

Factor-Augmented VARs with Noisy Factor Proxies*

Emanuel Moench^{†1} and Soroosh Soofi-Siavash^{‡2}

¹Frankfurt School of Finance & Management, CEPR

²Lietuvos Bankas, Vilnius University

September 2026

Abstract

Factor-augmented VARs (FAVARs) combine latent factors extracted from large panels with factors that are assumed to be perfectly observed. We relax this assumption by treating individual variables as noisy factor proxies. Our proposed estimation approach iterates between principal component extraction and reduced-rank regression and allows multiple proxies to inform factor estimation. The estimated factors are shown to be consistent in the presence of weak factors. A group-lasso extension enables sparse proxy selection, and a cross-validation scheme jointly selects model dimensions and tuning parameters. Monte Carlo evidence shows that the method substantially improves factor estimation and forecasting with weak factors, and reduces bias in structural impulse response estimation when observable factor proxies are measured with error. We document the usefulness of our approach in a recursive pseudo-out-of-sample bond return forecasting analysis.

Keywords: factor-augmented vector autoregression, noisy factor proxy, principal components, reduced-rank regression, sparse regression

*We thank participants at the Vienna Workshop on Economic Forecasting, the International Symposium on Forecasting, the International Panel Data Conference, the annual meetings of the International Association for Applied Econometrics, the Bank of Lithuania, and the Baltic Economic Conference for helpful comments and suggestions. We are particularly grateful to Peter Vaporakis for valuable discussions. The views expressed in this paper are those of the authors and do not necessarily reflect the views of Lietuvos Bankas or the Eurosystem.

[†]Corresponding author; E-mail: e.moench@fs.de

[‡]E-mail: ssoofi@lb.lt

1 Introduction

Since Bernanke, Boivin, and Elias (2005), factor-augmented vector autoregressions (FAVARs) have become a workhorse model in empirical macro-financial economics. A key assumption made in these models is that some factors are perfectly measured by observable time series while the other factors need to be extracted from a large panel of time series using principal components. However, the assumption of some factors being perfectly observed appears unnecessarily strong. Many economic time series are subject to substantial data revisions. Moreover, economic concepts such as inflation or economic activity are rarely proxied by only one variable and are thus likely observed with measurement error.

In this paper we treat observable factor proxies as noisy measures of the true underlying factors. We show that a FAVAR with more observable proxies than true underlying factors can be written as a dynamic factor model (DFM) with a reduced-rank restriction. We propose to estimate the model by an iterative procedure which alternates between principal components estimation and solving a reduced-rank regression model (Anderson 1951). We further extend the method by adding a group-lasso-type penalty (Yuan and Lin 2006; Chen and Huang 2012; Qian, Tanigawa, Li, Tibshirani, Rivas, and Hastie 2022) to leverage a sparsity assumption when a subset of factor proxies is believed to be particularly relevant to the prediction. Building on Cheng, Liao, and Schorfheide (2016), we propose a cross-validation procedure to jointly determine the number of common factors along with the rank and penalty parameters. We also derive the regularity conditions required for consistent estimation of the factor space allowing for weak factors under large N and T asymptotics.

We document the desirable finite-sample properties of the proposed method via two Monte Carlo analyses. The first uses a DFM design with strong and weak factors. The weak-factor cases we consider are motivated by previous Monte Carlo and empirical evidence suggesting that the choice of data used in factor estimation strongly affects the estimated factor space (see e.g., Boivin and Ng 2006; Bai and Ng 2008): if the factor is related to only a small share of all observed series or explains only a small fraction of their overall variation, the factor is weak

in the cross-section used for the analysis and the ability of standard principal components to estimate the factor space can deteriorate substantially in practice. In principle, the standard FAVAR framework is well-suited to tackle the empirical issues related to weak factors, as it allows for inclusion of observable series as proxies for factors which are not well captured by the cross-section. However, this advantage vanishes when the observable factors are measured with error. We document that our approach, which combines principal components with sparse reduced-rank regressions (RRR henceforth), considerably improves factor estimation and produces superior forecasts.

As a second Monte Carlo analysis, we use a data-based simulation design inspired by Stock and Watson (2016) to identify oil supply and demand shocks in the spirit of Kilian (2009). We extend their analysis by including two alternative measures of global economic activity due to Baumeister and Hamilton (2019) and Baumeister, Korobilis, and Lee (2022). We treat the three global activity measures as noisy proxies for a single underlying factor and show that our proposed estimator substantially reduces the bias induced by relying on noisy observable series as factors in estimating structural impulse response functions (SIRFs) with a FAVAR.

Finally, we document the usefulness of our approach in a recursive pseudo-out-of-sample forecast exercise for Treasury returns. This analysis is in the spirit of Ludvigson and Ng (2009), which serves as a natural benchmark. We also compare our method to the SOFAR approach by Uematsu and Yamagata (2023a), who also document strong predictive power for bond returns.

Our estimation approach builds on several strands of the literature. The first is the large body of work related to FAVAR estimation via iterative principal components. This method was first developed by Stock and Watson (2005) and Boivin, Giannoni, and Mihov (2009). We extend this estimator to include an additional reduced-rank regression step which closely follows Izenman (1975) and Velu and Reinsel (1998). We further introduce a group-lasso penalty for variable selection and regularization using the algorithm proposed by Chen and Huang (2012) and further extended by Qian, Tanigawa, Li, Tibshirani, Rivas, and Hastie (2022).

A second related literature studies DFMs in the presence of weak factors. Bai and Ng (2023) provide conditions on the sample size under which the principal components estimator remains consistent when the loadings are weak. They show that weak factors can be recovered by principal components, but at a slower rate. The standard principal components estimator assumes that the series used in the analysis have been chosen so that they contain useful information for factor estimation. However, this does not necessarily hold in large datasets. One approach to handle this issue in practice is to model factors as being related to subsets of the data in the context of hierarchical DFMs (see e.g., Moench, Ng, and Potter 2013). Another approach is to use screening methods to prune the data as in Bai and Ng (2008), a method developed further by Giglio, Xiu, and Zhang (2025) in an asset pricing context. Alternatively, one can use sparsity-inducing methods to select a subset of variables, for example by adding a lasso-type penalty (e.g., see Uematsu and Yamagata 2023a,b). All these methods use either *a priori* knowledge or data-driven approaches to screen out useless variables. In this paper, we instead use FAVARs to incorporate information from observable noisy proxies into the space of factors, and evaluate the properties of our estimator when some factors are weak. Using Monte Carlo simulations, we show that our reduced-rank principal components approach outperforms other methods including the one by Uematsu and Yamagata (2023a) in factor estimation and forecasting in the presence of weak factors.¹

Another related recent paper is Li, Plagborg-Møller, and Wolf (2024) who explore the properties of various estimators for SIRFs, including local projections and structural VARs. The authors use the DFM from Stock and Watson (2016) and Lazarus, Lewis, Stock, and Watson (2018) as their encompassing model, from which they form many lower dimensional DGPs. As in Li et al. (2024), we parameterize our second Monte Carlo design following Stock and Watson (2016), with the main difference being that we focus on FAVAR estimators of SIRFs in the presence of noisy observable factor proxies.

¹For other important work studying strong and weak factors and statistical methods for measuring factor strength, see e.g., Chudik, Pesaran, and Tosetti (2011), Onatski (2010, 2012), Kapetanios and Marcellino (2010) and Bailey, Kapetanios, and Pesaran (2021).

2 Methodology

In this section we begin with a discussion of the standard FAVAR framework. We then introduce several modifications to this framework and discuss our proposed estimation procedure.

2.1 Framework

To fix ideas, consider a FAVAR with q observed factors $\tilde{F}_t$ measured without error by the variables Y_t , and r unobserved factors stacked in the vector F_t . These factors drive the common variation in a set of observed time series variables stacked in a $N \times 1$ vector X_t . Both the observable variables X_t and the factor proxies Y_t are assumed to be second-order stationary and integrated of order zero. N is commonly assumed to be large. In particular, it may be greater than T , the number of time periods. It is also commonly assumed to be much greater than the total number of factors in the FAVAR model, i.e. $N \gg r + q$.

In typical applications, Y_t would, for example, include a policy rate, a measure of real economic activity, and a measure of inflation (e.g. Bernanke et al. 2005). Hence, the FAVAR model can be used to trace impulse responses to the observed variables Y in a data-rich environment, taking into account the joint dynamics of Y_t with a small number of factors F_t which span the information in a large number of additional variables X_t .

The FAVAR then takes the following form:

$$\begin{bmatrix} Y_t \\ X_t \end{bmatrix} = \begin{bmatrix} I_q & 0_{q \times r} \\ \Lambda_{X\tilde{F}} & \Lambda_{XF} \end{bmatrix} \begin{bmatrix} \tilde{F}_t \\ F_t \end{bmatrix} + \begin{bmatrix} 0 \\ e_{Xt} \end{bmatrix} \quad (1)$$

$$F_t^+ = \Phi(L) F_t^+ + \eta_t, \quad \text{where } F_t^+ = \begin{bmatrix} \tilde{F}_t \\ F_t \end{bmatrix} \quad (2)$$

where $\Lambda_{X\tilde{F}}$ and Λ_{XF} are the $N \times q$ and $N \times r$ loading matrices, $\Phi(L)$ is a lag polynomial of dimension $(r + q) \times (r + q)$, e_{Xt} is the $N \times 1$ vector of idiosyncratic errors and η_t is the $(r + q) \times 1$ vector of factor innovations. Equations (1) and (2) capture the main feature of FAVARs: Y_t and X_t are jointly driven by $\tilde{F}_t$ and F_t , and $\tilde{F}_t$ is perfectly observed by Y_t .

In many applications, the assumption of observed factors which are measured without error appears unduly strong. For example, there are different variables which are proxies for economic concepts such as economic activity and inflation. Moreover, macroeconomic variables are often observed with a lag and feature substantial data revisions. As a consequence, in relevant applications it seems more appealing to assume that observable variables are proxies for the true underlying factors and are measured with error.

In what follows, we will therefore interpret Y_t as observed proxies which may be noisy measures of the true underlying factors. To be concrete, suppose we observe m proxy variables Y_t . In principle, m could be greater than the number of underlying factors $\tilde{F}_t$ ($m \geq q$). We thus assume that (X_t, Y_t) admit the following factor model representation:

$$\begin{bmatrix} Y_t \\ X_t \end{bmatrix} = \begin{bmatrix} \Lambda_{Y\tilde{F}} & 0_{m \times r} \\ \Lambda_{X\tilde{F}} & \Lambda_{XF} \end{bmatrix} \begin{bmatrix} \tilde{F}_t \\ F_t \end{bmatrix} + \begin{bmatrix} T^{-\delta} e_{Yt} \\ e_{Xt} \end{bmatrix}, \quad (3)$$

$$F_t^+ = \Phi(L) F_t^+ + \eta_t, \quad \text{with} \quad F_t^+ = \begin{bmatrix} \tilde{F}_t \\ F_t \end{bmatrix}, \quad (4)$$

where $\Lambda_{Y\tilde{F}}$ is $m \times q$ with full-column rank, and $T^{-\delta} e_{Yt}$, with $\delta > 0$, is an $m \times 1$ vector of measurement errors with mean zero that is assumed to be uncorrelated with e_{Xt} .

Equations (3) and (4) make one important modification to (1) and (2): the variables Y_t measure the true factors with observation errors that diminish at rate $T^{-\delta}$, for some $\delta > 0$. In other words, the factors $\tilde{F}_t$ can be well-approximated by linear combinations of Y_t when T is large.² This permits rewriting Equation (3) as

$$X_t = \Lambda_{X\tilde{F}} B' (Y_t - T^{-\delta} e_{Yt}) + \Lambda_{XF} F_t + e_{Xt} = \Lambda_{X\tilde{F}} B' Y_t + \Lambda_{XF} F_t + u_{Xt}, \quad (5)$$

where $B = \Lambda_{Y\tilde{F}} (\Lambda_{Y\tilde{F}}' \Lambda_{Y\tilde{F}})^{-1}$ is of dimension $m \times q$, and the error term vector is $u_{Xt} = e_{Xt} - T^{-\delta} \Lambda_{X\tilde{F}} B' e_{Yt}$. Thus, Equation (5) obtains with one restriction: the $N \times m$ coefficient

²The special case $\delta = 1/2$ corresponds to the assumption in Parker and Sul (2016) for approximate factor proxies. Moreover, Schorfheide (2005) and Li et al. (2024) assume drifting DGPs in the context of forecasting and analyzing impulse responses where the sample is not generated from a VAR(p), but the degree of misspecification vanishes at rate $T^{-1/2}$.

matrix $\Lambda_{X\tilde{F}}B'$ has rank q , and its $N \cdot m$ elements are expressed in terms of the $N \cdot q + m \cdot q$ elements of $\Lambda_{X\tilde{F}}$ and B' . The main advantage of this representation is that the factors can be estimated using an iterative method that alternates between estimating B via reduced-rank regression for a given F_t and estimating F_t for a given estimate of B , see Section 2.2 below.

Note that our assumption that measurement errors diminish as $T \rightarrow \infty$ ensures that the part of u_{Xt} arising from proxy noise becomes negligible in large samples. As a result, Equation (5) asymptotically behaves as if the proxies Y recovered the factor space spanned by $\tilde{F}$ without error. The proxies may therefore be imperfect in finite samples, but their measurement errors are not large enough to change the large-sample representation of the model. In applied work, whether good proxies are available is ultimately an empirical question. We address this question in two ways. First, in Section 2.4, we discuss an extension of the estimation method by adopting a group lasso penalty on the rows of B to impose a sparsity assumption when a subset of factor proxies is believed to be relevant for the FAVAR estimation. Second, in Section 3.1 we introduce a model selection procedure via cross-validation that jointly determines the number of the unobserved factors r , the number of factor proxy combinations q , as well as the penalty parameter in the case with additional sparsity constraints. In Section 4, we show in a Monte Carlo analysis that both approaches perform very well specifically in the presence of weak noisy factors.

2.2 Estimation

If the factors $\tilde{F}_t$ are measured without error as in the standard FAVAR framework, the model can be estimated via the iterative principal components estimator proposed by Stock and Watson (2005) and Boivin, Giannoni, and Mihov (2009). However, under the rank constraint, Equation (5) cannot be estimated this way. Instead, we propose a method that combines iterative principal component estimation with an additional reduced-rank regression step (Izenman 1975; Velu and Reinsel 1998).

Specifically, denoting the matrix $[A_1, \dots, A_T]'$ by A for any time series of vectors $\{A_t\}_{t=1}^T$, the

factors can be estimated as the principal components of the filtered observations $(X - Y\hat{B}\hat{\Lambda}'_{X\hat{F}})$ given the parameter estimates $\hat{B}$ and $\hat{\Lambda}$. In turn, given the factor estimates $\hat{F}$, the coefficient matrices $\hat{B}$ and $\hat{\Lambda}$ can be estimated via a reduced-rank regression of $M_{\hat{F}}X$ on $M_{\hat{F}}Y$ where $M_{\hat{F}} = I_T - \hat{F}(\hat{F}'\hat{F})^{-1}\hat{F}'$. Specifically, the estimators of F , B , $\Lambda_{X\hat{F}}$ and Λ_{XF} jointly solve the minimization problem

$$\underset{F, B, \Lambda_{X\hat{F}}, \Lambda_{XF}}{\text{minimize}} \quad \|X - YB\Lambda'_{X\hat{F}} - F\Lambda'_{XF}\|^2, \quad (6)$$

subject to the normalization constraints

$$\Lambda'_{X\hat{F}}\Lambda_{X\hat{F}} = I_q, \quad F'F/T = I_r, \quad \Lambda'_{XF}\Lambda_{XF} \text{ diagonal}, \quad B'Y'M_FYB \text{ diagonal} \quad (7)$$

where $M_F = I_T - F(F'F)^{-1}F' = I_T - T^{-1}FF'$, and $\|\cdot\|$ denotes the Frobenius norm. This minimization can be done iteratively. Given a preliminary estimator of F , which is computed as the first r principal components of X , the estimators for $\Lambda_{X\hat{F}}$, B and Λ_{XF} are given by:

$$\begin{aligned} \hat{\Lambda}_{X\hat{F}} &= V, \quad \hat{B} = (Y'M_{\hat{F}}Y)^{-1}Y'M_{\hat{F}}XV, \\ \text{and } \hat{\Lambda}_{XF} &= \left(X - Y\hat{B}\hat{\Lambda}'_{X\hat{F}}\right)' \hat{F} (\hat{F}'\hat{F})^{-1} = \frac{1}{T} \left(X - Y\hat{B}\hat{\Lambda}'_{X\hat{F}}\right)' \hat{F}, \end{aligned} \quad (8)$$

where $V = [V_1, \dots, V_q]$ and V_j is the j th eigenvector of $X'M_{\hat{F}}Y(Y'M_{\hat{F}}Y)^{-1}Y'M_{\hat{F}}X$.³ Given estimates of B and $\Lambda_{X\hat{F}}$, we estimate F as the first r principal components of $(X - Y\hat{B}\hat{\Lambda}'_{X\hat{F}})$. Each step of this procedure reduces the sum of squares in Equation (6), and in practice we iterate until convergence.

2.3 Consistent Estimation of the Factor Space

The large-sample analysis of the estimation procedure builds on the principal-components results of Stock and Watson (2002) and Bai (2003), as well as the extension to weak factor loadings by Bai and Ng (2023). The additional complication is that part of the factor space is recovered through the reduced-rank regression step and may be weaker in the cross section than the component associated with F_t . We assume that the component associated with F_t

³See Velu and Reinsel (1998) for an extensive discussion on RRR models and their estimation.

is strong, with cross-sectional strength N , while the loadings on $\tilde{F}_t$ retain strength N^α , with $0 < \alpha < 1$, after partialling out the space spanned by the loadings of F_t . The formal regularity conditions on the factors, loadings, proxies, and errors are given in Online Appendix A.

For two matrices A_1 and A_2 , let

$$D(A_1, A_2) = \|P_{A_1} - P_{A_2}\|_2$$

denote the spectral norm distance between their column spaces. The following proposition summarizes the result for the factor estimates used in the FAVAR.

Proposition 1. *Let $N, T \rightarrow \infty$ with $T/N^{1-\alpha} \rightarrow \infty$, $c_{NT} = \min\{\sqrt{N}, \sqrt{T}\}$, and suppose Assumptions A.1–A.4 in Online Appendix A hold. Initialize the procedure with the first r principal components of X , denoted by $\hat{F}^{(0)}$, and let $\hat{B}^{(k+1)}$ denote the coefficient matrix obtained from the reduced-rank regression (8) after partialling out $\hat{F}^{(k)}$. Then, for every fixed integer $K \geq 0$,*

$$\max_{0 \leq k \leq K} D\left([Y\hat{B}^{(k+1)}, \hat{F}^{(k)}], [\tilde{F}, F]\right) = O_p\left(\sqrt{\frac{N^{1-\alpha}}{T}} + c_{NT}^{-1} + T^{-\delta}\right) = o_p(1). \quad (9)$$

The details of the proof are provided in Online Appendix A. Here, we offer a few remarks to provide intuition for the result. The terms $N^{(1-\alpha)/2}/\sqrt{T} + T^{-\delta}$ in the convergence rate reflect recovery of the less pervasive component using noisy proxies, while c_{NT}^{-1} is the standard principal-components rate. The proof first considers the space spanned by the leading r principal components of the factor component of X . The initial principal-components estimate consistently recovers this space. Given an estimate $\hat{F}$ of this principal-components space with a sufficiently small estimation error, the joint space spanned by $[Y\hat{B}, \hat{F}]$ consistently estimates the factor space. The estimation error of the principal-components estimate in the next step remains sufficiently small, allowing the argument to be repeated for any fixed number of iterations. Moreover, each reduced-rank regression and principal-components update weakly decreases the value of the objective function. In practice, we repeat the procedure until convergence.

The Monte Carlo evidence in Section 4 shows that the estimator also performs well in finite

samples. Specifically, we assess its usefulness for factor estimation and forecasting relative to the standard principal components estimator but also to the method proposed by Uematsu and Yamagata (2023a), who target weak factors by imposing sparsity on the factor loadings.

2.4 Modification with Sparsity Constraints

We now introduce an extension of our approach where a subset of factor proxies is believed to be particularly relevant for the estimation of the factors and FAVAR parameters. This can be achieved by assuming sparsity in the coefficient matrix B . To find a sparse estimate of the coefficients, we propose to add a group lasso penalty to the objective function (6) and solve the following penalized rank-constrained optimization problem

$$\underset{F, B, \Lambda_{X\tilde{F}}, \Lambda_{XF}}{\text{minimize}} \frac{1}{2} \|X - YB\Lambda'_{X\tilde{F}} - F\Lambda'_{XF}\|^2 + \lambda \sum_{j=1}^m \|B_{j\bullet}\| \quad (10)$$

subject to the same normalization (7) as above, where $B_{j\bullet}$ is the j th row of B , and $\lambda \geq 0$ is a penalty parameter. The penalty encourages row sparsity in B : if $\|B_{j\bullet}\| = 0$, proxy j is excluded from all proxy-factor combinations. Retained rows are shrunk jointly, but the penalty does not impose element-wise sparsity within a retained row.

This optimization problem can be solved using a similar iterative procedure as described in Section 2.2, except for one difference: we replace the RRR step with the alternating minimization method proposed by Chen and Huang (2012) and further extended by Qian et al. (2022). For a given estimate of factors $\hat{F}$ and value of λ , the estimate of the parameters B , $\Lambda_{X\tilde{F}}$ and Λ_{XF} are obtained by iterating two steps. Specifically, for the k th iteration:

1. Given an estimate $\hat{B}^{(k)}$ and $\hat{\Lambda}_{XF}^{(k)}$, we solve the orthogonal Procrustes problem

$$\underset{\Lambda_{X\tilde{F}}}{\text{minimize}} \|X - Y\hat{B}^{(k)}\Lambda'_{X\tilde{F}} - \hat{F}\hat{\Lambda}_{XF}^{(k)'}\|^2$$

subject to $\Lambda'_{X\tilde{F}}\Lambda_{X\tilde{F}} = I_q$. Let $(X - \hat{F}\hat{\Lambda}_{XF}^{(k)'})' Y\hat{B}^{(k)} = UDV'$ be the singular value decomposition with U , D and V having dimensions $N \times q$, $q \times q$ and $q \times q$, respectively.

The solution is given as $\hat{\Lambda}_{X\tilde{F}}^{(k+1)} = UV'$.

2. Given $\hat{\Lambda}_{X\hat{F}}^{(k+1)}$, we solve the group lasso problem

$$B^{(k+1)} = \underset{B}{\operatorname{argmin}} \frac{1}{2} \|M_{\hat{F}} X \hat{\Lambda}_{X\hat{F}}^{(k+1)} - M_{\hat{F}} Y B\|^2 + \lambda \sum_{j=1}^m \|B_{j\bullet}\|,$$

where $M_{\hat{F}} = I_T - T^{-1} \hat{F} \hat{F}'$ as per Equation (7) and $\hat{\Lambda}_{X\hat{F}}^{(k+1)} = T^{-1} \left(X - Y B^{(k+1)} \hat{\Lambda}_{X\hat{F}}^{(k+1)'} \right)' \hat{F}$.⁴

Starting values are obtained from the standard RRR estimation we have described in the previous section, and the procedure is iterated until convergence.

For a given estimate $\hat{F}$, each iteration alternates between an exact minimization of $\Lambda_{X\hat{F}}$ through an orthogonal Procrustes step and an exact minimization of B through a group-lasso step, so that the value of the penalized objective function is monotonically non-increasing across iterations. Since the penalized problem is non-convex, however, there is no general guarantee of convergence to a local or global minimum. Rather, following Qian et al. (2022), the resulting solution should be interpreted as a limiting point of the alternating minimization scheme. Nevertheless, the procedure is appealing in practice because it is stable and yields a monotone decrease in the objective function across iterations.

3 Practical Guidance for Implementation

We now discuss our approach to model selection and normalization of the factors in structural FAVARs for impulse response analysis.

3.1 Model Selection

In Sections 2.2-2.4, we laid out the estimation procedure for a given number of latent factors r , the rank q associated with the set of noisy factor proxies, and the regularization parameter λ in the group lasso approach. We now discuss how to determine these parameters in practice.

Our approach builds on the cross-validation procedure proposed by Cheng, Liao, and Schorfheide (2016). Using Monte Carlo analysis, these authors show that the procedure works

⁴We solve the group-lasso problem using `glmnet` (Friedman, Hastie, and Tibshirani 2010) with the `mgaussian` family.

well compared to the methods proposed by Bai and Ng (2002), Onatski (2010) and Ahn and Horenstein (2013) in standard factor models with common factors estimated via principal components. What distinguishes our setup is the existence of additional observable noisy factor proxies. We therefore adapt the cross-validation procedure of Cheng, Liao, and Schorfheide (2016) to jointly determine r and q . Moreover, in the case of sparsity constraints on B , we also let the procedure determine λ jointly with r and q .

Let $0 \leq r \leq r_{max}$ and $0 \leq q \leq q_{max}$. Hence, the smallest model that can be chosen does not include any common factors or noisy factor proxies Y . For a given r , let there be a decreasing sequence of values $\lambda_{max} = \lambda_1 > \lambda_2 > \dots > \lambda_L \geq 0$, where λ_{max} is defined such that it gives rise to the trivial solution of B where all the elements in B are zero and the remaining λ_i are selected from a grid that is equally spaced on the log scale. Following Qian et al. (2022), we set $\lambda_{max} = \max_{1 \leq j \leq m} \|Y'_j (X - \hat{F}\hat{\Lambda} * XF')\|$, where $Y * j$ is the $T \times 1$ vector of observations for the j th factor proxy, $\hat{F}$ is $\sqrt{T}$ times the matrix of eigenvectors corresponding to the r largest eigenvalues of XX' , and $\hat{\Lambda} * XF$ is the corresponding estimate of the factor loading matrix. In other words, the value $\lambda * max$ corresponds to the case where the noisy factor proxies Y carry no additional information about X over and above the factors $\hat{F}$.

A graphical illustration of the cross-validation procedure is given in Figure 1. We partition X and Y along the cross-sectional dimension into K disjoint folds, $X_{I_1}, X_{I_2}, \dots, X_{I_K}$ and $Y_{J_1}, Y_{J_2}, \dots, Y_{J_K}$, where $\bigcup_{k=1}^K I_k = \{1, 2, \dots, N\}$ and $\bigcup_{k=1}^K J_k = \{1, 2, \dots, m\}$. The k th validation fold therefore consists of the paired hold-out sets (X_{I_k}, Y_{J_k}) . Let $I_{-k} = \{1, \dots, N\} \setminus I_k$ and $J_{-k} = \{1, \dots, m\} \setminus J_k$ denote the corresponding estimation sets. For each combination (r, q, λ) , we estimate $\hat{F}^{+(-k)} = \left[Y_{J_{-k}} \hat{B}^{(-k)}, \hat{F}^{(-k)} \right]$ using only the retained variables $X_{I_{-k}}$ and retained proxies $Y_{J_{-k}}$. We then divide the time periods used for validation into rolling prediction blocks. For each block, the loadings of the held-out variables $i \in I_k$ and held-out proxies $j \in J_k$ are estimated by OLS over the preceding regression window. These loadings and the factor estimates $\hat{F}_t^{+(-k)}$ are subsequently used to predict both X_{it} and Y_{jt} over the corresponding prediction block. The overall cross-validation criterion averages these block-specific MSFEs over the rolling prediction blocks and cross-sectional folds. We select (r, q, λ)

by minimizing this criterion.

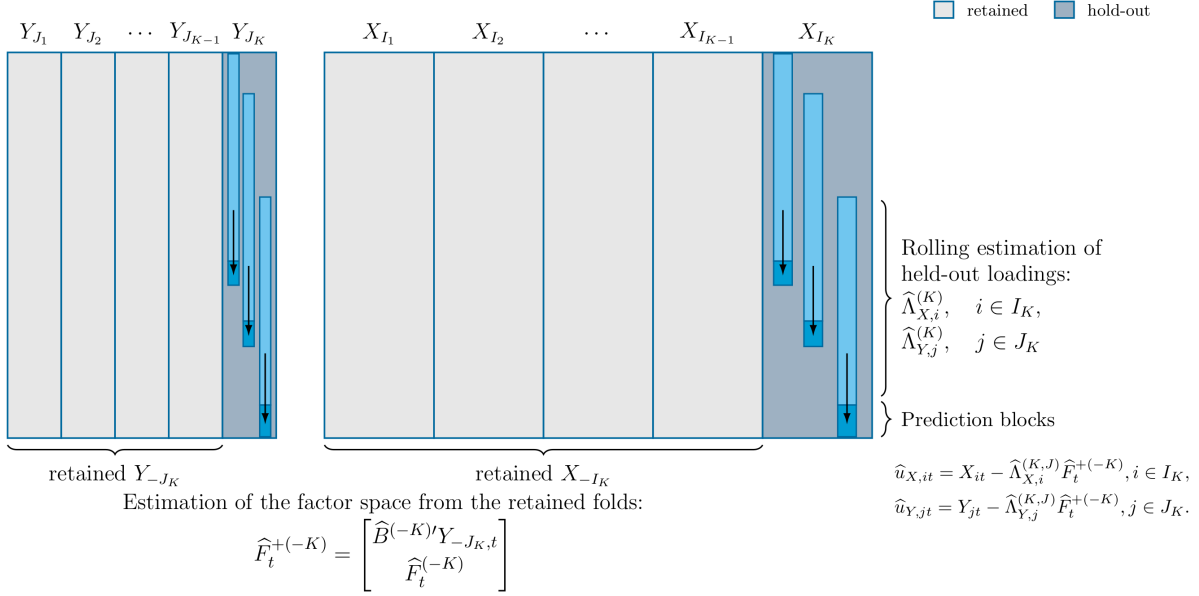


Figure 1: Graphical representation of the cross-validation procedure.

Note that paired partitioning of X and Y is feasible only if each estimation fold retains at least $q_{\max}$ proxies, that is, $q_{\max} \leq m - |J_k|$ for $k = 1, \dots, K$. Otherwise, only X is partitioned into K folds, whereas all proxies in Y are retained when estimating the factors, and the cross-validation loss is computed only from the held-out variables in X .

The proposed cross-validation adapts the procedure in Cheng et al. (2016) to our FAVAR setting with noisy factor proxies. As in their approach, we partition the sample along the cross-sectional and time-series dimensions and select the tuning parameters by minimizing pseudo-out-of-sample mean squared forecast errors. The main difference relative to Cheng et al. (2016) is that, in our setting, cross-validation must additionally select the number of factor proxy combinations q and, in the sparse case, the penalty parameter λ associated with the reduced-rank coefficient matrix. In Section 4, we document the strong performance of the method in conjunction with the cross-validation approach in a Monte Carlo analysis.

3.2 Impulse Responses and Named Factor Normalization

From the model described above, it is clear that only the space spanned by F_t and $B'Y_t$ is identified, whereas F_t and $B'Y_t$ themselves are not uniquely pinned down. Since the factors are unobserved, one commonly imposes normalization (7) which represents a mathematically convenient set of restrictions that solves the identification issue for typical applications such as forecasting. However, when a FAVAR model is used to identify structural shocks, the factors need to be assigned an economic interpretation. This requires a different set of identification restrictions. In what follows, we achieve this by adopting the “named-factor” normalization of Stock and Watson (2016). In this approach, each factor is aligned with a specific naming variable.⁵ Let

$$X_t^+ = \begin{bmatrix} Y_t \\ X_t \end{bmatrix}, \quad \hat{F}_t^+ = \begin{bmatrix} \hat{B}'Y_t \\ \hat{F}_t \end{bmatrix}, \quad \text{and } \hat{\Lambda} \text{ be the estimate of the loadings of } X_t^+ \text{ on } \hat{F}_t^+,$$

where $\hat{F}_t$ and $\hat{B}$ are the estimators produced via the procedure previously described. Moreover, order X_t^+ so that the first $(r + q)$ variables are the naming variables (including some of the elements of Y_t), and let $\hat{\Lambda}_{1:q+r}$ denote the $(r + q) \times (r + q)$ matrix of the loadings on the first $(r + q)$ variables. The named-factor normalization is

$$\tilde{\Lambda} = \begin{bmatrix} I_{q+r} \\ \tilde{\Lambda}_\bullet \end{bmatrix} = \begin{bmatrix} I_{q+r} \\ \hat{\Lambda}_\bullet \hat{\Lambda}_{1:q+r}^{-1} \end{bmatrix}, \quad \text{and} \quad \tilde{F}_t^+ = \hat{\Lambda}_{1:q+r} \hat{F}_t^+.$$

The above named-factor normalization aligns factors with naming variables and sets the magnitude of the loadings so that a one-unit innovation in factor i raises the i th naming variable by one unit on impact. As an example, suppose one of the naming variables is Kilian (2009)’s global activity measure. Then, the named-factor normalization associates the corresponding factor with the common component of Kilian (2009)’s measure.

⁵Bai and Ng (2013) label this the PC3 normalization.

4 Monte Carlo I: Factor Estimation and Forecasting

In this section, we discuss the finite-sample properties of the proposed estimator via simulations. The data are generated using a range of DFM specifications that feature serially correlated factors, and error terms that are serially and cross-sectionally correlated. Moreover, we study how the presence of weak factors affects the performance of the estimator for estimating the factor space and for forecasting. We refer to weak factors as factors for which only a small proportion of the variables feature nonzero loadings.

4.1 Experimental Design

Our experimental design is based on the following model

$$X_{it} = \lambda_{i1}F_{1t} + \lambda_{i2}F_{2t} + u_{it}, \quad (11)$$

$$\begin{bmatrix} F_{1t} \\ F_{2t} \end{bmatrix} = \begin{bmatrix} \phi_{11} & 0 \\ 0 & \phi_{22} \end{bmatrix} \begin{bmatrix} F_{1t-1} \\ F_{2t-1} \end{bmatrix} + \begin{bmatrix} \sqrt{\theta} & 0 \\ 0 & \sqrt{(1 - \phi_{22}^2)} \end{bmatrix} \begin{bmatrix} \eta_{1t} \\ \eta_{2t} \end{bmatrix}, \quad (12)$$

$$(1 - aL)u_{it} = \sqrt{\nu}e_{it}, \quad \text{and} \quad (13)$$

$$e_{it} = (1 + b^2)\epsilon_{it} + b\epsilon_{i-1t} + b\epsilon_{i+1t}, \quad (14)$$

where all factor loadings λ_{ik} and disturbances η_{kt} and ϵ_{it} are iid $N(0, 1)$. This model is a static form of a DFM with two factors following a VAR(1). We set $\phi_{11} = \phi_{22} = 0.5$. The first factor is assumed to be a common factor, i.e. there are no restrictions on the loadings. For the second factor, we consider a case with unrestricted loadings and two cases in which 50% or 90% of the loadings are set to zero. The parameter θ is chosen so that $\text{var}(F_{2t}) = \text{var}(F_{1t})$. We also consider a case where we do not impose any restrictions on factor loadings but set the parameter θ such that $\text{var}(F_{2t}) = 0.1 \times \text{var}(F_{1t})$. The parameters a and b govern serial and cross-sectional correlation in the errors and are assumed to take values equal to 0.5 and 1, respectively. The parameter ν is chosen to ensure that the variables X_{it} have an average $R^2 = 30\%$ on the factors. This specification for the error terms builds on Stock and Watson

(2002). Furthermore, the choice of average R^2 is consistent with common factor analyses of macroeconomic data (Stock and Watson 2016; McCracken and Ng 2016).

We use noisy proxies for the second factor. Specifically, we set $Y_{jt} = F_{2t} + T^{-\delta} \sigma_j \psi_{jt}$ where ψ_{jt} is iid $N(0, 1)$. For each value of δ , we choose σ_j so that proxy j has the R^2 specified below when $T = 100$. We use the same value of σ_j when $T = 200$. We consider five cases. In the first, we assume one proxy with an R^2 of 80% in the regression of Y_{jt} on F_{2t} . In the second, we add four additional proxies with an R^2 of 70%, 60%, 50% and 40%, respectively. The third and fourth cases, respectively, add five and ten additional proxies each having an R^2 of 30%. Modeling the proxies in this way allows us to consider cases where we start with the strongest proxy, but then add more and more proxies which are increasingly noisier measures of the true underlying factor. Finally, we also consider a case with 20 proxies where each proxy is generated to have an R^2 of 30%.

We experiment with $N = 100$, $T = (100, 200)$ and $\delta = (0, 1/2)$. The case $\delta = 1/2$ assumes that proxy noise declines with T , consistent with Proposition 1, while $\delta = 0$ checks robustness when proxy noise remains fixed. Each set of parameters uses 1000 repetitions.

4.2 Estimation

We examine five estimation approaches.

1. *PC*: We estimate the factors via principal components and determine r by the edge distribution (ED) method of Onatski (2010). Uematsu and Yamagata (2023a) stress the importance of using this approach for determining r in weak factor models and provide Monte Carlo evidence suggesting that the method works well relative to those proposed by Bai and Ng (2002) and Ahn and Horenstein (2013) in weak factor models.
2. *SOFAR*: We estimate the factors via the adaptive sparse orthogonal factor regression (SOFAR) estimator of Uematsu and Yamagata (2023a) and determine r by the ED method of Onatski (2010). Following Uematsu and Yamagata (2023a), we also use BIC to determine the regularization parameter for the adaptive SOFAR estimation.

3. $RPC_{(1,1)}$: We solve the optimization problem (6), setting $r = 1$ and $q = 1$. This is an infeasible benchmark that always combines one principal component with one combination of factor proxies.
4. RPC : We solve the optimization problem (6), and determine r and q using the cross-validation procedure detailed in Section 3.1.
5. $SRPC$: We solve the penalized optimization problem (10), and determine r , q and λ via the cross-validation procedure detailed in Section 3.1. This is a modification of RPC with sparsity constraints.

4.3 Summary Statistics

We use two summary statistics to assess the relative performance of the different estimation procedures. The first is the average R^2 of the regression of F on $\hat{F}^+$:

$$R_{F, \hat{F}^+}^2 = \hat{E} \left(\frac{1}{2} \sum_{i=1}^2 \frac{\|P_{\hat{F}^+} F_i\|^2}{\|F_i\|^2} \right), \quad (15)$$

where F_i is the $T \times 1$ vector of observations on factor i , $\hat{E}$ denotes the average over all the Monte Carlo repetitions, and $P_{\hat{F}^+} = \hat{F}^+ (\hat{F}^{+'} \hat{F}^+)^{-1} \hat{F}^{+'}$. The second statistic measures the out-of-sample forecast performance for the cross-section of variables, following Stock and Watson (2002). The out-of-sample forecast is $\hat{X}_{iT+1|T} = \hat{\beta}_i' \hat{F}_T^+$ where the OLS estimate $\hat{\beta}_i$ is obtained by regressing X_{it+1} on $\hat{F}_t^+$ for $t = 1, \dots, T-1$. The measure of relative forecast performance is

$$\frac{S_{(S)RPC}^2}{S_{PC}^2} = \frac{\sum_{i=1}^N \hat{E} \left(\hat{X}_{iT+1|T}^{(S)RPC} - X_{iT+1|T} \right)^2}{\sum_{i=1}^N \hat{E} \left(\hat{X}_{iT+1|T}^{PC} - X_{iT+1|T} \right)^2}. \quad (16)$$

Hence, we use the standard PC estimator as the benchmark, with r determined as in Onatski (2010). Thus, the statistic is less than one when the out-of-sample forecasting performance of a model combining principal components with combinations of factor proxies outperforms those using only principal components.

4.4 Results

Table 1 reports the results of the Monte Carlo simulations with $N = T = 100$. Panel A presents the results for the case in which there are two common factors. Panel B shows the corresponding results for the case where F_{1t} is a common factor, and 50% of the loadings are set to zero for F_{2t} , with $\text{var}(F_{2t}) = \text{var}(F_{1t})$. Panel C.1 shows the corresponding results for the case where F_{1t} is a common factor, F_{2t} is a weak factor with 90% of its loadings set to zero and $\text{var}(F_{2t}) = \text{var}(F_{1t})$. Panel C.2 presents the corresponding results for the case where F_{1t} is a common factor, F_{2t} is a weak factor without any restriction imposed on its loadings, but now $\text{var}(F_{2t}) = 0.1 \times \text{var}(F_{1t})$. We also report the average number of factors r determined via the ED method of Onatski (2010) as well as the average r and the rank q of factor proxies determined via the cross-validation procedure described in Section 3.1 under both the classical reduced-rank regression setting and its modification with sparsity constraints.

Consider first the results for the case of two common factors (panel A). Almost all procedures perform well in estimating the factor space with values of $R^2_{F, \hat{F}+}$ around 0.93 for most specifications. The exception is $\text{RPC}_{(1,1)}$. The $R^2_{F, \hat{F}+}$ value for $\text{RPC}_{(1,1)}$ is 0.86 when using one proxy. However, its performance quickly improves when adding more, albeit noisier proxies of the second factor. When using 20 proxies, the value of $R^2_{F, \hat{F}+}$ for $\text{RPC}_{(1,1)}$ equals 0.94.

Combining principal components with only one noisy factor proxy deteriorates the out-of-sample performance of the method. The average cross-validation estimate of q , $\hat{q}_{CV}$, is close to zero, suggesting the model does not include observed proxies in factor estimation in this case. However, this finding reverses when more noisy factor proxies are included in the estimation. The relative forecast performance quickly improves and the estimated values of $\hat{q}_{CV}$ approaching one indicate that in most simulations the procedure includes one common noisy factor.

Consider now the case in which 50% of the loadings are zero for the second factor. The results are summarized in Panel B. The values of $R^2_{F, \hat{F}+}$ for PC around 0.87 show that the standard PC approach performs slightly worse at estimating the factor space in the presence of

Table 1: Monte Carlo Results I

Case	#Proxies	$R^2_{F, \hat{F}^+}$		$S^2_{(S)RPC}/S^2_{PC}$				$\hat{r}_{ED}$	$\hat{r}_{CV}$	$\hat{q}_{CV}$	$\hat{r}_{CV}$	$\hat{q}_{CV}$	SRPC
		PC	SOFAR	RPC _(1,1)	RPC	SRPC	SOFAR	RPC _(1,1)	RPC	SRPC			
(A) Common F_2 : no restrictions imposed on F_2 loadings and $\text{var}(F_{2t}) = \text{var}(F_{1t})$													
1	1	0.93	0.92	0.86	0.93	0.93	0.92	1.59	0.91	0.91	2.04	1.93	0.08
2	5	0.93	0.92	0.92	0.93	0.93	1.00	0.98	0.95	0.95	2.04	1.62	0.40
3	10	0.93	0.92	0.93	0.93	0.93	0.95	0.88	0.88	0.89	2.03	1.50	0.52
4	20	0.93	0.92	0.94	0.93	0.93	0.96	0.80	0.83	0.81	2.06	1.27	0.74
5	20	0.93	0.92	0.92	0.93	0.93	0.94	0.99	0.92	0.98	2.03	1.72	0.29
(B) Somewhat weak F_2 : 50% of F_2 loadings set to zero and $\text{var}(F_{2t}) = \text{var}(F_{1t})$													
1	1	0.86	0.89	0.88	0.88	0.88	0.73	0.81	0.81	0.81	1.99	1.09	0.92
2	5	0.87	0.89	0.93	0.93	0.93	0.79	0.46	0.47	0.47	2.02	1.01	1.00
3	10	0.87	0.89	0.93	0.93	0.93	0.77	0.43	0.43	0.43	2.02	1.00	1.00
4	20	0.86	0.89	0.94	0.94	0.94	0.80	0.45	0.45	0.43	2.00	1.00	1.00
5	20	0.86	0.89	0.92	0.92	0.92	0.78	0.52	0.54	0.54	1.99	1.01	1.00
(C.1) Weak F_2 : 90% of F_2 loadings set to zero and $\text{var}(F_{2t}) = \text{var}(F_{1t})$													
1	1	0.51	0.56	0.88	0.69	0.69	0.79	0.33	0.54	0.54	1.27	1.01	0.50
2	5	0.51	0.56	0.91	0.76	0.82	0.78	0.28	0.46	0.37	1.30	1.01	0.69
3	10	0.51	0.56	0.88	0.77	0.86	0.79	0.35	0.49	0.37	1.30	1.01	0.79
4	20	0.51	0.56	0.80	0.77	0.88	0.85	0.55	0.55	0.36	1.31	1.00	0.87
5	20	0.51	0.57	0.78	0.73	0.87	0.79	0.54	0.54	0.34	1.33	1.00	0.79
(C.2) Weak F_2 : no restrictions imposed on factor loadings and $\text{var}(F_{2t}) = 0.1 \times \text{var}(F_{1t})$													
1	1	0.52	0.52	0.88	0.73	0.73	0.93	0.40	0.62	0.62	1.13	1.01	0.59
2	5	0.52	0.51	0.93	0.78	0.80	0.95	0.25	0.47	0.48	1.14	1.01	0.66
3	10	0.52	0.51	0.92	0.79	0.80	0.93	0.28	0.46	0.44	1.14	1.00	0.68
4	20	0.52	0.52	0.91	0.80	0.83	0.96	0.33	0.48	0.44	1.13	1.00	0.72
5	20	0.52	0.51	0.89	0.78	0.81	0.97	0.38	0.53	0.47	1.11	1.01	0.69

Note: This table reports the performance of the PC, SOFAR, and (S)RPC estimators described in Section 4.2. Five proxy designs are considered. In Cases 1–4, we start with the least noisy proxy and sequentially add noisier proxies; in Case 5, we include 20 similarly highly noisy proxies. The dataset dimensions are $N = T = 100$, and the number of simulation repetitions is 1000. The results are summarized by two statistics: $R^2_{F, \hat{F}^+}$ denotes the average R^2 of the regression of F onto $\hat{F}^+$; $S^2_{(S)RPC}/S^2_{PC}$ is the ratio of the squared difference between the out-of-sample forecasts from the indicated method relative to the infeasible forecast based on the true factors divided by the same object for the PC estimator. The last five columns report the average $\hat{r}$ from Onatski (2010)'s ED method and the average $\hat{r}$ and $\hat{q}$ selected by cross-validation in Section 3.1 under both RRR and its modification with sparsity constraints.

a mildly weak factor. The SOFAR method of Uematsu and Yamagata (2023a) implies slightly higher values of explained factor variation, which illustrates the usefulness of their approach in the presence of weak factors. That said, the $R_{F,\hat{F}+}^2$ values further increase for our RPC and SRPC methods when the number of noisy factor proxies is greater than one. Even more strikingly, the forecasting power, summarized by the values of $\frac{S_{(S)RPC}^2}{S_{PC}^2}$, improves dramatically for both methods in this case. When 20 proxies are used, both approaches essentially halve the forecast error relative to the standard PC method.

Finally, we consider the cases where the second factor is very weak, summarized in Panels C.1 and C.2. Panel C.1 shows that the performance of the standard PC method deteriorates substantially in the case where weakness of the second factor is introduced by sparse loadings. While the SOFAR method only leads to a slight improvement in factor space estimation, the RPC and SRPC methods substantially outperform in this dimension. Moreover, they considerably strengthen the forecast performance.

The last panel of Table 1 introduces weakness in the second factor not via sparse loadings but through a substantially smaller factor variance. As before, standard PC performs poorly, while SOFAR provides similar factor-space estimates and only modest forecasting gains. Turning to the RPC and SRPC methods, we see that both substantially improve the estimation of the factor space even for small numbers of noisy factor proxies. Moreover, they dramatically improve the forecasting performance of the model.

Tables C.1 and C.2 in the Online Appendix report the results for $T = 200$, with proxy noise held fixed ($\delta = 0$) and declining with the sample size ($\delta = 1/2$). The results reinforce the previous findings. In the weak-factor designs in Panels C.1 and C.2 in the respective tables, RPC and SRPC continue to yield substantial improvements over PC in both factor estimation and forecasting even when proxy noise remains fixed. The gains are generally larger when proxy noise declines with T . In Panel A of Table C.1, PC also benefits from the longer sample, reducing the relative gains of RPC and SRPC under fixed proxy noise. Finally, regardless of the choice of δ , the longer sample confirms the benefit of imposing sparsity in Panel C.1 when many noisy proxies are included and improves model selection in Panel C.2,

where the average selected total number of factors, $\hat{r}_{CV} + \hat{q}_{CV}$, is close to its true value of two.

In summary, these experiments show that the use of RPC and SRPC substantially improves factor estimation and forecast performance in the presence of weak factors. When there is no weak factor, the methods do not come at a cost relative to the standard PC approach.

5 Monte Carlo II: Structural IRFs

In this section, we use data-based simulations to illustrate the desirable properties of our proposed estimator for the identification of structural economic shocks using a FAVAR model in the presence of noisy observable factors. Specifically, we identify structural shocks driving factors when one factor is observed with measurement error. Since we know the true structural impulse responses, we can assess the degree to which measurement error in the factors induces a bias in FAVAR-based estimates of structural impulse responses.

For illustration, consider the following simple case with two proxies for a single unobserved factor. Suppose that the researcher observes two proxies of which one is measured without error while the other contains some measurement noise. The researcher does not know which one is the proxy without noise. Now estimate three FAVARs: (1) with the perfectly observed factor proxy; (2) with the noisy proxy; and (3) with a linear combination of the two proxies estimated using our reduced-rank approach.

Our object of interest is the structural impulse response to a shock underlying the factor. We assess the bias induced by the noise in the proxy, and show how the results differ across the three FAVAR specifications.

5.1 Data

Our data are quarterly observations on 209 time series taken from Stock and Watson (2016). This dataset covers the most important categories of U.S. economic activity, as well as variables representing the oil market and global economic activity. Each series was transformed to be approximately stationary by taking growth rates or first differences. A small number of outliers

were replaced by the median of the preceding five periods, and the long-run mean of each series was removed using a biweight low-pass filter with a bandwidth of 100 quarters. Note that the original Stock and Watson (2016) dataset consists of 207 series. We make three small changes to the data. First, we update Kilian (2009)’s global activity index to its more recent Kilian (2019) version. Moreover, we include the growth rate of IP for the OECD countries and the six largest emerging market economies from Baumeister and Hamilton (2019) and the global economic conditions indicator developed by Baumeister et al. (2022) as two additional proxies of global activity. The latter has been constructed as the leading principal component of 16 indicators covering real activity, commodity prices, financial variables, as well as variables related to transportation, weather, energy and consumer expectations. We follow Stock and Watson (2016) and use the sample from 1985Q1 to 2014Q4.

5.2 Experimental Design

Our Monte Carlo analysis is presented in the context of identifying oil and global demand shocks. We consider the structural impulse responses obtained by using the recursive identification scheme of Kilian (2009) which distinguishes between three shocks: an oil supply shock, a global demand shock, and an oil-specific demand shock. We use three FAVAR DGPs, and empirically calibrate them following the implementation of Stock and Watson (2016) of Kilian (2009). Each is a six-factor FAVAR containing global oil production, the real oil price, one global activity factor, and three latent factors. The designs differ only in which activity measure is treated as the true factor: the Kilian index, Baumeister and Hamilton (2019)’s world industrial production growth, or Baumeister et al. (2022)’s global economic conditions indicator. The factor loadings, factor VAR coefficients, and univariate idiosyncratic processes are estimated over 1985Q1–2014Q4 and then used to generate the simulated samples. Further details of the DGPs are provided in Online Appendix B.1.

Identification We use a recursive identification scheme to separately identify an oil supply, a global demand, and an oil-specific demand shock. To achieve this, we order global oil

production, a global demand measure and the real price of oil as the first three variables in the FAVAR and adopt a Cholesky factorization of the shocks. Moreover, we adopt the “unit-effect” normalization of Stock and Watson (2016), so that a unit increase in each shock increases the associated factor innovation by one unit.

Structural Impulse Response Functions We now define the structural impulse responses which are the main objective criterion in this Monte Carlo analysis. We formulate the responses in terms of the level of each series. This is accomplished by using cumulative impulse responses for those series that enter the model after first or second differencing. We assume that the factor innovations η_t are linear combinations of structural shocks:

$$\eta_t = H\epsilon_t. \quad (17)$$

The recursive identification with the unit-effect normalization implies $H = \text{Chol}(\Sigma_\eta)\Sigma_\epsilon^{-1/2}$, where $\Sigma_\epsilon = \text{diag}(\{\text{Chol}(\Sigma_\eta)_{jj}^2, j = 1, \dots, 6\})$ with Chol denoting the Cholesky factorization. Letting $\Phi(L)$ denote the lag polynomial matrix of the factor VAR, the structural impulse responses are given by

$$\text{SIRF}_{ij}(h) = \begin{cases} C_{ij,h} & \text{if } X_{it} \text{ is in levels} \\ \sum_{k=0}^h C_{ij,k} & \text{if } X_{it} \text{ is in first or second differences} \end{cases}, \quad \text{and} \quad C_h = \Lambda B_h H, \quad (18)$$

where B_h is the h th lag matrix of the polynomial $B(L) = (I - \Phi(L)L)^{-1}$.

Given the simulation draws of the data and the true SIRFs for each DGP, we assess the FAVAR estimator of the SIRFs using different measures of global activity. Specifically, we estimate four FAVARs which all have a similar specification to the above model of Stock and Watson (2016), but use different global activity measures as inputs:

1. *FAVAR-K*: the FAVAR in which the global activity measure is Kilian (2019)’s index,
2. *FAVAR-BH*: the FAVAR in which the global activity measure is Baumeister and Hamilton (2019)’s world IP growth,

3. *FAVAR-BKL*: the FAVAR in which the global activity measure is Baumeister et al. (2022)'s global economic conditions indicator,
4. *FAVAR-RPC*: the FAVAR in which the global activity measure is an estimated linear combination of the three measures obtained using RPC.

We then compare the SIRF estimates to the true SIRFs and assess the bias induced by using noisy measures of global activity. Regarding the lag order, we consider both fixing $p = 4$ lags and selecting the FAVAR lag length by the Akaike Information Criterion (AIC) with $p \leq 8$. Both cases delivered similar results, and in what follows we report those based on the lag order selected by AIC. For all the Monte Carlo results we conduct 1000 repetitions.

5.3 Results

This section presents the results of the Monte Carlo analysis. For each DGP and identification approach, we focus on the corresponding average SIRFs across all Monte Carlo repetitions and assess the bias with respect to their true counterparts. The average R^2 from regressions of selected variables on the estimated factors across all Monte Carlo repetitions, and the SIRFs for selected variables are reported in Online Appendix B.2. Here, we focus on average absolute bias across all variables.

Bias in SIRFs Columns A-C in Figure 2 assess the absolute bias in the FAVAR estimators of the SIRFs for the three shocks. Using the simulation draws of the data from each DGP, the figure shows the average (across all 209 series) of the absolute bias up to the three-year horizon. Specifically, letting $\widehat{\text{SIRF}}_{ij}(h)$ denote the estimated impulse response of variable i to shock j for a given FAVAR estimator, the absolute bias is constructed as $|\widehat{E}(\widehat{\text{SIRF}}_{ij}(h) - \text{SIRF}_{ij}(h))|$ where $\widehat{E}$ denotes the average over all the Monte Carlo repetitions. Each chart has three lines corresponding to three FAVARs we considered. Before taking the average across different variables, we scale the SIRFs following Li et al. (2024). This is achieved by dividing the absolute bias by $\sqrt{\frac{1}{h_{max}+1} \sum_{k=0}^{h_{max}} \text{SIRF}_{ij}^2(k)}$, the average root mean squared value of the true

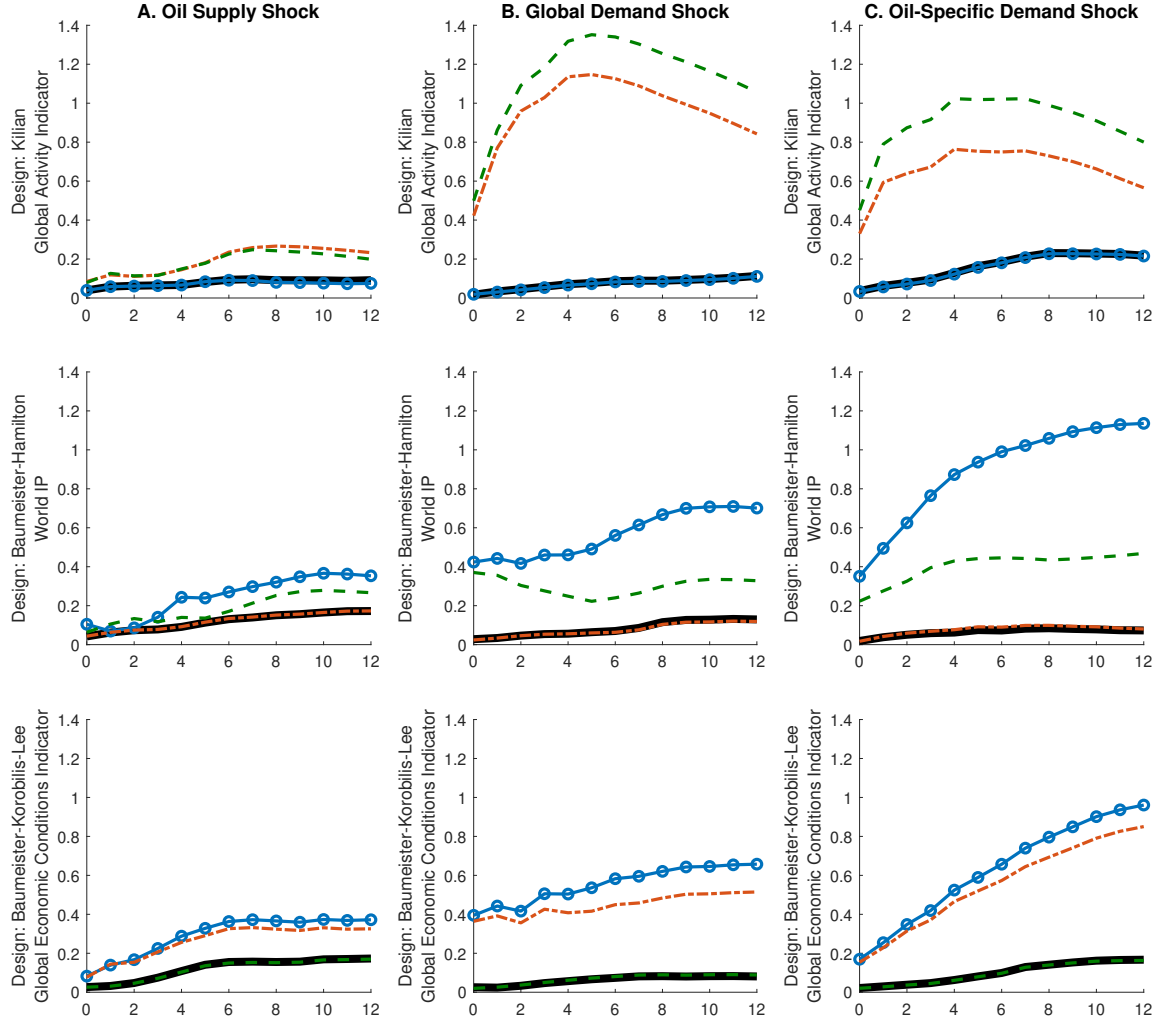


Figure 2: Average bias across 209 series. This figure plots, at each horizon, the average across 209 variables of the absolute bias in the FAVAR estimates of structural impulse response functions, normalized by $\sqrt{\frac{1}{13} \sum_{k=0}^{12} \text{SIRF}_{ij}^2(k)}$. The black solid line corresponds to FAVAR-RPC, the blue dotted line to FAVAR-K, the orange dash-dotted line to FAVAR-BH, and the green dashed line to FAVAR-BKL. Lower values indicate smaller average bias.

SIRFs. The figure shows that the FAVAR-RPC estimator has a substantially smaller absolute bias in all three designs we consider. For the design using the Kilian indicator as the true measure (first row), the FAVAR-RPC and FAVAR-K estimators produce similar results. The absolute biases increase with the horizon, and the bias of the FAVAR-BH and FAVAR-BKL estimators substantially exceeds that of FAVAR-RPC and FAVAR-K.

In the second design using the Baumeister and Hamilton (2019) world IP growth as the true global activity measure (second row), the average absolute biases of the FAVAR-RPC and

FAVAR-BH estimators are nearly indistinguishable: the reduced-rank PC estimator applied to a model with the three global activity measures as factor proxies gives rise to the same absolute bias as the FAVAR-BH, which is considerably smaller than that of the FAVAR-K and FAVAR-BKL estimators. Finally, for the design using the Baumeister et al. (2022) indicator as the global activity measure (third row), the FAVAR-RPC and FAVAR-BKL estimators produce identical results while the FAVAR-K and FAVAR-BH estimators give rise to SIRFs that show substantial bias.

In summary, the results in this section highlight that in the presence of noisy factor proxies, our reduced-rank PC estimator can be used to estimate structural impulse response functions which feature substantially less bias than approaches which assume the true factor proxies are known and observed without error.

6 Empirical Analysis

In this section, we document the usefulness of our approach in a recursive pseudo-out-of-sample forecast exercise for Treasury returns in the spirit of Ludvigson and Ng (2009), which serves as a natural benchmark.

6.1 Forecasts

Following Ludvigson and Ng (2009), we consider forecasts of 12-month-ahead log excess returns, $rx_{t+12}^{(n)}$, for maturities $n = 2, 3, 4, 5$, as well as their cross-sectional average, $\overline{rx}_{t+12}$. The excess return on an n -year zero-coupon bond is defined as $rx_{t+12}^{(n)} \equiv r_{t+12}^{(n)} - y_t^{(1)}$, where $r_{t+12}^{(n)}$ is the log holding-period return from buying an n -year bond at date t and selling it as an $(n-1)$ -year bond one year later, and $y_t^{(1)}$ is the one-year log yield. The first forecast is produced using data available through 1989M12, and the forecast evaluation period ends in 2024M12.

6.2 Forecasting models

We organize the forecasting models around the reduced-form factor representation in equation (5). Let X_t denote the $N \times 1$ vector of FRED-MD variables, Y_t an $m \times 1$ vector of observed factor proxies, and F_t an $r \times 1$ vector of unobserved factors. As observed factor proxies, we consider cross-sectionally differenced returns, following Crump and Gospodinov (2022), who propose to model yield co-movement to isolate the true underlying factor space from statistical artifacts. The factor model is $X_t = \Lambda_{X\bar{F}}B'Y_t + \Lambda_{XF}F_t + u_{Xt}$, where $\Lambda_{X\bar{F}}$, Λ_{XF} , and B have dimensions $N \times q$, $N \times r$, and $m \times q$, respectively. The observed proxies Y_t and the number of their combinations, q , vary across specifications.⁶ We consider the following forecasting models:

1. $RPC_{(q=1)}$: This specification uses one return-based factor. The observed proxy vector is $Y_t = (rx_t^{(2)}, dr_t^{(3)}, dr_t^{(4)}, dr_t^{(5)})'$, where $dr_t^{(n)} = rx_t^{(n)} - rx_t^{(n-1)}$ is the difference return between adjacent maturities. The return-based factor is $\hat{B}'Y_t$, and the forecast is obtained from $rx_{t+12} = \alpha + \beta\hat{B}'Y_t + \epsilon_{t+12}$. We drop the superscripts denoting maturity for brevity.
2. $RPC_{(\hat{q})}$: This specification is the same as $RPC_{(q=1)}$, except that it includes $\hat{q} \in \{1, 2\}$ combinations of $Y_t = (rx_t^{(2)}, dr_t^{(3)}, dr_t^{(4)}, dr_t^{(5)})'$. The number of combinations, $\hat{q}$, is selected by the cross-validation procedure described in Section 3.1.
3. LN : This is the Ludvigson–Ng benchmark. Following Ludvigson and Ng (2009), we construct $\hat{G}_t = (\hat{F}_{1t}, \hat{F}_{1t}^3, \hat{F}_{3t}, \hat{F}_{4t}, \hat{F}_{8t})'$, where the $\hat{F}_{it}$ are principal components extracted from the FRED-MD dataset. To address the challenges that arise when estimating factor models with data that include the COVID period, we hold the pre-COVID factor loadings fixed when constructing factors during the COVID period. This adjustment is motivated by Ng (2021), who argues that COVID-period observations create new challenges for estimating economic factors, and is related to the treatment of COVID-period factor estimation in Stock and Watson (2025).⁷

⁶Throughout the forecasting exercise, we set $r = 8$ to maintain comparability with the Ludvigson–Ng benchmark. The cross-validation procedure is used only to select the number of return-proxy combinations.

⁷Specifically, we estimate the factors and loadings by principal components over the pre-COVID sample

4. *LN-SOFAR*: This specification is the same as *LN*, except that the factors are estimated using the SOFAR estimator of Uematsu and Yamagata (2023a). Since SOFAR estimates depend on the scale of the variables, we scale each FRED-MD series by the standard deviation of its idiosyncratic component, estimated from the residuals of the eight-factor principal-component model. This avoids downweighting series with substantial variation attributable to the common factors. We then align the resulting SOFAR factors with the corresponding principal components using an orthogonal Procrustes alignment before including them in the forecasting model.⁸ This implementation of the SOFAR estimator delivers somewhat better forecast performance compared to a version where the FRED-MD series are scaled by the total standard deviation as in Uematsu and Yamagata (2023a).
5. *LN + RPC_(q=1)* and *LN + RPC_(q̂)*: These specifications augment the *LN* factors with the return-based factors used in *RPC_(q=1)* and *RPC_(q̂)*, respectively.
6. *LN + FR*: This specification augments the *LN* factors with the forward-rate vector $FR_t = (y_t^{(1)}, f_t^{(2)}, f_t^{(3)}, f_t^{(4)}, f_t^{(5)})'$, where $f_t^{(n)}$ is the log forward rate for the period from $t + 12(n - 1)$ to $t + 12n$.
7. *LN + RX⁽ⁿ⁾* and *LN + $\overline{RX}$* : These specifications augment the *LN* factors with excess returns. The model *LN + RX⁽ⁿ⁾* adds $rx_t^{(n)}$ as an additional predictor. The model *LN + $\overline{RX}$* instead adds the average excess return, $\overline{rx}_t = (rx_t^{(2)} + rx_t^{(3)} + rx_t^{(4)} + rx_t^{(5)})/4$.
8. *EH*: For comparison, we also report the expectations hypothesis benchmark, i.e., the constant-only model $rx_{t+12} = \alpha + \epsilon_{t+12}$ which assumes no time-varying predictability in excess bond returns.

until February 2020, obtaining the loading matrix $\hat{\Lambda}_{XF}$. Given these loadings, factors are computed as $\hat{F}_t = (\hat{\Lambda}'_{XF} \hat{\Lambda}_{XF})^{-1} \hat{\Lambda}'_{XF} X_t$. For COVID-period observations, we keep $\hat{\Lambda}_{XF}$ fixed and apply the same projection to X_t . The forecast is then obtained from $rx_{t+12} = \alpha + \beta' \hat{G}_t + \epsilon_{t+12}$.

⁸Letting $\hat{F}_{PC} = (\hat{F}_1, \hat{F}_3, \hat{F}_4, \hat{F}_8)$ denote the four PC factor estimates used in *LN* and $\hat{F}_{SOFAR}$ the eight normalized SOFAR factor estimates, we obtain $\hat{Q} = \arg \min_{Q'Q=I_4} \|\hat{F}_{PC} - \hat{F}_{SOFAR}Q\|_F^2$ using the pre-COVID observations and use $\hat{F}_{SOFAR}\hat{Q}$ as the aligned factor estimates.

Table 2: Relative MSFEs of Excess Bond Returns, 1989M12–2024M12

	A. Relative to EH benchmark					B. Relative to LN				
	$rx_{t+12}^{(2)}$	$rx_{t+12}^{(3)}$	$rx_{t+12}^{(4)}$	$rx_{t+12}^{(5)}$	$\overline{rx}_{t+12}$	$rx_{t+12}^{(2)}$	$rx_{t+12}^{(3)}$	$rx_{t+12}^{(4)}$	$rx_{t+12}^{(5)}$	$\overline{rx}_{t+12}$
RPC _(q=1)	0.815**	0.844**	0.855*	0.874*	0.852*	1.081	1.117	1.130	1.122	1.129
RPC _(q̂)	0.810**	0.837**	0.855*	0.879	0.852*	1.075	1.108	1.129	1.128	1.128
EH	1.000	1.000	1.000	1.000	1.000	1.327	1.323	1.321	1.284	1.324
LN	0.754**	0.756***	0.757***	0.779***	0.755***	1.000	1.000	1.000	1.000	1.000
LN-SOFAR	0.759**	0.750***	0.738***	0.756***	0.739***	1.007	0.992	0.976	0.970	0.979
LN + RPC _(q=1)	0.738**	0.719**	0.698***	0.711***	0.702***	0.979	0.951	0.923	0.913	0.929
LN + RPC _(q̂)	0.678**	0.678***	0.676***	0.702***	0.675***	0.899*	0.897*	0.893*	0.901	0.894*
LN + FR	0.886	0.895	0.894	0.901	0.897	1.175	1.185	1.181	1.157	1.187
LN + RX ⁽²⁾	0.791*	0.771**	0.751**	0.765**	0.755**	1.050	1.020	0.992	0.982	1.000
LN + RX ⁽³⁾	0.799*	0.784**	0.768**	0.784**	0.771**	1.060	1.037	1.015	1.007	1.021
LN + RX ⁽⁴⁾	0.806*	0.793**	0.780**	0.798**	0.783**	1.069	1.049	1.031	1.024	1.037
LN + RX ⁽⁵⁾	0.813*	0.801**	0.790**	0.808**	0.792**	1.078	1.060	1.043	1.038	1.048
LN + $\overline{RX}$	0.808*	0.794**	0.779**	0.796**	0.782**	1.072	1.050	1.030	1.022	1.036

Note: This table reports mean squared forecast errors (MSFEs) for the forecasting models listed in the rows. Panel A reports MSFEs relative to the expectations hypothesis (EH) benchmark. Panel B reports MSFEs relative to a benchmark model including only the Ludvigson–Ng (LN) factors. The columns report forecasts of 12-month-ahead excess bond returns, $rx_{t+12}^{(n)}$. The first out-of-sample forecast is produced using data through 1989M12, and the final forecast is for 2024M12. Bold entries indicate improvements relative to the LN benchmark. Gray-shaded entries indicate the best-performing specification among all alternative factor models. Statistical significance is assessed using the Diebold–Mariano test: ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively. The forecasting factors are described in the text.

6.3 Results

The results of the forecast exercise are summarized in Table 2. We report mean squared forecast errors relative to two benchmarks: the constant return expectations hypothesis (EH) model and the Ludvigson and Ng (2009) model, which only relies on macro factors. We assess statistical significance of the forecast improvement via the Diebold and Mariano (1995) statistic using HAC standard errors. The results show that the LN + RPC model outperforms both benchmarks significantly, reducing MSFEs by about one third relative to the EH model when the number of factor proxies is determined by cross-validation. Relative to the LN model, the improvement is a sizable 10% MSFE reduction. Among the competitor models, the SOFAR approach by Uematsu and Yamagata (2023a) performs the best. It achieves MSFE reductions of about 25% relative to the EH benchmark and modestly outperforms the LN benchmark.

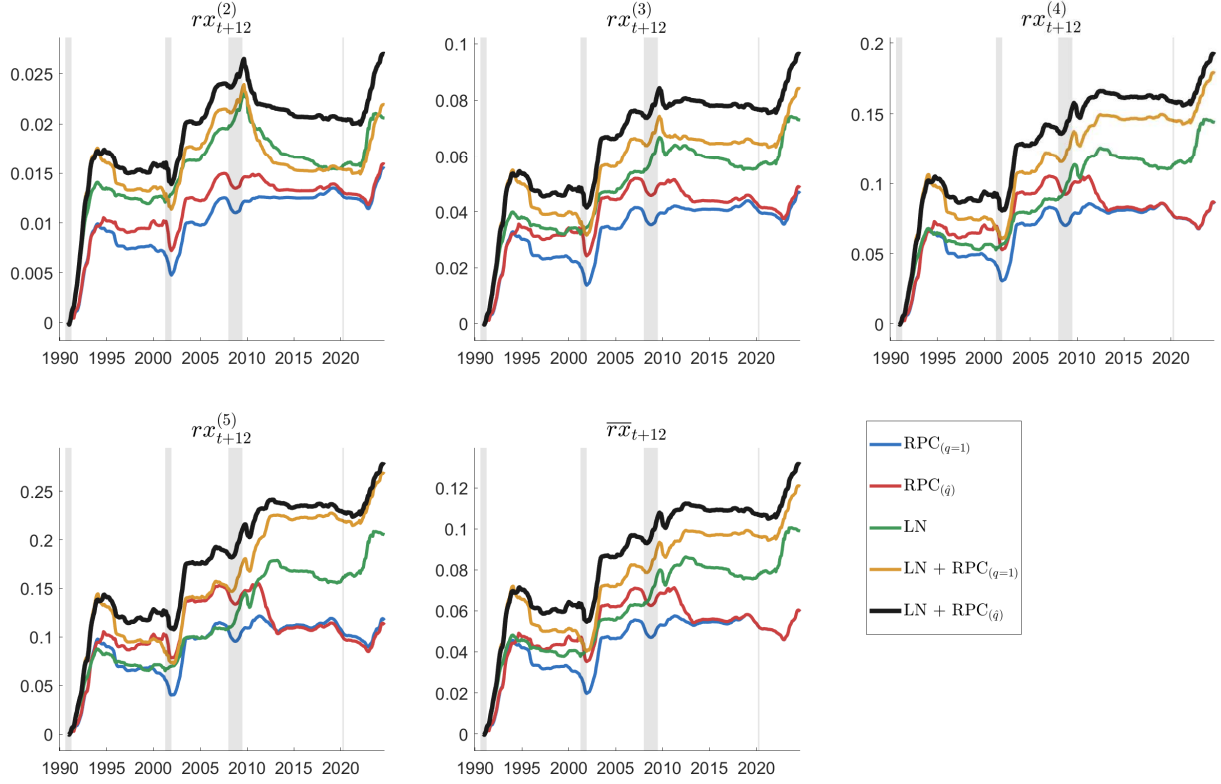


Figure 3: Cumulative sum of squared forecast error differentials. This figure shows the recursively calculated sum of squared forecast errors for the EH benchmark minus the sum of squared forecast errors based on forecasting factors for the 12-month-ahead excess bond returns, $rx_{t+12}^{(n)}$. The first out-of-sample forecast is produced using data through 1989M12, and the final forecast is for 2024M12. The forecasting factors are described in the text.

Figure 3 shows the cumulative sum of squared forecast error differentials relative to the EH benchmark. The charts underscore that the LN+RPC model with the number of factor proxies determined by cross-validation consistently outperforms the benchmark as well as all competitor models throughout the out-of-sample forecast period. In sum, this section highlights that our proposed estimation approach can considerably improve the forecasting ability in a pseudo-out-of-sample setting.

7 Conclusion

In this paper, we have introduced a method for extending FAVARs to include multiple observable time series as noisy proxies of some true underlying factors. Our approach builds on the observation that a FAVAR model can be represented with a rank restriction when there are more observable proxies than factors. We propose an estimator that alternates between principal components and solving a reduced-rank regression. We establish the regularity conditions under which our proposed estimator consistently recovers the factor space.

We further discuss an extension adding sparsity constraints on the loading matrix for proxy selection and regularization and propose a cross-validation procedure to jointly determine the number of factors along with the rank and penalty parameters.

In a series of Monte Carlo simulations, we document that combining principal components with (sparse) reduced-rank regressions improves factor estimation and forecasting, especially when factors are weak. It also substantially reduces the bias which arises in the presence of noisy factor proxies when estimating structural impulse response functions. Finally, we demonstrate the usefulness of our approach in a pseudo out-of-sample forecasting exercise for Treasury returns.

Our method allows researchers to include relevant information from observable proxies into the factor space without relying on the standard assumption made in FAVARs that some of the factors are perfectly observed.

References

- Ahn, S. C. and Horenstein, A. R. (2013), “Eigenvalue Ratio Test for the Number of Factors,” *Econometrica*, 81, 1203–1227.
- Anderson, T. W. (1951), “Estimating Linear Restrictions on Regression Coefficients for Multivariate Normal Distributions,” *The Annals of Mathematical Statistics*, 22, 327–351.
- Bai, J. (2003), “Inferential Theory for Factor Models of Large Dimensions,” *Econometrica*, 71, 135–171.
- Bai, J. and Ng, S. (2002), “Determining The Number of Factors in Approximate Factor Models,” *Econometrica*, 70, 191–221.
- (2008), “Forecasting Economic Time Series Using Targeted Predictors,” *Journal of Econometrics*, 146, 304–317.
- (2013), “Principal Components Estimation and Identification of Static Factors,” *Journal of Econometrics*, 176, 18–29.
- (2023), “Approximate factor models with weaker loadings,” *Journal of Econometrics*, 235, 1893–1916.
- Bailey, N., Kapetanios, G., and Pesaran, M. H. (2021), “Measurement of Factor Strength: Theory and Practice,” *Journal of Applied Econometrics*, 36, 587–613.
- Baumeister, C. and Hamilton, J. D. (2019), “Structural Interpretation of Vector Autoregressions with Incomplete Identification: Revisiting the Role of Oil Supply and Demand Shocks,” *American Economic Review*, 109, 1873–1910.
- Baumeister, C., Korobilis, D., and Lee, T. K. (2022), “Energy Markets and Global Economic Conditions,” *The Review of Economics and Statistics*, 104, 828–844.
- Bernanke, B. S., Boivin, J., and Elias, P. (2005), “Measuring the Effects of Monetary Policy: A Factor-Augmented Vector Autoregressive (FAVAR) Approach,” *The Quarterly Journal of Economics*, 120, 387–422.
- Boivin, J., Giannoni, M. P., and Mihov, I. (2009), “Sticky Prices and Monetary Policy: Evidence from Disaggregated US Data,” *American Economic Review*, 99, 350–384.
- Boivin, J. and Ng, S. (2006), “Are More Data Always Better for Factor Analysis?” *Journal of Econometrics*, 132, 169–194.
- Bunea, F., She, Y., and Wegkamp, M. H. (2011), “Optimal Selection of Reduced Rank Estimators of High-Dimensional Matrices,” *The Annals of Statistics*, 39, 1282–1309.
- Chen, L. and Huang, J. Z. (2012), “Sparse Reduced-Rank Regression for Simultaneous Dimension Reduction and Variable Selection,” *Journal of the American Statistical Association*, 107, 1533–1545.

- Cheng, X., Liao, Z., and Schorfheide, F. (2016), “Shrinkage Estimation of High-Dimensional Factor Models with Structural Instabilities,” *The Review of Economic Studies*, 83, 1511–1543.
- Chudik, A., Pesaran, M. H., and Tosetti, E. (2011), “Weak and Strong Cross-Section Dependence and Estimation of Large Panels,” *The Econometrics Journal*, 14, C45–C90.
- Crump, R. K. and Gospodinov, N. (2022), “On the Factor Structure of Bond Returns,” *Econometrica*, 90, 295–314.
- Diebold, F. X. and Mariano, R. S. (1995), “Comparing Predictive Accuracy,” *Journal of Business & Economic Statistics*, 13, 253–263.
- Friedman, J. H., Hastie, T., and Tibshirani, R. (2010), “Regularization Paths for Generalized Linear Models via Coordinate Descent,” *Journal of Statistical Software*, 33, 1–22.
- Giglio, S., Xiu, D., and Zhang, D. (2025), “Test Assets and Weak Factors,” *The Journal of Finance*, 80, 259–319.
- Izenman, A. J. (1975), “Reduced-Rank Regression for the Multivariate Linear Model,” *Journal of Multivariate Analysis*, 5, 248–264.
- Kapetanios, G. and Marcellino, M. (2010), “Factor-GMM Estimation with Large Sets of Possibly Weak Instruments,” *Computational Statistics & Data Analysis*, 54, 2655–2675, the Fifth Special Issue on Computational Econometrics.
- Kilian, L. (2009), “Not All Oil Price Shocks Are Alike: Disentangling Demand and Supply Shocks in the Crude Oil Market,” *American Economic Review*, 99, 1053–1069.
- (2019), “Measuring Global Real Economic Activity: Do Recent Critiques Hold Up to Scrutiny?” *Economics Letters*, 178, 106–110.
- Lazarus, E., Lewis, D. J., Stock, J. H., and Watson, M. W. (2018), “HAR Inference: Recommendations for Practice,” *Journal of Business & Economic Statistics*, 36, 541–559.
- Li, D., Plagborg-Møller, M., and Wolf, C. K. (2024), “Local Projections vs. VARs: Lessons from thousands of DGPs,” *Journal of Econometrics*, 244, 105722.
- Ludvigson, S. C. and Ng, S. (2009), “Macro Factors in Bond Risk Premia,” *The Review of Financial Studies*, 22, 5027–5067.
- McCracken, M. W. and Ng, S. (2016), “FRED-MD: A Monthly Database for Macroeconomic Research,” *Journal of Business & Economic Statistics*, 34, 574–589.
- Moench, E., Ng, S., and Potter, S. (2013), “Dynamic Hierarchical Factor Models,” *The Review of Economics and Statistics*, 95, 1811–1817.
- Ng, S. (2021), “Modeling Macroeconomic Variations After COVID-19,” Tech. rep., National Bureau of Economic Research.

- Onatski, A. (2010), “Determining the Number of Factors from Empirical Distribution of Eigenvalues,” *The Review of Economics and Statistics*, 92, 1004–1016.
- (2012), “Asymptotics of the Principal Components Estimator of Large Factor Models with Weakly Influential Factors,” *Journal of Econometrics*, 168, 244–258.
- Parker, J. and Sul, D. (2016), “Identification of Unknown Common Factors: Leaders and Followers,” *Journal of Business & Economic Statistics*, 34, 227–239.
- Qian, J., Tanigawa, Y., Li, R., Tibshirani, R., Rivas, M. A., and Hastie, T. (2022), “Large-Scale Multivariate Sparse Regression with Applications to UK Biobank,” *The Annals of Applied Statistics*, 16, 1891.
- Schorfheide, F. (2005), “VAR Forecasting under Misspecification,” *Journal of Econometrics*, 128, 99–136.
- Stock, J. H. and Watson, M. W. (2002), “Forecasting Using Principal Components from a Large Number of Predictors,” *Journal of the American Statistical Association*, 97, 1167–1179.
- (2005), “Implications of Dynamic Factor Models for VAR Analysis,” Working Paper 11467, National Bureau of Economic Research.
- (2016), “Dynamic Factor Models, Factor-Augmented Vector Autoregressions, and Structural Vector Autoregressions in Macroeconomics,” in *Handbook of Macroeconomics*, Elsevier, vol. 2, pp. 415–525.
- (2025), “Recovering from COVID,” *Brookings Papers on Economic Activity*, 2025, 297–346.
- Uematsu, Y. and Yamagata, T. (2023a), “Estimation of Sparsity-Induced Weak Factor Models,” *Journal of Business & Economic Statistics*, 41, 213–227.
- (2023b), “Inference in Sparsity-Induced Weak Factor Models,” *Journal of Business & Economic Statistics*, 41, 126–139.
- Velu, R. and Reinsel, G. C. (1998), *Multivariate Reduced-Rank Regression: Theory and Applications*, Springer New York, NY.
- Wedin, P.-Å. (1972), “Perturbation Bounds in Connection with Singular Value Decomposition,” *BIT*, 12, 99–111.
- Yuan, M. and Lin, Y. (2006), “Model Selection and Estimation in Regression with Grouped Variables,” *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 68, 49–67.

Online Appendix for

Factor-Augmented VARs with Noisy Factor Proxies

A Large-sample properties of the RPC estimator

This appendix establishes consistency of the factor space estimated by the iterative RPC procedure. Consider the following representation in matrix form:

$$X = \tilde{F}\Lambda'_{X\tilde{F}} + F\Lambda'_{XF} + e_X, \quad (\text{A.1})$$

$$Y = \tilde{F}\Lambda'_{Y\tilde{F}} + T^{-\delta}e_Y, \quad \delta > 0. \quad (\text{A.2})$$

The matrices $\tilde{F}$ and F have dimensions $T \times q$ and $T \times r$. The proxy dimension m and the factor dimensions q and r are fixed.

Let

$$F^+ = [\tilde{F}, F], \quad X_0 = \tilde{F}\Lambda'_{X\tilde{F}} + F\Lambda'_{XF}, \quad B = \Lambda_{Y\tilde{F}}(\Lambda'_{Y\tilde{F}}\Lambda_{Y\tilde{F}})^{-1}.$$

Thus $X = X_0 + e_X$, $\Lambda'_{Y\tilde{F}}B = I_q$, and $YB = \tilde{F} + T^{-\delta}e_YB$. For a matrix A , let P_A denote the projection onto its column space and let $M_A = I - P_A$. In the proof, $\bar{F}$ denotes the matrix containing the leading r left singular vectors of X_0 . For two matrices A_1 and A_2 , define

$$D(A_1, A_2) = \|P_{A_1} - P_{A_2}\|_2. \quad (\text{A.3})$$

Throughout, $\|\cdot\|$, $\|\cdot\|_2$ and $\|\cdot\|_*$ denote the Frobenius norm, spectral norm and nuclear norm, respectively. For a symmetric $d \times d$ matrix A , $\lambda_j(A)$ denotes its j th largest eigenvalue. For a

general $d_1 \times d_2$ matrix A , $\sigma_j(A)$ denotes its j th largest singular value. Also, κ , $\underline{\kappa}$ and $\bar{\kappa}$ denote generic positive constants which are independent of N and T .

A.1 Assumptions

Assumption A.1 (Factors). The factors satisfy

$$\mathbb{E}\|F_t^+\|^8 \leq \bar{\kappa}, \quad T^{-1}F^{+'}F^+ \xrightarrow{p} \Sigma_{F^+}, \quad \Sigma_{F^+} \succ 0. \quad (\text{A.4})$$

Assumption A.2 (Factor loadings). The loading matrix associated with F is pervasive, in the sense that

$$\underline{\kappa}N \leq \lambda_r(\Lambda'_{XF}\Lambda_{XF}) \leq \lambda_1(\Lambda'_{XF}\Lambda_{XF}) \leq \bar{\kappa}N. \quad (\text{A.5})$$

For some fixed $0 < \alpha < 1$, the loadings on $\tilde{F}$ have strength N^α after removing the loading space of F :

$$\underline{\kappa}N^\alpha \leq \lambda_q(\Lambda'_{X\tilde{F}}M_{\Lambda_{XF}}\Lambda_{X\tilde{F}}) \leq \lambda_1(\Lambda'_{X\tilde{F}}\Lambda_{X\tilde{F}}) \leq \bar{\kappa}N^\alpha. \quad (\text{A.6})$$

Assumption A.3 (Factor proxies). The matrix $\Lambda_{Y\tilde{F}}$ has full column rank and

$$\underline{\kappa} \leq \lambda_q(\Lambda'_{Y\tilde{F}}\Lambda_{Y\tilde{F}}) \leq \lambda_1(\Lambda'_{Y\tilde{F}}\Lambda_{Y\tilde{F}}) \leq \bar{\kappa}. \quad (\text{A.7})$$

Assumption A.4 (Idiosyncratic and measurement errors). The idiosyncratic component has the linear-process representation

$$e_{X,t} = \sum_{\ell=0}^{\infty} A_{\ell,N} \varepsilon_{t-\ell}, \quad \sum_{\ell=0}^{\infty} \|A_{\ell,N}\|_2 \leq \bar{\kappa}. \quad (\text{A.8})$$

The elements of ε_t are independent across i and t , have mean zero and variances bounded above, and have uniformly bounded eighth moments.

The measurement errors satisfy

$$e_{Y,t} = \sum_{\ell=0}^{\infty} D_{\ell} \nu_{t-\ell}, \quad \sum_{\ell=0}^{\infty} \|D_{\ell}\|_2 \leq \bar{\kappa}, \quad (\text{A.9})$$

where ν_t is centered, independent across t , and has uniformly bounded eighth moments. The covariance matrix $E(e_{Y,t} e'_{Y,t})$ is positive definite. The innovation sequences $\{\varepsilon_t\}$ and $\{\nu_t\}$ are mutually independent and are independent of $\{F_t^+\}$.

A.2 Main result

Let $\widehat{F}^{(0)}$ be the first r principal components of X . Given $\widehat{F}^{(k)}$, define

$$\widetilde{Y}_{\widehat{F}^{(k)}} = M_{\widehat{F}^{(k)}} Y. \quad (\text{A.10})$$

Let $\widehat{B}^{(k+1)}$ and $\widehat{\Lambda}_{X\widetilde{F}}^{(k+1)}$ be the corresponding rank- q RRR estimates, and write

$$\widehat{C}^{(k+1)} = \widehat{\Lambda}_{X\widetilde{F}}^{(k+1)} \widehat{B}^{(k+1)'}$$

The subsequent PC step produces $\widehat{F}^{(k+1)}$ from $X - Y\widehat{C}^{(k+1)'}$.

The result below concerns the RPC estimator for any fixed number of iterations, assuming that the true values of r and q are known.

Theorem A.1. *Let $N, T \rightarrow \infty$ with $T/N^{1-\alpha} \rightarrow \infty$, $c_{NT} = \min\{\sqrt{N}, \sqrt{T}\}$, $r_{NT} = \sqrt{\frac{N^{1-\alpha}}{T}} + c_{NT}^{-1} + T^{-\delta}$, and suppose Assumptions A.1–A.4 hold. Then, for every fixed integer $K \geq 0$,*

$$\max_{0 \leq k \leq K} D\left([Y\widehat{B}^{(k+1)}, \widehat{F}^{(k)}], F^+\right) = O_p(r_{NT}) = o_p(1). \quad (\text{A.11})$$

A.3 Proof

We first present the two standard results used throughout the proof.

Lemma A.1. (i) Let U_0 be a fixed-dimensional left singular subspace of A_0 , separated from its complement by γ_{NT} . If $A = A_0 + R$, $\|R\|_2/\gamma_{NT} = o_p(1)$, and U is the corresponding left singular subspace of A , then

$$D(U, U_0) = O_p\left(\frac{\|R\|_2}{\gamma_{NT}}\right). \quad (\text{A.12})$$

(ii) Consider the reduced-rank regression model

$$W = ZA + E, \quad \text{rank}(A) \leq q, \quad (\text{A.13})$$

and let

$$\hat{A} = \arg \min_{\text{rank}(C) \leq q} \|W - ZC\|^2. \quad (\text{A.14})$$

Then there exists a constant κ such that

$$\|Z\hat{A} - ZA\| \leq \kappa\sqrt{q} \|P_Z E\|_2. \quad (\text{A.15})$$

Proof. Part (i) follows from Wedin's perturbation theorem (Wedin [1972](#)). For part (ii), by the definition of $\hat{A}$, we have

$$\|W - Z\hat{A}\|^2 \leq \|W - ZA\|^2.$$

We plug in $W = ZA + E$. It yields

$$\begin{aligned} \|Z\hat{A} - ZA\|^2 &\leq 2 \text{tr} \left[E'(Z\hat{A} - ZA) \right] \\ &= 2 \text{tr} \left[(P_Z E)'(Z\hat{A} - ZA) \right] \\ &\leq 2 \|P_Z E\|_2 \|Z\hat{A} - ZA\|_* \\ &\leq 2\sqrt{2q} \|P_Z E\|_2 \|Z\hat{A} - ZA\|, \end{aligned}$$

where the second line uses the fact that $Z\hat{A} - ZA$ belongs to the column space of Z . The

third line follows from $\text{tr}(C'D) \leq \|C\|_2 \|D\|_*$, and the last uses

$$\text{rank}(Z\hat{A} - ZA) \leq \text{rank}(\hat{A} - A) \leq \text{rank}(\hat{A}) + \text{rank}(A) \leq 2q.$$

Therefore, we have

$$\|Z\hat{A} - ZA\| \leq 2\sqrt{2q} \|P_Z E\|_2,$$

which proves (A.15). A similar argument is provided in Bunea et al. (2011, Theorem 5). $\square$

Remark. In our application of part (ii), we set $W = M_{\hat{F}}X$ and $Z = M_{\hat{F}}Y$, and $Z\hat{A}$ is the fitted component obtained from reduced-rank regression. Thus, (A.15) controls the component of the regression error projected onto the proxies.

We next collect some auxiliary results in the following lemma which we need to prove the main results later.

Lemma A.2. *Under Assumptions A.1–A.4, the following statements hold.*

(i)

$$\|e_X\|_2 = O_p(\sqrt{N} + \sqrt{T}), \quad \|e_Y\|_2 = O_p(\sqrt{T}). \quad (\text{A.16})$$

(ii) *The projection of the idiosyncratic component onto the factor space satisfies*

$$\|P_{F^+} e_X\|_2 \leq \|P_{F^+} e_X\| = O_p(\sqrt{N}). \quad (\text{A.17})$$

(iii) *The first r singular values of X_0 are of order $\sqrt{TN}$ and its remaining q nonzero singular values are of order $\sqrt{TN^\alpha}$.*

(iv) *Let $\bar{F}$ contain the first r left singular vectors of X_0 and satisfy $\bar{F}'\bar{F}/T = I_r$. Set $\bar{G} = M_{\bar{F}}X_0$. Then*

$$[\bar{G}, \bar{F}] \text{ and } F^+ \text{ have the same column space,} \quad \sigma_q(\bar{G}) \geq \underline{\kappa}\sqrt{TN^\alpha}. \quad (\text{A.18})$$

(v)

$$M_{\bar{F}}\tilde{F} \text{ and } \bar{G} \text{ have the same column space,} \quad \sigma_q(M_{\bar{F}}\tilde{F}) \geq \underline{\kappa}\sqrt{T}. \quad (\text{A.19})$$

(vi) The initial PC estimate satisfies

$$D(\hat{F}^{(0)}, \bar{F}) = O_p(c_{NT}^{-1}). \quad (\text{A.20})$$

Proof. Let E_ε collect the independent innovations in (A.8). The independent-entry norm bound gives $\mathbb{E}\|E_\varepsilon\|_2 \leq \kappa(\sqrt{N} + \sqrt{T})$. The triangle inequality and absolute summability of $\{A_{\ell,N}\}$ therefore imply $\mathbb{E}\|e_X\|_2 \leq \kappa(\sqrt{N} + \sqrt{T})$. Markov's inequality then gives

$$\|e_X\|_2 = O_p(\sqrt{N} + \sqrt{T}).$$

Similarly, because m is fixed, bounded innovation variances and absolute summability of $\{D_\ell\}$ imply

$$\|e_Y\|_2 = O_p(\sqrt{T}),$$

which proves part (i).

Conditional on F^+ , independence of F^+ and e_X gives

$$\mathbb{E}(\|P_{F^+}e_X\|^2 \mid F^+) \leq \bar{\kappa}N \text{rank}(P_{F^+}) = O(N).$$

Part (ii) follows from Markov's inequality.

For the remaining parts, decompose the proxy-related loading matrix as

$$\Lambda_{X\tilde{F}} = M_{\Lambda_{XF}}\Lambda_{X\tilde{F}} + \Lambda_{XF}(\Lambda'_{XF}\Lambda_{XF})^{-1}\Lambda'_{XF}\Lambda_{X\tilde{F}}.$$

The two loading blocks $M_{\Lambda_{XF}}\Lambda_{X\tilde{F}}$ and Λ_{XF} are orthogonal, and

$$\|(\Lambda'_{XF}\Lambda_{XF})^{-1}\Lambda'_{XF}\Lambda_{X\tilde{F}}\|_2 = O(N^{(\alpha-1)/2}) = o(1).$$

Assumption A.2 therefore implies that their singular values are of orders $N^{\alpha/2}$ and $N^{1/2}$, respectively. Since

$$X_0 = F^+[\Lambda_{X\tilde{F}}, \Lambda_{XF}]' \quad \text{and} \quad T^{-1}F^{+'}F^+ \xrightarrow{p} \Sigma_{F^+} \succ 0,$$

the first r singular values of X_0 are of order $\sqrt{TN}$ and its remaining q nonzero singular values are of order $\sqrt{TN^\alpha}$. This proves part (iii).

Moreover, $\overline{F}$ contains the first r left singular vectors of X_0 , while $\overline{G} = M_{\overline{F}}X_0$ contains its remaining q directions. Since X_0 and F^+ have the same column space, $[\overline{G}, \overline{F}]$ and F^+ also span the same space. Furthermore, we have

$$\sigma_q(\overline{G}) = \sigma_{r+q}(X_0) \geq \underline{\kappa}\sqrt{TN^\alpha},$$

which proves part (iv).

To prove part (v), we write

$$X_0 = F\Lambda'_{XF} + \tilde{F}\Lambda'_{X\tilde{F}}.$$

The two terms have spectral orders $\sqrt{TN}$ and $\sqrt{TN^\alpha}$, respectively. Wedin's theorem therefore gives

$$D(\overline{F}, F) = O_p(N^{(\alpha-1)/2}) = o_p(1).$$

Therefore, we have

$$\begin{aligned} \lambda_q\left(T^{-1}\tilde{F}'M_{\overline{F}}\tilde{F}\right) &\geq \lambda_q\left(T^{-1}\tilde{F}'M_F\tilde{F}\right) - \left\|T^{-1}\tilde{F}'(M_{\overline{F}} - M_F)\tilde{F}\right\|_2 \\ &\geq \underline{\kappa} - T^{-1}\|\tilde{F}\|_2^2 D(\overline{F}, F) \\ &= \underline{\kappa} - o_p(1). \end{aligned}$$

Here the lower bound for $\lambda_q(T^{-1}\tilde{F}'M_F\tilde{F})$ follows directly from (A.4). This in turn implies $\sigma_q(M_{\overline{F}}\tilde{F}) \geq \underline{\kappa}\sqrt{T}$. Moreover, $\overline{G}$ and $M_{\overline{F}}\tilde{F}$ both have rank q and their column spaces are contained in the space spanned by $M_{\overline{F}}F^+$. Thus, they have the same column space, proving

part (v).

Finally, part (iii) and $\alpha < 1$ imply $\sigma_r(X_0) - \sigma_{r+1}(X_0)$ is of order $\sqrt{TN}$. Since $X = X_0 + e_X$, Wedin's theorem and part (i) give

$$\begin{aligned} D(\widehat{F}^{(0)}, \overline{F}) &= O_p\left(\frac{\|e_X\|_2}{\sqrt{TN}}\right) \\ &= O_p\left(\frac{\sqrt{N} + \sqrt{T}}{\sqrt{NT}}\right) = O_p(c_{NT}^{-1}), \end{aligned}$$

which proves part (vi). □

Lemma A.3. *Let $\widehat{F}$ be the PC estimate obtained at a given iteration and suppose that*

$$D(\widehat{F}, \overline{F}) = O_p(c_{NT}^{-1} + N^{(\alpha-1)/2}T^{-\delta}). \quad (\text{A.21})$$

Then

$$D(\widetilde{Y}_{\widehat{F}}B, M_{\overline{F}}\widetilde{F}) = O_p(c_{NT}^{-1} + T^{-\delta}), \quad (\text{A.22})$$

$$D([\widetilde{Y}_{\widehat{F}}B, \widehat{F}], F^+) = O_p(c_{NT}^{-1} + T^{-\delta}). \quad (\text{A.23})$$

Proof. Since $\Lambda'_{Y\widetilde{F}}B = I_q$, we have

$$\widetilde{Y}_{\widehat{F}}B - M_{\overline{F}}\widetilde{F} = \underbrace{(P_{\overline{F}} - P_{\widehat{F}})\widetilde{F}}_{\text{I}} + \underbrace{T^{-\delta}M_{\widehat{F}}e_YB}_{\text{II}}. \quad (\text{A.24})$$

By (A.4), (A.21), and (A.16), we know

$$\begin{aligned} \|\text{I}\|_2 &= O_p\left[\sqrt{T}\{c_{NT}^{-1} + N^{(\alpha-1)/2}T^{-\delta}\}\right], \\ \|\text{II}\|_2 &= O_p(\sqrt{T}T^{-\delta}). \end{aligned}$$

This in turn implies

$$\left\| \tilde{Y}_{\hat{F}} B - M_{\bar{F}} \tilde{F} \right\|_2 = O_p \left[\sqrt{T} (c_{NT}^{-1} + T^{-\delta}) \right]. \quad (\text{A.25})$$

The lower bound for $\sigma_q(M_{\bar{F}} \tilde{F})$ in (A.19) and Wedin's theorem therefore give (A.22).

For the space spanned by F^+ , we have

$$D\left([\tilde{Y}_{\hat{F}} B, \hat{F}], F^+\right) \leq D\left(\tilde{Y}_{\hat{F}} B, M_{\bar{F}} \tilde{F}\right) + D(\hat{F}, \bar{F}).$$

Equations (A.22) and (A.21) prove (A.23). $\square$

Lemma A.4. *Let $\hat{F}$ be the initial PC estimate or a PC estimate obtained after a fixed number of RPC iterations, and suppose that the conditions of Lemma A.3 hold. Let $\hat{B}$ and $\hat{\Lambda}_{X\hat{F}}$ be the rank- q RRR estimates and define*

$$\hat{C} = \hat{\Lambda}_{X\hat{F}} \hat{B}'.$$

Then

$$D\left([Y\hat{B}, \hat{F}], F^+\right) = O_p(r_{NT}). \quad (\text{A.26})$$

Proof. Define

$$G_{\hat{F}} = P_{\tilde{Y}_{\hat{F}} B} M_{\hat{F}} X_0, \quad E_{\hat{F}} = M_{\hat{F}} X - G_{\hat{F}}.$$

We first show that the smallest nonzero singular value of $G_{\hat{F}}$ is of order $\sqrt{TN^\alpha}$. Using (A.19), and adding and subtracting $P_{M_{\bar{F}} \tilde{F}} M_{\hat{F}} X_0$, we have

$$\begin{aligned} \|G_{\hat{F}} - \bar{G}\|_2 &\leq \left\| \left(P_{\tilde{Y}_{\hat{F}} B} - P_{M_{\bar{F}} \tilde{F}} \right) M_{\hat{F}} X_0 \right\|_2 + \left\| P_{M_{\bar{F}} \tilde{F}} (M_{\hat{F}} - M_{\bar{F}}) X_0 \right\|_2 \\ &= O_p \left[\sqrt{TN^\alpha} (c_{NT}^{-1} + T^{-\delta}) + \sqrt{TN} (c_{NT}^{-1} + N^{(\alpha-1)/2} T^{-\delta}) \right]. \end{aligned}$$

The first term on the right-hand side uses (A.22), while the second term uses (A.21) and the strong singular-value order of X_0 . We then have

$$\frac{\|G_{\hat{F}} - \bar{G}\|_2}{\sqrt{TN^\alpha}} = O_p \left[c_{NT}^{-1} + T^{-\delta} + \frac{N^{(1-\alpha)/2}}{c_{NT}} + T^{-\delta} \right]$$

$$= O_p\left(c_{NT}^{-1} + T^{-\delta} + \frac{N^{(1-\alpha)/2}}{c_{NT}}\right) = o_p(1).$$

Finally, (A.18) and Weyl's inequality give

$$\begin{aligned}\sigma_q(G_{\hat{F}}) &\geq \sigma_q(\bar{G}) - \|G_{\hat{F}} - \bar{G}\|_2 \\ &\geq (\underline{\kappa} - o_p(1)) \sqrt{TN^\alpha}.\end{aligned}\tag{A.27}$$

We next consider the projected error, $\|P_{\tilde{Y}_{\hat{F}}} E_{\hat{F}}\|_2$, entering the reduced-rank regression inequality. Let $B_\perp$ be an $m \times (m - q)$ matrix satisfying

$$B'_\perp B_\perp = I_{m-q}, \quad \Lambda'_{Y\tilde{F}} B_\perp = 0.$$

If $m = q$, the terms involving $B_\perp$ are absent and the argument below applies only to term I. When $m > q$, since $[B, B_\perp]$ is nonsingular and $Y B_\perp = T^{-\delta} e_Y B_\perp$, we can write

$$P_{\tilde{Y}_{\hat{F}}} = P_{\tilde{Y}_{\hat{F}}B} + P_{M_{[\hat{F}, \tilde{Y}_{\hat{F}}B]} e_Y B_\perp}.\tag{A.28}$$

Using (A.28) and adding and subtracting $P_{M_{[\hat{F}, M_{\hat{F}}\tilde{F}]} e_Y B_\perp} E_{\hat{F}}$, we have

$$\begin{aligned}\|P_{\tilde{Y}_{\hat{F}}} E_{\hat{F}}\|_2 &\leq \underbrace{\|P_{\tilde{Y}_{\hat{F}}B} E_{\hat{F}}\|_2}_I \\ &\quad + \underbrace{\|P_{M_{[\hat{F}, M_{\hat{F}}\tilde{F}]} e_Y B_\perp} E_{\hat{F}}\|_2}_{II} \\ &\quad + \underbrace{\left\| \left(P_{M_{[\hat{F}, \tilde{Y}_{\hat{F}}B]} e_Y B_\perp} - P_{M_{[\hat{F}, M_{\hat{F}}\tilde{F}]} e_Y B_\perp} \right) E_{\hat{F}} \right\|_2}_{III}.\end{aligned}\tag{A.29}$$

We consider these terms separately. For term I, we know

$$P_{\tilde{Y}_{\hat{F}}B} E_{\hat{F}} = P_{\tilde{Y}_{\hat{F}}B} e_X.$$

Therefore, we have

$$\begin{aligned} \text{I} &\leq \left\| P_{M_{\tilde{F}}\tilde{F}} e_X \right\|_2 + D\left(\tilde{Y}_{\tilde{F}} B, M_{\tilde{F}}\tilde{F}\right) \|e_X\|_2 \\ &= O_p\left[\sqrt{N} + (c_{NT}^{-1} + T^{-\delta})(\sqrt{N} + \sqrt{T})\right], \end{aligned}$$

where the second line follows from (A.17), (A.22), and (A.16).

For term II, let

$$W = M_{[\hat{F}, M_{\tilde{F}}\tilde{F}]} e_Y B_\perp$$

for brevity, and decompose

$$\begin{aligned} E_{\hat{F}} &= \underbrace{\left(P_{M_{\tilde{F}}\tilde{F}} - P_{\tilde{Y}_{\tilde{F}}B}\right) M_{\hat{F}} X_0}_{\Pi_1} + \underbrace{M_{\hat{F}} e_X}_{\Pi_2} \\ &\quad + \underbrace{M_{M_{\tilde{F}}\tilde{F}} M_{\hat{F}} X_0}_{\Pi_3}. \end{aligned} \tag{A.30}$$

Since

$$\tilde{Y}_{\tilde{F}} B - M_{\tilde{F}}\tilde{F} = T^{-\delta} M_{\tilde{F}} e_Y B, \tag{A.31}$$

Wedin's theorem yields

$$\begin{aligned} \|P_W \Pi_1\|_2 &\leq \|\Pi_1\|_2 \\ &\leq D\left(\tilde{Y}_{\tilde{F}} B, M_{\tilde{F}}\tilde{F}\right) \|M_{\hat{F}} X_0\|_2 \\ &= O_p\left(\frac{\sqrt{T} T^{-\delta}}{\sqrt{T}} \sqrt{TN^\alpha}\right) \\ &= O_p\left(\sqrt{TN^\alpha} T^{-\delta}\right). \end{aligned}$$

To derive the rates for the projections of Π_2 and Π_3 , we define

$$Y^{\text{nf}} = \tilde{F} \Lambda'_{Y\tilde{F}}.$$

For the initial estimate, set $\widehat{F}^{\text{nf}} = \widehat{F}^{(0)}$. For a later fixed iterate, let $\widehat{C}_{\text{old}}$ denote the coefficient matrix used to construct $\widehat{F}$ in the preceding PC step. Let $\widehat{C}_{\text{old}}^{\text{nf}}$ and $\widehat{F}^{\text{nf}}$ denote the corresponding quantities when Y is replaced by Y^{nf} in that and all preceding iterations. The superscript nf denotes the procedure without proxy noise, under which $\widehat{F}^{\text{nf}}$ is independent of e_Y ; this independence is used in the argument below.

At each iteration $k \geq 1$, the difference between the r th and $(r+1)$ st singular values of $X - Y^{\text{nf}} \widehat{C}_{\text{old}}^{\text{nf}'}$ is of order $\sqrt{TN}$. Wedin's theorem therefore gives

$$\begin{aligned} D(\widehat{F}, \widehat{F}^{\text{nf}}) &= O_p \left(\frac{\left\| (X - Y \widehat{C}_{\text{old}}') - (X - Y^{\text{nf}} \widehat{C}_{\text{old}}^{\text{nf}'}) \right\|_2}{\sqrt{TN}} \right) \\ &\leq O_p \left(\frac{\|(Y - Y^{\text{nf}}) \widehat{C}_{\text{old}}'\|_2 + \|Y^{\text{nf}} (\widehat{C}_{\text{old}} - \widehat{C}_{\text{old}}^{\text{nf}})'\|_2}{\sqrt{TN}} \right) \\ &= O_p \left(\frac{\sqrt{N} + \sqrt{TN}^\alpha T^{-\delta}}{\sqrt{TN}} \right) \\ &= O_p(T^{-1/2} + N^{(\alpha-1)/2} T^{-\delta}), \end{aligned}$$

where the second line follows by adding and subtracting $Y^{\text{nf}} \widehat{C}_{\text{old}}'$. For the third line, the first term uses $Y - Y^{\text{nf}} = T^{-\delta} e_Y$, while the second follows from the RRR coefficient formula, (A.15), and the rate for the factor estimates entering the preceding RRR step. Since the initial PC estimate is the same under the two procedures, repeating this argument gives the same rate after any fixed number of iterations.

Moreover, letting

$$W^{\text{nf}} = M_{[\widehat{F}^{\text{nf}}, M_{\widehat{F}^{\text{nf}}} \widetilde{F}]} e_Y B_\perp,$$

we notice that

$$D(W, W^{\text{nf}}) = O_p(D(\widehat{F}, \widehat{F}^{\text{nf}})).$$

This follows because the smallest nonzero singular value of W^{nf} is of order $\sqrt{T}$ and Wedin's

theorem applies to W and W^{nf} . In particular,

$$\begin{aligned}
D(W, W^{\text{nf}}) &= \|P_W - P_{W^{\text{nf}}}\|_2 \\
&= O_p\left(\frac{\|W - W^{\text{nf}}\|_2}{\sqrt{T}}\right) \\
&= O_p\left[\frac{1}{\sqrt{T}} \left\| \left(M_{[\hat{F}, M_{\hat{F}}\tilde{F}]} - M_{[\hat{F}^{\text{nf}}, M_{\hat{F}^{\text{nf}}}\tilde{F}]} \right) e_Y B_\perp \right\|_2 \right] \\
&= O_p\left[\frac{\|e_Y B_\perp\|_2}{\sqrt{T}} D([\hat{F}, M_{\hat{F}}\tilde{F}], [\hat{F}^{\text{nf}}, M_{\hat{F}^{\text{nf}}}\tilde{F}])\right] \\
&= O_p\left\{ \frac{\|e_Y B_\perp\|_2}{\sqrt{T}} \left[D(\hat{F}, \hat{F}^{\text{nf}}) + \frac{\|(M_{\hat{F}} - M_{\hat{F}^{\text{nf}}})\tilde{F}\|_2}{\sqrt{T}} \right] \right\} \\
&= O_p(D(\hat{F}, \hat{F}^{\text{nf}})) \\
&= O_p(T^{-1/2} + N^{(\alpha-1)/2}T^{-\delta}).
\end{aligned}$$

The fifth line uses the orthogonal decomposition $P_{[\hat{F}, M_{\hat{F}}\tilde{F}]} = P_{\hat{F}} + P_{M_{\hat{F}}\tilde{F}}$ and its noise-free counterpart, which in turn gives $D([\hat{F}, M_{\hat{F}}\tilde{F}], [\hat{F}^{\text{nf}}, M_{\hat{F}^{\text{nf}}}\tilde{F}]) \leq D(\hat{F}, \hat{F}^{\text{nf}}) + D(M_{\hat{F}}\tilde{F}, M_{\hat{F}^{\text{nf}}}\tilde{F})$. The next line follows from Wedin's theorem, since $\sigma_q(M_{\hat{F}}\tilde{F})$ and $\sigma_q(M_{\hat{F}^{\text{nf}}}\tilde{F})$ are of order $\sqrt{T}$.

We now give the rates for $\|P_W A\|_2$ for $A = \Pi_2$ or Π_3 . We have

$$\|P_W A\|_2 \leq \|P_{W^{\text{nf}}} A^{\text{nf}}\|_2 + \|A - A^{\text{nf}}\|_2 + D(W, W^{\text{nf}})\|A^{\text{nf}}\|_2, \quad (\text{A.32})$$

where A^{nf} denotes the corresponding term constructed using $\hat{F}^{\text{nf}}$. For the first term, we have

$$\|P_{W^{\text{nf}}} A^{\text{nf}}\|^2 \leq \lambda_{m-q} (W^{\text{nf}'} W^{\text{nf}})^{-1} \|W^{\text{nf}'} A^{\text{nf}}\|^2 = O_p(T^{-1} \|A^{\text{nf}}\|^2). \quad (\text{A.33})$$

To see this, notice that independence of e_Y from A^{nf} and $\hat{F}^{\text{nf}}$, together with absolute summability of $\{D_\ell\}$, implies

$$\mathbb{E}\left(\|W^{\text{nf}'} A^{\text{nf}}\|^2 \mid A^{\text{nf}}, \hat{F}^{\text{nf}}\right) \leq \bar{\kappa} \|A^{\text{nf}}\|^2.$$

Because $m - q$ is fixed, Markov's inequality then yields

$$\|W^{\text{nf}'} A^{\text{nf}}\|^2 = O_p(\|A^{\text{nf}}\|^2).$$

Moreover, under (A.9), positive definiteness of $E(e_{Y,t} e'_{Y,t})$ and the fixed rank of $[\widehat{F}^{\text{nf}}, M_{\widehat{F}^{\text{nf}}} \widetilde{F}]$ imply

$$\lambda_{m-q}(T^{-1} W^{\text{nf}'} W^{\text{nf}}) \geq \underline{\kappa} + o_p(1).$$

Thus, we can write

$$\begin{aligned} \|P_W A\|_2 &\leq \|P_{W^{\text{nf}}} A^{\text{nf}}\|_2 + \|A - A^{\text{nf}}\|_2 + D(W, W^{\text{nf}}) \|A^{\text{nf}}\|_2 \\ &= O_p[T^{-1/2} \|A^{\text{nf}}\| + \|A - A^{\text{nf}}\|_2 + (T^{-1/2} + N^{(\alpha-1)/2} T^{-\delta}) \|A^{\text{nf}}\|_2] \\ &= O_p[T^{-1/2} \sqrt{NT} + \sqrt{N} + \sqrt{TN^\alpha} T^{-\delta} + (T^{-1/2} + N^{(\alpha-1)/2} T^{-\delta}) \sqrt{TN}] \\ &= O_p(\sqrt{N} + \sqrt{TN^\alpha} T^{-\delta}). \end{aligned} \tag{A.34}$$

The second line uses (A.33) and the rate for $D(W, W^{\text{nf}})$. The third uses the fact that $\|A^{\text{nf}}\|$ and $\|A^{\text{nf}}\|_2$ are bounded by the corresponding norms of e_X or X_0 , and

$$\begin{aligned} \|A - A^{\text{nf}}\|_2 &\leq \kappa D(\widehat{F}, \widehat{F}^{\text{nf}}) \sqrt{TN} \\ &= O_p(\sqrt{N} + \sqrt{TN^\alpha} T^{-\delta}). \end{aligned} \tag{A.35}$$

Combining these results with the rate for II_1 gives

$$\text{II} = O_p(\sqrt{N} + \sqrt{TN^\alpha} T^{-\delta}).$$

Finally, for III , we have

$$\begin{aligned} \text{III} &\leq D(M_{[\widehat{F}, \widetilde{Y}_{\widehat{F}} B]} e_Y B_\perp, M_{[\widehat{F}, M_{\widehat{F}} \widetilde{F}]} e_Y B_\perp) \|E_{\widehat{F}}\|_2 \\ &= O_p\left[\frac{\|e_Y B_\perp\|_2}{\sqrt{T}} D(\widetilde{Y}_{\widehat{F}} B, M_{\widehat{F}} \widetilde{F}) (\sqrt{N} + \sqrt{T} + \sqrt{TN^\alpha})\right] \end{aligned}$$

$$= O_p \left[T^{-\delta} (\sqrt{N} + \sqrt{T}) + \sqrt{TN^\alpha} T^{-\delta} \right],$$

where the second line follows from (A.30) and Wedin's theorem, and the last uses $\|e_Y B_\perp\|_2 = O_p(\sqrt{T})$ together with (A.31).

Combining the rates for I–III yields

$$\left\| P_{\tilde{Y}_{\hat{F}}} E_{\hat{F}} \right\|_2 = O_p \left[\sqrt{N} + (c_{NT}^{-1} + T^{-\delta})(\sqrt{N} + \sqrt{T}) + \sqrt{TN^\alpha} T^{-\delta} \right]. \quad (\text{A.36})$$

We now turn to the main result of the lemma. Notice that the columns of $[Y \hat{B}, \hat{F}]$ and $[\tilde{Y}_{\hat{F}} \hat{C}', \hat{F}]$ span the same space since $\hat{C}' = \hat{B} \hat{\Lambda}'_{X\tilde{F}}$ and $\hat{\Lambda}_{X\tilde{F}}$ has full column rank q . Therefore, we have

$$\begin{aligned} D([Y \hat{B}, \hat{F}], F^+) &\leq \underbrace{D(\tilde{Y}_{\hat{F}} \hat{C}', G_{\hat{F}})}_{\text{I}} \\ &\quad + \underbrace{D([G_{\hat{F}}, \hat{F}], F^+)}_{\text{II}}. \end{aligned} \quad (\text{A.37})$$

Because $G_{\hat{F}}$ has rank q and $G_{\hat{F}}$ and $\tilde{Y}_{\hat{F}} B$ span the same space, Lemma A.3 gives

$$\text{II} = O_p(c_{NT}^{-1} + T^{-\delta}). \quad (\text{A.38})$$

For term I, Lemma A.1(ii) gives

$$\left\| \tilde{Y}_{\hat{F}} \hat{C}' - G_{\hat{F}} \right\| \leq \kappa \sqrt{q} \left\| P_{\tilde{Y}_{\hat{F}}} E_{\hat{F}} \right\|_2. \quad (\text{A.39})$$

Since (A.27) gives $\sigma_q(G_{\hat{F}}) \geq \underline{\kappa} \sqrt{TN^\alpha}$, Wedin's theorem and (A.39) imply

$$\text{I} = O_p \left(\frac{\|P_{\tilde{Y}_{\hat{F}}} E_{\hat{F}}\|_2}{\sqrt{TN^\alpha}} \right).$$

Inserting (A.36), we have

$$I = O_p\left(\sqrt{\frac{N^{1-\alpha}}{T}} + c_{NT}^{-1} + T^{-\delta}\right) = O_p(r_{NT}). \quad (\text{A.40})$$

Finally, combining (A.38) and (A.40) in (A.37) implies

$$D([Y\hat{B}, \hat{F}], F^+) = O_p(r_{NT}),$$

which proves (A.26). $\square$

Lemma A.5. *Under the conditions of Lemma A.4, let $\hat{F}_{\text{new}}$ be the first r principal components of $X - Y\hat{C}'$. Then*

$$D(\hat{F}_{\text{new}}, \bar{F}) = O_p(c_{NT}^{-1} + N^{(\alpha-1)/2}T^{-\delta}). \quad (\text{A.41})$$

Proof. We start from

$$D(\hat{F}_{\text{new}}, \bar{F}) \leq \underbrace{D(\hat{F}_{\text{new}}, \hat{F})}_I + \underbrace{D(\hat{F}, \bar{F})}_{II}. \quad (\text{A.42})$$

Term II is bounded by (A.21). We apply Wedin's theorem to term I. To this end, first notice that

$$\sigma_r\left(P_{\hat{F}}(X - Y\hat{C}')\right) \geq \underbrace{\sigma_r(P_{\hat{F}}X_0)}_{I_1} - \underbrace{\|P_{\hat{F}}e_X\|_2}_{I_2} - \underbrace{\|P_{\hat{F}}Y\hat{C}'\|_2}_{I_3}.$$

For the first term, we have

$$\begin{aligned} I_1 &\geq \sigma_r(P_{\bar{F}}X_0) - \|(P_{\hat{F}} - P_{\bar{F}})X_0\|_2 \\ &\geq \underline{\kappa}\sqrt{TN} - D(\hat{F}, \bar{F})\|X_0\|_2 \\ &= \underline{\kappa}\sqrt{TN} + o_p(\sqrt{TN}). \end{aligned}$$

Moreover, (A.16) gives

$$I_2 \leq \|e_X\|_2 = O_p(\sqrt{N} + \sqrt{T}) = o_p(\sqrt{TN}).$$

Finally, plugging in the RRR formula for $\hat{B}$ and using (A.19), (A.25), (A.39), and (A.36), we have

$$I_3 = O_p(\sqrt{TN^\alpha} + \sqrt{N}) = o_p(\sqrt{TN}).$$

Therefore, we have

$$\sigma_r(P_{\hat{F}}(X - Y\hat{C}')) \geq \underline{\kappa}\sqrt{TN} + o_p(\sqrt{TN}),$$

so $P_{\hat{F}}(X - Y\hat{C}')$ has rank r and its smallest nonzero singular value is of order $\sqrt{TN}$.

Consider term I. Wedin's theorem provides

$$\begin{aligned} I &= O_p\left(\frac{\|M_{\hat{F}}(X - Y\hat{C}'))\|_2}{\sigma_r(P_{\hat{F}}(X - Y\hat{C}'))}\right) \\ &= O_p\left(\frac{\|M_{\hat{F}}(X - Y\hat{C}'))\|_2}{\sqrt{TN}}\right) \\ &= O_p\left(\frac{\underbrace{\|M_{\hat{F}}e_X\|_2}_{I_1} + \underbrace{\|M_{\tilde{Y}_{\hat{F}}B}M_{\hat{F}}X_0\|_2}_{I_2} + \underbrace{\|G_{\hat{F}} - \tilde{Y}_{\hat{F}}\hat{C}'\|_2}_{I_3}}{\sqrt{TN}}\right). \end{aligned}$$

We know that the first term is $O_p(\sqrt{N} + \sqrt{T})$. For the second term, we use $\bar{G} = M_{\bar{F}}X_0$ and $M_{\bar{G}}\bar{G} = 0$. Adding and subtracting $M_{\tilde{Y}_{\hat{F}}B}M_{\bar{F}}X_0$ and $M_{\bar{G}}M_{\bar{F}}X_0$, we have

$$\begin{aligned} &M_{\tilde{Y}_{\hat{F}}B}M_{\hat{F}}X_0 \\ &= M_{\tilde{Y}_{\hat{F}}B}(M_{\hat{F}} - M_{\bar{F}})X_0 + (M_{\tilde{Y}_{\hat{F}}B} - M_{\bar{G}})M_{\bar{F}}X_0. \end{aligned}$$

Therefore,

$$\begin{aligned} \text{I}_2 &\leq D(\widehat{F}, \overline{F}) \|X_0\|_2 + D(\widetilde{Y}_{\widehat{F}} B, \overline{G}) \|\overline{G}\|_2 \\ &= O_p \left[\sqrt{TN} D(\widehat{F}, \overline{F}) + \sqrt{TN^\alpha} (c_{NT}^{-1} + T^{-\delta}) \right], \end{aligned}$$

where the last equality uses (A.22), (A.19), and Lemma A.2(iii)–(iv). Moreover, for the third term, we have

$$\text{I}_3 = O_p \left[\sqrt{N} + (c_{NT}^{-1} + T^{-\delta})(\sqrt{N} + \sqrt{T}) + \sqrt{TN^\alpha} T^{-\delta} \right].$$

This follows from (A.39) and (A.36).

Dividing the rates for I_1 – I_3 by $\sqrt{TN}$ and using (A.21) therefore gives

$$\text{I} = O_p \left[D(\widehat{F}, \overline{F}) + c_{NT}^{-1} + N^{(\alpha-1)/2} T^{-\delta} \right].$$

Combining this result with the bound for term II in (A.42) proves (A.41). $\square$

Proof of Theorem A.1. Lemma A.2(vi) establishes the required condition for the initial PC estimate $\widehat{F}^{(0)}$. Applying Lemma A.4 to this estimate gives

$$D([Y\widehat{B}^{(1)}, \widehat{F}^{(0)}], F^+) = O_p(r_{NT}).$$

The subsequent PC step satisfies, by Lemma A.5,

$$D(\widehat{F}^{(1)}, \overline{F}) = O_p(c_{NT}^{-1} + N^{(\alpha-1)/2} T^{-\delta}).$$

Thus, the condition required by Lemma A.4 is preserved after the first iteration. The same argument applies at subsequent iterations. In particular, if $\widehat{F}^{(k)}$ satisfies (A.21), then Lemma A.4 gives

$$D([Y\widehat{B}^{(k+1)}, \widehat{F}^{(k)}], F^+) = O_p(r_{NT}),$$

and Lemma A.5 ensures that $\widehat{F}^{(k+1)}$ satisfies (A.21). Therefore, for every fixed K ,

$$\max_{0 \leq k \leq K} D\left([Y\widehat{B}^{(k+1)}, \widehat{F}^{(k)}], F^+\right) = O_p(r_{NT}).$$

Finally, $T/N^{1-\alpha} \rightarrow \infty$, $N, T \rightarrow \infty$, and $\delta > 0$ imply $r_{NT} \rightarrow 0$, completing the proof. $\square$

B Monte Carlo II: Design and Selected-Variable Results

B.1 Experimental Design

We now turn to the description of the DGPs, which follow Stock and Watson (2016)'s implementation of the Kilian (2009) identification. Consider the FAVAR of Stock and Watson (2016) with six factors:

$$\begin{bmatrix} \text{Oil Production}_t \\ \text{Global Activity}_t \\ \text{Oil Price}_t^{\text{RAC}} \\ X_{\bullet t} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ & & \Lambda_{\bullet} & & & \end{bmatrix} \begin{bmatrix} \text{Oil Production}_t \\ F_t^{\text{Global Activity}} \\ \text{Oil Price}_t^{\text{RAC}} \\ F_{4:6,t} \end{bmatrix} + u_t, \quad (\text{B.1})$$

$$\begin{bmatrix} \text{Oil Production}_t \\ F_t^{\text{Global Activity}} \\ \text{Oil Price}_t^{\text{RAC}} \\ F_{4:6,t} \end{bmatrix} = \Phi(L) \begin{bmatrix} \text{Oil Production}_{t-1} \\ F_{t-1}^{\text{Global Activity}} \\ \text{Oil Price}_{t-1}^{\text{RAC}} \\ F_{4:6,t-1} \end{bmatrix} + \eta_t, \quad \text{Cov}(\eta_t) = \Sigma_{\eta}, \quad \text{and} \quad (\text{B.2})$$

$$u_{it} = a_i(L)u_{it-1} + e_{it} \quad i = 1, 2, \dots, 209, \quad (\text{B.3})$$

where all of the disturbances e_{it} are iid $N(0, \sigma_i^2)$, all of the factor innovations η_t are iid $N(0, \Sigma_{\eta})$, $\Phi(L)$ is a VAR(4) lag polynomial, $a_i(L)$ is an AR(4) lag polynomial, and the unit elements on the diagonal of the upper 6×6 partition of the loading matrix impose the named-factor normalization so that $F_t^{\text{Global Activity}}$ (the global activity factor) is the common component of the global activity index.

We consider three DGPs in the context of the FAVAR model summarized by Equations (B.1)–(B.3). While they all augment global oil production, a global activity measure and the real oil price with three unobserved factors, they differ in terms of the global activity measure included in the factor space. Specifically, in the first DGP, we equate $F_t^{\text{Global Activity}}$ to Kilian's global activity index. This is exactly the model of Stock and Watson (2016). In the

second and third DGPs, we use the world IP growth measure from Baumeister and Hamilton (2019) and the global economic conditions measure of Baumeister et al. (2022) as the global activity index, respectively. The parameters underlying the DGPs are estimated by implementing the models following Stock and Watson (2016). The estimated models are then used to simulate data by independently drawing $e_{it} \sim N(0, \hat{\sigma}_i^2)$ and $\eta_t \sim N(0, \hat{\Sigma}_\eta)$.

B.2 R^2 of Common Component and SIRFs for Selected Variables

Before turning to the impulse responses, we assess how well the alternative models recover the common components. Table B.1 reports the average R^2 from regressions of selected variables on the estimated factors across all Monte Carlo repetitions. The FAVAR-RPC specification delivers values that are very close to those obtained under the model using the correct global activity measure in each simulation design. This is particularly evident for the three activity proxies themselves: when the DGP is based on one of the three global activity indexes, the FAVAR-RPC approach attains the R^2 values that closely track those from the corresponding correct model. For the broader macroeconomic and oil-market variables, the R^2 values are very similar across specifications. This suggests that the main advantage of the RPC approach is a more accurate recovery of the factor space associated with the relevant global activity component. This is particularly important for the analysis as more accurate recovery of global activity innovations improves the performance of the FAVAR estimator for structural impulse responses.

Table B.1: R^2 of common component

Variable	A. Design: Kilian Global Activity Indicator				B. Design: Baumeister-Hamilton World IP				C. Design: Baumeister-Korobilis-Lee GECON			
	K	BH	BKL	RPC	K	BH	BKL	RPC	K	BH	BKL	RPC
Kilian Global Activity Indicator	1.00	0.49	0.50	0.99	1.00	0.49	0.45	0.52	1.00	0.42	0.48	0.49
Baumeister-Hamilton World IP	0.80	1.00	0.76	0.81	0.80	1.00	0.78	1.00	0.74	1.00	0.77	0.78
Baumeister-Korobilis-Lee GECON	0.83	0.81	1.00	0.83	0.81	0.82	1.00	0.82	0.82	0.79	1.00	1.00
GDP	0.63	0.62	0.61	0.63	0.60	0.60	0.58	0.60	0.58	0.57	0.60	0.60
Consumption	0.66	0.65	0.64	0.66	0.64	0.64	0.62	0.64	0.61	0.61	0.61	0.61
Fixed Investment	0.75	0.75	0.75	0.75	0.73	0.76	0.75	0.76	0.75	0.77	0.79	0.79
Employment	0.94	0.94	0.93	0.94	0.93	0.94	0.93	0.94	0.93	0.93	0.93	0.93
Unemployment Rate	0.88	0.88	0.87	0.88	0.87	0.87	0.87	0.87	0.87	0.87	0.87	0.87
Inflation	0.57	0.57	0.56	0.57	0.55	0.56	0.56	0.56	0.56	0.56	0.56	0.56
Core Inflation	0.08	0.08	0.08	0.08	0.08	0.09	0.08	0.09	0.08	0.08	0.09	0.09
Federal Funds Rate	0.58	0.58	0.58	0.58	0.58	0.58	0.58	0.58	0.57	0.57	0.57	0.57
Oil Prices (RAC)	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Oil Prices (PPI)	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96
Gasoline Prices (CPI)	0.72	0.72	0.72	0.72	0.72	0.72	0.72	0.72	0.72	0.72	0.72	0.72
Oil Production	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Note: This table reports the average R^2 of regressions of individual variables on the common component (over 1000 simulation repetitions). For each simulation design, the black boldface numbers indicate the models using the respective correct global activity measure, while the gray-shaded boldface number reports the corresponding average R^2 based on the FAVAR-RPC approach, which uses all three measures of global activity as noisy factor proxies. “K”, “BH”, “BKL”, and “RPC” refer to the FAVAR-K, FAVAR-BH, FAVAR-BKL, and FAVAR-RPC models, respectively. For more details, see Section 5.2.

SIRFs for Selected Variables We now turn to the average SIRFs with respect to the three shocks and discuss each DGP separately. Figure B.1 presents the results given the simulation draws of the data in the first design with the Kilian index as the true global activity measure. There are three columns showing the SIRFs with respect to an oil supply shock, a global demand shock, and an oil-specific demand shock, respectively. The blue dotted line shows the average SIRFs computed using the FAVAR-K approach, along with one-standard-error bands. Those computed using the FAVAR-BH, FAVAR-BKL and FAVAR-RPC are shown as the orange dash-dotted, green dashed and black solid lines, respectively. Note that here the global demand shock is always identified as the shock that raises the common component of the global activity index by one unit.

Starting with the oil supply shock shown in the first column, we see that the four models produce very similar results. This is also reflected by the correlations across the three identified oil supply shocks across our DGPs in Table B.2, which all exceed 0.94. Note that the similarity of the results is consistent with global oil production being notoriously difficult to predict (Stock and Watson 2016). Turning to the global demand shock in the second column, the FAVAR-RPC estimator produces results which are virtually indistinguishable from the FAVAR-K which is the true DGP in this analysis. This is in stark contrast to the FAVAR-BH and FAVAR-BKL models which generate biased SIRFs for all variables of interest. Specifically, the responses of oil production, global activity, real GDP and inflation are all much weaker relative to the true SIRFs. The same is true considering the oil-specific demand shock in the third column. In contrast, the responses using the FAVAR-RPC closely match those implied by the true DGP.

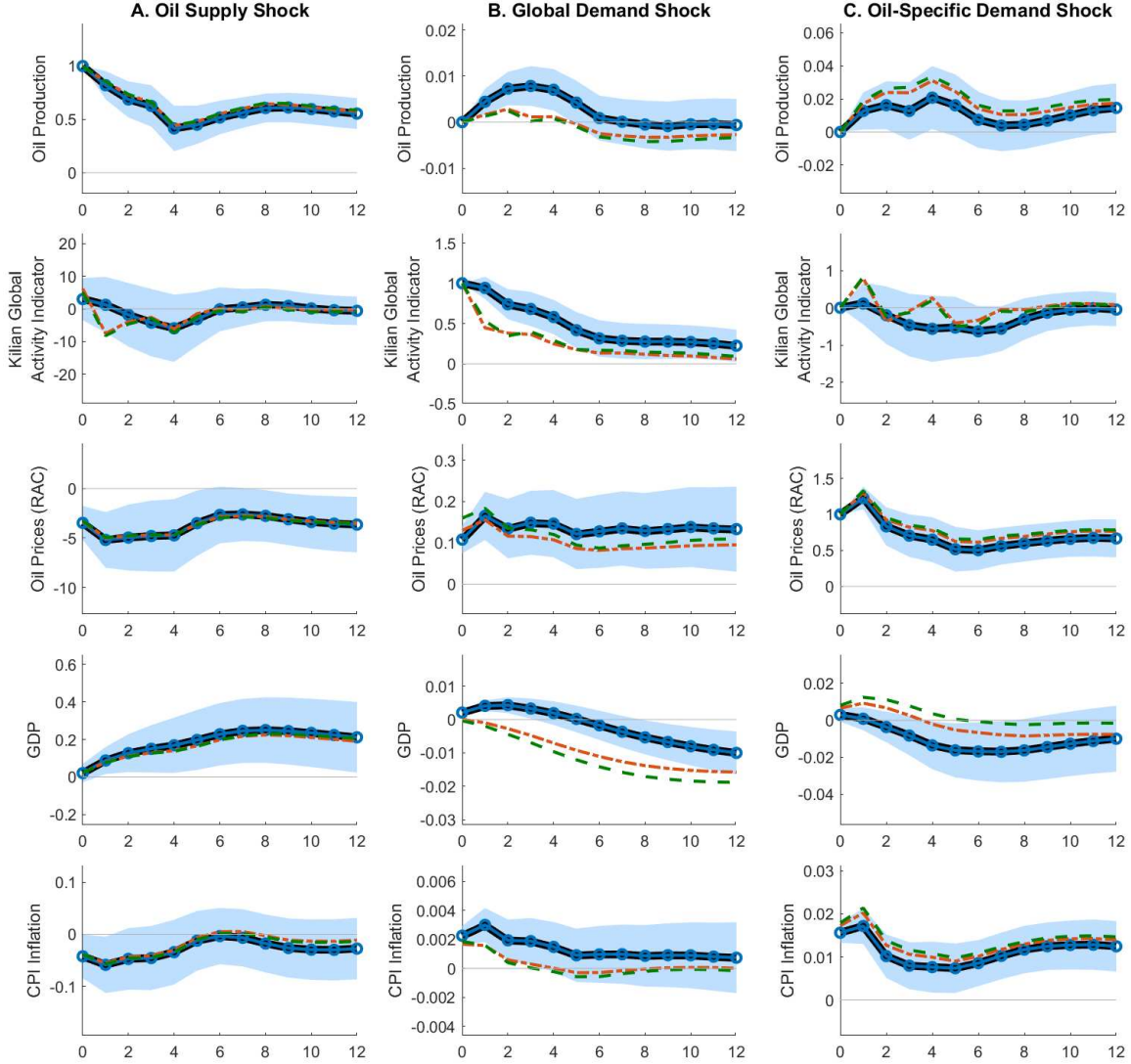


Figure B.1: SIRFs with design using Kilian global activity indicator. This figure reports the average, over 1000 simulation repetitions, of the structural impulse response functions from FAVAR-RPC, FAVAR-K, FAVAR-BH, and FAVAR-BKL for the oil supply shock (column A), the global demand shock (column B), and the oil-specific demand shock (column C). The black solid line corresponds to FAVAR-RPC, the blue dotted line with ± 1 standard error bands to FAVAR-K, the orange dash-dotted line to FAVAR-BH, and the green dashed line to FAVAR-BKL. The simulation draws are generated from the design in which Kilian's index is treated as the true global activity measure. The responses for Kilian's global activity indicator are reported in standard deviation units, and the responses for all other variables are reported in percentage points.

Table B.2: Correlations among estimated shocks

		OS ₁	OS ₂	OS ₃	GD ₁	GD ₂	GD ₃	OD ₁	OD ₂	OD ₃
Oil supply shock (OS)	FAVAR-K (1)	1.00								
	FAVAR-BH (2)	0.94	1.00							
	FAVAR-BKL (3)	0.94	0.94	1.00						
Global demand shock (GD)	FAVAR-K (1)	0.00	0.03	-0.01	1.00					
	FAVAR-BH (2)	0.00	0.00	0.00	0.36	1.00				
	FAVAR-BKL (3)	-0.01	0.01	0.00	0.43	0.43	1.00			
Oil-specific demand shock (OD)	FAVAR-K (1)	0.00	0.01	0.01	0.00	0.26	-0.03	1.00		
	FAVAR-BH (2)	0.00	0.00	0.00	0.21	0.00	-0.08	0.85	1.00	
	FAVAR-BKL (3)	-0.01	0.02	0.00	0.32	0.27	0.00	0.88	0.90	1.00

Note: This table reports the correlations among the shocks estimated using FAVAR-K, FAVAR-BH and FAVAR-BKL over the sample period 1985Q1 to 2014Q4. Shading denotes correlations for each one of the three shocks identified using Kilian (2009)’s identification scheme.

We now compare these results with the second simulation design in which the Baumeister and Hamilton (2019) index is treated as the true global activity measure. These are provided in Figure B.2. While all four models again produce similar results for the oil supply shock, they differ for the global demand shock and the oil-specific demand shock. Now the SIRFs based on the FAVAR-K and FAVAR-BKL models are biased, but the SIRFs for the FAVAR-RPC model are essentially identical to the true SIRFs based on the FAVAR-BH model. As in the previous simulation design, the bias is particularly pronounced for key variables of interest, particularly the real price of oil, real U.S. GDP and CPI inflation.

Finally, consider the case where the true global activity index is given as the global economic conditions indicator of Baumeister et al. (2022). The corresponding SIRFs are presented in Figure B.3. Across the four models, the SIRFs for the FAVAR-RPC model are essentially identical to the true SIRFs based on the FAVAR-BKL model, while the SIRFs based on the FAVAR-K and FAVAR-BH models are severely biased. It is worth pointing out that the positive impact of the global demand shock on U.S. real GDP based on the FAVAR-BKL peaks within a year and is much more persistent compared to those based on the FAVAR-K

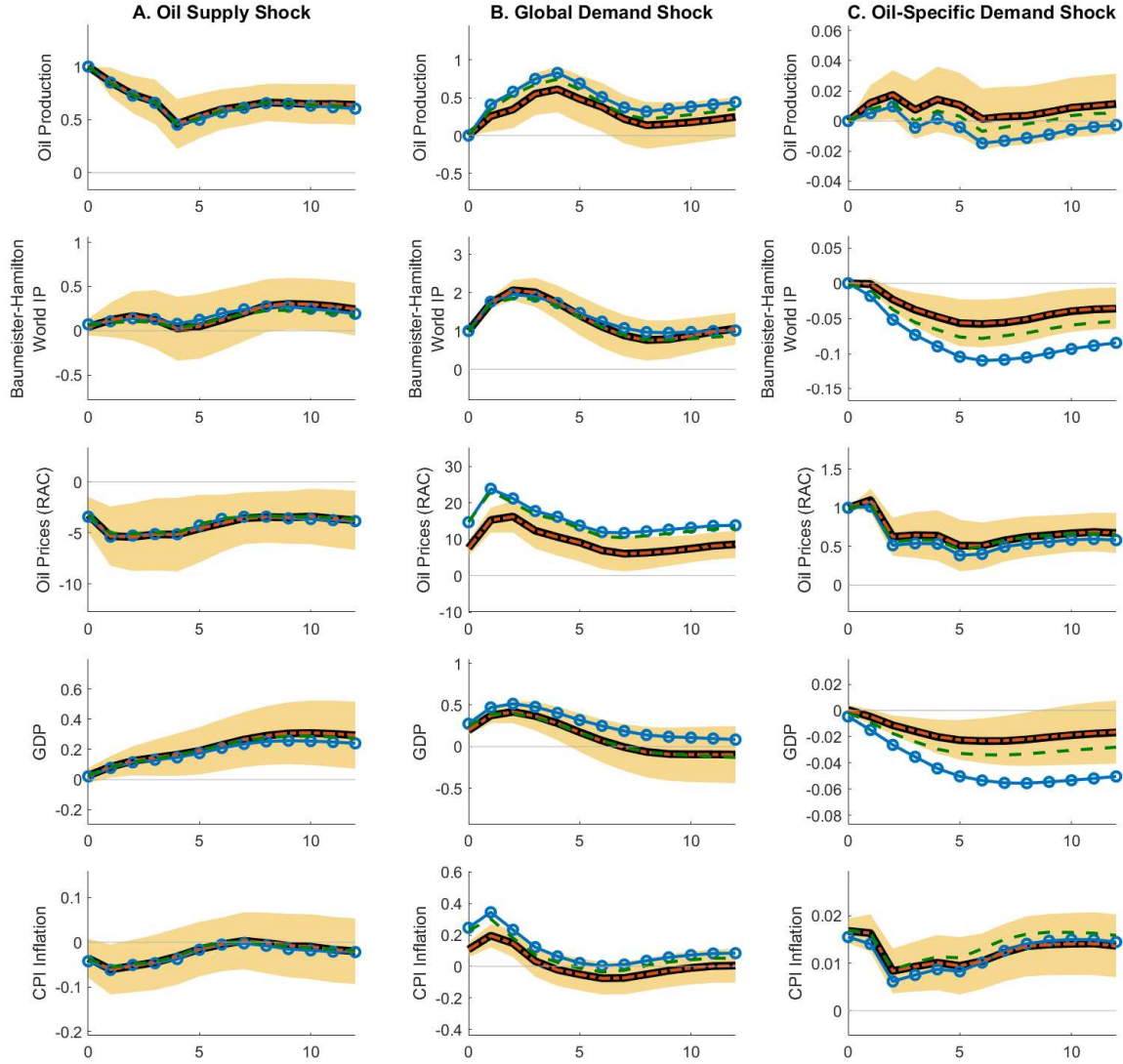


Figure B.2: SIRFs with design using Baumeister–Hamilton world IP growth. This figure reports the average, over 1000 simulation repetitions, of the structural impulse response functions from FAVAR-RPC, FAVAR-K, FAVAR-BH, and FAVAR-BKL for the oil supply shock (panel A), the global demand shock (panel B), and the oil-specific demand shock (panel C). The black solid line corresponds to FAVAR-RPC, the blue dotted line to FAVAR-K, the orange dash-dotted line with ± 1 standard error bands to FAVAR-BH, and the green dashed line to FAVAR-BKL. The simulation draws are generated from the design in which Baumeister and Hamilton (2019)’s world IP growth is treated as the true global activity measure. All responses are reported in percentage points.

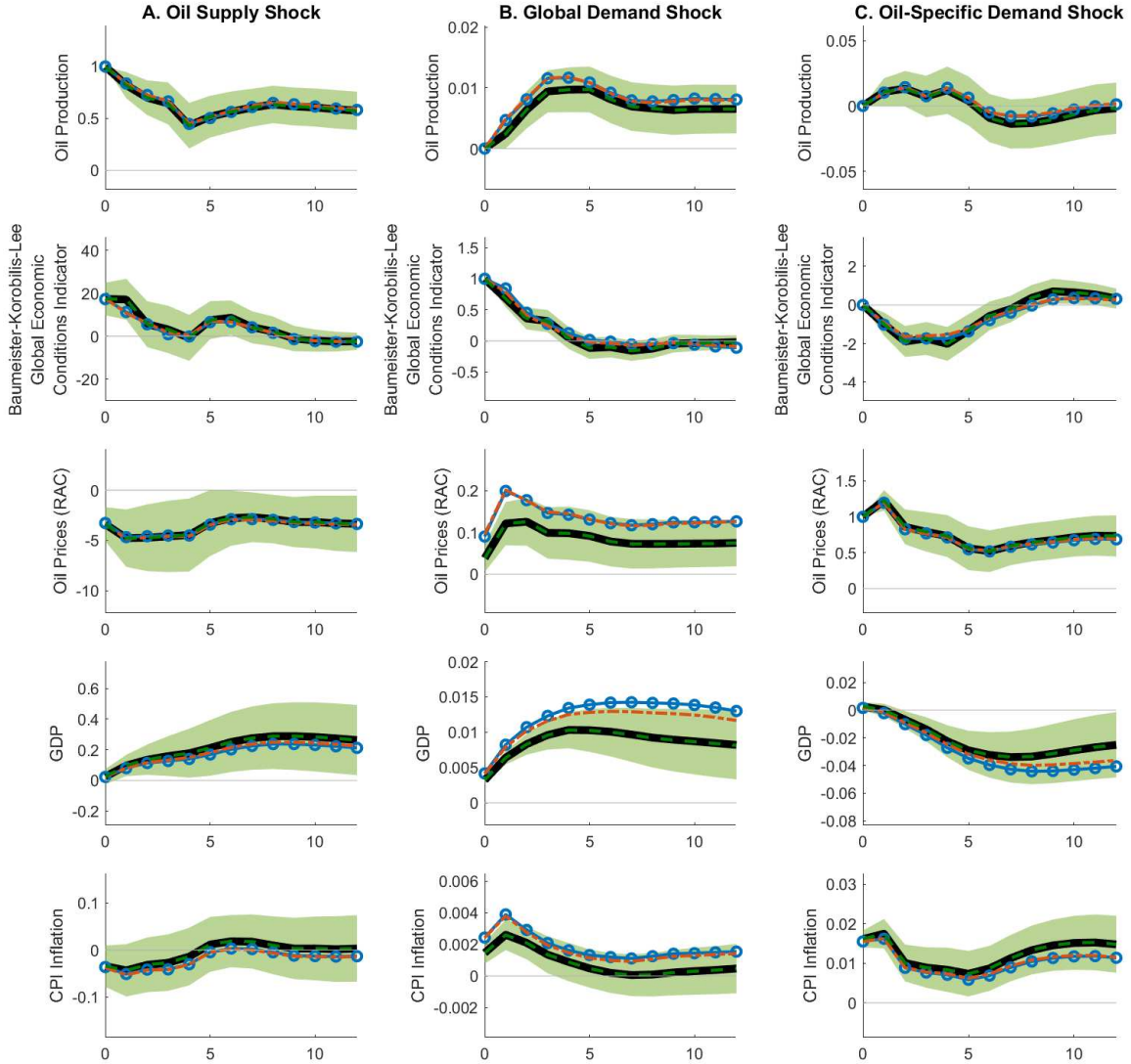


Figure B.3: SIRFs with design using Baumeister–Korobilis–Lee global economic conditions indicator. This figure reports the average, over 1000 simulation repetitions, of the structural impulse response functions from FAVAR-RPC, FAVAR-K, FAVAR-BH, and FAVAR-BKL for the oil supply shock (panel A), the global demand shock (panel B), and the oil-specific demand shock (panel C). The black solid line corresponds to FAVAR-RPC, the blue dotted line to FAVAR-K, the orange dash-dotted line to FAVAR-BH, and the green dashed line with ± 1 standard error bands to FAVAR-BKL. The simulation draws are generated from the design in which Baumeister et al. (2022)’s global economic conditions indicator is treated as the true global activity measure. The responses for the global economic conditions indicator are reported in standard deviation units, and the responses for all other variables are reported in percentage points.

and FAVAR-BH models. Indeed, aggregate demand shocks identified with the FAVAR-BKL model are an important force in real activity movements: as shown in Table B.3, the global demand shock estimated by FAVAR-BKL accounts for 42% of the variance of six-quarter-ahead forecast errors in real U.S. GDP, considerably larger than the corresponding estimated shares of 4% and 9% using FAVAR-K and FAVAR-BH.

Table B.3: Forecast error variance decompositions for six periods ahead forecasts of selected variables

Variable	A. Oil supply shock			B. Global demand shock			C. Oil-specific demand shock		
	K	BH	BKL	K	BH	BKL	K	BH	BKL
GDP	0.06	0.06	0.06	0.04	0.09	0.42	0.05	0.09	0.11
Inflation	0.02	0.02	0.01	0.09	0.07	0.08	0.20	0.23	0.28
Oil Prices (RAC)	0.71	0.67	0.67	0.09	0.11	0.23	0.04	0.02	0.02
Oil Production	0.11	0.14	0.09	0.22	0.37	0.15	0.52	0.45	0.65
Kilian Global Activity Indicator	0.01	0.04	0.04	0.89	0.14	0.02	0.02	0.04	0.07
Baumeister-Hamilton World IP	0.02	0.00	0.01	0.19	0.72	0.51	0.03	0.13	0.07
Baumeister-Korobilis-Lee GECON	0.06	0.07	0.09	0.22	0.28	0.52	0.05	0.07	0.24

Note: This table reports the fractions of the six periods ahead forecast error of the selected variables explained by the oil supply shock (columns A), global demand shock (columns B), and oil-specific demand shock (columns C). For each shock, “K”, “BH” and “BKL” denote the FAVAR models which use the Kilian global activity indicator, the Baumeister-Hamilton world IP growth and the Baumeister-Korobilis-Lee global economic conditions indicator as the measure of global activity, respectively.

C Additional Tables and Figures

Table C.1: Monte Carlo Results I: $N = 100$, $T = 200$, $\delta = 0$

Case	#Proxies	$R^2_{F, \hat{F}^+}$		$S^2_{(S)RPC}/S^2_{PC}$						$\hat{r}_{ED}$	RPC		SRPC		
		PC	SOFAR	RPC _(1,1)	RPC	SRPC	SOFAR	RPC _(1,1)	RPC	SRPC		$\hat{r}_{CV}$	$\hat{q}_{CV}$	$\hat{r}_{CV}$	$\hat{q}_{CV}$
(A) Common F_2 : no restrictions imposed on F_2 loadings and $\text{var}(F_{2t}) = \text{var}(F_{1t})$															
1	1	0.94	0.94	0.87	0.94	0.94	0.93	1.93	0.96	0.96	2.02	2.00	0.01	2.00	0.01
2	5	0.94	0.94	0.92	0.94	0.94	0.92	1.13	0.97	0.98	2.02	1.86	0.15	1.85	0.15
3	10	0.94	0.94	0.93	0.94	0.94	0.90	0.95	0.95	0.95	2.02	1.74	0.26	1.74	0.27
4	20	0.94	0.94	0.94	0.94	0.94	0.91	0.87	0.91	0.91	2.02	1.39	0.62	1.39	0.62
5	20	0.94	0.94	0.92	0.94	0.94	0.92	1.21	0.94	0.94	2.02	1.94	0.07	1.93	0.07
(B) Somewhat weak F_2 : 50% of F_2 loadings set to zero and $\text{var}(F_{2t}) = \text{var}(F_{1t})$															
1	1	0.90	0.92	0.88	0.88	0.88	0.79	0.97	0.93	0.93	2.03	1.15	0.85	1.15	0.85
2	5	0.90	0.92	0.93	0.93	0.93	0.79	0.61	0.63	0.63	2.03	1.01	0.99	1.01	0.99
3	10	0.90	0.92	0.94	0.94	0.94	0.86	0.55	0.56	0.56	2.02	1.01	1.00	1.01	1.00
4	20	0.90	0.92	0.95	0.95	0.95	0.84	0.50	0.50	0.50	2.02	1.00	1.00	1.00	1.00
5	20	0.90	0.92	0.93	0.93	0.93	0.86	0.69	0.70	0.70	2.02	1.01	0.99	1.01	0.99
(C.1) Weak F_2 : 90% of F_2 loadings set to zero and $\text{var}(F_{2t}) = \text{var}(F_{1t})$															
1	1	0.52	0.60	0.88	0.83	0.83	0.74	0.28	0.33	0.33	1.49	1.00	0.86	1.00	0.86
2	5	0.51	0.60	0.93	0.91	0.92	0.72	0.20	0.26	0.23	1.51	1.00	1.17	1.00	1.10
3	10	0.52	0.60	0.92	0.92	0.93	0.70	0.21	0.29	0.25	1.50	1.01	1.25	1.00	1.16
4	20	0.52	0.60	0.90	0.90	0.92	0.75	0.27	0.34	0.26	1.49	1.01	1.19	1.00	1.12
5	20	0.52	0.60	0.88	0.88	0.90	0.72	0.28	0.32	0.29	1.53	1.01	1.16	1.01	1.12
(C.2) Weak F_2 : no restrictions imposed on factor loadings and $\text{var}(F_{2t}) = 0.1 \times \text{var}(F_{1t})$															
1	1	0.57	0.56	0.88	0.87	0.87	0.96	0.41	0.42	0.42	1.30	1.01	0.98	1.01	0.98
2	5	0.58	0.57	0.93	0.93	0.93	0.96	0.28	0.29	0.29	1.32	1.00	0.99	1.00	0.99
3	10	0.58	0.57	0.94	0.94	0.94	0.95	0.26	0.26	0.26	1.34	1.00	1.00	1.00	1.00
4	20	0.57	0.57	0.94	0.94	0.94	0.95	0.28	0.29	0.28	1.31	1.00	1.00	1.00	1.00
5	20	0.57	0.56	0.92	0.92	0.92	0.98	0.34	0.35	0.35	1.31	1.01	1.00	1.00	1.00

Note: This table reports the performance of the PC, SOFAR, and (S)RPC estimators described in Section 4.2. Five proxy designs are considered. In Cases 1–4, we start with the least noisy proxy and sequentially add noisier proxies; in Case 5, we include 20 similarly highly noisy proxies. The dataset dimensions are $N = 100$ and $T = 200$, the proxy-noise exponent is $\delta = 0$, and the number of simulation repetitions is 1000. Proxy noise is $T^{-\delta} \sigma_j \psi_{jt}$. For each δ , σ_j is calibrated to the proxy R^2 targets at $T = 100$ and held fixed across sample sizes. The results are summarized by two statistics: $R^2_{F, \hat{F}^+}$ denotes the average R^2 of the regression of F onto $\hat{F}^+$; $S^2_{(S)RPC}/S^2_{PC}$ is the ratio of the squared difference between the out-of-sample forecasts from the indicated method relative to the infeasible forecast based on the true factors divided by the same object for the PC estimator. The last five columns report the average $\hat{r}$ from Onatski (2010)'s ED method and the average $\hat{r}$ and $\hat{q}$ selected by cross-validation in Section 3.1 under both RRR and its modification with sparsity constraints.

Table C.2: Monte Carlo Results I: $N = 100$, $T = 200$, $\delta = 1/2$

Case	#Proxies	$R^2_{F, \hat{F}^+}$		$S^2_{(S)RPC}/S^2_{PC}$						$\hat{r}_{ED}$	RPC		SRPC		
		PC	SOFAR	RPC _(1,1)	RPC	SRPC	SOFAR	RPC _(1,1)	RPC	SRPC		$\hat{r}_{CV}$	$\hat{q}_{CV}$	$\hat{r}_{CV}$	$\hat{q}_{CV}$
(A) Common F_2 : no restrictions imposed on F_2 loadings and $\text{var}(F_{2t}) = \text{var}(F_{1t})$															
1	1	0.94	0.94	0.91	0.93	0.93	0.93	1.29	1.02	1.02	2.02	1.76	0.24	1.76	0.24
2	5	0.94	0.94	0.94	0.94	0.94	0.92	0.82	0.84	0.84	2.02	1.22	0.78	1.23	0.78
3	10	0.94	0.94	0.95	0.95	0.95	0.91	0.71	0.74	0.73	2.03	1.11	0.89	1.11	0.89
4	20	0.94	0.94	0.95	0.95	0.95	0.91	0.72	0.73	0.73	2.02	1.02	0.99	1.02	0.99
5	20	0.94	0.94	0.94	0.94	0.94	0.92	0.87	0.85	0.86	2.03	1.25	0.75	1.25	0.76
(B) Somewhat weak F_2 : 50% of F_2 loadings set to zero and $\text{var}(F_{2t}) = \text{var}(F_{1t})$															
1	1	0.90	0.92	0.92	0.92	0.92	0.80	0.66	0.68	0.68	2.02	1.01	0.99	1.01	0.99
2	5	0.90	0.92	0.95	0.95	0.95	0.80	0.45	0.47	0.48	2.02	1.00	1.00	1.00	1.00
3	10	0.90	0.92	0.96	0.96	0.96	0.89	0.45	0.45	0.45	2.02	1.00	1.00	1.00	1.00
4	20	0.90	0.92	0.96	0.96	0.96	0.81	0.39	0.40	0.39	2.03	1.01	1.00	1.01	1.00
5	20	0.90	0.92	0.95	0.95	0.95	0.83	0.52	0.52	0.52	2.02	1.00	1.00	1.00	1.00
(C.1) Weak F_2 : 90% of F_2 loadings set to zero and $\text{var}(F_{2t}) = \text{var}(F_{1t})$															
1	1	0.52	0.60	0.93	0.88	0.88	0.76	0.19	0.23	0.23	1.51	1.00	0.90	1.00	0.90
2	5	0.51	0.60	0.95	0.94	0.95	0.68	0.16	0.24	0.19	1.50	1.00	1.17	1.00	1.06
3	10	0.52	0.60	0.94	0.94	0.95	0.71	0.18	0.24	0.19	1.53	1.01	1.25	1.01	1.10
4	20	0.52	0.61	0.91	0.91	0.94	0.73	0.24	0.31	0.23	1.51	1.01	1.25	1.01	1.13
5	20	0.52	0.60	0.90	0.90	0.93	0.72	0.24	0.30	0.23	1.52	1.01	1.22	1.02	1.12
(C.2) Weak F_2 : no restrictions imposed on factor loadings and $\text{var}(F_{2t}) = 0.1 \times \text{var}(F_{1t})$															
1	1	0.57	0.56	0.93	0.92	0.92	0.95	0.31	0.32	0.32	1.30	1.00	0.99	1.00	0.99
2	5	0.58	0.57	0.95	0.95	0.95	0.98	0.24	0.25	0.25	1.33	1.00	1.00	1.00	1.00
3	10	0.57	0.56	0.96	0.96	0.95	0.96	0.22	0.22	0.23	1.34	1.00	1.00	1.00	1.00
4	20	0.58	0.57	0.95	0.95	0.96	0.92	0.25	0.25	0.24	1.35	1.00	1.00	1.00	1.00
5	20	0.57	0.56	0.94	0.94	0.94	0.98	0.28	0.29	0.28	1.31	1.00	1.00	1.00	1.00

Note: This table reports the performance of the PC, SOFAR, and (S)RPC estimators described in Section 4.2. Five proxy designs are considered. In Cases 1–4, we start with the least noisy proxy and sequentially add noisier proxies; in Case 5, we include 20 similarly highly noisy proxies. The dataset dimensions are $N = 100$ and $T = 200$, the proxy-noise exponent is $\delta = 1/2$, and the number of simulation repetitions is 1000. Proxy noise is $T^{-\delta} \sigma_j \psi_{jt}$. For each δ , σ_j is calibrated to the proxy R^2 targets at $T = 100$ and held fixed across sample sizes. The results are summarized by two statistics: $R^2_{F, \hat{F}^+}$ denotes the average R^2 of the regression of F onto $\hat{F}^+$; $S^2_{(S)RPC}/S^2_{PC}$ is the ratio of the squared difference between the out-of-sample forecasts from the indicated method relative to the infeasible forecast based on the true factors divided by the same object for the PC estimator. The last five columns report the average $\hat{r}$ from Onatski (2010)'s ED method and the average $\hat{r}$ and $\hat{q}$ selected by cross-validation in Section 3.1 under both RRR and its modification with sparsity constraints.