2$ is defined as

$$\boldsymbol{p}_{2s} \equiv \hat{\mathbf{e}}_{ns}-\mathbf{A}_n\hat{\mathbf{C}}_n(s)\hat{\boldsymbol{\varepsilon}}_{ns} = T^{-1/2}\mathbf{A}_n\sum_{u=0}^{\infty}\mathbf{C}_n^{(u)}e^{-\mathrm{i}\omega_s u}\boldsymbol{d}_{2u}(s) \tag{A.11}$$

where

$$\boldsymbol{d}_{2u}(s) \equiv \sum_{\tau=1-u}^{\min(T-u,0)}\boldsymbol{\varepsilon}_{n\tau}e^{-\mathrm{i}\omega_s\tau}-\sum_{\tau=\max(T-u,0)+1}^{T}\boldsymbol{\varepsilon}_{n\tau}e^{-\mathrm{i}\omega_s\tau}.$$

The last equality in (A.11) follows by recalling that $\hat{\mathbf{C}}_n(s) = \sum_{u=0}^{\infty} \mathbf{C}_n^{(u)} e^{-\mathrm{i}\omega_s u}$ and $\hat{\boldsymbol{\varepsilon}}_{ns} = T^{-1/2} \sum_{\tau=1}^{T} \boldsymbol{\varepsilon}_{n\tau} e^{-\mathrm{i}\omega_s \tau}$ and noting that $\hat{\mathbf{e}}_{ns} = T^{-1/2}\, \mathbf{A}_n \sum_{t=1}^{T} \sum_{u=0}^{\infty} \mathbf{C}_n^{(u)} \boldsymbol{\varepsilon}_{n,t-u} e^{-\mathrm{i}\omega_s t}$, can be re-written as

$$\hat{\mathbf{e}}_{ns} = T^{-1/2}\, \mathbf{A}_n \sum_{u=0}^{\infty} \mathbf{C}_n^{(u)} \sum_{\tau=1-u}^{T-u} \boldsymbol{\varepsilon}_{n\tau} e^{-\mathrm{i}\omega_s(\tau+u)} = T^{-1/2}\, \mathbf{A}_n \sum_{u=0}^{\infty} \mathbf{C}_n^{(u)} e^{-\mathrm{i}\omega_s u} \sum_{\tau=1-u}^{T-u} \boldsymbol{\varepsilon}_{n\tau} e^{-\mathrm{i}\omega_s \tau}$$

First we consider

$$\begin{aligned}
\mathbb{E}\,\|\boldsymbol{p}_{2s}\|^2 &\le T^{-1} \|\,\mathbf{A}_n\,\|^2 \sum_{j=1}^{n} \mathbb{E} \left| \sum_{u,v=0}^{\infty} c_{ju} e^{-i\omega_s u} d_{2sj}(u) \right|^2 \\
&\le T^{-1} \|\,\mathbf{A}_n\,\|^2 \sum_{j=1}^{n} \sum_{u,v=0}^{\infty} |c_{ju} e^{-\mathrm{i}\omega_s v}| |c_{jv} e^{\mathrm{i}\omega_s u}|\, \left[\mathbb{E}\,|d_{2,sj}(u)|^2\, \mathbb{E}\,|d_{2,sj}(v)|^2\right]^{1/2} \\
&\le T^{-1} \|\,\mathbf{A}_n\,\|^2 \sum_{j=1}^{n} \sum_{u,v=0}^{\infty} |c_{ju}| u^{1/2} |c_{jv}| v^{1/2} = O\left(\frac{n}{T}\right) \qquad \text{(A.12)}
\end{aligned}$$

The last inequality follows along the line of (A.9), after noting that by Assumption 3 we have

$$\mathbb{E}\,|d_{2ju}(s)|^2 = \sum_{\tau=1-u}^{\min(T-u,0)} \mathbb{E}\,\varepsilon_{j\tau}^2 + \sum_{\tau=\max(T-u,0)+1}^{T} \mathbb{E}\,\varepsilon_{j\tau}^2 = \min(u,T), \quad \text{for} \quad j = 1,\ldots,n.$$

By Lemma 9 $\mathbb{E}\,\|\boldsymbol{P}_2\|^2 \le \sum_{s=1}^{m} \mathbb{E}\,\|\boldsymbol{p}_{2s}\|^2 = O\,(mn/T)$. Next, we consider the matrix $\boldsymbol{R}_2$. Note that, for $j = 1,\ldots,n$

$$|C_j(e^{-\mathrm{i}\omega_s u}) - C_j(e^{-\mathrm{i}\omega_\ell u})| \le \sum_{u=0}^{\infty} c_{ju} |e^{-\mathrm{i}\omega_s u} - e^{-\mathrm{i}\omega_0 u}| \le \frac{2\pi m}{T} \sum_{u=0}^{\infty} |c_{ju}| u$$

By Assumption 3

$$\left\| \hat{\mathbf{C}}_n(s) - \hat{\mathbf{C}}_n(\ell) \right\| \le \frac{2\pi m}{T} \left[\max_{0<j\le n} \sum_{u=0}^{\infty} |c_{ju}| u \right] = O\left(\frac{m}{T}\right)$$

uniformly in s. Moreover,

$$\mathbb{E}\,\|\hat{\boldsymbol{\varepsilon}}_{ns}\|^2 = \mathbb{E}\left[\sum_{j=1}^{n} \left| \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \varepsilon_{jt} e^{-\mathrm{i}\omega_s t} \right|^2 \right] = \sum_{j=1}^{n} \mathbb{E}\,\varepsilon_{jt}^2 = O(n).$$

Hence, $\sum_s \|\hat{\boldsymbol{\varepsilon}}_{ns}\|^2 = O_p(mn)$ and

$$\begin{aligned}
\|\boldsymbol{R}_2\|^2 &\le \sum_s \mathbb{E}\,\|\boldsymbol{r}_{2s}\|^2 \le \|\,\mathbf{A}_n\,\|^2 \max_s \left\| \hat{\mathbf{C}}_n(s) - \hat{\mathbf{C}}_n(\ell) \right\|^2 \sum_s \|\hat{\boldsymbol{\varepsilon}}_{ns}\|^2 \\
&= O(1) O\left(\frac{m^2}{T^2}\right) O_p(mn) = O(1) O(T^{-1}) O_p(mn) = O_p\left(\frac{mn}{T}\right) \qquad \text{(A.13)}
\end{aligned}$$

where we have used that, by Assumption 4, $m^2/T = O(1)$. The statement follows from (A.12) and (A.13) and the fact that $\|\hat{\mathbf{E}} - \tilde{\mathbf{E}}\|^2 \leq [\|\boldsymbol{P}_2\| + \|\boldsymbol{R}_2\|]^2 = O_p(mn/T)$. *Q.E.D.*

PROOF OF LEMMA 4: By Lemma 13 , for each $k = 1, \ldots, m$ there exists a positive real number δ_k satisfying (A.3) and

$$\mu_k\left(\frac{\tilde{\mathbf{E}}\tilde{\mathbf{E}}'}{n}\right) = \mu_k\left(\frac{\mathbf{A}_n\,\hat{\mathbf{C}}_n(\ell)\hat{\boldsymbol{\mathcal{E}}}\hat{\boldsymbol{\mathcal{E}}}'\hat{\mathbf{C}}_n(\ell)'\,\mathbf{A}_n'}{n}\right) = \delta_k\mu_k\left(\frac{\hat{\boldsymbol{\mathcal{E}}}\hat{\boldsymbol{\mathcal{E}}}'}{n}\right) \tag{A.14}$$

The $n \times m$ matrix $\hat{\boldsymbol{\mathcal{E}}}$ can be written as $\boldsymbol{\mathcal{E}}\,\mathbf{B}$, where $\mathbf{B}$ is a $T \times m$ matrix with s-th column

$$\boldsymbol{b}(\omega_s) = T^{-1/2}\left[e^{\mathrm{i}\omega_s}, e^{\mathrm{i}2\omega_s}, \ldots, e^{\mathrm{i}T\omega_s}\right]', \quad \text{for } s \in S.$$

The $T \times 1$ vectors $\boldsymbol{b}(\omega_j)$ constitute an orthonormal basis for $\mathbb{C}^T$, so $\mathbf{P}_B \equiv \mathbf{B}\,\mathbf{B}'$ is a projection matrix. Let $\mathbf{M}_B \equiv \boldsymbol{I}_T - \mathbf{P}_B$. By Lemma 11

$$\mu_k\left(\frac{\hat{\boldsymbol{\mathcal{E}}}\hat{\boldsymbol{\mathcal{E}}}'}{n}\right) = \mu_k\left(\frac{\boldsymbol{\mathcal{E}}\mathbf{P}_B\boldsymbol{\mathcal{E}}'}{n}\right) \leq \mu_k\left(\frac{\boldsymbol{\mathcal{E}}\mathbf{P}_B\boldsymbol{\mathcal{E}}'}{n} + \frac{\boldsymbol{\mathcal{E}}\mathbf{M}_B\boldsymbol{\mathcal{E}}'}{n}\right) = \mu_k\left(\frac{\boldsymbol{\mathcal{E}}\boldsymbol{\mathcal{E}}'}{n}\right) \leq \mu_1\left(\frac{\boldsymbol{\mathcal{E}}\boldsymbol{\mathcal{E}}'}{n}\right)$$

Moreover, by Lemma 12, for $k = 1, \ldots, m$

$$\mu_k\left(\frac{\hat{\boldsymbol{\mathcal{E}}}\hat{\boldsymbol{\mathcal{E}}}'}{n}\right) = \mu_k\left(\frac{\boldsymbol{\mathcal{E}}\mathbf{P}_B\boldsymbol{\mathcal{E}}'}{n}\right) \geq \mu_{k+(T-m)}\left(\frac{\boldsymbol{\mathcal{E}}\mathbf{P}_B\boldsymbol{\mathcal{E}}'}{n} + \frac{\boldsymbol{\mathcal{E}}\mathbf{M}_B\boldsymbol{\mathcal{E}}'}{n}\right) \geq \mu_T\left(\frac{\boldsymbol{\mathcal{E}}\boldsymbol{\mathcal{E}}'}{n}\right).$$

We conclude that, for $k = 1, \ldots, m$,

$$\mu_T\left(\frac{\boldsymbol{\mathcal{E}}\boldsymbol{\mathcal{E}}'}{n}\right) \leq \mu_k\left(\frac{\hat{\boldsymbol{\mathcal{E}}}\hat{\boldsymbol{\mathcal{E}}}'}{n}\right) \leq \mu_1\left(\frac{\boldsymbol{\mathcal{E}}\boldsymbol{\mathcal{E}}'}{n}\right)$$

By Lemma 17

$$\mu_1\left(\frac{\boldsymbol{\mathcal{E}}\boldsymbol{\mathcal{E}}'}{n}\right) = \frac{T}{n}\mu_1\left(\frac{\boldsymbol{\mathcal{E}}\boldsymbol{\mathcal{E}}'}{T}\right) \xrightarrow{a.s} \frac{1}{y}(1+\sqrt{y})^2 = \left(1 + \frac{1}{\sqrt{y}}\right)^2 \tag{A.15}$$

and

$$\mu_T\left(\frac{\boldsymbol{\mathcal{E}}\boldsymbol{\mathcal{E}}'}{n}\right) \xrightarrow{a.s} \left(1 - \frac{1}{\sqrt{y}}\right)^2, \tag{A.16}$$

The proof follows from equations (A.14), (A.15), and (A.16). *Q.E.D.*

PROOF LEMMA 5: Following Onatski (2009)[p. 1475], we define the $m \times m$ unitary matrix $\mathbf{V}$ such that $\hat{\mathbf{F}}\mathbf{V} = [\hat{\mathbf{F}}_1\ \mathbf{0}]$ and $\tilde{\mathbf{E}}\mathbf{V} = [\tilde{\mathbf{E}}_1\ \tilde{\mathbf{E}}_2]$ where $\tilde{\mathbf{E}}_1$ is $n \times q$ and $\hat{\mathbf{F}}_1$ is $q \times q$ and consider the decomposition

$$\frac{\tilde{\mathbf{X}}\tilde{\mathbf{X}}'}{mn} = \frac{(\hat{\boldsymbol{\Lambda}}_n(\ell)\hat{\mathbf{F}}_1 + \tilde{\mathbf{E}}_1)(\hat{\boldsymbol{\Lambda}}_n(\ell)\hat{\mathbf{F}}_1 + \tilde{\mathbf{E}}_1)'}{mn} + \frac{\tilde{\mathbf{E}}_2\tilde{\mathbf{E}}_2'}{mn} \tag{A.17}$$

Note that, for $j = 1, \ldots, m - q$

$$\mu_{q+j}\left[\frac{(\hat{\mathbf{\Lambda}}_n(\ell)\hat{\mathbf{F}}_1 + \tilde{\mathbf{E}}_1)(\hat{\mathbf{\Lambda}}_n(\ell)\hat{\mathbf{F}}_1 + \tilde{\mathbf{E}}_1)'}{mn}\right] = 0. \tag{A.18}$$

We first prove the upper bound of the inequality in (A.4). Lemma 10 and display (A.18) entail that, for $q + 1 + j \leq m$

$$\tilde{\nu}_{n,q+j}(\omega_\ell) \leq \mu_{q+1}\left(\frac{(\hat{\mathbf{\Lambda}}_n(\ell)\hat{\mathbf{F}}_1 + \tilde{\mathbf{E}}_1)(\hat{\mathbf{\Lambda}}_n(\ell)\hat{\mathbf{F}}_1 + \tilde{\mathbf{E}}_1)'}{mn}\right) + \mu_j\left(\frac{\tilde{\mathbf{E}}_2\tilde{\mathbf{E}}_2'}{mn}\right) = \mu_j\left(\frac{\tilde{\mathbf{E}}_2\tilde{\mathbf{E}}_2'}{mn}\right). \tag{A.19}$$

By Lemma 11,

$$\mu_j\left(\frac{\tilde{\mathbf{E}}_2\tilde{\mathbf{E}}_2'}{mn}\right) \leq \mu_j\left(\frac{\tilde{\mathbf{E}}_1\tilde{\mathbf{E}}_1' + \tilde{\mathbf{E}}_2\tilde{\mathbf{E}}_2'}{mn}\right) = \mu_j\left(\frac{\tilde{\mathbf{E}}\tilde{\mathbf{E}}'}{mn}\right). \tag{A.20}$$

Hence, by (A.19) and (A.20) we conclude that for every $j \geq 1$

$$\tilde{\nu}_{n,q+j}(\omega_\ell) \leq \mu_1\left(\frac{\tilde{\mathbf{E}}\tilde{\mathbf{E}}'}{mn}\right) \tag{A.21}$$

Next we derive the lower bound of (A.4). Using again Lemma 11, for any $1 \leq j \leq m - q$

$$\mu_{q+j}\left(\frac{\tilde{\mathbf{E}}_2\tilde{\mathbf{E}}_2'}{mn}\right) \leq \tilde{\nu}_{n,q+j}(\omega_\ell), \tag{A.22}$$

whereas, by Lemma 12, for $1 \leq j \leq m - 2q$,

$$\mu_{2q+j}\left(\frac{\tilde{\mathbf{E}}\tilde{\mathbf{E}}'}{mn}\right) \leq \mu_{q+1}\left(\frac{\tilde{\mathbf{E}}_1\tilde{\mathbf{E}}_1'}{mn}\right) + \mu_{q+j}\left(\frac{\tilde{\mathbf{E}}_2\tilde{\mathbf{E}}_2'}{mn}\right) = \mu_{q+j}\left(\frac{\tilde{\mathbf{E}}_2\tilde{\mathbf{E}}_2'}{mn}\right) \tag{A.23}$$

because $\mu_{q+1}(\tilde{\mathbf{E}}_1\tilde{\mathbf{E}}_1') = 0$ by construction. Equations (A.21) and (A.24) imply that

$$\mu_{2q+j}\left(\frac{\tilde{\mathbf{E}}\tilde{\mathbf{E}}'}{mn}\right) \leq \tilde{\nu}_{n,q+j}(\omega_\ell) \leq \mu_1\left(\frac{\tilde{\mathbf{E}}\tilde{\mathbf{E}}'}{mn}\right) \tag{A.24}$$

for $j = 1, \ldots, m - 2q$, concluding the proof. *Q.E.D.*

PROOF LEMMA 6: Follows from Lemmas 4 and 5. *Q.E.D.*

PROOF LEMMA 7: Using the decomposition in equation (A.17) we write

$$|\tilde{\nu}_{nk}(\omega_\ell) - \tilde{\nu}^{\chi}_{nk}(\omega_\ell)| = \left|\sigma_k^2\left(\frac{\tilde{\mathbf{X}}}{\sqrt{mn}}\right) - \sigma_k^2\left(\frac{\hat{\mathbf{\Lambda}}_n(\ell)\hat{\mathbf{F}}}{\sqrt{mn}}\right)\right|$$

$$\leq \left|\sigma_k^2\left(\frac{\tilde{\mathbf{X}}}{\sqrt{mn}}\right) - \sigma_k^2\left(\frac{\hat{\mathbf{\Lambda}}_n(\ell)\hat{\mathbf{F}}_1 + \tilde{\mathbf{E}}_1}{\sqrt{mn}}\right)\right| + \left|\sigma_k^2\left(\frac{\hat{\mathbf{\Lambda}}_n(\ell)\hat{\mathbf{F}}_1 + \tilde{\mathbf{E}}_1}{\sqrt{mn}}\right) - \sigma_k^2\left(\frac{\hat{\mathbf{\Lambda}}_n(\ell)\hat{\mathbf{F}}_1}{\sqrt{mn}}\right)\right| \tag{A.25}$$

We first show that, for $k = 1, \ldots, q$

$$\begin{aligned}&\left|\sigma_k^2\left(\frac{\tilde{\mathbf{X}}}{\sqrt{mn}}\right) - \sigma_k^2\left(\frac{\hat{\mathbf{\Lambda}}_n(\ell)\hat{\mathbf{F}}_1 + \tilde{\mathbf{E}}_1}{\sqrt{mn}}\right)\right| \\ \leq &\sigma_1\left(\frac{\tilde{\mathbf{E}}_2}{\sqrt{mn}}\right)\left[2\sigma_k\left(\frac{\hat{\mathbf{\Lambda}}_n(\ell)\hat{\mathbf{F}}_1 + \tilde{\mathbf{E}}_1}{\sqrt{mn}}\right) + \sigma_1\left(\frac{\tilde{\mathbf{E}}_2}{\sqrt{mn}}\right)\right] = O_p\left(\frac{1}{\sqrt{m}}\right).\end{aligned} \tag{A.26}$$

The first inequality follows from Lemma 16. From the inspection of the proof of Lemma 4, we have

$$\sigma_1^2\left(\frac{\tilde{\mathbf{E}}_2}{\sqrt{mn}}\right) = \mu_1\left(\frac{\tilde{\mathbf{E}}_2\tilde{\mathbf{E}}_2'}{mn}\right) \leq \frac{1}{m}\mu_1\left(\frac{\tilde{\mathbf{E}}\tilde{\mathbf{E}}'}{n}\right) = O_p\left(\frac{1}{m}\right). \tag{A.27}$$

By Lemma 2, for $k = 1, \ldots, q$,

$$\sigma_k^2\left(\frac{\hat{\mathbf{\Lambda}}_n(\ell)\hat{\mathbf{F}}_1}{\sqrt{mn}}\right) = \mu_k\left(\frac{\hat{\mathbf{\Lambda}}_n(\ell)\hat{\mathbf{F}}\hat{\mathbf{F}}'\hat{\mathbf{\Lambda}}_n'(\ell)}{mn}\right) = O_p(1), \tag{A.28}$$

because $\hat{\mathbf{F}}_1\hat{\mathbf{F}}_1' = \hat{\mathbf{F}}\hat{\mathbf{F}}'$. Hence, by Lemma 15

$$\left|\sigma_k\left(\frac{\hat{\mathbf{\Lambda}}_n(\ell)\hat{\mathbf{F}}_1 + \tilde{\mathbf{E}}_1}{\sqrt{mn}}\right) - \sigma_k\left(\frac{\hat{\mathbf{\Lambda}}_n(\ell)\hat{\mathbf{F}}_1}{\sqrt{mn}}\right)\right| \leq \sigma_1\left(\frac{\hat{\mathbf{E}}_1}{\sqrt{mn}}\right) = O_p\left(\frac{1}{\sqrt{m}}\right) \tag{A.29}$$

because, by Lemma 11 and equation A.27

$$\sigma_1^2\left(\frac{\tilde{\mathbf{E}}_1}{\sqrt{mn}}\right) \leq \mu_1\left(\frac{\tilde{\mathbf{E}}_1\tilde{\mathbf{E}}_1' + \tilde{\mathbf{E}}_2\tilde{\mathbf{E}}_2'}{mn}\right) = O_p\left(\frac{1}{m}\right). \tag{A.30}$$

The bound in (A.26) follows from (A.27)-(A.30).

Using similar arguments, it follows that

$$\begin{aligned}&\left|\sigma_k^2\left(\frac{\hat{\mathbf{\Lambda}}_n(\ell)\hat{\mathbf{F}}_1 + \tilde{\mathbf{E}}_1}{\sqrt{mn}}\right) - \sigma_k^2\left(\frac{\hat{\mathbf{\Lambda}}_n(\ell)\hat{\mathbf{F}}_1}{\sqrt{mn}}\right)\right| \\ \leq &\sigma_1\left(\frac{\tilde{\mathbf{E}}_1}{\sqrt{mn}}\right)\left[2\sigma_k\left(\frac{\hat{\mathbf{\Lambda}}_n(\ell)\hat{\mathbf{F}}_1}{\sqrt{mn}}\right) + \sigma_1\left(\frac{\tilde{\mathbf{E}}_1}{\sqrt{mn}}\right)\right] = O_p\left(\frac{1}{\sqrt{m}}\right).\end{aligned} \tag{A.31}$$

The proof follows from the right hand side of the inequality in (A.25), and the bounds derived in displays (A.30) and (A.31). *Q.E.D.*

PROOF LEMMA 8: By Lemma 16,

$$|\hat{\nu}_{nk}(\omega_\ell) - \tilde{\nu}_{nk}(\omega_\ell)| \leq \sigma_1\left(\frac{\hat{\mathbf{X}} - \tilde{\mathbf{X}}}{\sqrt{mn}}\right)\left[2\sigma_k\left(\tilde{\mathbf{X}}\right) + \sigma_1\left(\frac{\hat{\mathbf{X}} - \tilde{\mathbf{X}}}{\sqrt{mn}}\right)\right]$$

We first note that, by Lemma 14, Lemmas 1 and 3,

$$\sigma_1\left(\frac{\hat{\mathbf{X}}}{\sqrt{mn}}-\frac{\tilde{\mathbf{X}}}{\sqrt{mn}}\right)\leq\sigma_1\left(\frac{\hat{\boldsymbol{\chi}}-\hat{\boldsymbol{\Lambda}}_n(\ell)\hat{\mathbf{F}}}{\sqrt{mn}}\right)+\sigma_1\left(\frac{\hat{\mathbf{E}}-\tilde{\mathbf{E}}}{\sqrt{mn}}\right)=O_p\left(\frac{1}{\sqrt{T}}\right)$$

For $k\leq q$ the result follows by noting that, Lemmas 7 and 2 entail that $\sigma_k\left(\tilde{\mathbf{X}}\right)=O_p(1)$. For $k>q$, by Lemma 6 we conclude that $\sigma_k\left(\tilde{\mathbf{X}}\right)=O_p(m^{-1/2})$, completing the proof. *Q.E.D.*

APPENDIX B: ONLINE APPENDIX

B.1. *Auxiliary Lemmas*

This section introduces auxiliary results that prove the main Lemmas in Section A.2 and Theorem 1 in Section A.3.

LEMMA 9: *For any* $(n \times m)$ *matrix* $\mathbf{A}$, $\| \mathbf{A} \| = \sigma_1^2(\mathbf{A}) \leq \sum_{i=1}^{n} \sum_{j=1}^{m} |a_{ij}|^2$.

PROOF: See Horn and Johnson (1990), p. 421. *Q.E.D.*

LEMMA 10: *For* $\mathbf{A}, \mathbf{B}$ $(n \times n)$ *Hermitian matrices,* $\mu_{i+j-1}(\mathbf{A}+\mathbf{B}) \leq \mu_i(\mathbf{A}) + \mu_j(\mathbf{B})$, *for* $i + j \leq n$.

PROOF: See Bernstein (2009), Fact 8.4.11. p. 470. *Q.E.D.*

LEMMA 11: *For* $\mathbf{A}(n \times n)$ *Hermitian and* $\mathbf{B}(n \times n)$ *Hermitian positive-semi-definite,* $\mu_i(\mathbf{A}) \leq \mu_i(\mathbf{A}+\mathbf{B})$, *for* $i = 1, \ldots, n$.

PROOF: See Lütkepohl (1996), p. 135. *Q.E.D.*

LEMMA 12: *For* $\mathbf{A}, \mathbf{B}$ $(n \times n)$ *Hermitian, with rank*$(\mathbf{B}) \leq r$, $\mu_{i+r}(\mathbf{A}+\mathbf{B}) \leq \mu_i(\mathbf{A})$, *for* $i = 1, \ldots, n-r$.

PROOF: See Lütkepohl (1996), p. 135. *Q.E.D.*

LEMMA 13: *Let* $\mathbf{A}, \mathbf{B}$ $(n \times n)$ *matrices with* $\mathbf{A}$ *Hermitian and* $\mathbf{B}$ *nonsingular. Then, there exists a non-negative real number* δ_k, $k = 1, 2, \ldots n$, *such that* $\mu_n(\mathbf{B}\,\mathbf{B}') \leq \delta_k \leq \mu_1(\mathbf{B}\,\mathbf{B}')$ *and* $\mu_k(\mathbf{B}\,\mathbf{A}\,\mathbf{B}') = \delta_k \mu_k(\mathbf{A})$.

PROOF: See Theorem 4.5.9 in Horn and Johnson (1990) *Q.E.D.*

LEMMA 14: *Let* $\mathbf{A}, \mathbf{B}$ $(n \times m)$ *matrices, and* $r \equiv \min\{n, m\}$. *For decreasingly ordered singular values of* $\mathbf{A}$, $\mathbf{B}$ *and* $\mathbf{A}+\mathbf{B}$, $\sigma_1(\mathbf{A}+\mathbf{B}) \leq \sigma_1(\mathbf{A}) + \sigma_1(\mathbf{B})$

LEMMA 15: *For* $\mathbf{A}, \mathbf{B}, r$ *as in Lemma 14,* $|\sigma_i(\mathbf{A}+\mathbf{B}) - \sigma_i(\mathbf{A})| \leq \sigma_1(\mathbf{B})$, *for* $i = 1, \ldots, r$.

PROOF: See Lemma 3.3.16.(c) in Horn and Johnson (1991). *Q.E.D.*

LEMMA 16: *For* $\mathbf{A}, \mathbf{B}, r$ *as in Lemma 14,*

$$|\sigma_i^2(\mathbf{A}+\mathbf{B}) - \sigma_i^2(\mathbf{A})| \leq \sigma_1(\mathbf{B}) \left(2\sigma_i(\mathbf{A}) + \sigma_1(\mathbf{B})\right), \qquad i = 1, \ldots, r.$$

PROOF: The proof follows Onatski (2009), Proof of Lemma 3. By Lemma 15

$$\begin{aligned} |\sigma_i^2(\mathbf{A}+\mathbf{B}) - \sigma_i^2(\mathbf{A})| &\leq |\sigma_i(\mathbf{A}+\mathbf{B}) - \sigma_i(\mathbf{A})||\sigma_i(\mathbf{A}+\mathbf{B}) + \sigma_i(\mathbf{A})| \\ &\leq \sigma_1(\mathbf{B}) \left(2\sigma_i(\mathbf{A}) + |\sigma_i(\mathbf{A}+\mathbf{B}) - \sigma_i(\mathbf{A})|\right) \\ &\leq \sigma_1(\mathbf{B}) \left(2\sigma_i(\mathbf{A}) + \sigma_1(\mathbf{B})\right), \end{aligned}$$

proving the result. *Q.E.D.*

LEMMA 17: *Assume that the entries of* $\{y_{ij}\}$ *are a double array of iid complex random variables with mean zero, variance* σ^2*, and finite fourth moment. Let* $\boldsymbol{Y}_n = [y_{ij} : i \leq p, j \leq n]$ *be the* $p \times n$ *matrix of the upper-left corner of the double array. Let* $\boldsymbol{S}_n = n^{-1}\boldsymbol{Y}_n\boldsymbol{Y}_n'$ *and denote the eigenvalues of* $\boldsymbol{S}_n$ *by* $\mu_1 \geq \mu_2 \geq \cdots \geq \mu_p$*. Write* $\mu_{\max} = \mu_1(\boldsymbol{S}_n)$ *and*

$$\mu_{\min} = \begin{cases} \mu_p(\boldsymbol{S}_n), \text{ if } p \leq n \\ \mu_n(\boldsymbol{S}_n), \text{ if } p > n \end{cases}$$

If $p/n \to y \in (0, \infty)$*, we have a.s.*

$$\lim_{n\to\infty} \mu_{\min} = \sigma^2 \left(1 - \sqrt{y}\right)^2 \qquad \textit{and} \qquad \lim_{n\to\infty} \mu_{\max} = \sigma^2 \left(1 + \sqrt{y}\right)^2.$$

PROOF: See (Bai and Silverstein, 2010, Theorem 5.10) and Bai and Yin (1993) (Theorem 1 and Remark 1 therein). *Q.E.D.*

B.2. *Analogies with Angeletos et al. (2020)*

In this appendix we discuss the relationship of our methodology and empirical application with ACD's method and findings. First, let us emphasise that our methodology is complementary to the frequency-domain identification approach of ACD; that is, one can implement ACD, assuming that the world is GDFM, or alternatively our methodology can guide on how to implement their method after having estimated the number of shocks-per-frequency band. Second, we shed some light on the reasons our empirical results differ from ACD's findings.

B.2.1. *Analogies with ACD Methodology*

ACD proposes a frequency-domain identification method in the context of structural VAR models. The method allows for the identification of the shock, which maximizes the explained variance of a given variable on a specific frequency band, along with the corresponding IRFs. Our GDFM setup naturally couples with the ACD methodology, as long as the number of shocks q is greater than one. In particular, applying ACD's approach to (2.3), leads to identifying the column of the orthogonal matrix $\mathbf{Q}$, and thus the structural shock $\mathbf{f}_t^* = \mathbf{q}_{1k}'\mathbf{f}_t$ that explains the *largest* share of the variance of the k-th observable, relative to the frequency band $[\underline{\omega}, \bar{\omega}]$,$0 \leq \underline{\omega} \leq \bar{\omega} \leq \pi$. Precisely,

$$\mathbf{q}_{1k} = \underset{\mathbf{q}\ s.t.\mathbf{q}'\mathbf{q}=1}{\arg\max}\ \mathbf{q}' \left(\int_{\underline{\omega}}^{\bar{\omega}} \boldsymbol{\Lambda}_{nk}(e^{i\omega}) \boldsymbol{\Lambda}_{nk}'(e^{-i\omega}) d\omega \right) \mathbf{q}, \tag{B.1}$$

where $\boldsymbol{\Lambda}_{nk}(L)$ is the k-th row of the $n \times q$ matrix of filters $\boldsymbol{\Lambda}_n(L)$.[18] The solution to problem (B.1), as ACD explains, is to set $\mathbf{q}_{1k}$ equal to the eigenvector corresponding to the largest eigenvalue of the matrix in brackets in (B.1). This implementation of ACD's method differs from ACD in two respects. First, it only considers the common components, netting out the possibly distorting effect of the idiosyncratic component $\mathbf{e}_t$ in (2.2). Second, it is carried out over q shocks, where q is estimated by a consistent criterion, rather than a number of shocks equal to the number of variables included in the VAR, which in turn is based on a discretionary choice of the researcher.[19]

[18] Note that $\mathbf{q}_{1k} = \mathbf{q}_1(k, \underline{\omega}, \bar{\omega})$, where the suffix $1k$ indicates that $\mathbf{q}_{1k}$ is the eigenvector corresponding to the largest eigenvalue for the kth observed variable. We do not make this dependence explicit for notation convenience.

[19] Typically, estimation of the GDFM (2.2) leads to a small q, implying a problem whose dimension is smaller than the one for ACD.

B.2.2. *Analogies with ACD Empirics*

Let us now comment on the relation between ACD and ours empirical results. ACD finds that (i) a single shock, dubbed the main business cycle (MBC) shock, is enough to explain most of the business cycle fluctuations in real activity variables, and (ii) the MBC is disconnected with both real activity in the long term and inflation fluctuations. Result (i) is not in conflict with our finding that there are two business cycle shocks: as explained below, our shocks are designed to explain *all* macroeconomic variables, not just real activity variables. Indeed, also in the ACD setting at least two shocks are needed to explain both real activity and inflation, precisely because the MBC shock is disconnected from inflation. As for (ii), focusing on real activity variables, we find that one of the shocks explains the dynamics for up to a five-year horizon, but it dissipates over a longer horizon, which is where the other shock kicks in. In this respect, we also find a clean disconnection between short cycles, on one hand, and a frequency band including both long cycles and the long run, on the other hand. This result is not completely out of line with respect to ACD's findings. The substantial difference with respect to ACD is that we find a *close* connection between our main business cycle shock and inflation.

As explained above, we could in principle apply equation (B.1) to rotate our shocks and identify a shock maximizing the explained variance in GDP growth at cyclical frequencies. By mixing our two shocks, the resulting MBC shock would have smaller effects on inflation than our first shock, because the inflationary effects of the first shock and the deflationary effects of the second shock would partially cancel out, leading us to results closer to ACD's findings. However, then we should arguably loose the long-run disconnection. Alternatively, we could apply ACD's technique to identify a long-run shock as the one maximizing the explained variance of, say, GDP growth, on a frequency band close to frequency zero. By doing this, however, we would end up with a shock very similar to the supply shock found here, because our supply shock already maximizes the explained variance in GDP growth in the long run.

To better understand our results and the difference from ACD's findings, consider that our variance decomposition is based on the principal-component series Brillinger (1981) introduces, where standard PCA analysis is generalized to a dynamic setting. The principal-component series, like ordinary principal components, maximizes the sum of the explained variances for all variables in the dataset. Hence, unlike ACD's method, (a) dynamic principal components target *all* cross-sectional units (including inflation indexes), not just real activity variables; moreover, (b) dynamic principal components target *all* frequencies, not just business cycle frequencies.[20] Point (a) helps understand why we do not get the price disconnection. Point (b) helps understand why our first shock captures long-run inflation variance and why our second shock captures long-run growth.[21]

B.2.3. *Analyzing Subset of Variables*

We obtain further insights about our results and the relation with ACD findings by applying our method to some *specially selected* subsets of the observable variables. Our focus is on GDP growth and inflation (as measured by the log differences of the GDP deflator), which are therefore included in all subsets. In addition to GDP growth and inflation, the first subset

[20] Other obvious differences with ACD are the dataset, the data treatment (our variables are transformed to reach stationarity), and the fact that we do not have a VAR model, which imposes constraints on the shocks.

[21] As a robustness check, ACD performs a principal component analysis of their data and finds that their first (ordinary) principal component explains little of inflation. This difference with our results depends on the different datasets used: in ACD's dataset, inflation is just one out of 10 variables; in our dataset there are 46 inflation variables out of 216 variables. We repeat the same robustness exercise from ACD, but with our dataset, and find that the first ordinary principal component of our dataset explains about 80% of inflation variance.

includes only variables related to real economic activity.[22] The second includes only inflation indexes.[23] The third includes both real activity and inflation series.[24] It turns out that for both the real activity and the inflation datasets we have just one shock according to our criteria (not the same one, clearly). For the third subset we identify two shocks. We then examine the percentage of explained variance by the first dynamic principal component for the first two subsets and the percentage of explained variance by the first two principal components for the third dataset. Figure B.1 reports these percentages, decomposed by frequency.

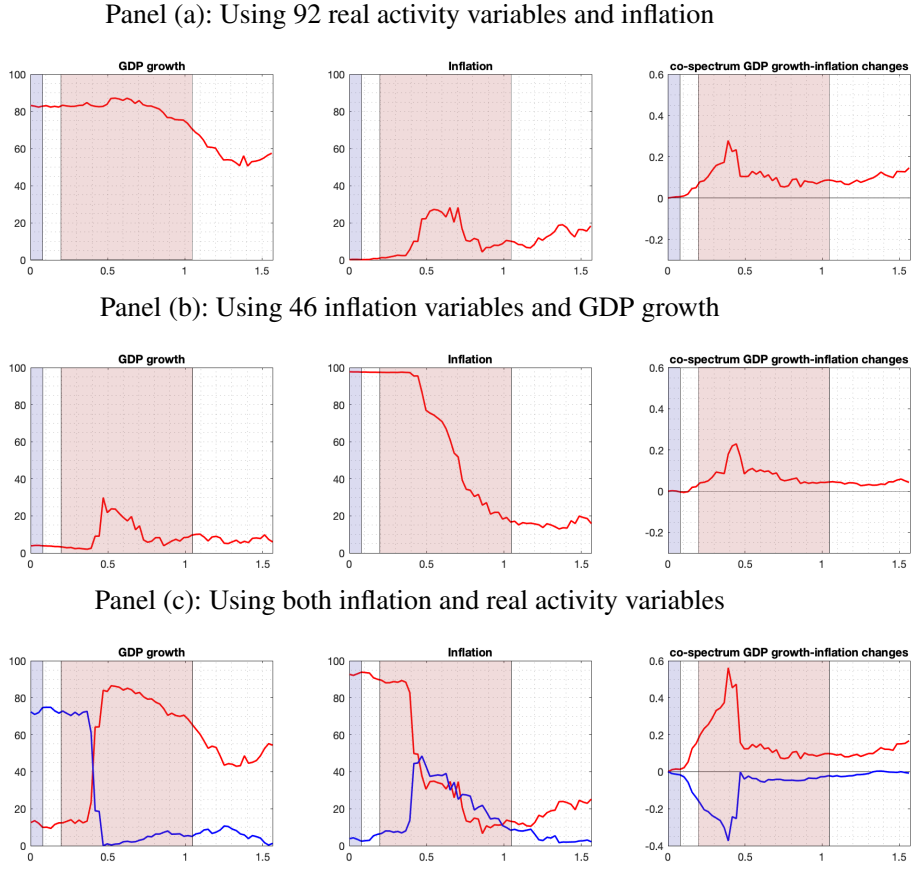

FIGURE B.1.—Top panels: percentage of spectra and co-spectra explained by the first dynamic principal component of real activity variables (subset one). Middle panels: percentage of spectra and co-spectra explained by the first dynamic principal component of inflation variables (subset two). Bottom panels: percentage of spectra and co-spectra explained by the first (red line) and the second (blue line) dynamic principal components of real activity and inflation variables (subset three). Lilac shadowed area: long-run frequencies; pink shadowed area: business cycle frequencies.

The top panels refer to real activity variables. By including only real activity variables we get close to ACD's method: the difference highlighted in point (a) above is removed. Here we get

[22] National account series, industrial production indexes, employment and unemployment series, sales and orders (92 variables out of 216). Including inflation, we have 93 variables.

[23] 46 variables out of 216. Including GDP growth, we have 47 variables.

[24] 138 variables out of 216, including GDP growth and inflation.

the price disconnection: the top-center panel shows that the real activity shock explains little of inflation at all frequencies. By contrast, the top-left panel demonstrates that the real activity shock explains a very high percentage of GDP growth at all frequencies.[25] This confirms ACD's finding that just a single shock captures most of real activity fluctuations, but it also shows that the same shock can explain both the business cycle and the long-run.

Why do we not get the long-run disconnection for this subset of variables? The basic reason is point (b) above: we are targeting all frequencies, not just the business cycle.[26]

By looking at the top-left panel of Figure B.1, one might interpret this shock as a real business cycle shock: a single shock explaining both the long-run and the business-cycle fluctuations of real activity variables. However, the top-right panel shows that the co-spectrum between GDP growth and inflation changes induced by this shock is positive. This result is not in line with the RBC interpretation; if the shock were a supply shock, the covariance should be negative. Below we provide an alternative interpretation of this shock.

The disconnection between prices and real activity is also evident in the middle panels of Figure B.1, which refers to the second (nominal) subset. The nominal shock explains most of inflation variance in the long-run and in the cyclical band, but little of GDP fluctuations (across all frequencies).

Let us now consider the third subset (bottom panels of Figure B.1). Because the nominal and real shocks appear almost orthogonal, it is not surprising that our criteria now detect two shocks for the third subset, which includes both real activity and inflation variables. This is the reason we have two lines: the red line refers to one shock, related to the first principal-component series; the blue line refers to the other shock. The picture is now very similar to the one in Figure 3, obtained with the whole dataset. Comparing the red and blue lines with the corresponding lines of the top and middle panels of the GDP and inflation columns, respectively, the GDP line is very much similar to the upper envelope of the blue and the red curves, whereas the inflation line is very much similar to the lower envelope. If we interpret the red line as demand and the blue line as supply, the *real* shock of the top panels is clearly a dynamic mixture of supply and demand, capturing mainly supply at long-run frequencies and mainly demand at all other frequencies. This is the reason the cospectrum of the top-right panel is positive when away from frequency zero.

B.3. *Frequency Domain Analysis of DSGE Models*

In Section 4.2 we use the dynamic-principal components and the related eigenvalues to estimate the number of shocks and analyze their effects on a large macroeconomic data set, in a Monte Carlo simulation setting. Here we depart from the large factor model framework and study two small data sets extracted from the data used in the empirical application of Section 4.1, corresponding to the ACD* and JPT models. Of course, in this context estimation of the number of shocks is less reliable; still, we can use dynamic-principal components as a descriptive statistic highlighting the frequency domain features of the empirical data: such features are often overlooked and cannot be studied with standard time-domain tools. In this

[25] It explains 58%, 82%, 87%, and 67% of business-cycle fluctuations of consumption, investment, unemployment and hours worked, respectively; as for the long run, the percentages are in order 67%, 78%, 80%, and 65%.

[26] To verify this, we apply ACD's method to the seven variables in Figure 3. By targeting unemployment cycle, as in ACD, we find that the MBC shock explains only 13% of long-run GDP growth, so that the long-run disconnection is there. By targeting GDP on the frequency band $[0, 2\pi/6]$, which includes both the long-run and the business cycle, we find that the MBC shock explains 67% of long-run growth and 70% of the GDP cycle: percentages not that much different from the ones in Figure B.1.

section we show how the dynamic principal components can verify whether and to what extent DSGE models replicate frequency-domain empirical evidence; we do so by comparing the explained variances of the dynamic-principal components from the variables generated by the model with those from their empirical counterparts. The number of observations for the simulated data is set to $T = 240$ (equal to the sample size of the empirical data).

The Justiniano et al. (2010) DSGE Model In the first exercise we consider eight variables in the JPT model: GDP, Consumption, Investment, Hours worked, Labor productivity, the nominal interest rate, inflation and TFP.

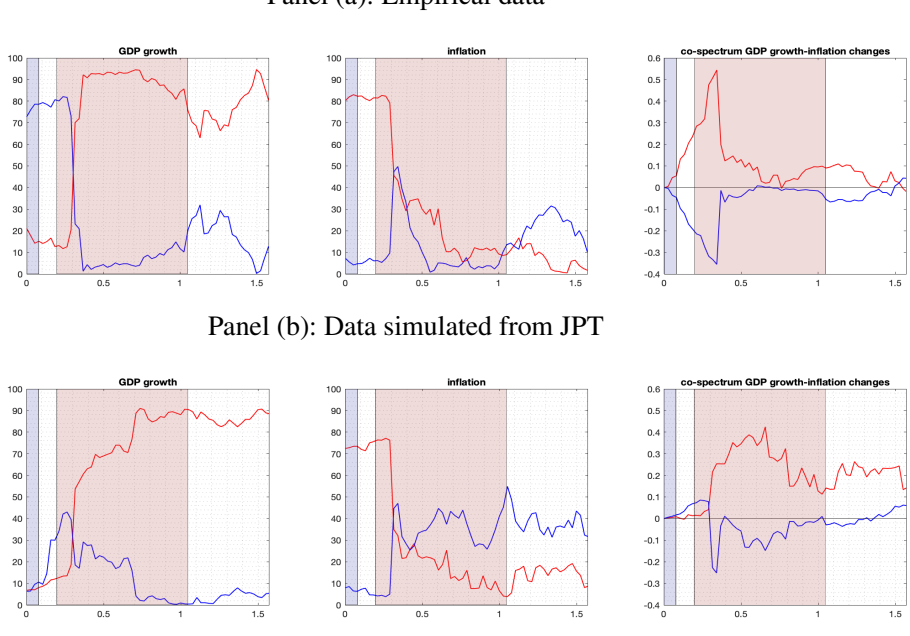

FIGURE B.2.—Top panels: Empirical data. Percentage of spectra of GDP growth and inflation explained by the first dynamic principal component (red line) and the second principal component (blue line). The top-right panel shows the co-spectrum of GDP growth and inflation changes explained by the two dynamic principal components. Bottom panels: JPT simulated data. Lilac shadowed area: long-run frequencies; pink shadowed area: business cycle frequencies.

Figure B.2 shows the percentage of the variance in GDP growth and inflation explained by the first two principal-component series. The right panels show the co-spectra of GDP growth and inflation changes driven by the first two principal components. For the empirical data (top panels), the picture is very much similar to the one for the large data set in Section 4.1 (Figure 3), and especially for the subset including both real activity and inflation variables in Section B.2.3 (see Figure B.1, bottom panels). The first principal component (red line) captures mainly transitory shocks inducing positive comovements between GDP changes and inflation changes; the second principal component (blue line) captures mainly permanent shocks inducing negative co-movements between GDP changes and inflation changes. This shows that the results in the main text are not an oddity, related to a particular large data set or to a particular treatment of variables; on the contrary, they can also be obtained with a reduced data set, similar to those commonly used in structural VAR and DSGE studies.

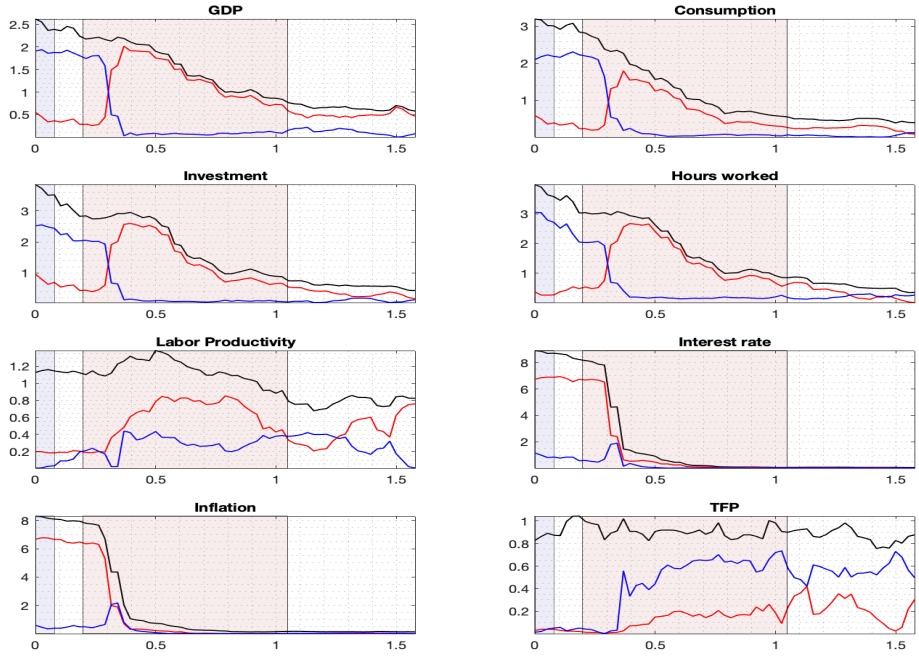

FIGURE B.3.—Real data. The estimated spectral density functions of the eight variables, namely GDP, Consumption, Investment, Hours worked, Labor productivity, the nominal interest rate, inflation and TFP (black line), and the components driven by the first principal component series (red line) and the second principal component series (blue line). Lilac shadowed area: long-run frequencies; pink shadowed area: business cycle frequencies.

As for the simulated data (bottom panels), the red line is similar to that of real data, whereas the blue line differ across the two rows of plots. The basic difference is that in the JPT model the blue line does not capture the long-run of GDP growth (which is instead explained by the third and the fourth principal components, not shown here). Our interpretation is that the JPT model is successful in reproducing the empirical features of demand sources of variation, but is less so in reproducing the supply side of the economy.

Figures B.3 and B.4 show the estimated spectral density functions of all variables, along with the components driven by the first and the second principal-component series (red line and blue line, respectively). Figure B.3 refers to empirical data; Figure B.4 refers to simulated data. The basic message is the same as in Figure B.2: the red lines of the two figures are similar to one another (albeit less so for hours and interest rate), whereas the blue lines are not as close (with the exception of investment).

The Angeletos et al. (2018) DSGE Model In the second exercise we compare real data with data simulated from the ACD* model. We consider seven variables: GDP, Consumption, Investment, Hours worked, Labor productivity, TFP, and the real interest rate. We do not include inflation, because the model is intended to fit real variables. Therefore, here we consider the first principal component only, because the second principal component is strongly related to inflation, which as said is not included in the data set.

Figure B.5 shows the percentage of the spectral density of GDP growth explained by the first dynamic principal component, frequency by frequency. The left panel refers to empirical data,

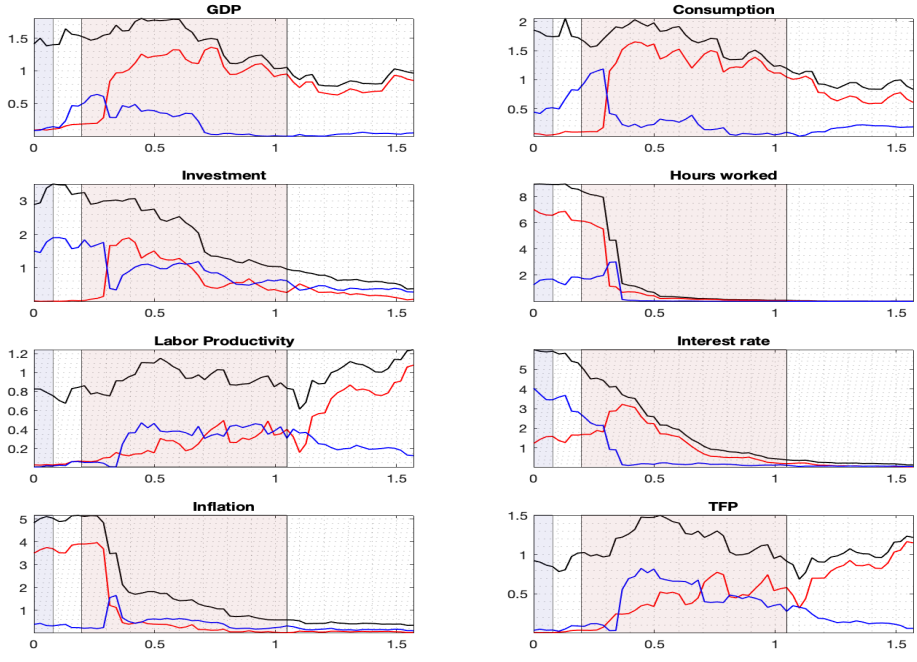

FIGURE B.4.—JPT simulated data. The estimated spectral density functions of eight variables, namely GDP, Consumption, Investment, Hours worked, Labor productivity, the nominal interest rate, inflation, and TFP (black line), and the components driven by the first principal component series (red line) and the second principal component series (blue line). Lilac shadowed area: long-run frequencies; pink shadowed area: business cycle frequencies.

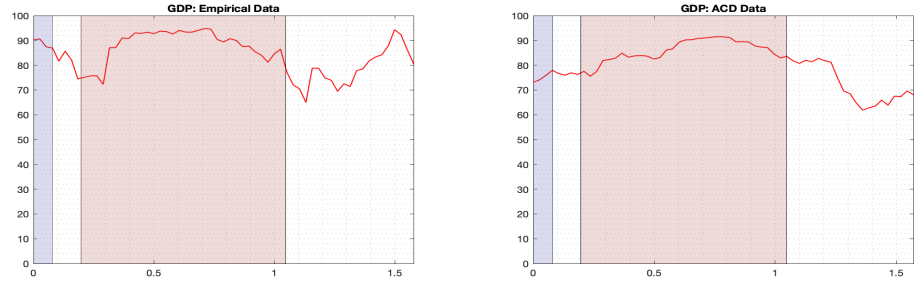

FIGURE B.5.—Percentage of the spectral density of GDP growth explained by the first dynamic principal component, frequency by frequency. Left panel: empirical data. Right panel: ACD* simulated data.

whereas the right panel refers to simulated data. Here the first principal component does not capture transitory sources of variation but arguably captures a mixture of long-run and short-run shocks: the left panel is very similar to the one in the main text obtained with the large data set (see Section B.2.3) including only real activity variables; see Figure B.1, top-left panel. This is not surprising, because here we do not have nominal variables. The basic message from the figure is that the right panel is very similar to the left panel, so that the ACD* model is very much in line with the data.

Figures B.6 and B.7 show the estimated spectral density functions of all variables (black lines) along with the components driven by the first principal-component series (red line). Figure B.6 refers to the empirical data; Figure B.7 refers to simulated data from ACD*. For GDP, consumption, investment, and labor productivity, the ACD* model is successful in reproducing the features of the data. In the empirical data the first principal component is poorly related to TFP and closely related to hours, whereas the opposite is true for the ACD* model. Another difference between the model and the data concerns the interest rate, which is not captured by the first principal component of the empirical data, whereas it appears well described for simulated data. Summing up, the model does a good job for GDP, consumption, investment, and labor productivity but less so for hours, TFP, and the interest rate.

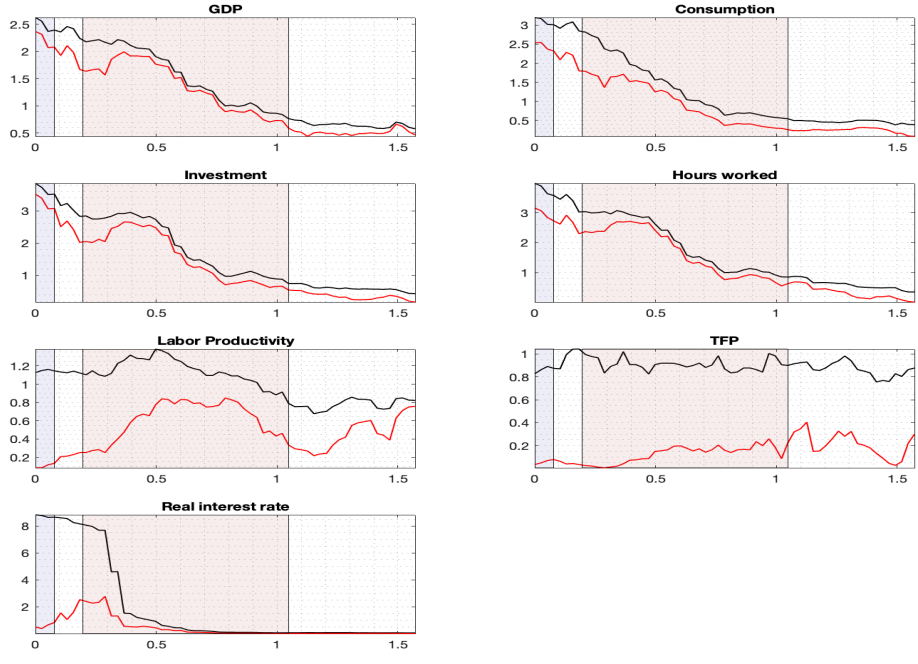

FIGURE B.6.—Empirical data. The estimated spectral density functions of seven variables (black line) and the components driven by the first principal component series (red line) and the second principal component series (blue line). Lilac shadowed area: long-run frequencies; pink shadowed area: business cycle frequencies.

B.4. *The DGPs of the DSGE Models*

For both models, we used the same DGPs that ACD use.[27] We transformed the variables produced by each model to get stationarity. To produce a large data set we proceeded as follows. Let the (transformed) variables produced by the model be the first n_1 common components $\boldsymbol{\chi}_{n_1 t}$. First, we standardized the entries of $\boldsymbol{\chi}_{n_1 t}$ and the shocks to get the r-dimensional vector

[27]The codes were kindly provided by Fabrice Collard.

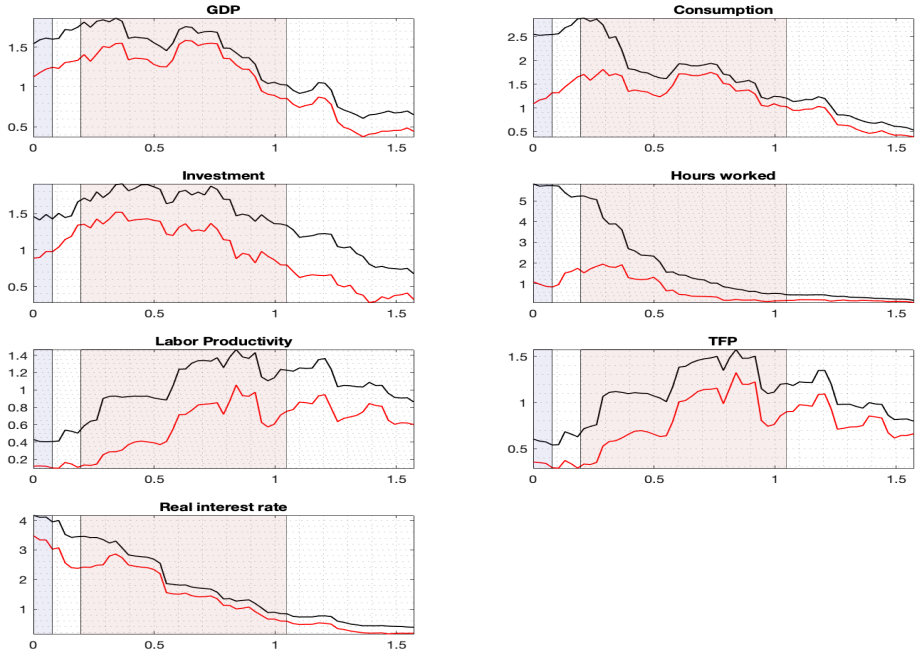

FIGURE B.7.—ACD* simulated data. The estimated spectral density functions of seven variables (black line) and the components driven by the first principal-component series (red line) and the second principal-component series (blue line). Lilac shadowed area: long-run frequencies; pink shadowed area: business cycle frequencies.

$\boldsymbol{F}_t$, $r = n_1 + q$. Second, we generated additional $n_2 = n - n_1$ common components (macro variables outside the model, sectoral or survey variables) as $\chi_{it} = a_{i1}F_{1t} + \cdots + a_{ir,}F_{rt}$, $i = n_1 + 1, \ldots, n$, where $a_{ik} \sim iid\, \mathcal{U}_{[-1,1]}$. Finally, we added the idiosyncratic components (measurement errors or sectoral components) e_{it}, generated according independent AR(1) models: $e_{it} = s(\sigma_{\chi_i}/\sigma_{\zeta_i})\zeta_{it}$, $\zeta_{it} = 0.5\zeta_{i,t-1} + \epsilon_{it}$, $\epsilon_{it} \sim \mathcal{N}(0,1)$. Here σ_{χ_i} and σ_{ζ_i} are the sample standard deviations of χ_{it} and ζ_{it}, respectively. In this way, s^2 is the idiosyncratic-to-common variance ratio of all variables. For both models we set $s^2 = 0.1, 0, 2$; (n, T)= $(60, 120)$.

To try also a different value of q, we shut down some of the shocks by setting their variance to zero. For ACD* we set $q = 4, 8$; for JPT we set $q = 4, 7$. $(120, 240)$. For all experiments we generated 500 artificial datasets and set $k_{\max} = 15$.

B.4.1. *JPT Model*

Let $\boldsymbol{\phi}_t$, be the vector whose entries are the theoretical variables in JPT. The DGP for these variables is given by

$$\boldsymbol{\phi}_{mt} = \boldsymbol{M}\boldsymbol{s}_t \tag{B.2}$$

$$\boldsymbol{s}_t = \boldsymbol{P}\boldsymbol{s}_{t-1} + \boldsymbol{J}\boldsymbol{S}\boldsymbol{f}_t, \tag{B.3}$$

where $\boldsymbol{\phi}_t$ is (11×1), $\boldsymbol{M}$ is (11×22), $\boldsymbol{s}_t$ is (22×1), $\boldsymbol{P}$ is (22×22), $\boldsymbol{s}_t$ is (22×1), $\boldsymbol{J}$ is (22×7), $\boldsymbol{S}$ is (7×7) and $\boldsymbol{f}_t$ is (7×1). Setting $\boldsymbol{P} = (\boldsymbol{P}_1 \; \boldsymbol{P}_2)$, the matrices are

$$\boldsymbol{M}' = \begin{pmatrix}
0.0031 & 0.0635 & -0.1940 & -0.1946 & 0.1977 & 0.0138 & -0.1839 & -0.0108 & -0.0308 & 1.0414 & 0.0265 \\
0.5672 & 0.7445 & -0.1453 & 0.5522 & 0.0150 & 0.0126 & -0.0024 & 0.1500 & 0.0179 & 85.9791 & 0.0937 \\
0.1588 & -0.0066 & 0.8010 & 0.1546 & 0.0042 & 0.0052 & 0.0010 & 0.0417 & 0.0041 & 37.5009 & 0.0262 \\
0.1693 & 0.0831 & 0.5005 & 0.1647 & 0.0046 & 0.0231 & 0.0185 & -0.1861 & 0.0295 & 75.9155 & 0.0275 \\
-0.5785 & -0.2839 & -1.7099 & -0.5628 & -0.0157 & -0.0791 & -0.0633 & 0.6360 & -0.1007 & -23.7064 & -0.0941 \\
-0.0671 & -0.0308 & -0.2031 & -0.0646 & -0.0024 & -0.1123 & -0.1099 & 0.0622 & 0.2103 & -116.0562 & -0.0087 \\
-0.0562 & -0.0402 & -0.1443 & -0.0597 & 0.0035 & 0.8220 & 0.8185 & 0.0207 & 0.0928 & -53.4098 & -0.0272 \\
0.0268 & 0.0131 & 0.0793 & 0.0261 & 0.0007 & 0.0040 & 0.0033 & -0.0093 & 0.0053 & 0.7339 & 0.0043 \\
0.0024 & 0.0012 & 0.0073 & 0.0024 & 0.0001 & -0.0016 & -0.0016 & -0.0986 & -0.0041 & -6.6639 & 0.0004 \\
0.0041 & 0.0020 & 0.0120 & 0.0040 & 0.0001 & 0.0002 & 0.0001 & -0.0259 & -0.0002 & 28.1988 & 0.0007 \\
-0.1693 & -0.0831 & -0.5005 & -0.1647 & -0.0046 & -0.0231 & -0.0185 & 0.1861 & -0.0295 & 0.6844 & -0.0275 \\
-0.0062 & -0.0044 & -0.0159 & -0.0066 & 0.0004 & 0.0904 & 0.0900 & 0.0023 & 0.0102 & 0.0938 & -0.0030 \\
-0.6313 & -0.7195 & -0.2780 & -0.4120 & -0.2193 & -0.8388 & -0.6196 & -0.0634 & -0.0916 & -0.0424 & -0.1125 \\
7.7656 & 12.5038 & -11.0450 & 7.5562 & 0.2093 & 0.7277 & 0.5184 & 2.4356 & 0.7781 & 1.9143 & 1.2704 \\
0.1057 & -0.0124 & 0.5639 & 0.1028 & 0.0028 & 0.0091 & 0.0063 & 0.0148 & 0.0079 & 0.0091 & 0.0173 \\
0.7968 & -0.2212 & -0.1591 & 0.7757 & 0.0211 & 0.0065 & -0.0146 & 0.0554 & 0.0224 & 0.0332 & 0.1318 \\
-3.2446 & -0.8070 & -12.5055 & -3.1277 & -0.1169 & -5.1797 & -5.0628 & 0.7111 & 4.0689 & -5.0282 & -0.4245 \\
-4.8425 & -6.9363 & 3.2748 & -4.7569 & -0.0856 & 6.9860 & 7.0716 & 1.4505 & 7.1071 & -11.3526 & -0.9544 \\
-0.8266 & -0.4055 & -2.4437 & -0.8041 & -0.0225 & -0.1146 & -0.0920 & 0.7358 & -0.1472 & 0.7564 & -0.1344 \\
2.4002 & 0.5494 & 9.4519 & 2.3164 & 0.0838 & 3.3851 & 3.3013 & -0.3291 & -2.4865 & -0.3291 & 0.3237 \\
4.4375 & 6.4319 & -3.3428 & 4.3507 & 0.0868 & -5.0126 & -5.0994 & -1.3220 & -6.4936 & -1.3220 & 0.8443 \\
1.0000 & 1.0000 & 1.0000 & 0.0000 & 1.0000 & 1.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 1.0000
\end{pmatrix},$$

$$\boldsymbol{P}_1 = \begin{pmatrix}
0.9646 & -0.0043 & 0.0238 & 0.0148 & -0.0507 & -0.0060 & -0.0043 & 0.0024 & 0.0002 & 0.0004 & -0.0148 \\
0.0635 & 0.7445 & -0.0066 & 0.0831 & -0.2839 & -0.0308 & -0.0402 & 0.0131 & 0.0012 & 0.0020 & -0.0831 \\
-0.1940 & -0.1453 & 0.8010 & 0.5005 & -1.7099 & -0.2031 & -0.1443 & 0.0793 & 0.0073 & 0.0120 & -0.5005 \\
0.0031 & 0.5672 & 0.1588 & 0.1693 & -0.5785 & -0.0671 & -0.0562 & 0.0268 & 0.0024 & 0.0041 & -0.1693 \\
-0.0108 & 0.1500 & 0.0417 & -0.1861 & 0.6360 & 0.0622 & 0.0207 & -0.0093 & -0.0986 & -0.0259 & 0.1861 \\
-0.0308 & 0.0179 & 0.0041 & 0.0295 & -0.1007 & 0.2103 & 0.0928 & 0.0053 & -0.0041 & -0.0002 & -0.0295 \\
0.0138 & 0.0126 & 0.0052 & 0.0231 & -0.0791 & -0.1123 & 0.8220 & 0.0040 & -0.0016 & 0.0002 & -0.0231 \\
-0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.9705 & -0.0201 & 0.0192 & 0.0000 \\
-0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0961 & 0.6560 & -0.0323 & 0.0000 \\
-0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0073 & -0.6780 & 0.6469 & 0.0000 \\
0.0000 & 0.0000 & 0.0000 & 0.0000 & -0.0000 & -0.0000 & 0.0000 & 0.0715 & 0.3866 & 0.1065 & -0.0000 \\
-0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 \\
0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 \\
0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 \\
0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 \\
0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 \\
0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 \\
0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 \\
0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 \\
0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000
\end{pmatrix},$$

$$\boldsymbol{P}_2 = \begin{pmatrix}
0.9646 & -0.0043 & 0.0238 & 0.0148 & -0.0507 & -0.0060 & -0.0043 & 0.0024 & 0.0002 & 0.0004 & -0.0148 \\
0.0635 & 0.7445 & -0.0066 & 0.0831 & -0.2839 & -0.0308 & -0.0402 & 0.0131 & 0.0012 & 0.0020 & -0.0831 \\
-0.1940 & -0.1453 & 0.8010 & 0.5005 & -1.7099 & -0.2031 & -0.1443 & 0.0793 & 0.0073 & 0.0120 & -0.5005 \\
0.0031 & 0.5672 & 0.1588 & 0.1693 & -0.5785 & -0.0671 & -0.0562 & 0.0268 & 0.0024 & 0.0041 & -0.1693 \\
-0.0108 & 0.1500 & 0.0417 & -0.1861 & 0.6360 & 0.0622 & 0.0207 & -0.0093 & -0.0986 & -0.0259 & 0.1861 \\
-0.0308 & 0.0179 & 0.0041 & 0.0295 & -0.1007 & 0.2103 & 0.0928 & 0.0053 & -0.0041 & -0.0002 & -0.0295 \\
0.0138 & 0.0126 & 0.0052 & 0.0231 & -0.0791 & -0.1123 & 0.8220 & 0.0040 & -0.0016 & 0.0002 & -0.0231 \\
-0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.9705 & -0.0201 & 0.0192 & 0.0000 \\
-0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0961 & 0.6560 & -0.0323 & 0.0000 \\
-0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0073 & -0.6780 & 0.6469 & 0.0000 \\
0.0000 & 0.0000 & 0.0000 & 0.0000 & -0.0000 & -0.0000 & 0.0000 & 0.0715 & 0.3866 & 0.1065 & -0.0000 \\
-0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 \\
0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 \\
0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 \\
0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 \\
0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 \\
0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 \\
0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 \\
0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000
\end{pmatrix},$$

$$\boldsymbol{J}' = \begin{pmatrix}
0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,1\,0\,0\,0\,0\,0\,0\,0\,0 \\
0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,1\,0\,0\,0\,0\,0\,0\,0 \\
0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,1\,0\,0\,0\,0\,0\,0 \\
0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,1\,0\,0\,0\,0\,0 \\
0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,1\,0\,0\,1\,0 \\
0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,1\,0\,0\,1 \\
0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,1\,0\,0
\end{pmatrix},$$

$\boldsymbol{S}$ is the diagonal matrix having on the diagonal the square roots of 0.88, 0.04, 6.03, 0.35, 0.14, 0.2, and 0.22, in that order. The entries of $\boldsymbol{f}_t$ are $\sim iid\, \mathcal{N}(0,1)$ so that $\boldsymbol{S}\boldsymbol{f}_t$ is a vector of serially and mutually independent Gaussian shocks with ordered variances 0.88, 0.04, 6.03, 0.35, 0.14, 0.2, and 0.22. Having $\boldsymbol{\phi}_t$, we take the first difference of the I(1) variables in $\boldsymbol{\phi}_t$ to get the vector

of stationary variables $\boldsymbol{\chi}_{n_1 t}$ and construct a large data set as explained above. The version of the model with $q = 4$ shocks is realized by setting to zero the last three diagonal entries of S.

B.4.2. *ACD* Model*

Let $\boldsymbol{\phi}_t$ be the vector whose entries are the theoretical variables in ACD*. The DGP for these common components is given by

$$\boldsymbol{\phi}_t = \boldsymbol{M}\boldsymbol{s}_t \tag{B.4}$$

$$\boldsymbol{s}_t = \boldsymbol{P}\boldsymbol{s}_{t-1} + \boldsymbol{J}\boldsymbol{S}\boldsymbol{f}_t, \tag{B.5}$$

where where $\boldsymbol{\phi}_t$ is (10×1), $\boldsymbol{M}$ is (10×11), $\boldsymbol{s}_t$ is (11×1), $\boldsymbol{P}$ is (11×11), $\boldsymbol{s}_t$ is (11×1), $\boldsymbol{J}$ is (11×8), $\boldsymbol{S}$ is (8×8) and $\boldsymbol{f}_t$ is (8×1). The matrices are

$$\boldsymbol{M} = \begin{pmatrix}
0.0134 & 0.4573 & 0.1389 & 0.5280 & 0.0535 & 0.2577 & 0.0147 & 0.0564 & -0.0642 & 0.1734 & 0.2772 \\
0.0527 & 0.8063 & -0.0164 & 0.2128 & -0.0063 & 0.2447 & -0.0191 & -0.0067 & 0.0076 & -0.0205 & 0.2063 \\
-0.1026 & -0.2052 & 0.8027 & 0.6831 & 0.3089 & 0.5783 & 0.5221 & 0.3259 & -0.3709 & -0.0753 & 0.8107 \\
-0.3344 & 0.6185 & 0.1879 & -0.6383 & 0.0723 & -1.0039 & 0.0198 & 0.0763 & -0.0868 & 0.2346 & 0.3749 \\
0.3477 & -0.1612 & -0.0490 & 1.1663 & -0.0188 & 1.2616 & -0.0052 & -0.0199 & 0.0226 & -0.0611 & -0.0977 \\
-0.1228 & 0.0569 & 0.0173 & 1.0657 & 0.0067 & 1.0321 & -0.1226 & 0.0070 & -0.0080 & 0.0216 & 0.0345 \\
0.0904 & 0.1244 & 0.1571 & -0.5029 & -0.7180 & -0.0607 & 0.3785 & -0.0524 & 0.1052 & 0.0400 & -0.2987 \\
0.0000 & 0.0000 & 0.0000 & 1.3524 & -0.0000 & -0.0000 & -0.3524 & -0.0000 & -0.0000 & 1.0000 & -0.0000 \\
-0.4711 & 0.2183 & 0.0663 & 0.2521 & 0.0255 & 0.1230 & -0.4704 & 0.0269 & -0.0306 & 0.0828 & 0.1324 \\
0.0134 & 0.4573 & 0.1389 & 0.5280 & 0.0535 & 0.2577 & 0.0147 & 0.0564 & -0.0642 & 0.1734 & 0.2772
\end{pmatrix},$$

$$\boldsymbol{P} = \begin{pmatrix}
0.9724 & -0.0051 & 0.0201 & 0.0171 & 0.0077 & 0.0145 & -0.0119 & 0.0331 & -0.0093 & -0.0019 & 0.0203 \\
0.0527 & 0.8063 & -0.0164 & 0.2128 & -0.0063 & 0.2447 & -0.0191 & -0.0067 & 0.0076 & -0.0205 & 0.2063 \\
-0.1026 & -0.2052 & 0.8027 & 0.6831 & 0.3089 & 0.5783 & -0.4779 & 0.3259 & -0.3709 & -0.0753 & 0.8107 \\
0.0000 & 0.0000 & 0.0000 & 1.0000 & 1.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 \\
0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.3088 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 \\
0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.3917 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 \\
0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 1.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 \\
0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.3643 & 0.0000 & 0.0000 & 0.0000 \\
0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.4792 & 0.0000 & 0.0000 \\
0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.7875 & 0.0000 \\
0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.6204
\end{pmatrix},$$

$$\boldsymbol{J}' = \begin{pmatrix}
0\,0\,0\,1\,0\,0\,0\,0\,0\,0\,0 \\
0\,0\,0\,0\,1\,0\,0\,0\,0\,0\,0 \\
0\,0\,0\,0\,0\,1\,0\,0\,0\,0\,0 \\
0\,0\,0\,0\,0\,0\,1\,0\,0\,0\,0 \\
0\,0\,0\,0\,0\,0\,0\,1\,0\,0\,0 \\
0\,0\,0\,0\,0\,0\,0\,0\,1\,0\,0 \\
0\,0\,0\,0\,0\,0\,0\,0\,0\,1\,0 \\
0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,1
\end{pmatrix},$$

and $\boldsymbol{S}$ is the diagonal matrix having on the diagonal the square roots of 0.3959, 0.3764, 0.3385, 0.8457, 5.9607, 0.6579, 1.6752, and 1.7975, in that order. The eight entries of $\boldsymbol{f}_t$ are $\sim iid\, \mathcal{N}(0,1)$ so that $\boldsymbol{S}\boldsymbol{f}_t$ is a vector of serially and mutually independent Gaussian shocks with ordered variances 0.3959, 0.3764, 0.3385, 0.8457, 5.9607, 0.6579, 1.6752, and 1.7975. Having $\boldsymbol{\phi}_t$, we take the first difference of the I(1) variables in $\boldsymbol{\phi}_t$ to get the vector of stationary variables $\boldsymbol{\chi}_{n_1 t}$ and construct a large data set as explained above. The version of the model with $q = 4$ shocks is realized by setting to zero the last four diagonal entries of S.

B.5. *Monte Carlo Simulations*

To evaluate the performance of the criteria defined in Section 2, we run several Monte Carlo experiments. In Section B.6 we use three different specifications of model (2.2) to compare the performances of DDR, DER, and DGR with each other and with HL and O. The conclusion is that DDR dominates DGR and DER, and it performs comparably to, or even better than, HL and O.

As already observed, we can evaluate DDR (like DER and DGR) at a specific frequency of interest or on a frequency band. Whereas Section B.6 looks at single frequencies, Section B.7 considers a fourth specification, with frequency bands, where data-generating processes (DGPs) exhibits the spectral density matrix of the common components with reduced rank at specific frequencies. We show that DDR, when evaluated at these frequencies, can detect the rank reduction with reasonable accuracy.

A further Monte Carlo experiment demonstrates the ability of our methodology to identify the number of structural shocks in DSGE models (see Section 4.2.1 in the manuscript).

B.6. *Simulations: Analysis without Frequency Bands*

B.6.1. *Simulation Design*

First experiment (HL). The first DGP follows Hallin and Liska (2007), Section 5. Precisely:

(i) The common shocks f_{jt}, $j = 1, \dots, q$, $t = 1, \dots, T$, $q \leq 3$, are $\sim iid\, \mathcal{N}(0, D_j)$, with $D_1 = 1$, $D_2 = .5$ and $D_3 = 1.5$.

(ii) The idiosyncratic components are of form $e_{it} = \sum_{l=0}^{4} \sum_{k=0}^{2} g_{i,l,k}\, \varepsilon_{i+l,t-k}$, where the ε_{it} are $\sim iid\, \mathcal{N}(0,1)$, and the $g_{i,l,k}$ are $\sim iid\, \mathcal{U}_{[1,1.5]}$, with $\mathcal{U}_{[a,b]}$ indicating a uniformly distributed random variable between a and b, where $i = 1, \dots, n$, $t = 1, \dots, T$, $l = 1, \dots, 4$, and $k = 0, 1, 2$. The ε_{it} and the $g_{i,l,k}$ are mutually independent and independent of f_{jt}. Hence, the idiosyncratic components are both autocorrelated and locally cross-correlated.

(iii) The filters $\lambda_{ij}(L)$, $i = 1, \dots, n$, $j = 1, \dots, q$, are randomly generated (independently from the f_{jt} and e_{it}) according to one of the following time-series structures: (a) MA(2) loadings: $\lambda_{ij}(L) = \lambda_{ij,0} + \lambda_{ij,1} L + \lambda_{ij,2} L^2$ with iid and mutually independent coefficients $(\lambda_{ij,0}, \lambda_{ij,1}, \lambda_{ij,2}) \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I}_3)$; (b) AR(2) loadings: $\lambda_{ij}(L) = m_{ij,0}(1 - m_{ij,1}L)^{-1}(1 - m_{ij,2}L)^{-1}$ with iid and mutually independent coefficients $m_{ij,0} \sim \mathcal{N}(0,1), m_{ij,1} \sim \mathcal{U}_{[0.8,0.9]}$ and $m_{ij,2} \sim \mathcal{U}_{[0.5,0.6]}$.

Finally, for each i, the variance of each idiosyncratic component e_{it} and that of the corresponding common component $\chi_{it} = \sum_{j=1}^{q} \lambda_{ij}(L) f_{jt}$ are normalized to 0.5.

The artificial samples were generated with $q = 2, 3$ and (n, T)= $(60, 100)$, $(100, 100)$, $(70, 120)$, $(120, 120)$, and $(150, 120)$. Notice that in this experiment we have both large and small factors (the variance of the second factor is one-half of that of the first factor and one-third of the variance of the third factor).

Second experiment (O). The second DGP is the one Onatski (2009) studies, Sections 5.1 and 5.3. The basic difference for the previous DGP is that here the number of common shocks is fixed to $q = 2$, whereas the variance of the idiosyncratic components takes different values.

Precisely:

(i) The common shocks f_{jt}, $k = 1, \dots, q$, $t = 1, \dots, T$, are $\sim iid\, \mathcal{N}(0,1)$.

(ii) The idiosyncratic components follow AR(1) processes both cross-sectionally and over time: $e_{it} = \rho_i\, e_{i,t-1} + v_{it}$, $v_{it} = \rho v_{i-1,t} + \epsilon_{it}$, where ρ_i is $\sim iid\, \mathcal{U}_{[-0.8,0.8]}$, $\rho = 0.2$ and ϵ_{it} is $\sim iid\, \mathcal{N}(0,1)$.

(iii) The filters $\lambda_{ij}(L)$, $i = 1, \dots, n$ and $j = 1, \dots, q$, are randomly generated (independently from f_{jt} and e_{it}) by one of the following devices: (a) MA(2) loadings: $\lambda_{ij}(L) = m_{ij,0}(1 + m_{ij,1}L)(1 + m_{ij,2}L)$ with iid and mutually independent coefficients $m_{ij,0} \sim \mathcal{N}(0,1), m_{ij,1} \sim \mathcal{U}_{[0,1]}$ and $m_{ij,2} \sim \mathcal{U}_{[0,1]}$; (b) AR(2) loadings, same as in the first experiment; $\lambda_{ij}(L) = m_{ij,0}(1 - m_{ij,1}L)^{-1}(1 - m_{ij,2}L)^{-1}$ with iid and mutually independent coefficients $m_{ij,0} \sim \mathcal{N}(0,1), m_{ij,1} \sim \mathcal{U}_{[0.8,0.9]}$ and $m_{ij,2} \sim \mathcal{U}_{[0.5,0.6]}$.

For each i, the idiosyncratic component e_{it} and the common component $\chi_{it} = \sum_{j=1}^{q} \lambda_{ij}(L) f_{jt}$ are normalized so that their variances equal σ^2 and 1, respectively. Hence, the idiosyncratic-to-common variance ratio is σ^2 for all i. Following Onatski (2009), we set $q = 2$ and (n, T, σ^2) equal to $(70, 70, 1)$, $(70, 70, 2)$, $(70, 70, 4)$, $(100, 120, 1)$, $(100, 120, 2)$, $(100, 120, 6)$, $(150, 500, 1)$, $(150, 500, 8)$, $(150, 500, 16)$.

Third experiment. The main feature of this experiment is that, unlike the previous ones, the idiosyncratic-common variance ratio differs across different cross-sectional units. The loadings are ARMA$(1, 2)$ filters, and the number of common shocks is larger than in the previous experiments ($q = 2, 4, 6$). Precisely, the third GDP consists of the following:

(i) The common shocks f_{jt}, $j = 1, \ldots, q$, $t = 1, \ldots, T$, are $iid\mathcal{N}(0, 1)$.

(ii) Same as in the second experiment, where the idiosyncratic components follow AR(1) processes both cross-sectionally and over time, $e_{it} = \rho_i e_{it-1} + v_{it}$, $v_{it} = \rho v_{i-1t} + \epsilon_{it}$, where $\rho_i \sim iid\mathcal{U}_{[-0.8,0.8]}$, $\rho = 0.2$ and $\epsilon_{it} \sim iid\mathcal{N}(0, 1)$.

(iii) The filters $\lambda_{ij}(L)$, $i = 1, \ldots, n$ and $j = 1, \ldots, q$, are randomly generated (independently from the f_{jt} and e_{it}) with ARMA(1,2) loadings: $\lambda_{ij}(L) = (m_{ij,0} + m_{ij,1}L + m_{ij,2}L^2)/a_{ij,0}(1 - a_{ij,1}L)$, where the coefficients are iid and mutually independent, $m_{ij,s} \sim \mathcal{U}_{[-1,1]}$, $s = 0, 1, 2$, and $a_{ij,r} \sim \mathcal{U}_{[-0.8,0.8]}$, $r = 0, 1$.

In experiments one and two, all variables in the cross-section have the same common-idiosyncratic variance ratio. This is not the case with real data. In this third experiment, we want to control for the size of the idiosyncratic components without forcing the variables to have the same ratio. To this end, for each artificial dataset we compute the average sample variance of the common and the idiosyncratic components, say σ_χ^2 and σ_e^2. Then we multiply all common components by $1/\sigma_\chi$ and all idiosyncratic components by s/σ_e, with s taking on two values: (a) $s = 0.5$ (small idiosyncratic components) and (b) $s = 1$ (large idiosyncratic components). Because variables do not have the same variance, we standardize them before estimation.

We set $q = 2, 4, 6$ and (n, T)= $(60, 80)$, $(120, 80)$, $(60, 240)$, $(120, 240)$, $(240, 480)$.

Variation of Third experiment. The DGP is the same for the third experiment. The parameters q, n, and T are kept fixed, with values $q = 3$, $n = 100$, $T = 100$. The parameter s, which governs the average idiosyncratic variance, varies between 0.3 and 1.2.

B.6.2. *Calibrating the Criteria*

To compute DER, DGR, and DDR, we use the periodogram smoothing estimator (2.5) with the bandwidth parameter $M_T = [0.75\sqrt{T}]$ and take the average of the eigenvalues evaluated in the frequency grid $\omega_\ell = 2\pi\ell/T$, $\ell = 1, \ldots, T - 1$. As explained, in order to avoid denominators very close to 0, we construct the denominator of DDR by taking, for each k, the maximum between $\hat{\mu}_{n,k+1} - \hat{\mu}_{n,k+2}$ and the smallest non-zero eigenvalue $\hat{\nu}_{n,m}$. In our simulation, such correction is active in about 5% of the ratios.

We compare our criteria with HL and O. With regard to HL, we use the log information criterion $IC_{2;n}^T$ with penalty $p_1(n, T)$ and the Bartlett lag window with parameter $M_T = [0.75\sqrt{T}]$, which yield the best performance in the simulations the authors show. The method requires evaluation of the loss function over a grid n_j, T_j, $j = 1, \ldots, J$. We stick to the one the authors proposed (i.e., $n_j = n - 10j$, $T_j = T - 10j$, $j = 0, 1, 2, 3$).

When dealing with O, we use the procedure described in Section 5.3 of Onatski (2009). We found that the results are sensitive to the choice of the parameter m (Onatski, 2009, footnote 7). For the second experiment, we stick to Onatski's choice, which is very effective ($m = 30$ for

$(n,T) = (70,70)$, $m = 40$ for $(n,T) = (100,120)$, $m = 65$ for $(n,T) = (150,500)$). For the first DGP, we use $m = 15$; for the third experiment, we use $m = 15, 20, 30$ for $T = 80, 240, 480$, respectively. These values produce better results than the larger ones Onatski (2009) suggests.

For the three experiments and estimators, we set $q_{\max} = 8$. For all experiments, we generate 500 artificial datasets. We evaluate the results in terms of the percentage of times we find the correct number of shocks.

TABLE B.I

q	n	T	HL	O	DER	DGR	DDR
MA loadings							
2	60	100	**99.6**	47.6	80.8	92.4	98.4
	100	100	**99.6**	71.2	87.8	96.6	99.2
	70	120	97.8	53.2	87.2	96.2	**99.6**
	120	120	99.4	78.4	96.6	99.4	**100.0**
	150	120	99.6	83.0	97.4	99.4	**100.0**
3	60	100	62.0	15.8	27.2	52.6	**77.0**
	100	100	**91.2**	24.6	29.6	52.0	87.8
	70	120	89.8	18.8	32.8	62.2	**90.8**
	120	120	**99.0**	25.8	43.0	70.4	97.0
	150	120	**99.8**	29.6	48.0	74.2	97.4
AR loadings							
2	60	100	96.8	81.2	84.4	92.8	**98.8**
	100	100	**99.4**	93.0	90.4	94.6	**99.4**
	70	120	**99.8**	88.8	89.2	95.8	99.4
	120	120	**100.0**	96.8	94.4	97.4	**100.0**
	150	120	**100.0**	96.8	96.4	98.0	**100.0**
3	60	100	31.2	33.4	41.2	54.6	**67.6**
	100	100	62.4	54.4	49.0	66.0	**87.0**
	70	120	69.4	49.8	49.6	63.2	**81.0**
	120	120	90.8	67.6	57.8	70.0	**91.6**
	150	120	**94.6**	74.4	63.2	75.8	93.6

Note: First experiment described in Section B.6. Percentage of correct outcomes over 500 replications. HL: Hallin and Liška (2007) estimator, O: Onatski (2009) estimator, DER: Dynamic Eigenvalue Ratio estimator, DGR: Dynamic Growth Ratio estimator; DDR: Dynamic Difference Ratio estimator. Boldface numbers denote the estimator that performs best in each row.

B.6.3. *Simulation Results*

Table B.I reports results for the first experiment. Boldface numbers denote the estimator(s) that perform best for each q, n, T configuration. Results for HL are very close to those in Hallin and Liška (2007). With MA loadings and AR loadings, $q = 2$, HL and DDR perform similarly and dominate the other estimators. With AR loadings, $q = 3$, DDR clearly outperforms HL, with the exception of the cases $(n,T) = (120,120)$ and $(n,T) = (150,120)$, in which results are similar. DDR uniformly outperforms DGR, which in turn does better than DER and O. In this experiment the second factor is small as compared to the first and the third ones. Hence, we conclude that DDR is reasonably able to detect small factors.

Table B.II reports results for the second experiment. Results for O are close to those in Onatski (2009). With MA loadings, O is the best method for all n, T, σ^2 configurations, except for $(n,T,\sigma^2) = (70,70,2)$, which DDR beats. DDR is tied for first with O in five cases, performs slightly better than O in the case above, and ranks second in the remaining three cases. Again, DDR uniformly dominates DGR, which in turn dominates DER. With AR loadings, DDR performs the best for all n, T, σ^2 configurations. As Onatski (2009) notes, HL works well

TABLE B.II

n	T	σ^2	HL	O	DER	DGR	DDR
MA loadings							
70	70	1	**100.0**	**100.0**	99.6	99.8	**100.0**
70	70	2	94.6	99.8	89.2	94.0	**100.0**
70	70	4	1.6	**89.0**	49.4	58.2	77.6
100	120	1	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**
100	120	2	**100.0**	**100.0**	99.6	**100.0**	**100.0**
100	120	6	3.0	**98.8**	54.0	62.2	81.4
150	500	1	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**
150	500	8	**100.0**	**100.0**	99.2	99.8	**100.0**
150	500	16	40.2	**97.4**	45.6	47.0	88.8
AR loadings							
70	70	1	99.0	98.8	90.8	96.0	**99.4**
70	70	2	85.6	89.4	78.4	87.8	**98.4**
70	70	4	14.4	64.6	61.0	69.4	**84.8**
100	120	1	**100.0**	99.0	99.6	**100.0**	**100.0**
100	120	2	**100.0**	99.4	95.8	98.2	**100.0**
100	120	6	44.8	77.6	75.4	82.0	**91.4**
150	500	1	**100.0**	98.4	**100.0**	**100.0**	**100.0**
150	500	8	**100.0**	97.8	99.6	**100.0**	**100.0**
150	500	16	99.6	91.6	91.4	93.0	**99.8**

Note: Second experiment described in Section B.6 ($q = 2$). Percentage of correct outcomes over 500 replications. HL: Hallin and Liška (2007) estimator, O: Onatski (2009) estimator, DER: Dynamic Eigenvalue Ratio estimator, DGR: Dynamic Growth Ratio estimator; DDR: Dynamic Difference Ratio estimator. Boldface numbers denote the estimator that performs best in each row.

for low values of σ^2, but fails dramatically for noisy data, except for the AR case with $T = 500$. Onatski (2009) argues that the grid of the calibration procedure does not work with high values of σ^2.

Table B.III reports results for the third experiment. In the upper panel (small idiosyncratic components), DDR and DGR perform similarly and dominate all other estimators, except for the configuration $(q, n, T) = (4, 120, 80)$, where HL performs slightly better. O does not perform well for $q = 4, 6$, and small T. In the lower panel (large idiosyncratic components), DDR dominates DGR, which in turn dominates DER. No criterion correctly captures the number of common shocks for the (q, n, T) configurations $(4, 60, 120)$, $(4, 120, 80)$, $(6, 60, 120)$, $(6, 120, 80)$, and $(6, 60, 240)$. In the remaining rows, DDR works best in all cases except the $(6, 120, 240)$ configuration, for which HL performs slightly better. Again, O works reasonably well for $q = 2$ but not for $q = 4, 6$.

Surprisingly, HL fails for large n, T configurations, particularly for $q = 2$, which should be the simplest one. The fact that the performance deteriorates as n, T gets larger suggests that the calibration procedure for the penalty function does not work in these cases. Paradoxically, the procedure works much better when the idiosyncratic components are larger (lower panel). Here, the problem is not that the data is noisy; on the contrary, it is not noisy enough. To investigate this aspect we run the modification of the third experiment, illustrated above.

Table B.IV shows results for HL and DDR.[28] The table reports not only the percentage of correct outcomes but also the percentages of cases for which $\hat{q} < q$ and $\hat{q} > q$. HL performs very well for intermediate values of s but has problems for DGP's with both small and large idiosyncratic components. When s is very small, the number of common shocks is overestimated, whereas when s is very large the number of common shocks is underestimated. On the

[28] We do not include O, because our aim here is to verify whether HL fails for small values of s, as suggested by Table B.III.

TABLE B.III

q	n	T	HL	O	DER	DGR	DDR
Small idiosyncratic components							
2	60	120	99.2	98.8	**100.0**	**100.0**	**100.0**
	120	80	88.6	90.8	**100.0**	**100.0**	99.8
	60	240	99.6	**100.0**	**100.0**	**100.0**	**100.0**
	120	240	35.6	**100.0**	**100.0**	**100.0**	**100.0**
	240	480	0.6	**100.0**	**100.0**	**100.0**	**100.0**
4	60	120	97.6	64.8	88.8	97.0	**98.6**
	120	80	**89.2**	36.2	67.2	86.6	88.8
	60	240	**100.0**	94.6	**100.0**	**100.0**	**100.0**
	120	240	91.6	99.2	**100.0**	**100.0**	**100.0**
	240	480	73.4	**100.0**	**100.0**	**100.0**	**100.0**
6	60	120	8.6	23.8	17.8	56.4	**57.0**
	120	80	0.8	16.4	5.0	**26.6**	21.8
	60	240	96.8	54.0	96.4	99.2	**99.8**
	120	240	98.6	87.0	**100.0**	**100.0**	**100.0**
	240	480	91.6	99.8	**100.0**	**100.0**	**100.0**
Large idiosyncratic components							
2	60	120	97.2	87.2	85.8	93.4	**99.4**
	120	80	99.0	93.4	85.4	94.2	**99.8**
	60	240	**100.0**	99.0	98.8	99.4	**100.0**
	120	240	99.8	99.8	99.8	**100.0**	**100.0**
	240	480	99.8	**100.0**	**100.0**	**100.0**	**100.0**
4	60	120	0.0	16.8	2.0	7.6	**30.0**
	120	80	0.0	**15.2**	2.0	7.4	11.8
	60	240	16.8	32.4	29.4	55.8	**80.8**
	120	240	99.8	73.4	92.0	97.8	**100.0**
	240	480	96.8	**100.0**	**100.0**	**100.0**	**100.0**
6	60	120	0.0	**8.2**	0.0	0.0	0.8
	120	80	0.0	**9.0**	0.0	0.0	0.0
	60	240	0.0	**7.8**	0.0	0.4	7.6
	120	240	**48.0**	20.8	5.4	11.2	47.0
	240	480	94.8	94.8	**100.0**	**100.0**	**100.0**

Note: Third experiment described in Section B.6. Percentage of correct outcomes over 500 replications. HL: Hallin and Liška (2007) estimator, O: Onatski (2009) estimator, DER: Dynamic Eigenvalue Ratio estimator, DGR: Dynamic Growth Ratio estimator; DDR: Dynamic Difference Ratio estimator. Boldface numbers denote the estimator that performs best in each row.

other hand, DDR does not lose accuracy for small idiosyncratic components; when data becomes noisier, its performance deteriorates, but less so than HL. More specifically, DDR and HL have similar performances in the interval $0.50 \leq s \leq 0.80$; for all other values of s, DDR outperforms HL, and for extreme values of s the difference is very large. We conjecture that, by changing the grid n_j, T_j, $j = 1, \ldots, J$, over which to evaluate the loss function, HL performs better. However, in practice we do not know in advance how noisy the data is, so the choice of grid is a possible source of error.

Summing up, for experiments one to three, DDR dominates DGR and DER for almost all DGPs. In Experiment 1, with AR loadings, Experiment 2, with AR loadings, and Experiment 3, with large idiosyncratic components, DDR has the best performance for almost all parameter configurations. For Experiment 2, with MA loading, O has the best performance, but DDR ranks either first or second in all cases. In Table B.IV, DDR generally outperforms HL. For Experiment 1, with MA loadings, DDR and HL perform similarly and dominate the other estimators. For Experiment 3, with small idiosyncratic components, DDR and DGR perform similarly and dominate the other estimators. We conclude that DDR is preferable to DGR and DER and has an excellent performance compared to the best existing estimators. DGR, how-

TABLE B.IV

	HL				DDR			
s	$\hat{q} < q$	$\hat{q} = q$	$\hat{q} > q$	Total	$\hat{q} < q$	$\hat{q} = q$	$\hat{q} > q$	Total
0.30	0.0	**12.2**	87.8	100.0	0.0	**99.8**	0.2	100.0
0.35	0.0	**41.4**	58.6	100.0	0.0	**99.6**	0.4	100.0
0.40	0.0	**69.4**	30.6	100.0	0.0	**99.8**	0.2	100.0
0.45	0.0	**88.8**	11.2	100.0	0.2	**99.4**	0.4	100.0
0.50	0.0	**98.4**	1.6	100.0	0.0	**99.8**	0.2	100.0
0.55	0.0	**99.8**	0.2	100.0	0.2	**99.6**	0.2	100.0
0.60	0.0	**100.0**	0.0	100.0	0.0	**100.0**	0.0	100.0
0.65	0.0	**100.0**	0.0	100.0	0.0	**100.0**	0.0	100.0
0.70	0.0	**100.0**	0.0	100.0	0.0	**100.0**	0.0	100.0
0.75	0.6	**99.4**	0.0	100.0	0.8	**99.2**	0.0	100.0
0.80	0.4	**99.6**	0.0	100.0	0.2	**99.8**	0.0	100.0
0.85	2.6	**97.4**	0.0	100.0	1.2	**98.6**	0.2	100.0
0.90	6.6	**93.4**	0.0	100.0	3.0	**97.0**	0.0	100.0
0.95	15.0	**85.0**	0.0	100.0	3.8	**96.2**	0.0	100.0
1.00	28.2	**71.8**	0.0	100.0	5.4	**94.6**	0.0	100.0
1.05	45.0	**55.0**	0.0	100.0	11.0	**89.0**	0.0	100.0
1.10	63.2	**36.8**	0.0	100.0	15.2	**84.8**	0.0	100.0
1.15	77.4	**22.6**	0.0	100.0	18.0	**82.0**	0.0	100.0
1.20	89.4	**10.6**	0.0	100.0	26.8	**73.2**	0.0	100.0

Note: Variation of the third experiment described in Section B.6, with $q = 3$, $n = 100$, $T = 100$ and $s = 0.3 + 0.05i$, $i = 0, 1, \ldots, 18$. Percentage of underestimated outcomes, correct outcomes, and overestimated outcomes over 500 replications. HL: Hallin and Liška (2007) estimator, DDR: Dynamic Difference Ratio estimator. Boldface numbers denote the percentage of correct outcomes.

ever, performs remarkably well when the idiosyncratic components are small, as is likely the case with most macroeconomic data sets.

B.7. *Simulations: Analysis with Frequency Bands*

B.7.1. *Simulation Design*

In this section, we run a fourth Monte Carlo experiment. We use two DGPs, which we call the Trend-Cycle Model and the Stop-Band Model. In both models, the spectral density matrix of the common components has reduced rank at a specific frequency. In the Trend-Cycle Model, one of the common shocks is loaded by all variables with IRFs that vanish for $L = 1$, so that the spectrum has rank $q - 1$ at frequency zero. Hence, assuming that x_t is the first difference of the integrated vector y_t, this shock has transitory effects on all variables in y_t. In the Stop-Band Model, one of the common shocks is loaded by all variables with filters whose frequency response vanishes at frequency $\pi/6$, which can be interpreted as a cyclical frequency, in that it corresponds to a period of three years with quarterly data.

Fourth experiment (reduced rank spectral density). The data are generated as follows:

(i) We have two common shocks f_{jt}, $j = 1, 2$, $t = 1, \ldots, T$, which are $\sim iid\, \mathcal{N}(0, 1)$.

(ii) The idiosyncratic components are mutually independent white noises: $e_{it} = g_i \epsilon_{it}$, where $\epsilon_{it} \sim iid\, \mathcal{N}(0, 1)$ and the g_i are $\sim iid\, \mathcal{U}_{[-1,1]}$.

(iii) Trend-Cycle Model. For the first factor, the permanent shock, we have the AR(1) loadings $\lambda_{i1}(L) = a_{i1,0}/(1 - a_{i1,1}L)$, where the coefficients are iid, mutually independent, $a_{i1,0} \sim \mathcal{U}_{[-1,1]}$, $a_{i1,1} \sim \mathcal{U}_{[-0.8,0.8]}$. For the second factor, the transitory shock, we have the ARMA(1,1) loadings $\lambda_{i2}(L) = a_{i2,0}(1 - L)/(1 - a_{i2,1}L)$, where the coefficients are iid, mutually independent, $a_{i2,0} \sim \mathcal{U}_{[-1,1]}$, $a_{i2,1} \sim \mathcal{U}_{[0,0.7]}$.

(iv) Stop-Band model. For the first factor we have the same AR(1) loadings as in the Trend-cycle model: $\lambda_{i1}(L) = a_{i1,0}/(1 - a_{i1,1}L)$, $a_{i1,0} \sim \mathcal{U}_{[-1,1]}$, $a_{i1,1} \sim \mathcal{U}_{[-0.8,0.8]}$. For the second factor we use here a rough stop-band filter whose frequency response vanishes at frequencies $\pi/6$ and $-\pi/6$. Precisely, we use the ARMA (2,1) loadings $\lambda_{i2}(L) = a_{i2,0}(1 - e^{-\mathrm{i}\pi/6}L)(1 - e^{\mathrm{i}\pi/6}L)/(1 - a_{i2,1}L)$, where $a_{i2,0} \sim \mathcal{U}_{[-0.5,0.5]}$, $a_{i2,1} \sim \mathcal{U}_{[0.8,0.9]}$.

In the Trend-Cycle model, we have $\lambda_{i2}(1) = 0$ for all i, so that there is just one factor affecting the variables at frequency 0. Hence $q = 2$ for $\omega \neq 0$ and $q = 1$ for $\omega = 0$. To get an economic interpretation of the model, assume that the variables are growth rates of some macroeconomic series related to real economic activity, such as GDP and its components, employment variables, industrial production indexes, hours worked, and so on. Then the first factor is a long-run, permanent shock, driving the common trends of the series, expressed in log levels. For instance, it can be a technology shock or a generic supply shock. On the other hand, the second factor has transitory effects and can be interpreted as a demand shock driving a common cycle. Both shocks may affect variables at business cycle frequencies.

In the Stop-Band model, we have $\lambda_{i2}(e^{\mathrm{i}\pi/6}) = \lambda_{i2}(e^{-\mathrm{i}\pi/6}) = 0$ for all i, so that there is just one factor affecting the variables at frequency $\pi/6$, which corresponds to a period of three years with quarterly data. As a result, we have $q = 1$ for $\omega = \pm\pi/6$. The filters $\lambda_{i2}(L)$, $i = 1, \ldots, n$, can be regarded as rough stop-band filters toning down cyclical frequencies.

As in the third experiment, for each artificial dataset we compute the average sample variance of the common and the idiosyncratic components, say σ_χ^2 and σ_e^2. Then we multiply all common components by $1/\sigma_\chi$ and all idiosyncratic components by s/σ_e, with s taking on two values: (i) $s = 0.6$ (small idiosyncratic components) and (ii) $s = 1.2$ (large idiosyncratic components). Again, we standardize the variables before estimation. For this exercise, we set $n = 120$ and $T = 240$. We evaluate DDR at the points $\omega = 0$ and $\omega = \pi/6$. Moreover, we evaluate it in the long-run frequency band $[0, 2\pi/80]$, corresponding to cycles of 20 years or more with quarterly data; in the cyclical band $[2\pi/32, 2\pi/6]$, corresponding to cycles between 18 months and eight years; and in the short-run band $[2\pi/8, \pi]$, corresponding to cycles between six and 18 months. Finally, we evaluate DDR in the whole interval $[0, \pi]$.

B.7.2. *Simulation Results*

Table B.V reports the percentage of outcomes $\hat{q} = 1$, $\hat{q} = 2$ and $\hat{q} > 2$, over 500 replications. Boldface numbers denote the percentage of correct outcomes. On the long-run band $0 \leq \omega \leq 2\pi/80$ (which corresponds to periodicity greater than 20 years with quarterly data) the true number of common shocks is two for both models, but, for the Trend-Cycle model, the contribution of the transitory shock to total variance is negligible, so that we consider correct the outcome $\hat{q} = 1$. On the cyclical band $2\pi/32 \leq \omega \leq 2\pi/8$ the true number of common shocks is two for both models, but, for the Stop-Band model, the contribution of the non-cyclical shock to total variance is very small, so that we consider correct the outcome $\hat{q} = 1$.

In most cases, DDR can detect the correct number of common shocks. At a first sight, the worst outcome from the Stop-Band model, small idiosyncratic components (upper-right panel), for the cyclical band, where the result is $\hat{q} = 2$ in almost 40% of the cases. Indeed, we have already observed that the true number of common shocks is in fact two. With small idiosyncratic components, in several cases the criterion can detect the presence of a second factor, even though its effect is very small.

B.8. *Calibrating the Window Size*

To better calibrate the window size of DDR, DER, and DGR for the macroeconomic dataset used in the empirical application, we run a further simulation exercise.

TABLE B.V

Frequency band	Trend-Cycle Model				Stop-Band Model			
	$\hat{q}=1$	$\hat{q}=2$	$\hat{q}>2$	Total	$\hat{q}=1$	$\hat{q}=2$	$\hat{q}>2$	Total
Small idiosyncratic components								
$\omega=0$	**82.6**	13.8	3.6	100.0	1.8	**95.2**	3.0	100.0
$0\leq\omega\leq 2\pi/80$	**89.6**	10.2	0.2	100.0	1.6	**98.4**	0.0	100.0
$\omega=2\pi/12$	0.0	**100.0**	0.0	100.0	**99.8**	0.2	0.0	100.0
$2\pi/32\leq\omega\leq 2\pi/8$	0.0	**100.0**	0.0	100.0	**60.6**	39.4	0.0	100.0
$2\pi/8\leq\omega\leq\pi$	0.0	**99.4**	0.6	100.0	0.0	**96.4**	3.6	100.0
$0\leq\omega\leq\pi$	0.0	**99.8**	0.2	100.0	0.0	**98.0**	2.0	100.0
Large idiosyncratic components								
$\omega=0$	**93.2**	3.6	3.2	100.0	9.6	**82.0**	8.4	100.0
$0\leq\omega\leq 2\pi/80$	**98.4**	1.2	0.4	100.0	11.2	**87.4**	1.4	100.0
$\omega=2\pi/12$	3.8	**96.2**	0.0	100.0	**100.0**	0.0	0.0	100.0
$2\pi/32\leq\omega\leq 2\pi/8$	4.2	**95.8**	0.0	100.0	**94.6**	5.4	0.0	100.0
$2\pi/8\leq\omega\leq\pi$	0.0	**99.8**	0.2	100.0	0.6	**98.4**	1.0	100.0
$0\leq\omega\leq\pi$	0.0	**99.8**	0.2	100.0	1.4	**98.2**	0.4	100.0

Note: DGP: Trend-cycle model (left panel) and Stop-band model (right panel) described in Section B.7, with $q=2$, $n=120$, $T=240$ and $s=0.6$ (small idiosyncratic components), $s=1.2$ (large idiosyncratic components). Percentage of outcomes $\hat{q}=1$, $\hat{q}=2$ and $\hat{q}>2$, over 500 replications, obtained with the DDR estimator, evaluated at selected frequencies or frequency bands. Boldface numbers denote the percentage of correct outcomes. In the Trend-cycle Model, one of the two common shocks has zero effect in the long run for all variables so that the true value of q at frequency 0 is 1. On the long-run band $0\leq\omega\leq 2\pi/80$ (which corresponds to periodicity greater than 20 years with quarterly data), the true number of factors is two, but the contribution of the transitory shock to total variance is negligible so that we consider correct the outcome $\hat{q}=1$. In the Stop-band Model, one of the two common shocks has zero effect at frequency $\pi/6$ for all variables so that the true value of q at this frequency is 1. On the cyclical band $2\pi/32\leq\omega\leq 2\pi/8$ (which corresponds to cycles between two and eight years with quarterly data), the true number of factors is two, but the contribution of the transitory shock to total variance is very small, so we consider correct the outcome $\hat{q}=1$.

TABLE B.VI

q	a	$T=120$			$T=160$			$T=240$		
		DER	DGR	DDR	DER	DGR	DDR	DER	DGR	DDR
1	0.5	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
	0.75	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
	1.00	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
	1.25	100.0	100.0	48.6	100.0	100.0	91.8	100.0	100.0	100.0
2	0.50	96.0	99.6	99.8	100.0	100.0	100.0	100.0	100.0	100.0
	0.75	99.6	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
	1.00	99.6	99.8	100.0	100.0	100.0	100.0	100.0	100.0	100.0
	1.25	99.4	99.8	99.8	100.0	100.0	100.0	100.0	100.0	100.0
3	0.50	48.2	73.0	37.2	88.2	96.8	89.6	99.8	100.0	100.0
	0.75	87.2	96.8	98.8	98.2	99.6	100.0	100.0	100.0	100.0
	1.00	93.8	98.0	99.8	100.0	100.0	100.0	100.0	100.0	100.0
	1.25	91.4	96.0	98.4	99.0	99.6	100.0	100.0	100.0	100.0
4	0.50	1.2	11.6	0.0	19.8	49.8	0.8	86.8	95.6	81.0
	0.75	40.0	65.4	49.2	81.4	94.4	91.2	100.0	100.0	100.0
	1.00	63.6	83.0	95.6	94.4	98.0	99.6	99.8	100.0	100.0
	1.25	64.6	81.2	93.4	94.6	98.2	99.8	100.0	100.0	100.0

Note: Calibration of the bandwidth parameter for the empirical application. The DGP is the model of the third experiment described in Section B.5, large idiosyncratic components, with $q=1,2,3,4$, $n=216$, $T=120,160,240$. The bandwidth parameter is $M_T=[a\sqrt{T}]$ with $a=0.5,0.75,1,1.25$. The table reports the percentage of correct outcomes over 500 replications. DER: Dinamic Eigenvalue Ratio estimator, DGR: Dynamic Growth Ratio estimator; DDR: Dynamic Difference Ratio estimator.

We set $n=216$ (the number of series in the dataset) and use three values for T, namely $T=240$ (the time dimension of the whole sample), $T=120$, and $T=160$ (the sizes of the sub-samples used in the application). For the bandwidth parameter M_T, we use $M_T=[a\sqrt{T}]$,

with $a = 0.5, 0.75, 1, 1.25$. The DGP is the one used in the third experiment, Subsection B.6, in the version with large idiosyncratic components.

Table B.VI reports the percentage of correct outcomes over 500 replications. Let us consider first the case $T = 240$, reported in the right part of the table. With $q = 1, 2, 3$, all bandwidths perform well. With $q = 4$, the bandwidth $M_T = [0.5\sqrt{T}]$ performs poorly, particularly for DDR. The other bandwidths perform well for all estimators. Hence for $T = 240$ we stick to the bandwidth used in the simulations (i.e., $M_T = [0.75\sqrt{T}]$). Coming to the sample size $T = 120$ reported in the left part of the table, when $q = 1$, $a = 1.25$ performs poorly for DDR. With $q = 2$, all values of a perform well. With $q = 3$, $a = 0.5$ have a bad performance for all estimators. For $q = 4$, both $a = 0.5$ and $a = 0.75$ perform poorly. The results for $T = 160$ are qualitatively similar. Hence for the subsamples with $T = 120$ and $T = 160$ we choose $a = 1$, corresponding to $M_T = [\sqrt{T}]$.

B.9. *Empirical Variables and Transformations*

The data are from FRED-QD. We retain the 216 series starting in 1960Q1. We report here the ID number and the mnemonic. For the description of each variable, see McCracken and Ng (2020). Transformation codes: 1 = no transformation; 2 = first difference; 5 = log difference; 7 = first difference of the of percentage variation.

ID number	FRED-QD ID number	Transf. code	FRED-QD Mnemonic	ID number	FRED-QD ID number	Transf. code	FRED-QD Mnemonic
1	1	5	GDPC1	52	54	5	CES9091000001
2	2	5	PCECC96	53	55	5	CES9092000001
3	3	5	PCDGx	54	56	5	CES9093000001
4	4	5	PCESVx	55	57	5	CE16OV
5	5	5	PCNDx	56	58	2	CIVPART
6	6	5	GPDIC1	57	59	2	UNRATE
7	7	5	FPIx	58	60	2	UNRATESTx
8	8	5	Y033RC1Q027SBEAx	59	61	2	UNRATELTx
9	9	5	PNFIx	60	62	2	LNS14000012
10	10	5	PRFIx	61	63	2	LNS14000025
11	11	1	A014RE1Q156NBEA	62	64	2	LNS14000026
12	12	5	GCEC1	63	65	5	UEMPLT5
13	13	1	A823RL1Q225SBEA	64	66	5	UEMP5TO14
14	14	5	FGRECPTx	65	67	5	UEMP15T26
15	15	5	SLCEx	66	68	5	UEMP27OV
16	16	5	EXPGSC1	67	73	5	LNS12032194
17	17	5	IMPGSC1	68	74	5	HOABS
18	18	5	DPIC96	69	76	5	HOANBS
19	19	5	OUTNFB	70	77	1	AWHMAN
20	20	5	OUTBS	71	79	1	AWOTMAN
21	22	5	INDPRO	72	80	1	HWIx
22	23	5	IPFINAL	73	81	5	HOUST
23	24	5	IPCONGD	74	82	5	HOUST5F
24	25	5	IPMAT	75	83	5	PERMIT
25	26	5	IPDMAT	76	84	5	HOUSTMW
26	27	5	IPNMAT	77	85	5	HOUSTNE
27	28	5	IPDCONGD	78	86	5	HOUSTS
28	29	5	IPB51110SQ	79	87	5	HOUSTW
29	30	5	IPNCONGD	80	88	5	CMRMTSPLx
30	31	5	IPBUSEQ	81	89	5	RSAFSx
31	32	5	IPB51220SQ	82	90	5	AMDMNOx
32	34	1	CUMFNS	83	92	5	AMDMUOx
33	35	5	PAYEMS	84	95	5	PCECTPI
34	36	5	USPRIV	85	96	5	PCEPILFE
35	37	5	MANEMP	86	97	5	GDPCTPI
36	38	5	SRVPRD	87	98	5	GPDICTPI
37	39	5	USGOOD	88	99	5	IPDBS
38	40	5	DMANEMP	89	100	5	DGDSRG3Q086SBEA
39	41	5	NDMANEMP	90	101	5	DDURRG3Q086SBEA
40	42	5	USCONS	91	102	5	DSERRG3Q086SBEA
41	43	5	USEHS	92	103	5	DNDGRG3Q086SBEA
42	44	5	USFIRE	93	104	5	DHCERG3Q086SBEA
43	45	5	USINFO	94	105	5	DMOTRG3Q086SBEA
44	46	5	USPBS	95	106	5	DFDHRG3Q086SBEA
45	47	5	USLAH	96	107	5	DREQRG3Q086SBEA
46	48	5	USSERV	97	108	5	DODGRG3Q086SBEA
47	49	5	USMINE	98	109	5	DFXARG3Q086SBEA
48	50	5	USTPU	99	110	5	DCLORG3Q086SBEA
49	51	5	USGOVT	100	111	5	DGOERG3Q086SBEA
50	52	5	USTRADE	101	112	5	DONGRG3Q086SBEA
51	53	5	USWTRADE	102	113	5	DHUTRG3Q086SBEA

ID number	FRED-QD ID number	Transf. code	FRED-QD Mnemonic	ID number	FRED-QD ID number	Transf. code	FRED-QD Mnemonic
103	114	5	DHLCRG3Q086SBEA	160	187	5	EXCAUSx
104	115	5	DTRSRG3Q086SBEA	161	188	1	UMCSENTx
105	116	5	DRCARG3Q086SBEA	162	190	2	B020RE1Q156NBEA
106	117	5	DFSARG3Q086SBEA	163	191	2	B021RE1Q156NBEA
107	118	5	DIFSRG3Q086SBEA	164	194	5	IPMANSICS
108	119	5	DOTSRG3Q086SBEA	165	195	5	IPB51222S
109	120	5	CPIAUCSL	166	196	5	IPFUELS
110	121	5	CPILFESL	167	197	1	UEMPMEAN
111	122	5	WPSFD49207	168	198	1	CES0600000007
112	123	5	PPIACO	169	199	5	TOTRESNS
113	124	5	WPSFD49502	170	200	7	NONBORRES
114	125	5	WPSFD4111	171	201	1	GS5
115	126	5	PPIIDC	172	202	1	TB3SMFFM
116	127	5	WPSID61	173	203	1	T5YFFM
117	129	5	WPU0561	174	204	1	AAAFFM
118	130	5	OILPRICEx	175	205	5	WPSID62
119	132	5	CES2000000008x	176	206	5	PPICMM
120	133	5	CES3000000008x	177	207	5	CPIAPPSL
121	135	5	COMPRNFB	178	208	5	CPITRNSL
122	136	5	RCPHBS	179	209	5	CPIMEDSL
123	138	5	OPHNFB	180	210	5	CUSR0000SAC
124	139	5	OPHPBS	181	211	5	CUSR0000SAD
125	140	5	ULCBS	182	212	5	CUSR0000SAS
126	142	5	ULCNFB	183	213	5	CPIULFSL
127	143	5	UNLPNBS	184	214	5	CUSR0000SA0L2
128	144	1	FEDFUNDS	185	215	5	CUSR0000SA0L5
129	145	1	TB3MS	186	216	5	CES0600000008
130	146	1	TB6MS	187	217	5	DTCOLNVHFNM
131	147	1	GS1	188	218	5	DTCTHFNM
132	148	1	GS10	189	219	5	INVEST
133	150	1	AAA	190	220	1	HWIURATIOx
134	151	1	BAA	191	221	5	CLAIMSx
135	152	1	BAA10YM	192	222	5	BUSINVx
136	154	1	TB6M3Mx	193	223	1	ISRATIOx
137	155	1	GS1TB3Mx	194	224	1	CONSPIx
138	156	1	GS10TB3Mx	195	225	1	CP3M
139	157	1	CPF3MTB3Mx	196	226	1	COMPAPFF
140	158	5	BOGMBASEREALx	197	227	5	PERMITNE
141	160	5	M1REAL	198	228	5	PERMITMW
142	161	5	M2REAL	199	229	5	PERMITS
143	162	5	MZMREAL	200	230	5	PERMITW
144	163	5	BUSLOANSx	201	231	5	NIKKEI225
145	164	5	CONSUMERx	202	234	5	TLBSNNCBx
146	165	5	NONREVSLx	203	235	1	TLBSNNCBBDIx
147	166	5	REALLNx	204	236	5	TTAABSNNCBx
148	168	5	TOTALSLx	205	237	5	TNWMVBSNNCBx
149	170	5	TABSHNOx	206	238	2	TNWMVBSNNCBBDIx
150	171	5	TLBSHNOx	207	239	5	TLBSNNBx
151	172	5	LIABPIx	208	240	1	TLBSNNBBDIx
152	173	5	TNWBSHNOx	209	241	5	TABSNNBx
153	174	1	NWPIx	210	242	5	TNWBSNNBx
154	175	5	TARESAx	211	243	2	TNWBSNNBBDIx
155	176	5	HNOREMQ027Sx	212	244	5	CNCFx
156	177	5	TFAABSHNOx	213	245	5	S&P 500
157	184	5	EXSZUSx	214	246	5	S&P: indust
158	185	5	EXJPUSx	215	247	1	S&P div yield
159	186	5	EXUSUKx	216	248	5	S&P PE ratio