Nonlinear non-Gaussian dynamic factor models with interacting location and scale factors

Geert Mesters,[(a)] *Siem Jan Koopman,*[(b)] *Bernd Schwaab*[(c)*]

[(a)] Universitat Pompeu Fabra & Barcelona GSE

[(b)] VU University Amsterdam & Tinbergen Institute

[(c)] European Central Bank, Financial Research

February 15, 2024

*****Preliminary conference submission*****

Abstract

We propose a new class of nonlinear non-Gaussian dynamic factor models for panel time series data and provide approximate filtering and smoothing recursions for its analysis. The model has latent common factor structures underlying both locations and scales and, subject to only identification restrictions, both types of factors can interact freely in an unrestricted vector autoregression. The approximate filtering recursions require augmenting the Kalman filter with weighted least squares regressions, and yield the approximate marginal likelihood as a by-product. The quality of the approximation can be assessed through diagnostic procedures. Simulation experiments confirm that our method accurately recovers all factors and that the approximation errors do not accumulate over time. An empirical application to G7 countries' real GDP growth rates illustrates the relevance of interacting factors for in-sample analysis and out-of-sample forecasting.

JEL classification: C32, C43.

Keywords: factor model, unobserved components, approximate filter and smoother, common volatility, uncertainty.

*Corresponding author: Bernd Schwaab, email: `bernd.schwaab@ecb.europa.eu`. We thank Max J. Roos for excellent research assistance, and Massimiliano Marcellino and Ivan Petrella for comments. The views expressed in this paper are those of the authors and they do not necessarily reflect the views or policies of the European Central Bank.

1 Introduction

The presence of a factor structure underlying the conditional *means* of many interesting macroeconomic and financial panel data sets has by now been firmly established; see, for example, Stock and Watson (2002), Bai and Ng (2002), Bai and Li (2012), Ahn and Horenstein (2013), and Stock and Watson (2016). Moreover, recent studies have further suggested a factor structure for the *volatilities* of such panels; see, for example, Jurado et al. (2015), Barigozzi and Hallin (2016), Barigozzi and Hallin (2017) and Gorodnichenko and Ng (2017). Given this evidence, it is natural to inquire about the interaction between such mean (location) and volatility (scale) factors. Specifically, in macroeconomic time series panels of widespread interest, do shocks to the location factors also drive the common scale factors? Or vice versa? Are factor interactions contemporaneous, or more pronounced at a lag? To what extent is joint inference sharpened when both location and scale factors are inferred simultaneously? Are factor interactions empirically relevant according to e.g. likelihood ratio tests? Do they improve point and density forecasts? Unfortunately, and despite its theoretical appeal and obvious practical relevance, a comprehensive statistical framework to address such first-order questions in macroeconomics and finance is currently missing.

This paper introduces a novel nonlinear and non-Gaussian dynamic factor model in state space form. In this model, the location and scale factors can interact freely within an unrestricted vector autoregression, subject only to identification restrictions. We propose an estimation method for the model's deterministic parameters, along with approximate filtering and smoothing recursions that enable joint inference on all latent factors. The approximate filter is straightforward to implement, involving the augmentation of the Kalman filter with weighted least squares regressions. Additionally, it conveniently yields the approximate marginal likelihood as a by-product.

To our knowledge, we are the first to present a comprehensive, tractable, non-simulation-based methodology for the statistical analysis of nonlinear non-Gaussian dynamic factor models with interacting factor structures. We focus on modeling data with a low-to-moderate

cross-sectional dimension N. In other aspects, the statistical model is a natural generalization of the dynamic factor model for the conditional mean; see e.g. Stock and Watson (2016) for a review.

In our proposed model, the observed data follow a Student's t density, parameterized by two signals: one representing the location and the other the (log-)scale. Deviating from the linear Gaussian assumption is crucial for analyzing macroeconomic and financial time series panels, where heavy tails and volatility clustering are prevalent characteristics. Within our framework, both the location and scale signals exhibit factor structures, which can now be interdependent. Notably, our approach captures the empirical regularity that macroeconomic time series' locations and scales often move in tandem, exhibiting either cyclical or countercyclical behavior; see e.g. Jurado et al. (2015), Adrian et al. (2019), Caldara et al. (2021), and Adrian et al. (2022).

While the concept of a state space model with interacting location and scale factors is intuitively appealing, its statistical analysis necessitates the development of novel methodology. Unfortunately, existing methods are not readily applicable for this purpose. Traditional estimation techniques for factor models, such as principal components analysis (PCA) and maximum likelihood methods based on the Kalman filter, would require substantial adjustments owing to the model's non-Gaussian nature and the non-linear entry of both location and scale factors into the model; see e.g. Durbin and Koopman (2012). Moreover, the contemporaneous dependence between location and scale factors renders standard Gibbs sampling routines inadmissible, as observed in previous work by Chib et al. (2006) and Kastner et al. (2017). Simulation-based approaches, including particle filtering, Markov Chain Monte Carlo, and Importance Sampling, also present significant challenges: they are not straightforward to implement, can be computationally expensive, and may ultimately still fall short of providing convincing results in practice.

In our approximate filtering and smoothing algorithms, we sequentially replace the infeasible true conditional density with an approximate density from the Gaussian family. The

mean and variance of this approximating density are optimally chosen to minimize the divergence between the target and the approximating density. Our criterion for optimal fit can be seen as an online adaptation of the efficient importance sampling criterion discussed in Richard and Zhang (2007) and Koopman et al. (2014). The resulting "minimum variance" approximating densities can be computed using numerical integration methods, which can be implemented through weighted least squares regressions. As a valuable by-product, the filtering algorithm provides an accurate approximation for the marginal likelihood. This, in turn, offers a straightforward way to obtain quasi-maximum likelihood (QML) estimates for the model's deterministic parameters.

Our approach draws inspiration from a recent surge in approximate inference methods. Notable examples include Variational Inference (as discussed e.g. by Bishop (2006) and Blei et al. (2017)), Expectation Propagation (explored by Minka (2001)), the Integrated Nested Laplace Approach (as presented by Rue et al. (2009)), and Approximate Bayesian Computation methods (studied by Marin et al. (2012)). Our approximate filter retains the key advantages of these approximate inference techniques: First, the method is relatively straightforward to implement, requiring an augmentation of standard Kalman filter-type recursions with weighted least squares regressions. Second, by avoiding the need for sampling, our algorithm outperforms traditional Markov Chain Monte Carlo approaches in terms of speed. This efficiency makes it well-suited for complex models such as the nonlinear and non-Gaussian model under consideration. Finally, our approach remains flexible. For instance, different observation densities (including non-differentiable ones) could in principle be accommodated with minor adjustments. The main distinctions between our approach and the aforementioned methods lie in the sequential determination of the approximate posterior, the use of a distinct divergence measure leading to computationally tractable optimization problems, and our estimation of deterministic model parameters via quasi-maximum likelihood. In contrast, the cited approaches typically operate within a Bayesian framework where all parameters are treated as latent states.

We evaluate our parameter estimation and signal extraction method in a comprehensive simulation study. This study is centered around four fundamental questions: Can our method effectively differentiate between lagged and contemporaneous factor dependencies in finite samples of realistic size? Does our approach inadvertently identify factor interactions that do not truly exist? Can our methodology yield more precise estimates for all factors compared to simpler, non-parametric alternatives, such as two-stage PCA? Finally, is our method susceptible to approximation errors that accumulate over time? Our findings suggest that the answers to the first and third questions are affirmative and otherwise negative.

To demonstrate the utility of our approach for structural analysis and forecasting, we examine end-of-quarter real GDP growth rates for the Group of Seven (G7) countries: the United States, Japan, Germany, Great Britain, France, Italy, and Canada. The dataset ranges from 1961Q3 to 2022Q4, resulting in $N = 7$ countries and $T = 246$ observations. In our model, we select two location factors and one scale factor to fit these data. The factor loadings are restricted such that the two location factors can be interpreted economically: one represents a global level factor, while the other captures an additional European group factor.

The model's parameter estimates reveal statistically (and economically) significant factor interactions, both contemporaneously and at a lag. Non-zero factor interactions are further supported by likelihood ratio tests and impulse response function estimates. The scale factor is counter-cyclical, increasing macroeconomic uncertainty at the onset of and during a recession (Jurado et al., 2015) and making the bottom quantiles much more volatile than the upper quantiles (Adrian et al., 2019, Adrian et al., 2022). Suppressing factor interactions also has costs in terms of out-of-sample forecasting performance, leading to reduced accuracy in point and density forecasts up to two years ahead. Diagnostic checks confirm a generally tight fit between the intractable target densities and the approximating Gaussian densities. Our empirical data are moderately fat-tailed; this is intuitive since the Covid-19 pandemic recession is part of the estimation sample.

We proceed as follows. Section 2 introduces our nonlinear dynamic factor model, and presents our estimation methodology for its latent factors and deterministic parameters. Section 3 discusses simulation results. Section 4 provides an empirical illustration. Section 5 concludes. A web appendix provides further technical and empirical results.

2 Statistical model

2.1 The dynamic factor model in state space form

We introduce our nonlinear non-Gaussian dynamic factor model with interacting location and scale factors for the $N \times 1$ observation vectors $y_t = (y_{1t}, \ldots, y_{Nt})'$, where index $i = 1, \ldots, N$ indicates a cross-sectional unit, and $t = 1, \ldots, T$ denotes time. We have in mind a finite and low-dimensional N, such as, say, $N \leq 15$. Two sets of common factors are important in the development below: (i) the $r \times 1$ vector of location factors $f_t = (f_{1t}, \ldots, f_{rt})'$ and (ii) the $q \times 1$ vector of scale factors $h_t = (h_{1t}, \ldots, h_{qt})'$. We collect these two sets of factors in the $(r+q) \times 1$ state vector $\alpha_t = (f_t', h_t')'$.

The factor model's observation (measurement) "equation" for y_{it} is then given by

$$\begin{aligned} y_{it}|\alpha_t &\overset{i.d.}{\sim} p(y_{it}|\mu_{it}, \sigma_{it}^2, \nu), \\ \mu_{it} &= c_i + \lambda_i' f_t, \qquad \sigma_{it}^2 = \exp(l_i' h_t), \end{aligned} \tag{1}$$

where $i.d.$ stands for independently distributed across i and t, and $p(y_{it}|\mu_{it}, \sigma_{it}^2, \nu)$ denotes the observation density with location μ_{it} and scale σ_{it}^2. The inclusion of ν in the observation density allows it to depend on additional parameters, such as a degrees-of-freedom parameter. The common factors f_t influence the location parameters μ_{it} via the $r \times 1$ loading vectors λ_i. Similarly, the common factors h_t influence the log-scales $\ln \sigma_{it}^2$ through the $q \times 1$ loading vectors l_i.

Important examples for the observation density $p(y_{it}|\mu_{it}, \sigma_{it}^2, \nu)$ include symmetric den-

sities such as the normal density $y_{it}|\alpha_t \overset{i.d.}{\sim} \mathcal{N}(\mu_{it}, \sigma_{it}^2)$ and the Student's t density with ν degrees of freedom $y_{it}|\alpha_t \overset{i.d.}{\sim} t(\mu_{it}, \sigma_{it}^2, \nu)$. In principle, skewed densities could also be considered within our framework. Skewness, however, is more instructively modeled in our setting via a contemporaneous and/or lagged dependence between the location and scale factors. For example, a negative (positive) dependence between f_t and h_t will cause multi-step ahead forecasts of the data to be negatively (positively) skewed. We consider non-zero factor interactions for simulated data in Section 3, and for actual data in Section 4.

The common location and scale factors are stacked in the state vector α_t and are jointly modeled as a first-order VAR. The state (transition) equation is given by

$$\alpha_t = (f_t', h_t')' = d + \mathbf{\Phi}\alpha_{t-1} + \eta_t, \qquad \eta_t \overset{i.d.}{\sim} \mathcal{N}(0, \mathbf{\Sigma}), \tag{2}$$

with $(r+q) \times 1$ conditional mean d, an $(r+q) \times (r+q)$ autoregressive coefficient matrix $\mathbf{\Phi}$, and an $(r+q) \times 1$ vector of Gaussian shocks η_t from the multivariate normal density $\mathcal{N}(0, \mathbf{\Sigma})$ with mean zero and positive-definite variance matrix $\mathbf{\Sigma}$. The first-order VAR in (2) is without loss of generality, as higher-order VAR processes can be cast into companion form. For convenience, we collect all unknown deterministic parameters in the vector θ.

The key novelty of model (1) – (2) lies in the joint and unrestricted treatment of its interdependent factors. An independent volatility model, as in e.g. Koopman and Bos (2004), is obtained when restricting $\mathbf{\Phi}$ and $\mathbf{\Sigma}$ to be block diagonal, see also Kastner et al. (2017) and references therein. By contrast, no a-priori restrictions are placed on the dynamic interaction between the location and scale factors in our framework. The scale of the data is thus fully endogenous, and likelihood ratio hypothesis tests can be conducted to assess the appropriateness of the exogenous volatility assumption.

As usual, the state vector α_t needs to be initialized. If the VAR in (2) is stationary, then the state vector's stationary distribution can be used, such that

$$\alpha_1 \sim \mathrm{N}\Big(\delta, (\mathbf{I}_{r+q} - \mathbf{\Phi}\mathbf{\Phi}')^{-1}\mathbf{\Sigma}\Big), \tag{3}$$

where $\delta = (\mathbf{I}_{r+q} - \mathbf{\Phi})^{-1} d$. Of course, other initial conditions not depending on system matrices δ, $\mathbf{\Phi}$, and $\mathbf{\Sigma}$ can also be adopted, depending on the application and data at hand.

2.2 Identification restrictions

Without further restrictions the deterministic parameters θ are not identified. Several strategies to identify dynamic factor models exist, and in many cases the identifying restrictions can be formulated such that the latent factors have an economic interpretation; see e.g. Bai and Li (2012) and Stock and Watson (2016). In Section 4 below, we restrict at least r^2 elements of the location factor loadings $\Lambda = (\lambda_1, \ldots, \lambda_N)'$, and at least q^2 elements of the scale factor loadings $L = (l_1, \ldots, l_N)'$ to ensure statistical identification. In addition, we set $c = (c_1, \ldots, c_N)'$ to the unconditional means of the data, and restrict the first r elements of δ to zero. Over-identifying restrictions can be imposed to help enhance the economic interpretability of the factors, and/or to facilitate the numerical estimation of the remaining unrestricted deterministic parameters.

2.3 Approximate filtering recursions

The estimation of the statistical model (1) –(3)'s latent factors and deterministic parameters is non-trivial because the model is non-Gaussian and both location and scale factors enter the model in a nonlinear way. This implies that standard filtering and smoothing methods based on the Kalman filter cannot be used. To see this immediately, note that the conditional mean of α_t is, by definition,

$$\mathbb{E}_p(\alpha_t|Y_t;\theta) = \int_{\alpha_t} \alpha_t p(\alpha_t|Y_t;\theta) \mathrm{d}\alpha_t, \tag{4}$$

where $Y_t = \{y_1, \ldots, y_t\}$. To obtain conditional mean estimates, we therefore need to know the conditional density $p(\alpha_t|Y_t;\theta)$. Unfortunately, this density is not available owing to the nonlinear features of the model. The same issue holds when we consider other conditional

mean functions, such as $\mathbb{E}_p(g(\alpha_t)|Y_t;\theta)$ for some function $g(\cdot)$; such estimates also require knowing the conditional density $p(\alpha_t|Y_t;\theta)$.

Our approximation method is based on replacing the conditional density $p(\alpha_t|Y_t;\theta)$ by an approximating conditional density, $\hat{p}(\alpha_t|Y_t;\theta)$. This approximating density needs to be computationally convenient – in the sense that the integral in (4) can be solved quickly and reliably. In addition, the approximating density should be as close as possible to the true conditional density.

We derive the approximate filtering algorithm as follows. Let $p(\alpha_t|Y_t;\theta)$ and $p(\alpha_{t+1}|Y_t;\theta)$ denote the filtered and predictive conditional densities, and let $\hat{p}(\alpha_t|Y_t;\theta)$ and $\hat{p}(\alpha_{t+1}|Y_t;\theta)$ denote their approximating counterparts. The approximating densities are chosen to belong to the Gaussian family, such that

$$\hat{p}(\alpha_t|Y_t;\theta) \equiv \mathcal{N}(\hat{a}_{t|t}, \hat{\mathbf{V}}_{t|t}) \qquad \text{and} \qquad \hat{p}(\alpha_{t+1}|Y_t;\theta) \equiv \mathcal{N}(\hat{a}_{t+1|t}, \hat{\mathbf{V}}_{t+1|t}),$$

where $\hat{a}_{t+s|t} = \mathbb{E}(\alpha_{t+s}|Y_t;\theta)$ and $\hat{\mathbf{V}}_{t+s|t} = \mathrm{Var}(\alpha_{t+s}|Y_t;\theta)$ for $s = 0, 1$. The goal is to compute $\hat{a}_{t+s|t}$ and $\hat{\mathbf{V}}_{t+s|t}$ recursively, for $t = 1, \ldots, T$, such that they are close to the true conditional means and variances.

We first discuss the computation of $\hat{p}(\alpha_{t+1}|Y_t;\theta)$ given that $\hat{p}(\alpha_t|Y_t;\theta)$ is available. Since the distribution of the common factors α_t is Gaussian, we take $\hat{p}(\alpha_{t+1}|\alpha_t) = p(\alpha_{t+1}|\alpha_t)$ and note that $p(\alpha_{t+1}|\alpha_t) \equiv \mathcal{N}((\mathbf{I}_{r+q} - \mathbf{\Phi})\delta + \mathbf{\Phi}\alpha_{t-1}, \mathbf{\Sigma})$. The corresponding predictive distribution conditional on the data is

$$\hat{p}(\alpha_{t+1}|Y_t;\theta) = \int_{\alpha_t} \hat{p}(\alpha_{t+1}|\alpha_t)\hat{p}(\alpha_t|Y_t;\theta)\mathrm{d}\alpha_t = \mathcal{N}\left(d + \mathbf{\Phi}\hat{a}_{t|t}, \mathbf{\Phi}\hat{\mathbf{V}}_{t|t}\mathbf{\Phi}' + \mathbf{\Sigma}\right),$$

which follows directly from the properties of the Gaussian distribution.

Next, we consider constructing $\hat{p}(\alpha_t|Y_t;\theta)$ given that $\hat{p}(\alpha_t|Y_{t-1};\theta)$ is known. From Bayes rule we know that $p(\alpha_t|Y_t;\theta) \propto p(y_t|\alpha_t;\theta)p(\alpha_t|Y_{t-1};\theta)$. As a result, once $p(\alpha_t|Y_{t-1};\theta)$ has been approximated, we only need to find a good approximation for $p(y_t|\alpha_t;\theta)$ to approximate

$p(\alpha_t|Y_t;\theta)$.

We consider the following Gaussian density approximation for the observation density

$$\hat{p}(y_t|\alpha_t;\theta) = \prod_{i=1}^{n} \hat{p}(y_{it}|z_{it};\theta), \qquad \hat{p}(y_{it}|z_{it};\theta) = \exp\left(\omega_{it} + b_{it}'z_{it} - \frac{1}{2}z_{it}'\mathbf{C}_{it}z_{it}\right), \tag{5}$$

where $z_{it} = (\mu_{it}, \ln\sigma_{it}^2)'$, and the 2×1 vector b_{it} and the 2×2 matrix $\mathbf{C}_{it}$ are chosen to ensure that $\hat{p}(y_t|\alpha_t;\theta)$ and $p(y_t|\alpha_t;\theta)$ are close. The term $\omega_{it} = \left(\ln|\mathbf{C}_{it}| - \ln 2\pi - b_{it}'\mathbf{C}_{it}^{-1}b_{it}\right)/2$ is chosen such that the density always integrates to one.

The coefficients b_{it} and $\mathbf{C}_{it}$ are obtained from the minimization problem

$$\arg\min_{\tilde{b}_{it},\tilde{\mathbf{C}}_{it}} \int_{z_{it}} \left[\ln\left(p(y_{it}|z_{it};\theta)/\hat{p}(y_{it}|z_{it};\theta)\right)\right]^2 \hat{p}(z_{it}|Y_t;\theta)\mathrm{d}z_{it}. \tag{6}$$

Since $\ln\hat{p}(y_{it}|z_{it}^{l,k};\theta)$ is a linear function of the coefficients in b_{it} and $\mathbf{C}_{it}$, the complicated-looking minimization problem (6) boils down to running a standard (weighted) least squares regression. We discuss the precise regression specification in Section 2.7 when discussing diagnostic checks of approximation quality.

The criterion function in (6) can be interpreted in at least two ways. First, it chooses the coefficients b_{it} and $\mathbf{C}_{it}$ such that the variance of the scalar random variable $\ln p(y_{it}|z_{it};\theta)/\hat{p}(y_{it}|z_{it};\theta)$ is minimized. (The variance is considered with respect to the approximating density $\hat{p}(z_{it}|Y_t;\theta)$.) Second, (6) can be seen as an online version of the efficient importance sampling criterion function considered in the importance sampling literature. There, this criterion function is used to select an accurate importance density, see e.g. Richard and Zhang (2007), Koopman et al. (2014), and Koopman et al. (2017).

The optimal coefficients $b_{it} = (b_{it,1}, b_{it,2})$ and $\mathbf{C}_{it} = [c_{it,11}, c_{it,12}; c_{it,21}, c_{it,22}]$ are stacked into the $2n\times 1$ vector $b_t = (b_{1t,1},\ldots,b_{nt,1},b_{1t,2},\ldots,b_{nt,2})'$, and the $2n\times 2n$ coefficient matrix

$$\mathbf{C}_t = \begin{bmatrix} \mathbf{C}_{t,11} & \mathbf{C}_{t,12} \\ \mathbf{C}_{t,21} & \mathbf{C}_{t,22} \end{bmatrix}, \qquad \text{where } \mathbf{C}_{t,lk} = \mathrm{diag}(c_{1t,lk},\ldots,c_{nt,lk}),$$

for $l, k = 1, 2$. We note that $\mathbf{C}_t$ is sparse, containing $4N \leq 4N^2$ non-zero elements. It is reasonably fast to invert as a result.

Since $\ln p(y_{it}|z_{it};\theta)$ is nonlinear in z_{it}, the integral in (6) cannot be solved analytically. However, this integral is of low dimension (two). This low dimension makes it attractive to use numerical integration methods to evaluate the integral. We follow the exposition in Monahan (2001) and use Gaussian-quadrature for the evaluation of the integral; see also Koopman et al. (2014) and Koopman et al. (2017).

To briefly sketch the numerical integration procedure, we consider a set of abscissa $\{v_l\}_{l=1}^L$ with associated Gauss-Hermite weights $h(v_l)$ for $l = 1, \ldots, L$. Typical values of L are 10, 20, or 30. The numerical implementation of the minimization problem (6) is given by

$$\{\hat{b}_{it}, \hat{\mathbf{C}}_{it}\} = \arg\min_{\tilde{b}_{it}, \tilde{\mathbf{C}}_{it}} \sum_{l,k=1}^{L} w_{l,k} \left[\ln\left(p(y_{it}|z_{it}^{l,k};\theta)/\hat{p}(y_{it}|z_{it}^{l,k};\theta)\right)\right]^2 \tag{7}$$

where $w_{l,k} = h(v_l)h(v_k)\exp(\frac{1}{2}v_l^2)\exp(\frac{1}{2}v_k^2)$, $z_{it}^{l,k} = \mathbf{R}_i\hat{a}_{t|t} + (\mathbf{R}_i\hat{\mathbf{V}}_{t|t}\mathbf{R}_i')^{1/2}v_{l,k}$, and $v_{l,k} = (v_l, v_k)'$. $\mathbf{R}_i = \text{diag}(\lambda_i', l_i')$ is a $2 \times (r+q)$ combined loading matrix, where $\text{diag}(\cdot,\cdot)$ denotes the diagonal concatenation of two matrices.

After solving the minimization problem (7), the estimated coefficients $\hat{b}_{it}$ and $\hat{\mathbf{C}}_{it}$ replace b_{it} and $\mathbf{C}_{it}$ in (5). Now $\hat{p}(y_t|\alpha_t;\theta)$ is completely known, and together with $\hat{p}(\alpha_t|Y_{t-1};\theta)$ we can compute $\hat{p}(\alpha_t|Y_t;\theta)$. Since both $\hat{p}(y_t|\alpha_t;\theta)$ and $\hat{p}(\alpha_t|Y_{t-1};\theta)$ are Gaussian, also $\hat{p}(\alpha_t|Y_t;\theta)$ is Gaussian and known in closed form. This completes the construction of the approximating conditional filtering density. The following algorithm summarizes the approximate filter.

Algorithm 1: Approximate Filter

1. Initialize: For example, set $t = 1$, $\hat{a}_{1|0} = \delta$ and $\hat{\mathbf{V}}_{1|0} = (\mathbf{I}_k - \mathbf{\Phi}\mathbf{\Phi}')^{-1}\mathbf{\Sigma}$.

2. Filter: set $\hat{b}_{it} = 0$ and $\hat{\mathbf{C}}_{it}^{-1} = \mathbf{I}_2$ for $i = 1, \ldots, n$ and iterate between (a) and (b) until convergence.

(a) Update:

$$\hat{a}_{t|t} = \hat{a}_{t|t-1} + \hat{\mathbf{V}}_{t|t-1}\mathbf{R}'(\mathbf{R}\hat{\mathbf{V}}_{t|t-1}\mathbf{R}' + \hat{\mathbf{C}}_t^{-1})^{-1}(\hat{\mathbf{C}}_t^{-1}\hat{b}_t - \mathbf{R}\hat{a}_{t|t-1})$$
$$\hat{\mathbf{V}}_{t|t} = \hat{\mathbf{V}}_{t|t-1} - \hat{\mathbf{V}}_{t|t-1}\mathbf{R}'(\mathbf{R}\hat{\mathbf{V}}_{t|t-1}\mathbf{R}' + \hat{\mathbf{C}}_t^{-1})^{-1}\mathbf{R}\hat{\mathbf{V}}_{t|t-1}$$

where $\mathbf{R} = \text{diag}(\mathbf{\Lambda}, \mathbf{L})$, with $\mathbf{\Lambda} = (\lambda_1, \ldots, \lambda_n)'$ and $\mathbf{L} = (l_1, \ldots, l_n)'$.

(b) Approximate: obtain $\hat{b}_{it}$ and $\hat{\mathbf{C}}_{it}^{-1}$ by solving the approximation problem (7) for $i = 1, \ldots, N$

3. Predict:

$$\begin{aligned} \hat{a}_{t+1|t} &= d + \mathbf{\Phi}\hat{a}_{t|t} \\ \hat{\mathbf{V}}_{t+1|t} &= \mathbf{\Phi}\hat{\mathbf{V}}_{t|t}\mathbf{\Phi}' + \mathbf{\Sigma} \end{aligned}$$

4. Increase $t = t + 1$ and go to step (ii)

Two issues deserve comment. First, the approximate filter's updating step (ii) needs to iterate between obtaining $\hat{a}_{t|t}$ and $\hat{V}_{t|t}$ and solving for $\{\hat{b}_{it}, \hat{\mathbf{C}}_{it}\}$. This is because the minimization problem (7) depends on $\hat{a}_{t|t}$ and $\hat{V}_{t|t}$, not $\hat{a}_{t|t-1}$ and $\hat{V}_{t|t-1}$. Typically, only a few iterations (less than five) are sufficient.

Second, upon closer inspection, the recursions above look similar to the standard Kalman filter recursions as presented in e.g. Durbin and Koopman (2012, Chapter 4). The important difference lies in the filtering step (ii), where the non-Gaussian observation density is approximated by the closest density in the Gaussian class, and the filtering is based on this approximating density. Specifically, $y_{it}^* = \mathbf{C}_{it}^{-1}b_{it}$ plays the role of a *bivariate* pseudo-observation for the *scalar* observation y_{it}, $\hat{\mathbf{C}}_t^{-1}\hat{b}_t - \mathbf{R}\hat{a}_{t|t-1}$ is a $[2N \times 1]$ prediction error, with one-step-ahead prediction error variance $\mathbf{R}\hat{\mathbf{V}}_{t|t-1}\mathbf{R}' + \hat{\mathbf{C}}_t^{-1}$. Matrix $\mathbf{C}_t^{-1}$ plays the role of a measurement error variance, and the expression $\hat{\mathbf{V}}_{t|t-1}\mathbf{R}'(\mathbf{R}\hat{\mathbf{V}}_{t|t-1}\mathbf{R}' + \hat{\mathbf{C}}_t^{-1})^{-1}$ is the approximate filter's Kalman gain matrix.

2.4 Approximate smoothing recursions

Let $\hat{a}_{t|T} = \mathbb{E}_{\hat{p}}(\alpha_t|Y_T;\theta)$ and $\hat{V}_{t|T} = \mathrm{Var}_{\hat{p}}(\alpha_t|Y_T;\theta)$. These full-sample quantities can be computed by the backward recursions based on the output of the approximate filter. The following algorithm summarizes the approximate smoother for model (1) – (3).

Algorithm 2: Smoothing recursions

1. Initialize; set $t = T$, $\hat{\mathbf{N}}_T = 0$ and $\hat{r}_T = 0$

2. Smooth:

$$\begin{aligned} r_{t-1} &= \mathbf{R}'(\mathbf{R}\hat{\mathbf{V}}_{t|t-1}\mathbf{R}' + \hat{\mathbf{C}}_t^{-1})^{-1}(\hat{\mathbf{C}}_t^{-1}\hat{b}_t - \mathbf{R}\hat{a}_{t|t-1}) + \mathbf{L}_t' r_t \\ \mathbf{N}_{t-1} &= \mathbf{R}'(\mathbf{R}\hat{\mathbf{V}}_{t|t-1}\mathbf{R}' + \hat{\mathbf{C}}_t^{-1})^{-1}\mathbf{R} + \mathbf{L}_t'\mathbf{N}_t\mathbf{L}_t \\ \hat{a}_{t|T} &= \hat{a}_{t|t-1} + \hat{\mathbf{V}}_{t|t-1} r_{t-1} \\ \hat{\mathbf{V}}_{t|T} &= \hat{\mathbf{V}}_{t|t-1} - \hat{\mathbf{V}}_{t|t-1}\mathbf{N}_{t-1}\hat{\mathbf{V}}_{t|t-1} \end{aligned}$$

   where $\mathbf{R} = \mathrm{diag}(\mathbf{\Lambda}, \mathbf{L})$, with $\mathbf{\Lambda} = (\lambda_1, \ldots, \lambda_n)'$ and $\mathbf{L}_t = \mathbf{\Phi} - \hat{\mathbf{V}}_{t|t-1}\mathbf{R}'(\mathbf{R}\hat{\mathbf{V}}_{t|t-1}\mathbf{R}' + \hat{\mathbf{C}}_t^{-1})^{-1}\mathbf{R}$.

3. Decrease $t = t - 1$ and go to step (ii).

The derivation follows along the lines of Durbin and Koopman (2012, Chapter 4). We again note a close correspondence between the standard Kalman Smoother and the backward smoothing recursions presented in Algorithm 2.

2.5 ML estimation of the deterministic parameters

The approximate filter provides a convenient prediction error decomposition from which an approximate marginal likelihood can be constructed. We use this marginal likelihood to obtain quasi-maximum likelihood estimates of the deterministic parameters θ. This is a common estimation approach for dynamic factor models when the levels of the variables are

driven by common factors, see, for example, Doz et al. (2012), Bai and Li (2012), Jungbacker and Koopman (2014), Banbura and Modugno (2014) and Bai and Li (2016).

We start by considering the following decomposition of the marginal likelihood

$$\begin{aligned} p(y;\theta) &= \prod_{t=1}^{T}\prod_{i=1}^{n} p(y_{it}|Y_{t-1};\theta) \\ &= \prod_{t=1}^{T}\prod_{i=1}^{n} \int_{z_{it}} p(y_{it}|z_{it};\theta)p(z_{it}|Y_{t-1};\theta)\ \mathrm{d}z_{it}, \end{aligned} \tag{8}$$

where again $p(y_{it}|z_{it};\theta)$ is nonlinear and $p(z_{it}|Y_{t-1};\theta)$ is not known in closed form. We proceed as with the approximate filter, replacing $p(z_{it}|Y_{t-1};\theta)$ by its approximated counterpart $\hat{p}(z_{it}|Y_{t-1};\theta)$. In addition, the integral in (8) is again low-dimensional (two). This means we can use numerical integration techniques to obtain an accurate estimate. We consider

$$\hat{p}(y;\theta) = \prod_{t=1}^{T}\prod_{i=1}^{n}\sum_{l=1}^{L}\sum_{k=1}^{L} w_{l,k}p(y_{it}|z_{it}^{l,k};\theta) \tag{9}$$

with $z_{it}^{l,k}$ and $v_{l,k} = (v_l, v_k)'$ as defined below (7). The marginal (log-)likelihood is optimized with respect to the deterministic parameters using numerical optimization methods, such as fminunc (in Matlab) or maxBFGS (in OxMetrics). The marginal likelihood approximation (9) takes into account the non-linearity in the observation density by considering $p(y_{it}|z_{it}^{l,k};\theta)$, which is known exactly in closed form.

2.6 Asymptotic behaviour

This section discusses the asymptotic properties of the QML estimator (QMLE) of the non-linear non-Gaussian model's deterministic parameters. We consider the case where N is fixed and $T \to \infty$. We adopt a QML perspective for two reasons. First, for any real-world data at hand, the statistical model, despite its considerable flexibility, is still unlikely to exactly coincide with the true DGP. Second, the log-likelihood is approximate since we replaced intractable densities such as $p(z_{it}|Y_{t-1};\theta)$ in (8) by their approximating counterparts.

Proposition 1 (Consistency): Under regularity conditions (see e.g. Gourieroux and Monfort, 2002, Property 8.1), and assuming that the approximation of $p(\alpha_t|Y_{t-1};\theta)$ by $\hat{p}(\alpha_t|Y_{t-1};\theta)$ is close such that the approximation errors have finite second moments, the QMLE converges almost surely to its pseudo-true value, i.e. $\hat{\theta}_{\text{QMLE}} \xrightarrow{a.s.} \theta_0^*$ as $T \to \infty$.

Proposition 2 (Asymptotic normality): Under the above regularity conditions and approximation quality assumption, the QMLE is also asymptotically normally distributed, i.e. $\sqrt{T}\left(\hat{\theta}_{\text{QMLE}} - \theta_0^*\right) \xrightarrow{d} \mathcal{N}\left(0, \Omega_I^{-1}\right)$ as $T \to \infty$.

We refer to the discussion in Gourieroux and Monfort (2002, Ch. 8.4). The WLS regression errors need to be sufficiently well-behaved such that a LLN and CLT can be applied to (10).

2.7 Diagnostic checks of approximation quality

While the filtering algorithm and accompanying parameter estimation method are intuitively appealing, they do not imply that the error introduced by replacing $p(\alpha_t|Y_t;\theta)$ by $\hat{p}(\alpha_t|Y_t;\theta)$ is always negligible. This section outlines a diagnostic check to assess approximation quality. The check effectively measures the distance between $\ln p(\alpha_t|Y_t;\theta)$ and $\ln \hat{p}(\alpha_t|Y_t;\theta)$ over time and over cross-sectional units.

We start by considering the criterion function (7). If this function were exactly equal to zero, then the approximation would be perfect and the filter exact. In particular, the random variable $\min_{\tilde{b}_{it},\tilde{\mathbf{C}}_{it}} \int_{z_{it}} \left[\ln\left(p(y_{it}|z_{it};\theta)/\hat{p}(y_{it}|z_{it};\theta)\right)\right]^2 \hat{p}(z_{it}|Y_t;\theta)\mathrm{d}z_{it}$, can be thought of as a squared disturbance term. The criterion function (7) can be restated in discretized form as

$$\hat{\sigma}_{it}^2 = \sum_{l,k=1}^{L} w_{l,k} \left(\ln p(y_{it}|z_{it}^{l,k};\theta) - \omega_{it} - \hat{b}_{it}' z_{it}^{l,k} + \frac{1}{2} z_{it}^{l,k\prime} \hat{\mathbf{C}}_{it} z_{it}^{l,k} \right)^2,$$

where the Gauss-Hermite integration weights w_{it} are defined below (7). This expression can

be viewed as the sum of squared residuals corresponding to the least squares regression

$$\tilde{y}_{it}^{l,k} = \tilde{x}_{it}^{l,k\prime}\beta_{it} + u_{it}^{l,k} \qquad l,k = 1,\ldots,L, \tag{10}$$

where $\tilde{y}_{it}^{l,k} = \sqrt{w_{l,k}}\ln p(y_{it}|z_{it};\theta)$, $\tilde{x}_{it}^{l,k} = \sqrt{w_{l,k}}(1, z_{it}^{l,k\prime}, \text{vech}(-\frac{1}{2}z_{it}^{l,k}z_{it}^{l,k\prime})')'$, $\beta_{it} = (\omega_{it}, b_{it}', \text{vech}(\mathbf{C}_{it})')'$ and $u_{it}^{l,k}$ are the disturbance terms of interest.

Two issues deserve comment. First, (10) is a least squares regression with L^2 pseudo-observations for each scalar y_{it}. There are thus $N \times T$ such regressions for each evaluation of the log-likelihood. The computational speed associated with the least squares problem (10) is thus welcome and essential. Second, the right-hand side regressors can be badly conditioned for some trial values of θ. In such cases, a Ridge-type regression can be used, mildly shrinking b_{it} towards zero and C_{it} towards an appropriate non-singular matrix.

For each regression (10), a corresponding R-squared is given by

$$\hat{R}_{it}^2 = 1 - \Big[\sum_{l,k=1}^{L}\Big(\hat{u}_{it}^{l,k}\Big)^2\Big] \Big/ \Big[\sum_{l,k=1}^{L}\Big(\tilde{y}_{it}^{l,k} - \bar{\tilde{y}}_{it}\Big)^2\Big], \tag{11}$$

where $\bar{\tilde{y}}_{it}$ is the average $\tilde{y}_{it}^{l,k}$. We suggest plotting these $\hat{R}_{it}^2$ over time, and studying their descriptive statistics. The closer the $\hat{R}_{it}^2$ are to one, the closer the fit between $\ln p(\alpha_t|Y_t;\theta)$ and $\ln \hat{p}(\alpha_t|Y_t;\theta)$.

3 Simulation study

We are particularly interested in addressing four questions with simulation experiments: First, is our method able to distinguish lagged from contemporaneous factor dependence in finite samples of realistic size? Second, is it maybe inappropriately pointing to factor interactions that are not there? Third, is it able to provide more precise estimates of f_t and h_t than simpler alternatives such as two-stage PCA? Finally, is it possibly subject to approximation errors that accumulate over time? We establish that the answers to questions

one and three are affirmative and otherwise negative.

3.1 Simulation design

Our simulation setup is similar to that of Doz et al. (2012), but now considers multiple ($r = 2$) location factors and one ($q = 1$) scale factor. The data dimensions are chosen as $N = 8$ and $T \in \{250, 500\}$, approximately in line with the dimensions of the macro panel data set studied in Section 4. The data density is a Student's t density with $\nu \in \{10, 50\}$, where $\nu = 10$ allows for outliers beyond those implied by a stochastic scale, and $\nu = 50$ proxies the Gaussian case. This yields four data generating processes (DGPs) in total.

The true $\alpha_t = (f_{1t}, f_{2t}, h_t)'$ always follows a stationary VAR with one lag. The first location factor f_{1t} interacts with h_t in a contemporaneous and lagged way, while the second location factor f_{2t} does not interact with h_t at all. We use the same θ_0 to generate $S = 100$ $[N \times T]$ data sets $\{y_{it}\}^{(s)}$, $s = 1, \ldots, S$. Parameters θ_0 are chosen with the above in mind and otherwise approximately in line with those reported in Section 4. We then obtain $\hat{\theta}_{ML}^{(s)}$, $\hat{\alpha}_{t|t}^{(s)}$, and $\hat{\alpha}_{t|T}^{(s)}$. To improve estimation speed we do not estimate back all parameters in θ, but focus on 12 parameters of particular interest; their respective true values are indicated by vertical lines in Figure 1 below.

3.2 Simulation results

We highlight three key findings.

First, the median maximum likelihood estimates are close to their true values, and the sampling distribution of most parameters is approximately Gaussian. This is fortunate, and also implies that our method correctly identifies factor interactions present in the DGP without suggesting spurious interactions.

Figure 1 presents the sampling distributions of $S = 100$ maximum likelihood estimates $\hat{\theta}_{ML}^{(s)}$ for the DGP with $T = 250$ and $\nu = 50$. The median parameter estimates are indicated by a vertical dashed blue line, while the corresponding true values of θ_0 are indicated by a

Table 1: Filtering accuracy

R^2 statistics from a regression of the true factors on a constant and estimated factors. $\hat{f}_t$ refers to location factors estimates (columns 3 – 5) and $\hat{h}_t$ refers to scale factor estimates (columns 6 – 8). Rows refer to four different simulation settings, with $T \in \{250, 500\}$ and $\nu \in \{10, 50\}$. R^2 statistics are reported for the approximate filter, the approximate smoother, and two-stage PCA (2SPCA) as a point of comparison.

T	ν	$\hat{f}^{\text{OENI}}_{t\|t}$	$\hat{f}^{\text{OENI}}_{t\|T}$	$\hat{f}^{2SPCA}$	$\hat{h}^{\text{OENI}}_{t\|t}$	$\hat{h}^{\text{OENI}}_{t\|T}$	$\hat{h}^{2SPCA}$
250	10	0.855	0.853	0.849	0.336	0.375	0.210
500	10	0.860	0.861	0.850	0.373	0.416	0.216
250	50	0.871	0.872	0.870	0.330	0.354	0.211
500	50	0.877	0.880	0.871	0.356	0.387	0.222

red line. The lines are visibly close for both the location factor loading in Panel [1,1], the scale factor loading in Panel [1,2], and all elements of the state transition matrix $\mathbf{\Phi}$. The sampling distributions of the on-diagonal elements of $\mathbf{\Sigma}$ are somewhat skewed and resemble those of (positive) standard deviation parameters. Parameters $\mathbf{\Phi}_{3,1}$ and δ_h are estimated to the right and left of their respective true values, and thus exhibit some finite-sample bias.

Parameters $\mathbf{\Phi}_{3,1}$ and $\mathbf{\Sigma}_{1,3}$ are associated with (present) interactions between f_{1t} and h_t and are estimated close to their true non-zero values. Parameters $\mathbf{\Phi}_{3,2}$ and $\mathbf{\Sigma}_{2,3}$ are associated with (absent) interactions between f_{2t} and h_t and are also estimated close to their true zero values.

Second, the approximate smoother is more accurate than the approximate filter, and both are always more accurate than two-stage PCA. Table 1 presents the R^2 statistics across four DGPs. R^2 statistics are reported for the approximate filter, the approximate smoother, and two-stage PCA. Figure 2 presents R^2 statistics from a regression of the true factors on (a constant and) the estimated location factors and scale factor, respectively. The R^2 statistics

thus address the question of what share of the true common variation in the data is recovered by the estimated factors. While the improvement of estimation accuracy remains modest for the common location factors, it becomes more pronounced for scale factors. Notably, for $T = 500$ and $\nu = 10$, the accuracy associated with $\hat{h}_{t|T}^{OENI}$ is approximately twice as high as that for $\hat{h}_{t|T}^{2SPCA}$. Despite the impact of outliers, the approximate filter and approximate smoother consistently maintain their superiority.

Finally, our analysis reveals no evidence of deterioration in approximation fit over time. Such deterioration could arise from approximation errors that accumulate in undesirable ways. Figure 3 presents R-squared statistics (computed using (11)) corresponding to the approximating WLS regressions (as described in (10)). We divide the approximation errors into five equally-sized time bins: 1-50, 51-100, 101-150, 151-200, and 201-250 (for a total sample size of $T = 250$). Remarkably, the approximation fit remains remarkably stable, with R^2 statistics hovering around 99.1% in all bins.

Figure 1: Sampling distribution of $\hat{\theta}_{ML}^{(s)}$

Sampling distributions of maximum likelihood parameter estimates $\hat{\theta}_{ML}^{(s)}$, $s = 1, \ldots, 100$. The simulation settings are $N = 8$, $T = 250$, and $\nu = 50$.

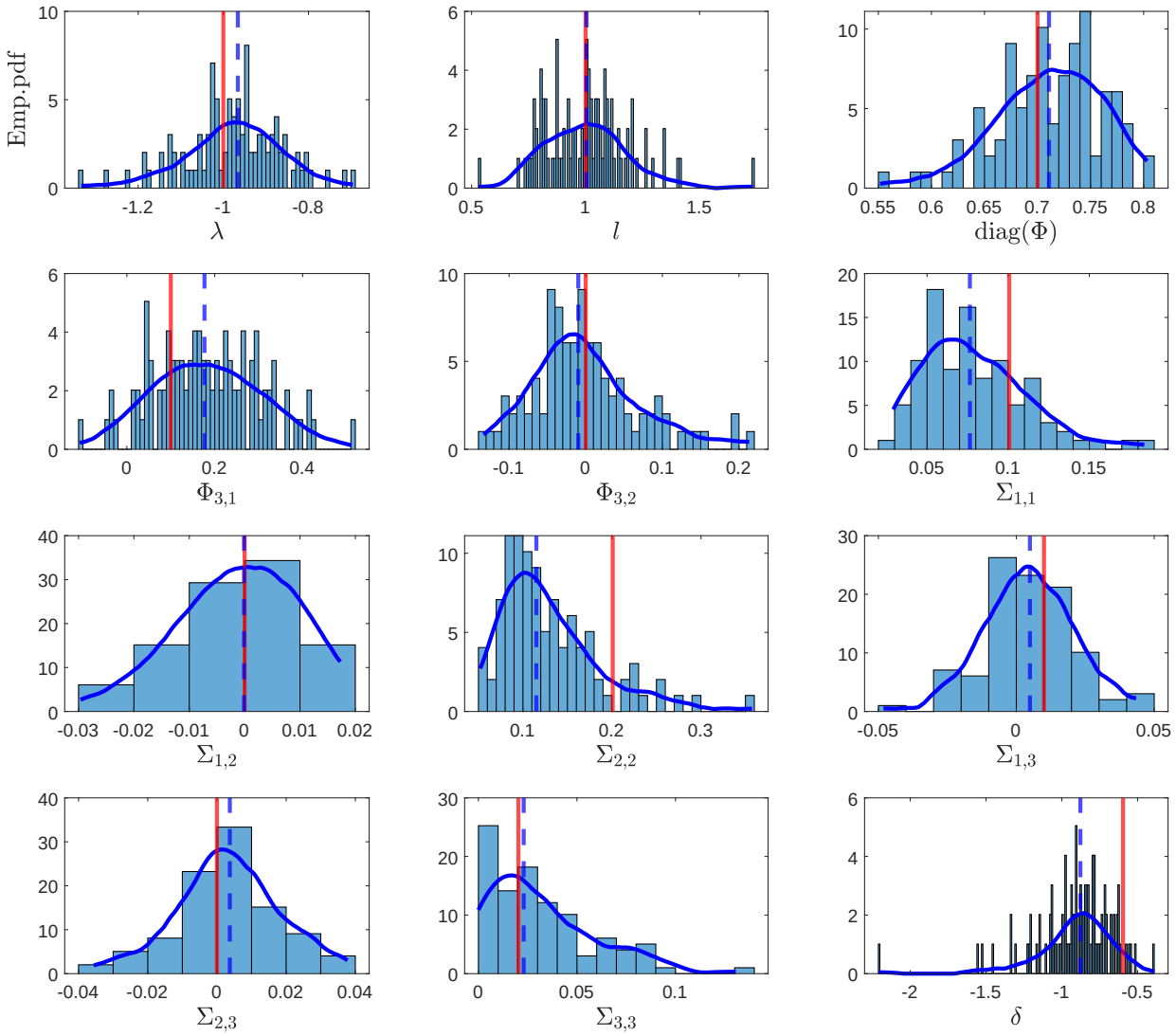


Figure 2: Filtering accuracy

R-squared statistics from a regression of the true on the estimated location factors (top panel) and scale factors (bottom panel). The panels in each row refer to the approximate filter (left), the approximate smoother (middle), and two-stage PCA (right). The simulation settings are $N = 8$, $T = 250$, $\nu = 50$, $S = 100$.

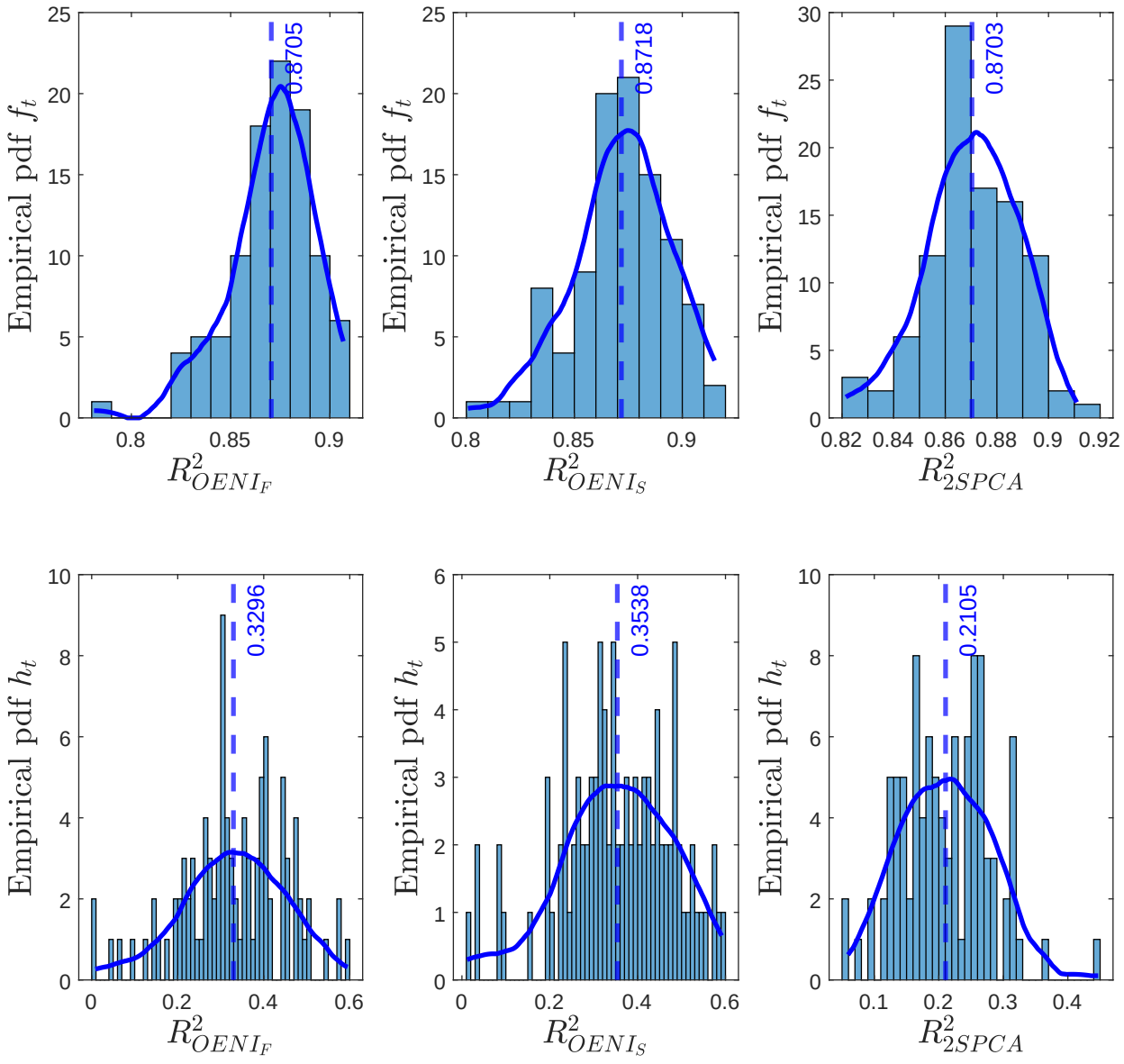
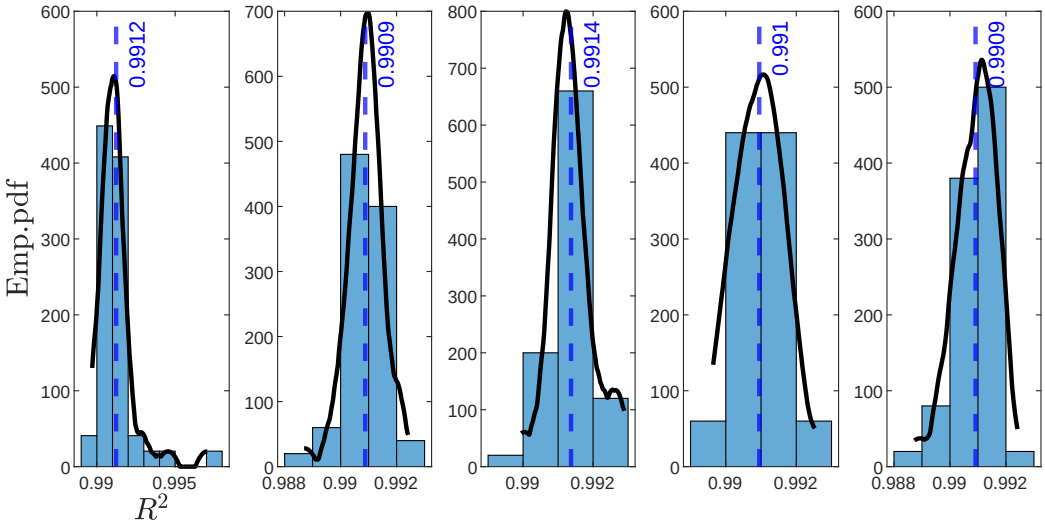


Figure 3: Approximation quality over time

R-squared statistics (11) corresponding to the approximating (weighted least squares) regressions (10) over time. The approximation errors are divided horizontally into five equally-sized time-bins, e.g. 1-50, 51-100, 101-150, 151-200, and 201-250 for $T = 250$. The simulation settings are $N = 8$, $T = 250$, $\nu = 50$, $S = 100$.



4 Empirical illustration: Real GDP growth rates

4.1 Data

To illustrate our approach, we obtain end-of-quarter real GDP growth rates for G7 countries: United States (USA), Japan (JPN), Germany (DEU), Great Britatin (GBR), France (FRA), Italy (ITA), and Canada (CAN). The data range from 1961Q3 to 2022Q4, yielding $N = 7$ and $T = 243$. The countries are ordered in y_{it} according to end-of-sample real GDP. All time series are easily and publicly available from the OECD website. We consider non-annualized log-changes $y_{it} = 100 \times (\ln Y_{it} - \ln Y_{i,t-1})$, where Y_{it} is country i's real GDP, before applying our methodology.

Figure A.1 in Web Appendix A provides a time series plot. For all countries, real GDP growth takes particularly low values during the 1970s owing to global oil price shocks, during the global financial crisis between 2008 and 2009, and during the Covid-19 pandemic recession in 2020.

4.2 Model specification

We choose $r = 2$ location factors and $q = 1$ scale factor for our data, following a preliminary principal components analysis and a study of Ahn and Horenstein (2013)'s information criteria.

The loadings matrices Λ and l need to be restricted to ensure factor identification. To identify f_{1t}, we set $\Lambda_{USA,1} = 1$ and leave the other parameters in $\Lambda_{*,1}$ unrestricted. To identify f_{2t}, we set $\Lambda_{DEU,2} = \Lambda_{FRA,2} = \Lambda_{ITA,2} = 1$ along with $\Lambda_{USA,2} = \Lambda_{JAP,2} = \Lambda_{CAN,2} = 0$ and leave $\Lambda_{GBR,2}$ unrestricted. This strategic placement of 1s and 0s in Λ implies that f_{1t} can be interpreted economically as a global level factor, and f_{2t} as an additional European (group) factor.

For the scale factor loading matrix l, we restrict $l_{USA} = 1$ and leave the other loadings unrestricted. This restriction identifies h_t as the first country's idiosyncratic term's stochastic

scale. The remaining elements are elasticities with respect to that baseline.

Finally, we model demeaned data $\tilde{y}_{it} = y_{it} - \bar{y}_{i,\cdot}$, such that $c_i = 0$. The state vector's unconditional mean δ is subsequently restricted such that $\mathbb{E}[f_t] = \delta_f = 0$. The scale factor's unconditional mean $\mathbb{E}[h_t] = \delta_h$ remains unrestricted.

4.3 Parameter and factor estimates

This section discusses our empirical parameter and factor estimates.

Table 2 presents maximum likelihood estimates of the deterministic parameters of our statistical model (1) – (3). Confidence intervals for each parameter are reported a 95% significance level.

Two findings deserve comment. First, estimates of the state equation's transition matrix $\mathbf{\Phi}$ and scale matrix $\mathbf{\Sigma}_\eta$ point to significant factor interactions, both at a lag (via a full matrix $\mathbf{\Phi}$) and contemporaneously (via a full $\mathbf{\Sigma}_\eta$). Some off-diagonal elements in $\mathbf{\Phi}$ are significantly different from zero, with coefficient $\mathbf{\Phi}_{3,2}$ particularly large. The estimate of $\mathbf{\Sigma}_\eta$, and its implied correlation matrix $\boldsymbol{R}_\eta$, suggests a negative contemporaneous correlation between both location factors and the scale factor. The implied correlation coefficients are $R_{\eta,3,1} = R_{\eta,3,2} \approx -0.2$.

Second, the data are moderately fat-tailed at $\nu \approx 10$. This is the case even though the model already features stochastic volatility. This is intuitive, however, since the Covid-19 pandemic recession with its associated extreme observations is part of the estimation sample.

Figure 4 reports filtered and full-sample estimates of f_t and h_t. Shaded areas indicate U.S. recession dates for f_{1t}, euro area recession dates for f_{2t}, and both types of recession dates for h_t. Both location factors take negative values during recessions. This is as expected, and intuitive. The scale factor h_t is counter-cyclical and takes particularly high values during the 2008 global financial crisis and the 2020 Covid-19 pandemic recession. Both location and scale factors are fairly precisely estimated, and the filtered and full-sample estimates are relatively close.

4.4 A closer look at factor interactions

This section discusses likelihood-ratio (LR) hypothesis tests to study whether location and scale factors are partially or fully independent. In addition, it discusses impulse response function estimates to shed light on shock propagation and persistence.

Table 3 summarizes our LR test outcomes. It distinguishes four models: First, a full model (F) for which $\mathbf{\Phi}$ and $\mathbf{\Sigma}$ are unrestricted. Second, a restricted model (R1) that does not allow for lagged dependence. Here, $\mathbf{\Sigma}$ is unrestricted but $\mathbf{\Phi}_{3,1} = \mathbf{\Phi}_{3,2} = \mathbf{\Phi}_{1,2} = \mathbf{\Phi}_{1,3} \equiv 0$. Third, a restricted model (R2) that does not allow for any contemporaneous dependence. Here, $\mathbf{\Phi}$ is unrestricted but $\mathbf{\Sigma}_{3,1}^{1/2} = \mathbf{\Sigma}_{3,2}^{1/2} \equiv 0$. Finally, a block-diagonal fully restricted model (R3), in which the location and scale factors interact neither contemporaneously nor with a lag.

The message from Table 3 is clear. All p-values are below 0.01. The RGDP data unequivocally prefer an unrestricted VAR for all factors.

Figure 5 plots impulse response function estimates for factors f_{1t}, f_{2t}, and h_t to a one standard deviation shock to the global factor f_{1t}. Shock identification is based on Choleski timing restrictions, implying that f_{1t} cannot contemporaneously respond to a shock to f_{2t} and h_t, and f_{2t} cannot contemporaneously respond to a shock to h_t. Factor interactions are evident: A positive shock to f_{1t}, representing a much improved global economic outlook, has a delayed and long-lasting positive effect on the four European countries via f_{2t}, and significantly reduces macroeconomic uncertainty (i.e., the scale of each country's idiosyncratic component) over the following eight quarters.

Table 2: Parameter estimates

Maximum likelihood parameter estimates for the nonlinear dynamic factor model. G7 countries' real GDP growth rates are considered between 1961Q3 to 2022Q4. Countries are ordered according to size, as United States, Japan, Germany, United Kingdom, France, Italy, and Canada. Confidence intervals are reported a 95% significance level.

Par.	estimates	Par.	estimates

$$\nu = \begin{bmatrix} 9.70 \\ [6.06,17.62] \end{bmatrix}$$

$$\Lambda = \begin{bmatrix} 1 & 0 \\ [-] & [-] \\ 2.00 & 0 \\ [1.38,2.59] & [-] \\ -0.03 & 1 \\ [-1.09,1.08] & [-] \\ 0.98 & -0.27 \\ [0.48,1.44] & [-0.51,-0.04] \\ -0.12 & 1 \\ [-1.14,0.97] & [-] \\ 0.27 & 1 \\ [-0.77,1.37] & [-] \\ 1.20 & 0 \\ [0.84,1.56] & [-] \end{bmatrix} \qquad l = \begin{bmatrix} 1.00 \\ [-] \\ 0.56 \\ [0.31,0.79] \\ 0.75 \\ [0.51,0.97] \\ 1.43 \\ [1.01,1.86] \\ 1.91 \\ [1.48,2.35] \\ 1.54 \\ [1.16,1.91] \\ 1.12 \\ [0.82,1.45] \end{bmatrix}$$

$$\delta = \begin{bmatrix} 0 \\ [-] \\ 0 \\ [-] \\ -2.26 \\ [-3.09,-1.39] \end{bmatrix} \qquad \Phi = \begin{bmatrix} 0.88 & -0.06 & 0.01 \\ [0.71,1.04] & [-0.17,0.06] & [-0.01,0.03] \\ 0.43 & 0.57 & 0.01 \\ [0.07,0.79] & [0.47,0.67] & [-0.02,0.03] \\ -0.71 & 0.78 & 0.88 \\ [-1.64,0.26] & [0.43,1.14] & [0.81,0.94] \end{bmatrix}$$

$$\Sigma_\eta = \begin{bmatrix} 0.04 & 0.00 & -0.02 \\ [0.02,0.07] & [-0.04,0.05] & [-0.07,0.02] \\ 0.00 & 0.03 & -0.02 \\ [-0.04,0.05] & [0.02,0.09] & [-0.06,0.01] \\ -0.02 & -0.02 & 0.37 \\ [-0.07,0.02] & [-0.06,0.01] & [0.22,0.64] \end{bmatrix} \qquad R_\eta = \begin{bmatrix} 1 & 0.07 & -0.19 \\ [-] & [-0.81,0.82] & [-0.47,0.13] \\ 0.07 & 1 & -0.23 \\ [-0.81,0.82] & [-] & [-0.43,0.10] \\ -0.19 & -0.23 & 1 \\ [-0.47,0.13] & [-0.43,0.10] & [-] \end{bmatrix}$$

Figure 4: Filtered & full-sample common factor estimates

Filtered and full-sample (smoothed) estimates of two common location factors f_{1t} (top panel), f_{2t} (middle panel), and one common scale factor h_t (bottom panel). Shaded areas indicate either U.S. recessions according to the NBER (top panel), euro area recessions according to the euro area business cycle dating committee (middle panel), or both (bottom panel). The estimation sample ranges from 1961Q3 to 2022Q4.

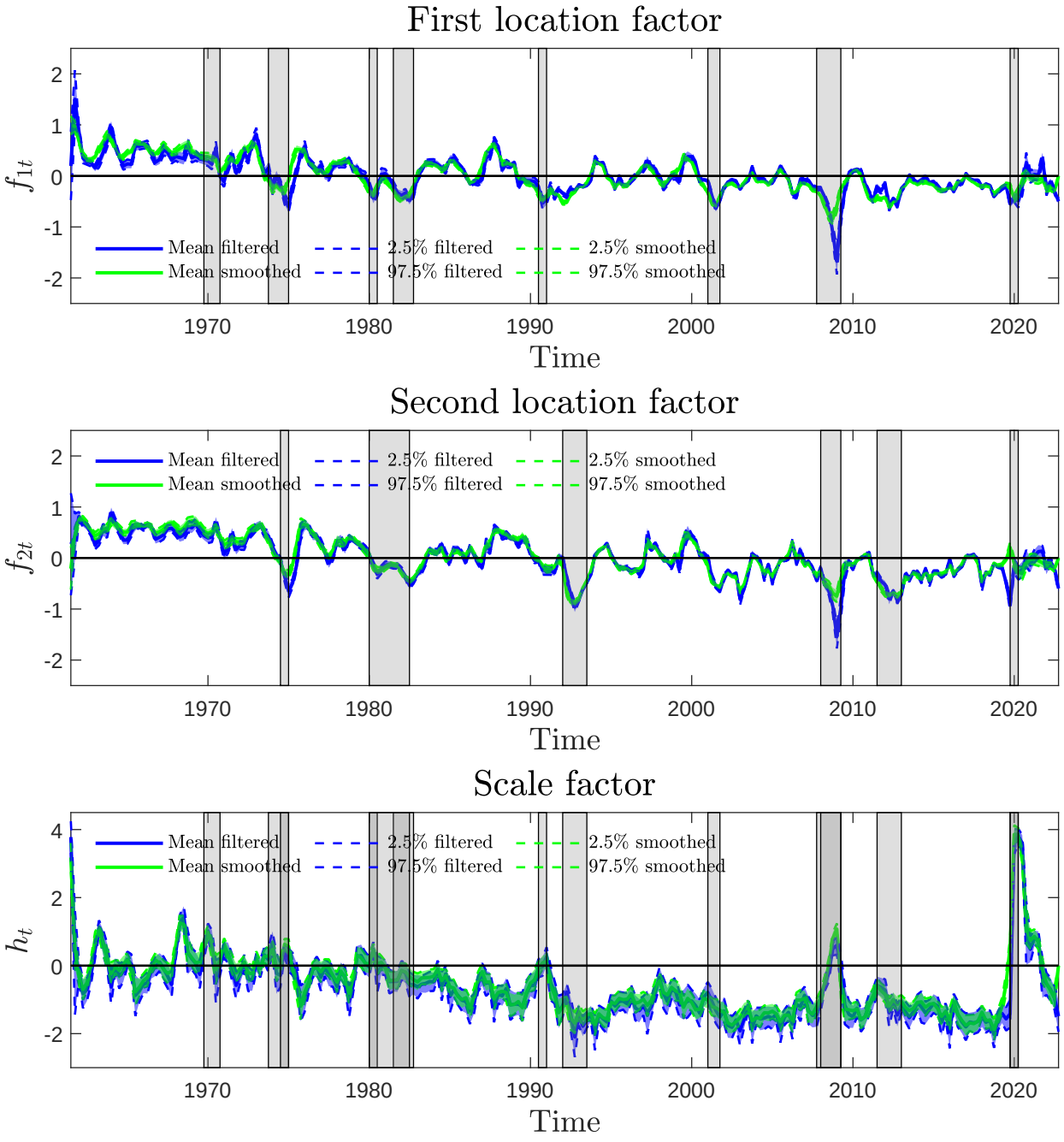


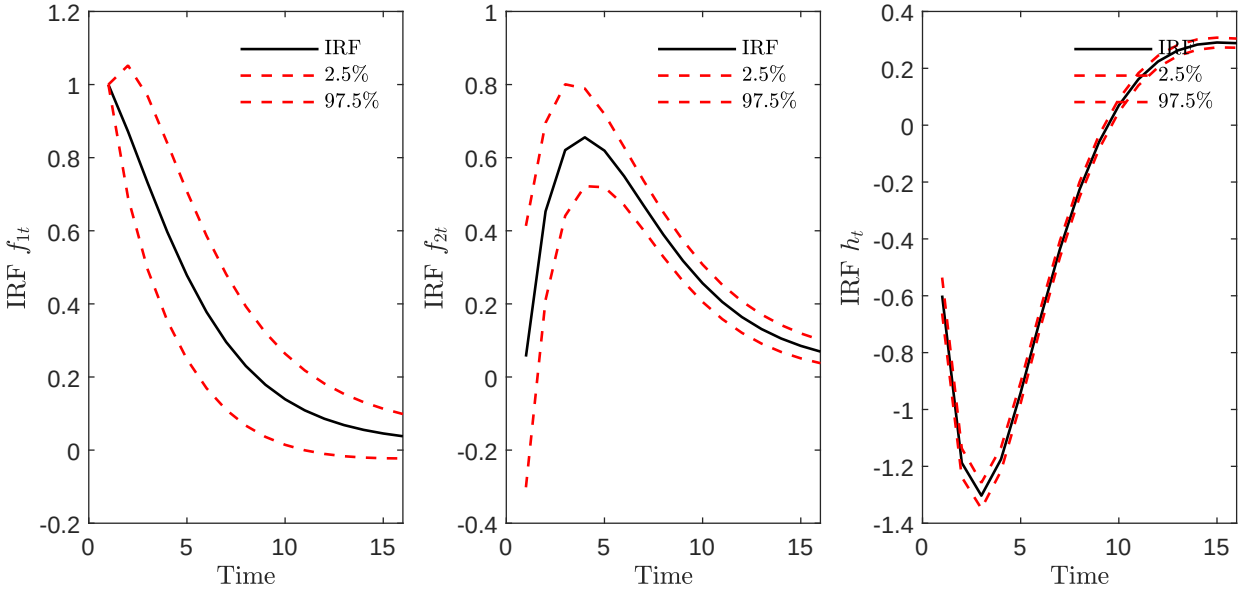
Table 3: LR-tests outcomes

Maximal log-likelihood values and LR test outcomes are reported for four model specifications. First row: a full model (F), where $\mathbf{\Phi}$ and $\mathbf{\Sigma}$ are unrestricted. Second row: a restricted model (R1), where $\mathbf{\Phi}_{3,1} = \mathbf{\Phi}_{3,2} = \mathbf{\Phi}_{1,2} = \mathbf{\Phi}_{1,3} = 0$. Third row: a restricted model (R2), where $\mathbf{\Sigma}_{3,1}^{1/2} = \mathbf{\Sigma}_{3,2}^{1/2} = 0$. Fourth row: a fully restricted model (R3), where the location and scale factors are fully independent.

Model	Restriction on	maxLL	LR stat	p-val
F	-	-9119		
R1	$\mathbf{\Phi}$	-9139	39.46	0.00
R2	$\mathbf{\Sigma}$	-9129	19.22	0.00
R3	$\mathbf{\Phi}, \mathbf{\Sigma}$	-9140	41.80	0.00

Figure 5: Impulse response functions

Impulse response of f_{1t} (left), f_{2t} (middle), and h_t (right) to a one standard deviation shock to the global location factor f_{1t}. Shock identification is based on Choleski timing restrictions. Standard error bands are bootstrapped.



4.5 Approximation quality

This section assesses the approximation quality underlying the approximate filter by studying the R^2 statistics (11). When these R^2s are close to one, then $\ln p(\alpha_t|Y_t;\theta)$ and $\ln \hat{p}(\alpha_t|Y_t;\theta)$ are virtually indistinguishable.

Table 4 reports descriptive statistics for the R^2s (11) and suggests a tight approximation. Typical (median) R^2 statistics are 0.91 for France at the low end and 0.997 for Germany at the top end. The R^2s are negatively skewed, as they are bounded from above (by one), but rarely drop below 0.6 (the 10th percentile for France). We conclude that, overall, the error introduced by replacing $p(\alpha_t|Y_t;\theta)$ by $\hat{p}(\alpha_t|Y_t;\theta)$ is modest for our empirical data at hand.

Figure 6 provides a density plot of the approximating models' residuals u_{it}^{lk} (left panel) and a Hill plot to study their extreme tail behavior (right panel). The approximating residuals are centered around zero by construction and appear approximately symmetric. The Hill plot suggests that first two, possibly three, moments of u_{it}^{lk} exist, and that a law of large numbers and a central limit theorem can be applied to these residuals as a result.

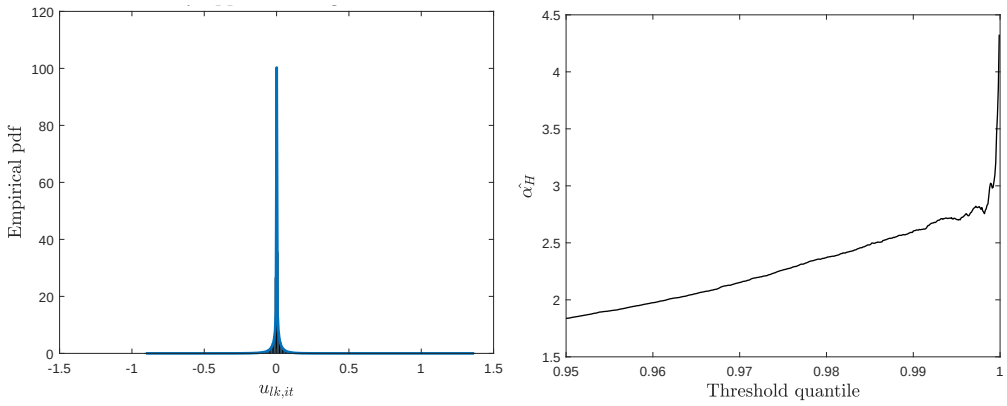
Table 4: Approximation quality

Descriptive information on R^2 statistics (11) from the approximating WLS regressions (10) evaluated at $\hat{\theta}_{ML}$. There are T R^2 statistics per country, and $N \times T$ R^2 statistics in total. P10 and P90 refer to the 10% and 90% percentile.

	Mean	Std	Median	Min	Max	P10	P90
USA	0.954	0.064	0.984	0.708	1.000	0.865	0.999
JPN	0.984	0.019	0.992	0.857	1.000	0.957	1.000
DEU	0.990	0.021	0.997	0.802	1.000	0.977	1.000
GBR	0.888	0.138	0.949	0.495	1.000	0.625	0.999
FRA	0.855	0.151	0.909	0.403	1.000	0.600	0.993
ITA	0.889	0.144	0.958	0.447	1.000	0.633	0.998
CAN	0.923	0.100	0.969	0.587	1.000	0.775	0.999

Figure 6: Approximation errors visual diagnostics

Left panel: Kernel density of approximating models' residuals u_{it}^{lk}. Right panel: Hill plot for u_{it}^{lk}.



4.6 Out-of-sample forecasting

This section studies the out-of-sample forecasting performance of the nonlinear non-Gaussian dynamic factor model (1) – (2). The model's deterministic parameters are estimated by QML using (9), and the factors are forecast using the approximate filter and treating the future data as missing.

We consider both point and density forecasting accuracy, and relate the dynamic factor model's forecasts to three points of comparison. First, we consider a restricted model with fully independent factors, shutting off any factor interactions. Both the full and restricted models are estimated for $r = q = 1$ for simplicity and to focus on the strongest factor for each moment. Second, we consider Student's t density-based "static" forecasts. Here, time-invariant country-specific location, scale, and degrees-of-freedom parameters are fitted to the estimation sample via maximum likelihood. Finally, we consider AR(1)-GARCH(1,1)-based forecasts, where parameters are estimated by maximum likelihood to fit each country's univariate time series.

Table 5 reports the forecasting outcomes. For each model, the deterministic parameters are estimated once on data between 1961Q3 and 2010Q4 and are not re-estimated as the forecasting origin moves towards the end of the sample 2022Q4. As a result, the estimation sample contains the volatile 1970s and the global financial crisis, but excludes the euro area sovereign debt crisis between 2011 and 2012 and the Covid-19 pandemic in 2020. We consider 1, 1–2, 1–4, 5–8, and 9–12 quarters ahead forecasts. Point forecasts are evaluated based on root mean squared forecasting error (RMSFE) statistics; density forecasts are evaluated based on log score and censored likelihood (CSL) score statistics. The CSL is computed to focus on the predictive densities' fit in the left tail, below the 0.25 empirical quantile associated with each country's data.

Table 5: Out-of-sample forecasting accuracy

We consider both point (RMSFE) and density (log score, CSL score) forecasting accuracy. Forecasting horizons range between 1 and 12 quarters ahead. The estimation sample ranges from 1961Q3 to 2010Q4. The evaluation sample ranges from 2011Q1 to 2022Q4. The forecasting models are explained in the main text.

Model		RMSFE					Log scores					CSL scores				
	h	1	1-2	1-4	5-8	9-12	1	1-2	1-4	5-8	9-12	1	1-2	1-4	5-8	9-12
Full	USA	0.44	1.03	1.52	2.01	2.02	-0.65	-0.93	-1.24	-1.57	-1.60	-0.42	-0.60	-0.67	-0.67	-0.66
	JPN	1.08	1.47	1.75	2.01	2.00	-1.55	-1.73	-1.87	-1.95	-1.93	-0.99	-1.14	-1.18	-1.19	-1.17
	DEU	0.61	1.22	1.75	2.30	2.29	-1.07	-1.25	-1.48	-1.77	-1.75	-0.38	-0.57	-0.65	-0.77	-0.69
	GBR	0.57	2.28	3.49	4.78	4.80	-0.61	-1.00	-1.45	-2.11	-2.17	-0.35	-0.65	-0.77	-0.86	-0.82
	FRA	1.08	1.91	2.88	3.86	3.87	-1.04	-1.35	-1.74	-2.15	-2.16	-0.99	-1.20	-1.27	-1.26	-1.19
	ITA	1.16	1.85	2.59	3.32	3.31	-1.35	-1.65	-1.98	-2.22	-2.13	-1.05	-1.26	-1.31	-1.16	-0.94
	CAN	0.61	1.36	1.94	2.53	2.53	-0.87	-1.16	-1.48	-1.81	-1.80	-0.58	-0.80	-0.88	-0.89	-0.81
	G7	0.79	1.59	2.27	2.97	2.97	-1.02	-1.29	-1.61	-1.94	-1.93	-0.68	-0.89	-0.96	-0.97	-0.90
Restr.	USA	0.45	1.04	1.53	2.01	2.02	-0.65	-0.98	-1.31	-1.65	-1.68	-0.48	-0.64	-0.70	-0.71	-0.72
	JPN	1.11	1.49	1.76	2.01	2.01	-1.59	-1.77	-1.92	-2.01	-2.00	-1.03	-1.19	-1.23	-1.24	-1.23
	DEU	0.61	1.22	1.75	2.30	2.29	-1.03	-1.26	-1.53	-1.86	-1.84	-0.40	-0.60	-0.69	-0.82	-0.75
	GBR	0.58	2.29	3.49	4.77	4.80	-0.55	-1.05	-1.57	-2.32	-2.38	-0.41	-0.71	-0.84	-0.97	-0.94
	FRA	1.09	1.92	2.89	3.86	3.87	-1.17	-1.47	-1.88	-2.33	-2.36	-1.17	-1.31	-1.35	-1.35	-1.31
	ITA	1.19	1.87	2.60	3.32	3.31	-1.50	-1.76	-2.10	-2.37	-2.30	-1.26	-1.38	-1.39	-1.25	-1.05
	CAN	0.61	1.37	1.95	2.53	2.53	-0.90	-1.23	-1.56	-1.91	-1.91	-0.65	-0.85	-0.91	-0.95	-0.89
	G7	0.80	1.60	2.28	2.97	2.98	-1.06	-1.36	-1.70	-2.06	-2.07	-0.77	-0.95	-1.02	-1.04	-0.98
Static	USA	0.52	1.09	1.54	2.00	2.02	-0.79	-0.89	-1.02	-1.20	-1.27	-0.44	-0.54	-0.57	-0.58	-0.61
	JPN	1.11	1.47	1.72	1.95	1.95	-1.51	-1.59	-1.67	-1.73	-1.72	-0.89	-0.97	-0.99	-1.00	-1.00
	DEU	0.67	1.26	1.76	2.30	2.29	-1.10	-1.19	-1.33	-1.55	-1.52	-0.41	-0.53	-0.59	-0.70	-0.63
	GBR	0.61	2.31	3.50	4.77	4.80	-0.69	-0.82	-0.99	-1.37	-1.43	-0.33	-0.47	-0.52	-0.61	-0.59
	FRA	1.13	1.93	2.88	3.85	3.86	-0.92	-1.03	-1.23	-1.50	-1.53	-0.74	-0.87	-0.93	-1.01	-0.97
	ITA	1.26	1.89	2.60	3.30	3.29	-1.26	-1.38	-1.55	-1.71	-1.65	-0.82	-0.95	-1.00	-0.92	-0.73
	CAN	0.69	1.42	1.96	2.52	2.52	-0.97	-1.10	-1.27	-1.51	-1.52	-0.54	-0.67	-0.73	-0.79	-0.74
	G7	0.86	1.63	2.28	2.96	2.96	-1.03	-1.14	-1.30	-1.51	-1.52	-0.60	-0.71	-0.76	-0.80	-0.75
AR(1)-GARCH(1,1)	USA	0.58	0.61	0.64	0.73	0.78	-1.34	-1.43	-1.64	-1.89	-1.92	-1.21	-1.30	-1.28	-1.15	-1.08
	JPN	0.74	0.78	0.85	1.03	1.14	-1.98	-2.11	-2.27	-2.27	-2.11	-1.55	-1.68	-1.61	-1.34	-1.21
	DEU	0.93	1.01	1.07	1.15	1.15	-1.10	-1.36	-1.72	-2.14	-2.14	-0.40	-0.60	-0.70	-0.82	-0.76
	GBR	0.55	0.57	0.59	0.63	0.66	-1.00	-1.18	-1.48	-2.78	-3.18	-1.02	-1.20	-1.24	-1.57	-1.57
	FRA	0.58	0.67	0.81	1.24	1.55	-1.37	-1.44	-1.79	-2.47	-2.66	-1.38	-1.36	-1.37	-1.39	-1.31
	ITA	0.69	0.79	0.92	1.18	1.22	-2.06	-2.04	-2.24	-2.47	-2.47	-1.74	-1.68	-1.57	-1.32	-1.15
	CAN	0.57	0.61	0.66	0.77	0.84	-1.46	-1.57	-1.80	-2.20	-2.16	-1.28	-1.40	-1.38	-1.36	-1.23
	G7	0.66	0.72	0.79	0.96	1.05	-1.47	-1.59	-1.85	-2.32	-2.38	-1.23	-1.32	-1.31	-1.28	-1.19

We highlight three findings. First, suppressing factor interactions is costly, not only in terms of in-sample fit, see Section 4.4, but also in terms of out-of-sample point and density forecasting accuracy. The full model's RMSFE statistics are typically lower than those of the restricted model, particularly at shorter horizons of up to one year. The full model's log scores and CSL scores are uniformly better as well than the those of the restrictive model.

Second, the full model achieves more accurate point forecasts than the static model for forecasting horizons up to one year ahead. The static model's density forecasts are competitive or better, possibly because of country-specific degrees-of-freedom parameters that are not present in the factor model. Low degrees-of-freedom do well given some of the extreme observations encountered in the evaluation sample.

Finally, the fully dynamic factor model performs worse than AR(1)-GARCH(1,1)-based forecasts in terms of point forecasting accuracy (RMSFE), but performs considerably better in terms of density forecasting accuracy (log scores and CSL scores). The latter's Gaussian conditional density is ill-suited to describe some of the evaluation sample's extreme observations.

5 Concluding discussion

We introduced a novel nonlinear non-Gaussian dynamic factor model that allows location and scale factors to interact freely within an unrestricted VAR. By relaxing the assumption of no interactions between these two types of factors, we gain a more flexible and possibly much more realistic representation of the underlying DGP. We provided approximate filtering and smoothing recursions to analyze the statistical model, allowing us to simultaneously estimate all latent factors along with all the model's deterministic parameters in a single step. An empirical illustration using G7 countries' real GDP growth rates illustrates the utility of our approach for in-sample analysis and out-of-sample forecasting. By relying on our new recursions, we extracted meaningful information from noisy observations and enhanced our

understanding of the system's nonlinear dynamics.

The paper presents several possible routes for future research. First, the approximate filter could be adapted to study an even more general class of nonlinear non-Gaussian state space models, and its empirical performance in such cases would be of interest. For example, the observation density $p(y_{it}|f_t, h_t, \theta)$ could denote a (non-differentiable) Asymmetric Laplace density, allowing researchers to model the time-varying quantiles associated with panels of real GDP growth rates, inflation rates, or financial asset returns. Second, the set of diagnostic checks could be extended. For example, new diagnostic procedures could include statistical tests of the null hypothesis that the approximating model is in fact exact. Finally, the framework could be extended to deal with a high-dimensional cross section, allowing our framework to be applied in various financial asset pricing settings.

References

Adrian, T., N. Boyarchenko, and D. Giannone (2019). Vulnerable growth. *American Economic Review 109*(4), 1263–89.

Adrian, T., F. Grinberg, N. Liang, and S. Malik (2022). The term structure of growth-at-risk. *American Economic Journal: Macroeconomics 14*(3), 283–323.

Ahn, S. C. and A. R. Horenstein (2013). Eigenvalue Ratio Test for the Number of Factors. *Econometrica 80*, 1203–1227.

Bai, J. and K. Li (2012). Statistical Analysis of Factor Models of High Dimension. *The Annals of Statistics 40*, 436–465.

Bai, J. and K. Li (2016). Maximum Likelihood Estimation and Inference for Approximate Factor Models of High Dimension. *The Review of Economics and Statistics 98*, 298–309.

Bai, J. and S. Ng (2002). Determining the Number of Factors in Approximate Factor Models. *Econometrica 70*, 191–221.

Banbura, M. and M. Modugno (2014). Maximum Likelihood Estimation of Factor Models on Datasets with Arbitrary Pattern of Missing Data. *Journal of Applied Econometrics 29*, 133–160.

Barigozzi, M. and M. Hallin (2016). Generalized dynamic factors models and volatilities: recovering the market volatility shocks. *Econometrics Journal 19*, 33–60.

Barigozzi, M. and M. Hallin (2017). Generalized dynamic factors models and volatilities estimation and forecasting. *Journal of Econometrics 201*, 307–321.

Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. New York: Springer.

Blei, D. M., A. Kucukelbir, and J. D. McAuliffe (2017). Variational inference: A review for statisticians. *Journal of the American Statistical Association 112*(518), 859–877.

Caldara, D., C. Scotti, and M. Zhong (2021). Macroeconomic and financial risks: A tale of volatility. *Federal Reserve Board International Finance discussion papers*.

Chib, S., F. Nardari, and N. Shephard (2006). Analysis of high dimensional multivariate stochastic volatility models. *Journal of Econometrics 134*, 341–371.

Doz, C., D. Giannone, and C. Reichlin (2012). A Quasi Maximum Likelihood Approach for Large Approximate Dynamic Factor Models. *Review of Economics and Statistics 94*, 1014–1024.

Durbin, J. and S. J. Koopman (2012). *Time Series Analysis by State Space Methods*. Oxford: Oxford University Press.

Gorodnichenko, Y. and S. Ng (2017). Level and volatility factors in macroeconomic data. *Journal of Monetary Economics*. forthcoming.

Gourieroux, C. and A. Monfort (2002). *Statistics and Econometric Models*. Cambridge: Cambridge University Press.

Jungbacker, B. and S. J. Koopman (2014). Likelihood-based Analysis for Dynamic Factor Analysis for Measurement and Forecasting. *Econometrics Journal 17*, forthcoming.

Jurado, K., S. C. Ludvigson, and S. Ng (2015). Measuring uncertainty. *American Economic Review 105*, 1117–1216.

Kastner, G., S. Fruhwirth-Schnatter, and H. F. Lopes (2017). Efficient bayesian inference for multivariate factor stochastic volatility models. *Journal of Computational and Graphical Statistics 0*(0), 1–13.

Koopman, S. J. and C. S. Bos (2004). State space models with a common stochastic variance. *Journal of Business and Economic Statistics 22*, 346–357.

Koopman, S. J., R. Lit, and A. Lucas (2017). Intraday stochastic volatility in discrete price changes: The dynamic skellam model. *Journal of the American Statistical Association 112*(520), 1490–1503.

Koopman, S. J., A. Lucas, and M. Scharth (2014). Numerically Accelerated Importance Sampling for Nonlinear Non-Gaussian State Space Models. *Journal of Business and Economic Statistics 33*, 114–127.

Marin, J.-M., P. Pudlo, C. P. Robert, and R. J. Ryder (2012, Nov). Approximate bayesian computational methods. *Statistics and Computing 22*(6), 1167–1180.

Minka, T. (2001). A family of algorithms for approximate bayesian inference. PhD Dissertation. Massachusetts Institute of Technology, Cambridge.

Monahan, J. F. (2001). *Numerical Methods of Statistics*. Cambridge: Cambridge University Press.

Richard, J. F. and W. Zhang (2007). Efficient High-Dimensional Importance Sampling. *Journal of Econometrics 141*, 1385–1411.

Rue, H., S. Martino, and N. Chopin (2009). Approximate bayesian inference for latent gaussian models by using integrated nested laplace approximations. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 71*(2), 319–392.

Stock, J. H. and M. W. Watson (2002). Forecasting Using Principal Components From a Large Number of Predictors. *Journal of the American Statistical Association 97*, 1167–1179.

Stock, J. H. and M. W. Watson (2016). Factor Models and Structural Vector Autoregressions in Macroeconomics. In J. B. Taylor and H. Uhlig (Eds.), *Handbook of Macroeconomics*. Amsterdam, North Holland: Elsevier.

Web Appendix to

"Nonlinear non-Gaussian dynamic factor models with interacting location and scale factors"

Web Appendix A: Data plots and eigenvalue analysis

Figure A.1 plots G7 countries' non-annualized real GDP growth rates at a quarterly frequency. Countries are ordered according to size (end-of-sample GDP): USA, Japan, Germany, Great Britain, France, Italy, and Canada.

Figure A.2 reports PCA eigenvalues for the conditional mean factors. We choose $r = 2$, capturing approximately 72% of the panel's total variance.

Figure A.1: G7 countries' real GDP growth rates

G7 countries' non-annualized real GDP growth rates at a quarterly frequency. Countries are ordered according to size (end-of-sample GDP): USA, Japan, Germany, Great Britain, France, Italy, and Canada. The sample ranges from 1961Q3 to 2022Q4. The top panel's vertical axis is capped at +/- 8 percent to improve visibility.

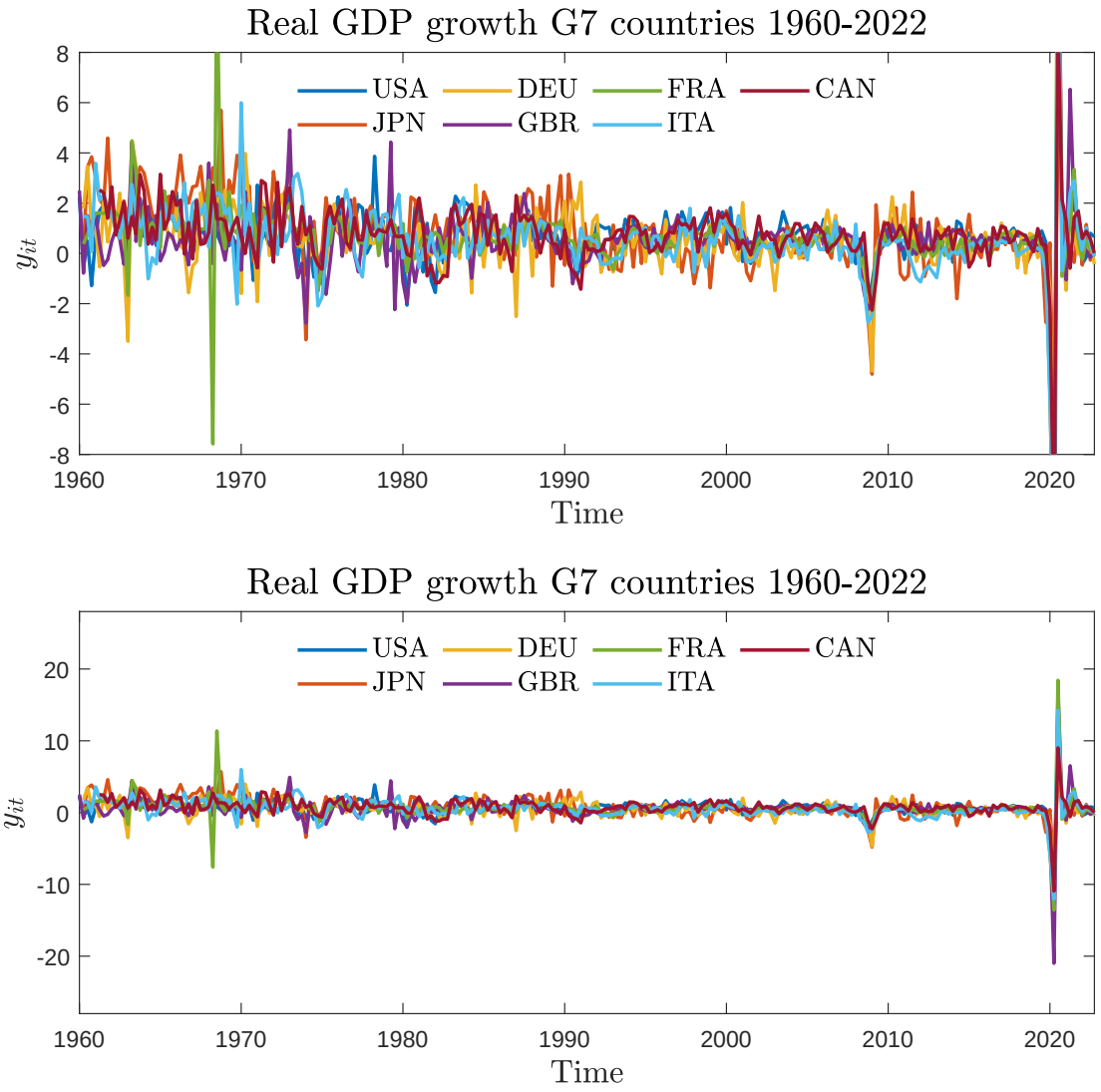


Figure A.2: Eigenvalues

Left panel: PCA eigenvalues (scree plot) for conditional mean factors. Right panel: cumulative share of total variance explained as the number of factors increases.