Identification and estimation of dynamic factor models

Jushan Bai and Peng Wang

Columbia University, Hong Kong University of Science and Technology

28. April 2012

Identification and Estimation of Dynamic Factor Models

Jushan Bai[*]and Peng Wang[†]

April, 2012

Abstract

We consider a set of minimal identification conditions for dynamic factor models. These conditions have economic interpretations, and require fewer number of restrictions than when putting in a static-factor form. Under these restrictions, a standard structural vector autoregression (SVAR) with or without measurement errors can be embedded into a dynamic factor model. More generally, we also consider overidentification restrictions to achieve efficiency. General linear restrictions, either in the form of known factor loadings or cross-equation restrictions, are considered. We further consider serially correlated idiosyncratic errors with heterogeneous coefficients. A numerically stable Bayesian algorithm for the dynamic factor model with general parameter restrictions is constructed for estimation and inference. A square-root form of Kalman filter is shown to improve robustness and accuracy when sampling the latent factors. Confidence intervals (bands) for the parameters of interest such as impulse responses are readily computed. Similar identification conditions are also exploited for multi-level factor models, and they allow us to study the spill-over effects of the shocks arising from one group to another.

Key Words: dynamic factor models, multi-level factor models, impulse response function, spill-over effects

[*]Department of Economics, Columbia University, New York, NY, 10027. Email: jb3064@columbia.edu

[†]Department of Economics, The Hong Kong University of Science and Technology, Clear Water Bay, NT, Hong Kong. Email: pwang@ust.hk

1

1 Introduction

Dynamic factor models of high dimension are increasingly used in data rich environments. This is particularly the case in economics and finance where common shocks drive the comovements of economic variables. As a result, economists are increasingly relying on these models for policy analysis. Examples include Bernanke et al. (2005), Stock and Watson (2005), and Kose et al. (2003), among others. This paper considers identification issues for a general class of dynamic factor models and their Bayesian estimation.

Two sources of dynamics are considered. One is that the latent factor is dynamic. An example would be $x_{it} = \lambda_i' f_t + e_{it}$, where f_t is a vector autoregressive process. While f_t is dynamic, the relationship between x_{it} and f_t is static. The other source of dynamics is that f_t affects x_{it} dynamically, for example, $x_{it} = \gamma_i' f_t + \tau_i' f_{t-1} + e_{it}$ so that the lags of f_t directly impacts x_{it}. We shall refer to the first example as a static factor model in spite of f_t being dynamic. The second example can be put into a static factor framework. Let $F_t = (f_t', f_{t-1}')'$ and $\lambda_i' = (\gamma_i', \tau_i')$, then $x_{it} = \lambda_i' F_t + e_{it}$. There exists a large literature for analyzing static factor models. There are drawbacks by treating a dynamic factor model as a static one. For example, the number of static factors is doubled here. To fix the rotational indeterminacy, the number of normalization restrictions would be $(2q)^2 = 4q^2$, quadrupling the number of restrictions that are actually needed for identification.

We propose a set of simple identification schemes for the dynamic factor models. These identification restrictions have economic interpretations. They allow a structural vector autoregression model (SVAR) to be analyzed through the factor analysis framework, while permitting measurement errors. We then consider the Bayesian estimation of dynamic factor models under these identification schemes. The Bayesian method proposed here has a few desirable features. First, we allow cross-equation restrictions as well as sign restrictions on the factor loadings, which is not available in the principal component analysis of factor models as in Stock and Watson (1998) and Bai and Ng (2002). Second, we employ the Jeffreys priors to account for the lack of a priori information about model parameters. Third, we combine the structural restrictions with a square-root form of the Kalman filter to improve the numerical robustness and accuracy when sampling the latent factors. Fourth, to make random draws from the posterior of the factor loadings and the latent factors, we provide a robust algorithm that guarantees the variance of the conditional normal distribution to be numerically positive definite when it should be.

The Bayesian analysis framework proposed in this paper is able to perform estimation and hypothesis testing of large (N, T) dynamic factor models with rich dynamic structures in both the measurement equations and the law of motion for the dynamic factors. We will also consider another source of dynamics, that is, the idiosyncratic errors are serially correlated. We show how this dynamics is naturally handled within the Bayesian framework.

Our method is closely related to the factor-augmented vector autoregression (FAVAR) approach of Bernanke et al. (2005), and the dynamic factor model of Stock and Watson (2005). The model of Bernanke et al. (2005) features a static factor

model for the measurement equation plus a VAR specification for the latent factors and a few observed aggregate macro variables. Stock and Watson (2005) further consider a setup featuring lags of dynamic factors and serially correlated error terms in the measurement equation, which in turn implies a high dimensional VAR with a large number of restrictions on the autoregressive coefficients. Both papers consider a two-step approach which applies either a direct principal components method or an iterated principal components method to estimate the static factor space, and then uses the estimates to obtain information about dynamics of either static factors or dynamic factors. Bernanke et al. (2005) also consider a Bayesian approach utilizing the Gibbs sampling method for state-space models. They impose conjugate priors for model parameters, and derive the posterior under the identifying assumption for conventional static factor models.

Our method differs from that of Bernanke et al. (2005) or Stock and Watson (2005) in several important aspects. Most importantly, we consider a minimal set of identifying restrictions to directly identify the dynamic factors instead of the static factors or the associated factor space. Secondly, different from most existing literature, our identifying restrictions allow the dynamic factors to be cross-correlated, and thus allow the study of impulse response function of one factor to the innovation of another factor. Thirdly, we develop a Bayesian approach to analyzing dynamic factor model with general cross-equation restrictions along with sign restrictions on factor loadings; the Jeffreys priors are considered for the model parameters. Finally, we also consider the Bayesian estimation of serial correlations in the idiosyncratic errors, which is important for economic time series. The outcome of the Bayesian analysis is the posterior distribution of model parameters along with the latent dynamic factors. The impulse responses and the associated confidence bands are natural by-products.

To conduct model comparison so as to determine the number of dynamic factors, we use the partial Bayes factor, in which the first few observations are used to form a proper prior distribution for computing the marginal likelihood of the latter sample. This is in line with O'Hagan (1995)'s fractional Bayes factors or Berger and Pericchi (1996)'s intrinsic Bayes factors. The Bayesian method proposed in this paper can also incorporate proper prior distributions so as to derive the conventional Bayes factors or the marginal likelihood of the data for the purpose of model comparison.

Another related paper is by Del Negro and Otrok (2006), which conducts Bayesian analysis of dynamic factor models with time-varying factor loadings and stochastic volatilities. Their main focus is the time-varying feature of the model. The identifying restrictions assume that factors are mutually independent. The measurement equations are static. Under their identifying assumptions, the factors are separately identified subject to a sign change. Our identification scheme targets the dynamic factors themselves instead of the static factors, and the dynamic factors are identified without independence assumption. The time-varying parameters version of the dynamic factor model is an important model capable of characterizing the evolution of business cycles. To fix idea, we do not explore that direction in this paper.

In the literature, the maximum likelihood estimation of dynamic factor models has been considered by many authors, such as Watson and Engle (1983), Quah and

Sargent (1993), Jungbacker and Koopman (2008), and Doz et al. (2011). The maximization can be implemented through the expectation-maximization (EM) method as in Doz et al. (2011). The EM method iterates between the expectation step and the maximization step until convergence so as to achieve the local maximum of the likelihood function. The generalized least squares estimator for non-dynamic factor models is studied by Breitung (2011) and Choi (2012).

The Bayesian MCMC approach provides a more complete picture regarding the sampling distribution of the objective of interest, the outcome of which is the posterior joint distribution of the model parameters and the latent factors. Our focus is on the analysis of dynamic factor models with structural restrictions, in which factors are potentially correlated with each other. The existing literature focuses on estimating a rotation of the underlying factors. The identification conditions in the context of dynamic factor models has not been well understood in the literature, an issue to be examined in the next section.

2　Dynamic factor models

We consider a dynamic factor model featuring a dynamic factor representation of the observables similar to Geweke (1977) and Forni et al. (2000) and a law of motion for the factors given by a $VAR(h)$ process. Using the same notation as in Bai and Ng (2007), the model is given by (1) and (2):

$$X_t = \Lambda_0 f_t + \Lambda_1 f_{t-1} + \cdots + \Lambda_s f_{t-s} + e_t \tag{1}$$

where the $q \times 1$ vector f_t is the latent dynamic factor, the $N \times 1$ vector X_t is the observable at time t, the $N \times q$ matrix Λ_j is the dynamic factor loading for f_{t-j}, $j = 0, 1, ..., s$. To fix idea, we first assume that e_t is i.i.d. $N(0, R)$, and will relax this assumption later when we consider serially correlated error terms. The dynamic factor follows a $VAR(h)$ process

$$f_t = \Phi_1 f_{t-1} + \cdots + \Phi_h f_{t-h} + \varepsilon_t, \tag{2}$$

where ε_t is i.i.d. $N(0, Q)$. We assume that $\{e_t\}_{t=1}^{T}$ is independent of $\{\varepsilon_t\}_{t=1}^{T}$. The above dynamic factor model has two different aspects of dynamics. First, there are s lagged factors entering equation (1), representing a dynamic relationship between f_t and the observable X_t. Second, the dynamics of the factors is explicitly parametrized by a $VAR(h)$ process. This paper emphasizes the dynamics in equation (1), it is this dynamics that makes a major distinction between dynamic and static factor analysis. When $s = 0$ (no lags entering the measurement equations), the model will be regarded as a static factor model, as is customary in the existing literature, even though f_t is dynamic. The number of dynamic factors is the dimension of f_t, i.e., q, irrespective of s and h.

Model (1) can be put into a static factor form. Let

$$\Lambda = (\Lambda_0, \Lambda_1, ..., \Lambda_s), \quad F_t = (f_t', f_{t-1}', ..., f_{t-s}')'$$

Then
$$X_t = \Lambda F_t + e_t. \tag{3}$$
There are $q(s+1)$ number of static factors, which is the dimension of F_t. Because of rotational indeterminacy, factor models, either in its dynamic form or in static form, are not identifiable without additional restriction. Classical factor analysis (e.g., Anderson and Rubin (1956), or Lawley and Maxwell (1971)) focuses on static factor models. Classical factor analysis assumes a fixed N, and the model is often estimated by the maximum likelihood method. Under large N, the principal components (PC) method is often used to estimate (3). The PC method would need to impose $q^2(s+1)^2$ restrictions to uniquely fix the rotational indeterminacy. These restrictions are discussed by Bai and Ng (2010). Let $F = (F_1, ..., F_T)'$. The restrictions in static factor analysis takes the form

    (i) $F'F/T = I_r$, where $r = q(s+1)$, and $\Lambda'\Lambda$ is diagonal, or

    (ii) $F'F/T = I_r$, and the first $r \times r$ block of Λ is lower triangular, or

    (iii) F_t is unrestricted, and the first $r \times r$ block of Λ is an identity matrix.

In each form, there are $q^2(s+1)^2$ restrictions. These restrictions do not necessarily structurally identify F_t (or f_t), but a rotation of F_t. The reason is that these restrictions are normalization restrictions, and they may not have economic interpretation in general. For some special cases, for example, when $s = 0$, (iii) allows the identification of f_t and Λ. In general, we only estimate rotations of F_t and Λ under the static factor framework.

The identification method to be introduced requires only q^2 restrictions, far fewer than the static counterpart. We impose restrictions on the innovations ε_t, instead of f_t. Because f_t is a VAR process, the components of f_t are mutually correlated, even when the components of innovations are uncorrelated. So it is appropriate not to assume f_t to have uncorrelated components.

2.1   Identification of the dynamic factor model

It is well know that the dynamic factor model defined in (1) and (2) is not identified without further restrictions. One can always pre-multiply the dynamic factor f_t with an arbitrary full rank $q \times q$ matrix to define a new model. The newly defined model will be observationally equivalent to the original dynamic factor model. We consider two types of identifying restrictions and discuss their implications for structural factor analysis as well as structural VAR analysis. The key identification result is that we only need q^2 restrictions. Let

$$\Lambda_0 = \begin{bmatrix} \Lambda_{01} \\ \Lambda_{02} \end{bmatrix}$$

where Λ_{01} is $q \times q$. Two different sets of identification restrictions, referred to as DFM1 and DFM2, are given below.

DFM1: $var\,(\varepsilon_t) = I_q$, and the $q \times q$ matrix Λ_{01} is a lower-triangular matrix and its diagonal elements are all strictly positive.

This is sufficient to identify the dynamic factor model. There are a total of q^2 restrictions in DFM1 irrespective of the number of lags s and h in the model. The

requirement $var(\varepsilon_t) = I_q$ imposes $q(q + 1)/2$ restrictions and the requirement that Λ_{01} is lower triangular imposes $q(q - 1)/2$ restrictions. The latter restriction says that the first variable X_{1t} is affected contemporaneously by the first dynamic factor, and the second variable X_{2t} is affected contemporaneously by the first two dynamic factors, and so on. In practice, the choice of the first q variables is important.

It is important to recognize that no restrictions are imposed on $\Lambda_1, ..., \Lambda_s$. Moreover, it makes more sense to assume the innovations ε_t to be mutually uncorrelated than to assume f_t to be mutually uncorrelated. This leaves the autoregressive coefficient matrix $\Phi_j, j = 1, ..., h$ in (2) unrestricted, allowing the dynamic factors to interact with each other.

Under DFM1, we shall say Λ_0 is lower triangular for ease of exposition.

Proposition 1. Consider the dynamic factor model as in (1) and (2). Under DFM1, the dynamic factors f_t and the dynamic factor loadings $\Lambda_j, j \in \{0, 1, ..., s\}$ are uniquely identified.

The proof is in Appendix A. Similar to Proposition 1, if ε_t has an identity covariance matrix and Λ_j is lower-triangular for some $j \in \{0, 1, ..., s\}$, then the dynamic factor model in (1) and (2) is identified up to a sign change. We summarize the result in Corollary 1.

Corollary 1. Consider the dynamic factor model as in (1) and (2). Assume that $var(\varepsilon_t) = I_q$, Λ_j is a lower-triangular matrix for some $j = 0, 1, ..., s$, and the diagonal elements of the first $q \times q$ block of Λ_j are all strictly positive. Then the dynamic factors f_t and the dynamic factor loadings $\Lambda_j, j \in \{0, 1, ..., s\}$ are uniquely identified.

Notice that in DFM1, other normalizations on the factor loadings also work, as long as they imply $q(q - 1)/2$ non-redundant restrictions. For example, the normalization that $\Lambda'_j \Lambda_j$ being diagonal for some $j \in \{0, 1, ..., s\}$ also works, which is similar to the normalization in principal component analysis of factor models (see Anderson (1984), Connor and Korajzcyk (1986), Stock and Watson (2002), Bai and Ng (2002), Bai (2003), etc.). A variation to DFM1 is that $var(\varepsilon_t)$ is diagonal but Λ_{01} has 1's on the diagonal.

We next consider a different set of identification conditions.

DFM2: The $q \times q$ block of Λ_{01} is an identity matrix, that is,

$$\Lambda_0 = \begin{bmatrix} I_q \\ * \end{bmatrix}.$$

This is also sufficient to identity the factor model. Again, there are q^2 restrictions, and no restrictions are needed for $\Lambda_1, ..., \Lambda_s$. In DFM2, the identifying restrictions are on the dynamic factor loadings only, which leave the $VAR(h)$ dynamics of f_t completely unrestricted. In DFM2, the innovations to factors ε_t can be linked to structural shocks that are implied by economic theory, which allows a structural VAR representation of the dynamic factors; see Section 7. Both DFM1 and DFM2

6

differ from the conventional practice which restricts the covariance matrix of the dynamic factors f_t or the static factors F_t to be an identity matrix.

Proposition 2. Consider the dynamic factor model as in (1) and (2). Under the normalization that the upper $q \times q$ block of Λ_0 is an identity matrix (DFM2), the dynamic factors f_t and the dynamic factor loadings $\Lambda_j, j \in \{0, 1, ..., s\}$ are uniquely identified.

Similar to Proposition 2, if the upper $q \times q$ block of Λ_j is an identity matrix for some $j \in \{0, 1, ..., s\}$, then the dynamic factor model in (1) and (2) is identified. We summarize the result in Corollary 2.

Corollary 2. Consider the dynamic factor model as in (1) and (2). Under the normalization that the upper $q \times q$ block of Λ_j is an identity matrix for some $j \in \{0, 1, ..., s\}$, then the dynamic factors f_t and the dynamic factor loadings $\Lambda_j, j \in \{0, 1, ..., s\}$ are uniquely identified.

Using a state space representation of the factor model, Doz et al. (2011) explicitly considers the dynamics in factors. Although they consider a vector autoregression in the factors, the factors considered therein are still "static factors". Only the current factors f_t enter the measurement equation at time t. Their specification also differs from Amengual and Watson (2007) in that their shocks to factors have a full rank covariance matrix. Stock and Watson (2005) considers a more general version of dynamic factor models. They allow the error terms in the observation equation to be serially correlated. Like Doz et al. (2011), their identification is also based on restricting the static factors, so that the principal component analysis is embedded into an iterative least squares estimation. Our identification scheme is directly on the dynamic factors f_t, not on the static factors $F_t = (f_t', f_{t-1}', ..., f_{t-s}')'$.

Our estimation of the model is based on a likelihood approach, using the Bayesian estimation of a restricted linear state space system. The likelihood estimation is feasible and computationally efficient due to the fact that $q \ll N$. An alternative likelihood approach is to use the expectation-maximization algorithm, which is not considered in this paper. We choose the Bayesian approach mainly because it allows a coherent hypothesis testing, and the Bayes factors allow a simple model comparison procedure to determine the number of dynamic factors and the number of lags in model (1) and (2).

3 Implications for multi-level factor models

The dynamic factor model with a multi-level factor structure has been increasingly applied to study the comovement of economic quantities at different levels (See Gregory and Head (1999), Kose et al. (2003), Crucini et al. (2011), and Moench et al. (2011), etc.). When researchers confront the model with data, they usually find ways to impose economic meaning of factors. For example, Kose et al. (2003) and a number of subsequent papers consider a dynamic factor model with a multi-level factor structure to characterize the comovement of international business cycles on

the global level, regional level, and country level, respectively. This section considers similar identification conditions for multi-level factor models.

In a typical setup, consider C countries, each having a $n_c \times 1$ vector of country variables X_t^c, $t = 1, ..., T, c = 1, ..., C$. One may model X_t^c as being affected by a world factor f_t^W, a regional factor f_t^r, $r = 1, ..., R$, and a country factor f_t^c, $c = 1, ..., C$, all factors being latent,

$$X_t^c = \Lambda_W^c f_t^W + \Lambda_R^c f_t^{r_c} + \Lambda_C^c f_t^c + e_t^c, \tag{4}$$

where $\Lambda_W^c, \Lambda_R^c, \Lambda_C^c$ are the matrices of factor loadings of country c, e_t^c is the vector of idiosyncratic error terms (possibly serially correlated) for country c's variables, and r_c denotes the region that country c belongs to.

The data generating process for factors is given by a vector autoregression (VAR) with diagonal autoregressive coefficient matrix and independent error terms. Let F_t^R be a vector collecting all the regional factors, and F_t^C be a vector collecting all the country factors. An example of VAR(1) specification of factors is given by,

$$\begin{bmatrix} f_t^W \\ F_t^R \\ F_t^C \end{bmatrix} = \begin{bmatrix} \rho_W & 0 & 0 \\ 0 & \rho_R & 0 \\ 0 & 0 & \rho_C \end{bmatrix} \begin{bmatrix} f_{t-1}^W \\ F_{t-1}^R \\ F_{t-1}^C \end{bmatrix} + \begin{bmatrix} u_t^W \\ U_t^R \\ U_t^C \end{bmatrix}, \tag{5}$$

where $\{\rho_W, \rho_R, \rho_C\}$ are also diagonal matrices conformable to the dimension of corresponding factors. The innovation to factors $\left[u_t^W, U_t^R, U_t^C \right]'$ is independent of e_t^c at all leads and lags and is i.i.d. normal,

$$\begin{bmatrix} u_t^W \\ U_t^R \\ U_t^C \end{bmatrix} \overset{i.i.d.}{\sim} N \left(0, \begin{bmatrix} \sigma_W^2 & 0 & 0 \\ 0 & \sigma_R^2 & 0 \\ 0 & 0 & \sigma_C^2 \end{bmatrix} \right),$$

where σ_R^2, σ_C^2 are themselves diagonal matrices. Given some sign restriction, this special VAR specification allows one to separately identify the factors at different levels up to a scale normalization. To achieve identification up to a sign normalization, it is sufficient to assume the variance of $\left[u_t^W, U_t^R, U_t^C \right]'$ is an identity matrix.

3.1   Motivating a new identification scheme

Observe that it is sufficient but not necessary to assume the independence among $\left\{ f_t^W, f_t^r, f_t^c, r = 1, ..., R, c = 1, ..., C \right\}$ in order to achieve identification of the factors. Equation (5) rules out the possibility that different factors are related. In practice, factors might be correlated through spill-over effects. Take the world factor for example. It might contain the exogenous oil shock as well as the spill-over effects from individual countries or regions. A shock that is originated from a large economy like US might have an impact on other economies. This implies that the innovation to an individual country might have a spill-over effect on the rest of the world, and thus possibly has a effect on the dynamics of the world factor.

According to the identification scheme given in Proposition 1 (DFM1), our new identification scheme allows the autoregressive coefficient matrix in equation (5) to be unrestricted. Under our assumption and using the same notation as (5), we have

$$
\begin{bmatrix} f_t^W \\ F_t^R \\ F_t^C \end{bmatrix} = \begin{bmatrix} \rho_{11} & \rho_{12} & \rho_{13} \\ \rho_{21} & \rho_{22} & \rho_{23} \\ \rho_{31} & \rho_{32} & \rho_{33} \end{bmatrix} \begin{bmatrix} f_{t-1}^W \\ F_{t-1}^R \\ F_{t-1}^C \end{bmatrix} + \begin{bmatrix} u_t^W \\ U_t^R \\ U_t^C \end{bmatrix}, \tag{6}
$$

where the innovation to factors $\left[u_t^W, U_t^R, U_t^C \right]'$ is independent of e_t^c in equation (4) at all leads and lags and is i.i.d. normal with a diagonal covariance matrix, as described earlier.

This implies that conditional on the history $\left\{ f_{t-s}^j, s = 1, 2, ..., t-1, j = W, R, C \right\}$, f_t^W, f_t^R, f_t^C are independent. However, unconditionally different factors can be correlated with each other.

Equation (6) allows us to separately identify different factors, and more importantly, it allows us to evaluate the impulse responses of different factors to the factor innovations $\left[u_t^W, U_t^R, U_t^C \right]'$. For example, a shock to U_t^C not only affects the country factor f_t^C, but also affects both the world factor and the regional factor. This allows us to evaluate the spill-over effects of country-level shocks as well as region-level shocks to other economies.

In addition, equation (6) nests the conventional specification of Kose et al. (2003) in equation (5) as a special case if we further restrict the autoregressive coefficient matrix in (6) to be diagonal. When confronting the model with data, we will be able to test equation (6) against equation (5) using a Wald-test in classical estimation, or using the Bayes factors in Bayesian estimation.

3.2 Identification: the case with two levels

We will focus on a simplified specification with only two levels of factors: a world factor and a country-specific factor. Consider C countries, each having a $n_c \times 1$ vector of country variables X_t^c, $t = 1, ..., T, c = 1, ..., C$. One may model X_t^c as being affected by a world factor f_t^W and a country factor $f_t^c, c = 1, ..., C$,

$$
X_t^c = \Lambda_W^c f_t^W + \Lambda_C^c f_t^c + e_t^c. \tag{7}
$$

Stacking the observations from all countries, we can write (7) in matrix form as,

$$
\begin{bmatrix} X_t^1 \\ \vdots \\ X_t^C \end{bmatrix} = \begin{bmatrix} \Lambda_W^1 & \Lambda_C^1 & & \\ \vdots & & \ddots & \\ \Lambda_W^C & & & \Lambda_C^C \end{bmatrix} \begin{bmatrix} f_t^W \\ f_t^1 \\ \vdots \\ f_t^C \end{bmatrix} + \begin{bmatrix} e_t^1 \\ \vdots \\ e_t^C \end{bmatrix}. \tag{8}
$$

Under the the same notation as (5) , we have

$$
\begin{bmatrix} f_t^W \\ F_t^C \end{bmatrix} = \begin{bmatrix} \rho_{11} & \rho_{12} \\ \rho_{21} & \rho_{22} \end{bmatrix} \begin{bmatrix} f_{t-1}^W \\ F_{t-1}^C \end{bmatrix} + \begin{bmatrix} u_t^W \\ U_t^C \end{bmatrix}, \tag{9}
$$

where the innovation to factors $\left[u_t^W, U_t^C \right]'$ is independent of e_t^c at all leads and lags and is i.i.d. normal,

$$\begin{bmatrix} u_t^W \\ U_t^c \end{bmatrix} \overset{i.i.d.}{\sim} N\left(0, \begin{bmatrix} \sigma_W^2 & 0 \\ 0 & \sigma_C^2 \end{bmatrix} \right). \tag{10}$$

We assume that the dimension of f_t^W and f_t^c are $k \times 1$ and $k_c \times 1$ respectively, $c = 1, ..., C$.

Proposition 3. Define a dynamic factor model as in (8), (9), and (10). Assume that $\sigma_W^2 = I_k$, $\sigma_C^2 = I_{k_1+\cdots+k_C}$, Λ_W^1 and $\{\Lambda_C^c\}_{c=1}^C$ are all lower-triangular matrices with strictly positive diagonal terms, and $[\Lambda_W^c, \Lambda_C^c]$ is of full column rank ($c = 1, \cdots, C$), then the factor model is uniquely identified.

The proof is in Appendix A. Proposition 3 can be easily extended to the case with more than two levels. A key feature with multi-level factor models is that there are many zero blocks in the loading matrix. The zero blocks are not sufficient to prevent the rotational indeterminacy, additional restrictions are needed. For efficient estimation, we will take into account the zero restrictions. Finally note that equation (7) is static. The presence of lags of f_t^W and f_t^c will not affect the identification conditions, and the proposition still holds, similar to the first two propositions. A similar set of identification conditions, which does not require sign restrictions on the factor loadings, is to require both σ_W^2 and σ_C^2 to be diagonal, and Λ_W^1 and $\{\Lambda_C^c\}_{c=1}^C$ to be lower-triangular matrices with 1's on the diagonal.

4 Bayesian estimation using Gibbs sampling

This section deals with Bayesian estimation of the dynamic factor models in Section 2 implemented by the Gibbs sampling. The method largely follows Carter and Kohn (1994)'s multimove Gibbs-sampling algorithm for estimation of the state space models. To avoid numerical singularities due to round-off errors, we will use the square root form of the Kalman filter (Bierman, 1977) and apply the Sherman-Morrison-Woodbury identity to guarantee that the covariance matrices are always numerically positive definite (see Appendix). We will first discuss the case with no serial correlation in the error terms. In Section 6, we consider the case with serially correlated error terms.

Consider the dynamic factor model defined in equations (1) and (2),

$$X_t = \Lambda_0 f_t + \Lambda_1 f_{t-1} + \cdots + \Lambda_s f_{t-s} + e_t$$
$$f_t = \Phi_1 f_{t-1} + \cdots + \Phi_h f_{t-h} + \varepsilon_t.$$

Assume $\{e_t, t = 1, ..., T\}$ is independent of $\{\varepsilon_t, t = 1, ..., T\}$, $e_t \sim N(0, R)$, $\varepsilon_t \sim N(0, Q)$, and are iid over t. Define $k = \max\{s + 1, h\}$ and let $\Phi_{h+1} = \cdots = \Phi_k = 0$ if $k = s + 1 > h$. Let $F_t^* = [f_t', ..., f_{t-k+1}']'$, we obtain a state space representation of the dynamic factor model, with the state equation governed by

$$F_t^* = \Phi^* F_{t-1}^* + u_t,$$

10

$$u_t = G\,\varepsilon_t,$$

where the $qk \times qk$ matrix Φ and the $qk \times q$ matrix G are given below

$$\Phi^* = \begin{pmatrix} \Phi_1 & \Phi_2 & \cdot & \cdot & \cdot & \Phi_k \\ I_q & 0 & 0 & \cdot & \cdot & 0 \\ 0 & I_q & \cdot & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & \cdot & \cdot & I_q & \cdot & 0 \end{pmatrix}, \; G = \begin{pmatrix} I_q \\ 0 \\ \cdot \\ \cdot \\ 0 \end{pmatrix}.$$

When $k = h > s$, $F_t = (f_t', f_{t-1}', \ldots f_{t-s}')'$ is a subvector of F_t^* and the measurement equation is given by

$$X_t = [\Lambda, 0]\, F_t^* + e_t,$$

with the AR coefficient matrix in the state equation given by

$$\Phi = \begin{pmatrix} \Phi_1 & \Phi_2 & \cdot & \Phi_{h-1} & \Phi_h \\ I_q & 0 & 0 & 0 & 0 \\ 0 & I_q & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & 0 & \cdot \\ 0 & \cdot & \cdot & I_q & 0 \end{pmatrix}.$$

To facilitate the analysis, we proceed with the case $k = h \geq s + 1$ and consider the identification scheme DFM2, in which the upper $q \times q$ block of Λ_0 is I_q.

The basic idea is to treat both the model parameters and the factors as missing data. The Gibbs sampling consists of the following two steps:

- Conditional on the model parameters and the observed data, generate the factors.

- Conditional on the factors and the observed data, generate the model parameters.

The second step is essentially the same as regression analysis with parameter restrictions. We next examine the first step.

4.1   Sampling the factors conditional on parameters

We use the notation $X^T = [X_T, X_{T-1}, \ldots, X_1]'$ to denote the history of X up to time T. For

$$F_t^* = \begin{bmatrix} f_t \\ \vdots \\ f_{t-h+1} \end{bmatrix},$$

only the last q elements of F_t^* are unknown, namely f_{t-h+1}, once F_{t+1}^* is given. We want to sample F_t^* from the joint distribution

$$p\left(F_T^*, \ldots, F_1^* | X^T\right) = p\left(F_T^* | X^T\right) \prod_{t=1}^{T-1} p\left(F_t^* | F_{t+1}^*, \ldots, F_T^*, X^T\right)$$

11

$$= p\left(F_T^*|X^T\right) \prod_{t=1}^{T-1} p\left(F_t^*|F_{t+1}^*, X^t\right), \tag{11}$$

where the second equality is derived from the Markov structure of the state space system. According to (11), it is the same as sampling F_t^*, $t = T, T-1, ..., 1$, using a backward-sampling algorithm, in which we first sample F_T^* from

$$p\left(F_T^*|X^T\right) = N\left(F_{T|T}^*, P_{T|T}\right), \tag{12}$$

where $F_{T|T}^* = E\left(F_T^*|X^T\right)$ and $P_{T|T} = Var\left(F_T^*|X^T\right)$ and they are obtained from the Kalman filter. Then we may sample F_t^* from

$$p\left(F_t^*|F_{t+1}^*, X^t\right) = N\left(h_t, H_t\right), \quad t = T-1, ..., 1.$$

Given F_{t+1}^*, only the last q elements of F_t^* are random, which can be drawn from

$$p\left(f_{t-h+1}|F_{t+1}^*, X^t\right), \quad t = T-1, ..., h. \tag{13}$$

In the linear Gaussian specification of the dynamic factor model, the distribution of f_{t-h+1} conditional on F_{t+1}^* and X^t is normal. So we just need to derive the conditional mean and the conditional variance of f_{t-h+1}. Based on West and Harrison (1997), we proceed according to the following sampling scheme.

First, notice that the kernel of the conditional density $p\left(f_{t-h+1}|F_{t+1}^*, X^t\right)$ can be written as

$$
\begin{aligned}
p\left(f_{t-h+1}|F_{t+1}^*, X^t\right) &= p\left(f_{t-h+1}\Big|f_{t+1}, f_t, ..., f_{t-h+2}, X^t\right) \\
&\propto p\left(f_{t-h+1}, f_{t+1}|f_t, ..., f_{t-h+2}, X^t\right) \\
&= p\left(f_{t-h+1}|f_t, ..., f_{t-h+2}, X^t\right) p\left(f_{t+1}|f_t, ..., f_{t-h+2}, f_{t-h+1}, X^t\right).
\end{aligned}
$$

The distribution of f_{t-h+1} conditional on $(f_t, ..., f_{t-h+2}, X^t)$ can be easily derived from the joint distribution of $F_t^* \equiv (f_t, ..., f_{t-h+2}, f_{t-h+1})'$ conditional on X^t from the Kalman filter. To see this, let

$$F_t^*|X^t \sim N\left(\begin{bmatrix} \mu_{1t} \\ \mu_{2t} \end{bmatrix}, \begin{bmatrix} P_{11,t} & P_{12,t} \\ P_{21,t} & P_{22,t} \end{bmatrix}\right),$$

where μ_{1t} and μ_{2t} are of dimension $qh \times 1$ and $q \times 1$ respectively. The dimension of the blocks in the variance matrix is defined accordingly. The moments of this multivariate normal distribution are derived from the standard Kalman filter. Then

$$f_{t-h+1}|\left(f_t, ..., f_{t-h+2}, X^t\right) \sim N\left(\mu_t, \Sigma_t\right), \tag{14}$$

where[1]

$$\mu_t = \mu_{2t} + P_{21,t}P_{11,t}^{-1}\left((f_t, ..., f_{t-h+2})' - \mu_{1t}\right),$$

---

[1] In the Appendix, we show how we may apply the Sherman-Morrison-Woodbury identity to guarantee that Σ_t is always numerically positive definite.

$$\Sigma_t = P_{22,t} - P_{21,t}P_{11,t}^{-1}P_{12,t}.$$

Next, from the law of motion for the dynamic factors $f_{t+1} = \Phi_1 f_t + \cdots + \Phi_h f_{t-h+1} + \varepsilon_{t+1}$, and $\varepsilon_{t+1} \sim N(0, Q)$, the distribution $p(f_{t+1}|f_t, ..., f_{t-h+2}, f_{t-h+1}, X^t)$ is given by

$$f_{t+1}|\left(f_t, ..., f_{t-h+2}, f_{t-h+1}, X^t\right) \sim N\left(\Phi_1 f_t + \cdots + \Phi_h f_{t-h+1}, Q\right) \qquad (15)$$

Based on (14) and (15), we may derive the kernel of our target density

$$p\left(f_{t-h+1}|F_{t+1}^*, X^t\right) \propto N\left(\mu_t, \Sigma_t\right) \cdot N\left(\Phi_1 f_t + \cdots + \Phi_h f_{t-h+1}, Q\right)$$
$$\propto exp\left\{-0.5\left(f_{t-h+1} - \mu_t\right)' \Sigma_t^{-1}\left(f_{t-h+1} - \mu_t\right)\right\} \cdot$$
$$exp\left\{-0.5\left(m_{t+1} - \Phi_h f_{t-h+1}\right)' Q^{-1}\left(m_{t+1} - \Phi_h f_{t-h+1}\right)\right\},$$

where $m_{t+1} = f_{t+1} - \Phi_1 f_t - \cdots - \Phi_{h-1}f_{t-h+2}$. The kernel can be further simplified as

$$p\left(f_{t-h+1}|F_{t+1}^*, X^t\right) \propto exp\left\{-0.5\left(f_{t-h+1} - \mu_t^*\right)' \Omega_t^{-1}\left(f_{t-h+1} - \mu_t^*\right)\right\},$$

where

$$\Omega_t = \left(\Sigma_t^{-1} + \Phi_h' Q^{-1}\Phi_h\right)^{-1},$$
$$\mu_t^* = \Omega_t\left(\Sigma_t^{-1}\mu_t + \Phi_h' Q^{-1}m_{t+1}\right).$$

In sum, conditional on the model's parameters, the following recursion describes how to draw from $p\left(F_T^*, ..., F_1^*|X^T\right)$.

- STEP 1. Apply Kalman filter to obtain $F_{T|T}^* = E\left(F_T^*|X^T\right)$ and $P_{T|T} = Var\left(F_T^*|X^T\right)$, and then draw F_T^* from

$$p\left(F_T^*|X^T\right) = N\left(F_{T|T}^*, P_{T|T}\right). \qquad (16)$$

- STEP 2. For $t = T - 1, ..., h$, draw f_{t-h+1} from

$$p\left(f_{t-h+1}|F_{t+1}^*, X^t\right) = N\left(\mu_t^*, \Omega_t\right). \qquad (17)$$

- The parameters of the normal distribution are determined as follows. Use the notation from the Kalman filter procedure and partition $F_t^* = [*, f_{t-h+1}']'$, let

$$F_t^*|X^t \sim N\left(\begin{bmatrix} \mu_{1t} \\ \mu_{2t} \end{bmatrix}, \begin{bmatrix} P_{11,t} & P_{12,t} \\ P_{21,t} & P_{22,t} \end{bmatrix}\right).$$

Define

$$\mu_t = \mu_{2t} + P_{21,t}P_{11,t}^{-1}\left((f_t, ..., f_{t-h+2})' - \mu_{1t}\right),$$
$$\Sigma_t = P_{22,t} - P_{21,t}P_{11,t}^{-1}P_{12,t},$$

$$m_{t+1} = f_{t+1} - \Phi_1 f_t - \cdots - \Phi_{h-1} f_{t-h+2},$$
$$\Omega_t = \left(\Sigma_t^{-1} + \Phi_h' Q^{-1} \Phi_h\right)^{-1},$$
$$\mu_t^* = \Omega_t \left(\Sigma_t^{-1} \mu_t + \Phi_h' Q^{-1} m_{t+1}\right).$$

Steps 1 and 2 finish one round of sampling from $p\left(F_T^*, ..., F_1^* | X^T\right).$

The above algorithm considers the case where $h \geq 2$. In the special case where $h = 1$, $f_{t-h+1} = f_t = F_t^*$. Then we may simply sample from

$$p\left(f_t | f_{t+1}, X^t\right) \propto p\left(f_{t+1} | f_t, X^t\right) p\left(f_t | X^t\right), t = T - 1, ..., 1.$$

Note that

$$f_{t+1} | \left(f_t, X^t\right) \sim N\left(\Phi_1 f_t, Q\right),$$
$$f_t | X^t \sim N\left(E\left(f_t | X^t\right), Var\left(f_t | X^t\right)\right).$$

So

$$p\left(f_t | f_{t+1}, X^t\right) \propto N\left(\mu_t, \Sigma_t\right) \cdot N\left(\Phi_1 f_t, Q\right)$$
$$\propto exp\left\{-0.5\left(f_t - \mu_t\right)' \Sigma_t^{-1}\left(f_t - \mu_t\right)\right\} \cdot$$
$$exp\left\{-0.5\left(f_{t+1} - \Phi_1 f_t\right)' Q^{-1}\left(f_{t+1} - \Phi_1 f_t\right)\right\},$$

and the kernel can be further simplified as

$$p\left(f_t | f_{t+1}, X^t\right) \propto exp\left\{-0.5\left(f_t - \mu_{ft}\right)' \Omega_{ft}^{-1}\left(f_t - \mu_{ft}\right)\right\},$$

where

$$\Omega_{ft} = \left(\Sigma_t^{-1} + \Phi_1' Q^{-1} \Phi_1\right)^{-1},$$
$$\mu_{ft} = \Omega_{ft}\left(\Sigma_t^{-1} \mu_t + \Phi_1' Q^{-1} f_{t+1}\right).$$

4.2 Sampling model parameters conditional on the factors

Let θ be the collection of all the model parameters

$$\theta = \{\Lambda_0, \cdots, \Lambda_s, \Phi_1, \cdots, \Phi_h, R, Q\}.$$

Consider the dynamic factor model defined in equations (1) and (2),

$$X_t = \Lambda_0 f_t + \Lambda_1 f_{t-1} + \cdots + \Lambda_s f_{t-s} + e_t$$
$$f_t = \Phi_1 f_{t-1} + \cdots + \Phi_h f_{t-h} + \varepsilon_t.$$

Assume $\{e_t, t = 1, ..., T\}$ is independent of $\{\varepsilon_t, t = 1, ..., T\}$; e_t are iid $N\left(0, R\right)$ and ε_t are iid $N\left(0, Q\right)$. We will use Jeffreys diffuse priors (Zellner, 1971) for θ.

4.2.1 Sampling $\{\Phi_j, j = 1, ..., h\}$ and Q conditional on the factors

Let

$$H_t = [f'_{t-1}, ..., f'_{t-h}],$$

$$H = \begin{bmatrix} H_{h+1} \\ \vdots \\ H_T \end{bmatrix},$$

$$G = \begin{bmatrix} f'_{h+1} \\ \vdots \\ f'_T \end{bmatrix}.$$

Let

$$A = (\Phi_1, ..., \Phi_h)', \quad \alpha = vec(A), \quad \varepsilon = (\varepsilon_{h+1}, ..., \varepsilon_T)'.$$

Then the $AR(h)$ process for the dynamic factors can be written as

$$G = HA + \varepsilon, \text{ or}$$
$$vec(G) = (I_q \otimes H)\alpha + vec(\varepsilon),$$

where $vec(\varepsilon) \sim N(0, Q \otimes I_{T-h})$.

- Use Jeffreys diffuse priors

$$p(A, Q) = p(A)p(Q),$$
$$p(A) = const.$$
$$p(Q) \propto |Q|^{-(q+1)/2}.$$

- The corresponding posterior takes the form (Zellner, 1971):

$$\alpha|Q, G \sim N\left(\hat{\alpha}, Q \otimes (H'H)^{-1}\right),$$
$$Q^{-1}|G \sim Wishart\left(\Omega^{-1}, v\right),$$

where[2]

$$\hat{A} = (H'H)^{-1} H'G,$$
$$\hat{\alpha} = vec\left(\hat{A}\right),$$
$$\Omega = \left(G - H\hat{A}\right)'\left(G - H\hat{A}\right),$$
$$v = T - h - hq.$$

---

[2]$Q|G \sim |Q|^{-\frac{T-h-hq+q+1}{2}} exp\left\{-\frac{1}{2}tr\left(Q^{-1}\Omega\right)\right\}$ is inverse Wishart parametrized by $(\Omega, v = T - h - hq)$. The mean of Q is given by $\frac{\Omega}{v-q-1}$. Equivalently, $Q^{-1}|G$ is Wishart parametrized by $(\Omega^{-1}, v = T - h - hq)$.

4.2.2 Formulating the restrictions on factor loadings

In additional to the minimal identifying restrictions as in DFM1 or DFM2, there may be other overidentifying restrictions. For example, the multi-level factor model has many zero blocks. Cross-equation restrictions may also be present. We show how these restrictions can be imposed in a systematic way.

We start with a static factor representation of (1)

$$X_t = \Lambda F_t + e_t,$$

where $F_t = [f_t, f_{t-1}, ..., f_{t-s}]'$, and $\Lambda = [\Lambda_0, \cdots, \Lambda_s]$. Let X be the $(T - s - 1) \times N$ data matrix, F be the $(T - s - 1) \times qs$ matrix of the static factors, then we have a matrix representation of the factor model

$$X = F\Lambda' + E, \text{ or}$$
$$vec(X) = (I_N \otimes F)\lambda + vec(E), \tag{18}$$

where $\lambda = vec(\Lambda')$ and $vec(E) \sim N(0, R \otimes I_{T-s-1})$.

The identifying restrictions in DFM2 impose linear restrictions on the factor loadings. We discuss the Gibbs sampling step in case of linear restrictions on the factor loadings. For $\lambda = vec(\Lambda')$, consider the following restriction

$$\lambda = B\delta + C, \tag{19}$$

where δ is a vector of free parameters with $dim(\delta) \leq dim(\lambda)$. In general, B and C are known matrices and vectors that are defined by either identifying restrictions or other structural model restrictions. We give an example how B and C in (19) can be obtained from DFM2.

- Example: In DFM2, the upper-left $q \times q$ block of Λ is I_q. So B is a $Nq(s+1) \times [Nq(s+1) - q^2]$ selection matrix consisting of zeros and ones. C is a vector of zeros and ones. δ is a $[Nq(s+1) - q^2] \times 1$ vector of free parameters. There are q ones and $q^2 - q$ zeros in λ, for which the corresponding rows of B are given by zeros and the corresponding elements of C are given by either one or zero. Combining all other rows of B we obtain an identity matrix $I_{Nq(s+1)-q^2}$. For this particular case, one only needs to sample the free parameters in Λ equation-by-equation due to the absence of cross-equation restrictions.

In view of (19), we may rewrite the factor model (18) as

$$vec(X) = (I_N \otimes F)\lambda + vec(E)$$
$$= (I_N \otimes F)(B\delta + C) + vec(E),$$

or as follows

$$vec(X) - (I_N \otimes F)C = [(I_N \otimes F)B]\delta + vec(E).$$

Let $y = vec(X) - (I_N \otimes F)C$ and $Z = [(I_N \otimes F)B]$. With restrictions on factor loadings, the dynamic factor model is rewritten as a regression model

$$y = Z\delta + vec(E), \quad vec(E|Z) \sim N(0, R \otimes I_{T-s}).$$

4.2.3 Sampling $\{\Lambda_0, \cdots, \Lambda_s, R\}$ conditional on the factors and the restrictions

We extend Zellner (1971)'s Bayesian multivariate regression analysis to the case with linear restrictions on regression coefficients. We consider diffuse priors for the factor loadings and R. Notice that the likelihood can be written as

$$p(y|Z, \delta, R) \propto |R|^{-\frac{T-s}{2}} exp\left\{-\frac{1}{2}(y - Z\delta)'(R \otimes I_{T-s})^{-1}(y - Z\delta)\right\}$$
$$= |R|^{-\frac{T-s}{2}} exp\left\{-\frac{1}{2}tr\left[(X - F\Lambda')'(X - F\Lambda')R^{-1}\right]\right\},$$

where Λ is subject to linear restrictions that are parametrized by δ. Noting that

$$(X - F\Lambda')'(X - F\Lambda') = \left(X - F\hat{\Lambda}'\right)'\left(X - F\hat{\Lambda}'\right) + \left(\Lambda - \hat{\Lambda}\right)F'F\left(\Lambda - \hat{\Lambda}\right)'$$
$$= S + \left(\Lambda - \hat{\Lambda}\right)F'F\left(\Lambda - \hat{\Lambda}\right)',$$

where

$$\hat{\Lambda} = X'F(F'F)^{-1},$$
$$S = \left(X - F\hat{\Lambda}'\right)'\left(X - F\hat{\Lambda}'\right).$$

We may further write the likelihood as

$$p(y|Z, \delta, R) \propto |R|^{-\frac{T-s}{2}} exp\left\{-\frac{1}{2}tr\left[SR^{-1}\right] - \frac{1}{2}tr\left[\left(\Lambda - \hat{\Lambda}\right)F'F\left(\Lambda - \hat{\Lambda}\right)'R^{-1}\right]\right\}.$$

Impose the Jeffreys prior for δ and R:

$$p(\delta) = const, \ p(R) \propto |R|^{-(N+1)/2}$$
$$p(\delta, R) = p(\delta)p(R).$$

This also implies that the prior on R^{-1} is $p(R^{-1}) \propto |R^{-1}|^{-(N+1)/2}$. We obtain the following joint posterior for δ and R:

$$p(\delta, R|X, F) \propto |R|^{-\frac{T-s+N+1}{2}} exp\left\{-\frac{1}{2}tr\left[SR^{-1}\right] - \frac{1}{2}tr\left[\left(\Lambda - \hat{\Lambda}\right)F'F\left(\Lambda - \hat{\Lambda}\right)'R^{-1}\right]\right\}.$$

We may write

$$p(\delta, R|X, F) = p(\delta|R, X, F)p(R|X, F).$$

Here

$$p(R|X, F) \propto |R|^{-v/2} exp\left(-\frac{1}{2}tr\left[R^{-1}S\right]\right), v = T - s + N + 1 - r.$$
$$p(\delta|R, X, F) \propto |R|^{-r/2} exp\left\{-\frac{1}{2}tr\left[\left(\Lambda(\delta) - \hat{\Lambda}\right)F'F\left(\Lambda(\delta) - \hat{\Lambda}\right)'R^{-1}\right]\right\} \quad (20)$$
$$\propto |R|^{-r/2} exp\left\{-\frac{1}{2}tr\left[\left(\lambda(\delta) - \hat{\lambda}\right)'\left[R^{-1} \otimes F'F\right]\left(\lambda(\delta) - \hat{\lambda}\right)\right]\right\}$$

17

Given the linear restriction $\lambda(\delta) = B\delta + C$, and let $\mu_\delta = B\left(B'B\right)^{-1}\left(\hat{\lambda} - C\right)$, we have

$$p\left(\delta|R, X, F\right) \propto |R|^{-r/2} exp\left\{-\frac{1}{2}tr\left[\left(B\delta + C - \hat{\lambda}\right)'\left[R^{-1} \otimes F'F\right]\left(B\delta + C - \hat{\lambda}\right)\right]\right\}$$

$$\propto |R|^{-r/2} exp\left\{-\frac{1}{2}tr\left[\left(\delta - \mu_\delta\right)'\left[B'\left(R^{-1} \otimes F'F\right)B\right]\left(\delta - \mu_\delta\right)\right]\right\},$$

which implies that

$$\delta|R, X, F \sim N\left(B\left(B'B\right)^{-1}\left(\hat{\lambda} - C\right), \left(B'\left(R^{-1} \otimes F'F\right)B\right)^{-1}\right). \qquad (21)$$

Thus we may draw δ according to (21) and construct the associated loading matrix $\Lambda(\delta)$. Note that in the special case where there is no cross-equation restrictions and R is diagonal, $B'\left(R^{-1} \otimes F'F\right)B$ becomes a block-diagonal matrix, and the factor loading can be drawn equation-by-equation. From (20), we may draw R according to a inverse-Wishart distribution

$$R|X, F \sim invWishart\left(S, T - s + N + 1 - r\right).$$

4.2.4 Restrictions on R

If the cross-section dimension N is small in (1), we may allow $R = Var\left(e_t\right)$ to be an unrestricted covariance matrix. Given our knowledge of the model structure, the factors and the factor loadings can be identified through the autocorrelation function implied by the dynamic model[3]. However, if N is large, it is practically infeasible to make inference on an unrestricted R. On the other hand, for large N, the majority of the cross-correlation in the data matrix is accounted for by the dynamic factors. Although the identification of the dynamic factor model allows an unrestricted R, to preserve a parsimonious model structure, we may consider restricting R to be diagonal when N is large.

Other plausible restrictions include making R a block-diagonal matrix. For example, in the multi-level factor model, we may allow the error term e_t to be cross-correlated within the same group, but uncorrelated between groups.

We consider a diagonal $R = Var\left(e_t\right)$ given as follows

$$R = \begin{bmatrix} \sigma_1^2 & & \\ & \ddots & \\ & & \sigma_N^2 \end{bmatrix}.$$

Impose the diffuse independent prior for δ (free parameters in loadings) and R:

$$p(\delta) = const, \ p(\sigma_n^2) \propto 1/\sigma_n^2$$

---

[3]In the static factor models, identification usually requires a diagonal R so as to achieve identification through the covariance of the data. In the dynamic factor models, however, one can relax the assumption of a diagonal R and can still achieve identification through the autocovariances of the data.

$$p(\delta, R) = p(\delta) \left[\Pi_{n=1}^{N} p(\sigma_n^2)\right].$$

We obtain the following joint posterior for δ and R:

$$p(\delta, R|X, F) \propto |R|^{-\frac{T-s+2}{2}} exp\left\{-\frac{1}{2}tr\left[SR^{-1}\right] - \frac{1}{2}tr\left[\left(\Lambda - \hat{\Lambda}\right)F'F\left(\Lambda - \hat{\Lambda}\right)'R^{-1}\right]\right\},$$

in which R is a diagonal matrix. We may write

$$p(\delta, R|X, F) = p(\delta|R, X, F)p(R|X, F).$$

Here the density $p(\delta|R, X, F)$ is the same as (21). The density $p(R|X, F)$ is given by

$$p(R|X, F) \propto |R|^{-v/2} exp\left(-\frac{1}{2}tr\left[R^{-1}S\right]\right)$$

$$= \left[\Pi_{n=1}^{N}\sigma_n^2\right]^{-v/2} exp\left(-\frac{1}{2}tr\left[R^{-1}S\right]\right).$$

Also notice that $S = \hat{e}'\hat{e}$ where $\hat{e} = [\hat{e}_1, ..., \hat{e}_N] = X - F\Lambda'$. We have

$$exp\left(-\frac{1}{2}tr\left[R^{-1}S\right]\right) = exp\left(-\frac{1}{2}\left[\frac{\hat{e}_1'\hat{e}_1}{\sigma_1^2} + \cdots + \frac{\hat{e}_N'\hat{e}_N}{\sigma_N^2}\right]\right)$$

$$= \prod_{n=1}^{N} exp\left(-\frac{1}{2}\left[\frac{\hat{e}_n'\hat{e}_n}{\sigma_n^2}\right]\right).$$

In sum

$$p(R|X, F) \propto \prod_{n=1}^{N} \left(\sigma_n^2\right)^{-(v/2-1)-1} exp\left(-\frac{\hat{e}_n'\hat{e}_n/2}{\sigma_n^2}\right).$$

Thus we may obtain a draw for σ_n^2 one-by-one according to the following inverse-Gamma distribution

$$\sigma_n^2|X, F \sim invGamma\left(v/2 - 1, \hat{e}_n'\hat{e}_n/2\right), v = T - s + 2 - r.$$

Or by noticing that $p(1/\sigma_n^2) = p(\sigma_n^2) \cdot (\sigma_n^2)^2$, we may drawn the precision $1/\sigma_n^2$ according the Gamma distribution[4]

$$\frac{1}{\sigma_n^2}|X, F \sim Gamma\left(v/2 - 1, (\hat{e}_n'\hat{e}_n/2)^{-1}\right).$$

4.2.5 Sampling $\{\Lambda_1, ..., \Lambda_s\}$ given a diagonal R

When the cross-section dimension N is large, we may restrict R to be diagonal or block diagonal. If R is diagonal, then the dynamic factor loadings can be sampled equation-by-equation when there is no cross-equation restriction on the factor loading. Note that

$$vec\left(\Lambda'\right)|R, X, F \sim N\left(\hat{\lambda}, R \otimes (F'F)^{-1}\right),$$

---

[4]We follow the notation from Matlab, which defines Gamma pdf as $p(x|\alpha, \beta) \propto x^{\alpha-1}e^{-x/\beta}$.

where

$$R \otimes (F'F)^{-1} = \begin{bmatrix} \sigma_1^2 (F'F)^{-1} & & \\ & \ddots & \\ & & \sigma_N^2 (F'F)^{-1} \end{bmatrix}.$$

Let $\Lambda' = [\lambda_1, ..., \lambda_N]$, then conditional on $\{R, X, F\}$, the columns of Λ' are independent. If the restrictions on the factor loadings are on an equation-by-equation basis, we are able to sample the factor loadings equation-by-equation:

$$\delta_i | R, X, F \sim N \left(B_i (B_i'B_i)^{-1} \left(\hat{\lambda}_i - C_i \right), \sigma_i^2 (B_i'F'FB_i)^{-1} \right), \quad (22)$$

where the restriction for equation i is given by $\lambda_i = B_i \delta_i + C_i$ and $\hat{\lambda}_i = (F'F)^{-1} F'X_i$ is the OLS estimator of regressing X_i on F.

4.2.6 Prior on the initial state vector

When sampling the factors, we utilize the Kalman filter which requires sampling from the prior

$$p \left(F_h^* | X_h \right)$$

conditional on the model parameters. Here $F_h^* = [f_h', ..., f_1']'$ is the initial observation of the state vector. We assume that this prior is a normal distribution whose moments are independent of the model parameters. The simplifying assumption allows us to avoid the use of Metropolis-Hasting algorithm to sample from a more complicated prior distribution, which might depend on the model parameters in a nonlinear way.

When sampling the autoregressive coefficient matrices Φ_j, we do not impose any stationarity restrictions. In the literature, usually an indicator function, $I[s(\Phi)]$, is multiplied to the prior density of Φ_j so as to restrict the roots of the AR process to be outside the unit circle (see Kim and Nelson (1999), Cogley and Sargent (2005)). We do not make such restrictions on the belief that if stationarity of the factors is valid it should be supported and justified by the posterior distribution. On the other hand, the posterior distribution allows us to make inference about the stationarity of the factors.

5 Model comparison with diffuse priors

Under the minimal identification scheme DFM1 or DFM2, we are able to discriminate between models with different lags and number of dynamic factors using Bayes factors. To fix idea, we will impose the identification scheme DFM2 (the upper $q \times q$ block of Λ_0 being identity) throughout. Let M_1 and M_2 be two possible competing models for the time series $X \equiv \{X_1, ..., X_T\}$, both being dynamic factor models but with different lags h, s and number of factors q. The Bayes factor for model M_2 against M_1 from data X is the ratio

$$B_{21} = \frac{p(X|M_2)}{p(X|M_1)}.$$

Note that if we multiply the Bayes factors by the ratio of priors for both models, we obtain the posterior odds of the two models. A key objective of interest is the marginal likelihood of the data

$$
\begin{aligned}
p\left(X|M_j\right) &= \int p\left(X, \theta|M_j\right) d\theta \\
&= \int p\left(X|\theta, M_j\right) \pi\left(\theta\right) d\theta
\end{aligned}
$$

where $p\left(X|\theta, M_j\right)$ is the likelihood under model M_j and $\pi\left(\theta\right)$ is the prior. Note that if we use the diffuse prior, $p\left(X|M_j\right)$ in general is not well defined unless θ has a compact support. Possible alternatives to Bayes factors in case of diffuse priors include fractional Bayes factors (O'Hagan (1995)) and intrinsic Bayes factors (Berger and Pericchi (1996)). To fix idea, we will adopt a version of the fractional Bayes factors for the purpose of comparing different dynamic factor models in case of diffuse priors. Note that the intrinsic Bayes factors can be obtained in a similar way.

Given a fraction b of the data $X^0 \equiv \{X_1, ..., X_{[Tb]}\}$, calculate the posterior based on the training sample

$$
p\left(\theta|M_j, X_1, ..., X_{[Tb]}\right) \equiv p\left(\theta|M_j, X^0\right).
$$

Then calculate the marginal likelihood for $X^1 \equiv \{X_{[Tb]+1}, ..., X_T\}$ conditional on the training sample

$$
\begin{aligned}
p\left(X^1|M_j, X^0\right) &= \int p\left(X^1, \theta|M_j, X^0\right) d\theta \\
&= \int p\left(X^1|\theta, M_j, X^0\right) p\left(\theta|M_j, X^0\right) d\theta.
\end{aligned}
$$

The fractional Bayes factor is then given by

$$
B_{21}^F = \frac{p\left(X^1|M_2, X^0\right)}{p\left(X^1|M_1, X^0\right)}.
$$

Note that

$$
\begin{aligned}
p\left(\theta|X^1, X^0\right) &\propto p\left(X^1|\theta, X^0\right) p\left(\theta|X^0\right) \\
&\propto p\left(X^1, X^0|\theta\right) \pi\left(\theta\right)
\end{aligned}
$$

where $\pi\left(\theta\right)$ is the diffuse prior. This suggests two equivalent sampling schemes to sample from the posterior $p\left(\theta|X^1, X^0\right)$. One is to treat $p\left(\theta|X^0\right)$ as the prior and $p\left(X^1|\theta, X^0\right)$ as the likelihood. The other is to treat $\pi\left(\theta\right)$ as the prior and $p\left(X^1, X^0|\theta\right)$ as the likelihood. Thus, to calculate the fractional Bayes factor, we may use the same algorithm as in the previous section to obtain posterior draws from $p\left(\theta|X^1, X^0\right)$.

The algorithm consists the following three steps.

Step 1. Assume Jeffreys priors on θ. Conduct a Bayesian analysis of the training sample X^0 to obtain a sequence of posterior draws $\{\theta^{(l)}, F^{(l)}\}_{l=1}^{L}$. Obtain the posterior density given X^0

$$p\left(\theta|M^j, X^0\right) = \int p\left(\theta, F|M^j, X^0\right) dF$$

$$= \int p\left(\theta|M^j, F, X^0\right) p\left(F|M^j, X^0\right) dF$$

$$\approx \frac{1}{L}\sum_{l=1}^{L} p\left(\theta|M^j, F^{(l)}, X^0\right)$$

Note that $p\left(\theta|M^j, F^{(l)}, X^0\right)$ has a closed-form solution.

Step 2. Repeat Step 1 for the whole sample $X = \{X^0, X^1\}$ to obtain another sequence of posterior draws $\{\theta^{(l)}, F^{(l)}\}_{l=1}^{L}$. Obtain the posterior density given X

$$p\left(\theta|M^j, X\right) \approx \frac{1}{L}\sum_{l=1}^{L} p\left(\theta|M^j, F^{(l)}, X\right).$$

Step 3. Calculate the fractional marginal likelihood for X^1 conditional on X^0

$$p\left(X^1|M_j, X^0\right) = \int p\left(X^1|\theta, M_j, X^0\right) p\left(\theta|M_j, X^0\right) d\theta$$

$$= \int \frac{p\left(X^1|\theta, M_j, X^0\right) p\left(\theta|M_j, X^0\right)}{p\left(\theta|M_j, X\right)} p\left(\theta|M_j, X\right) d\theta$$

$$\approx \frac{1}{L}\sum_{l=1}^{L} \frac{p\left(X^1|\theta^{(l)}, M_j, X^0\right) p\left(\theta^{(l)}|M_j, X^0\right)}{p\left(\theta^{(l)}|M_j, X\right)}.$$

A convenient alternative to Step 3 is to calculate the fractional marginal likelihood for X^1 conditional on X^0 according to the identity (Chib (1995))

$$p\left(X^1|M_j, X^0\right) = \frac{p\left(X^1|\theta, M_j, X^0\right) p\left(\theta|M_j, X^0\right)}{p\left(\theta|M_j, X^1, X^0\right)}.$$

Thus we may obtain the logarithm of the fractional marginal likelihood

$$log\left(p\left(X^1|M_j, X^0\right)\right) = log\left(p\left(X^1|\theta^*, M_j, X^0\right)\right) + log\left(p\left(\theta^*|M_j, X^0\right)\right)$$
$$- log\left(p\left(\theta^*|M_j, X^1, X^0\right)\right),$$

where θ^* is the posterior mean for θ.

Note that given the recursive structure of the dynamic factor model, it is easy to calculate the conditional likelihood

$$p\left(X^1|\theta, M_j, X^0\right)$$

which is multivariate normal:

$$p\left(X^1|\theta, M_j, X^0\right) = \Pi_{t=[Tb]+1}^{T} p\left(X_t|X_{t-1}, ..., X_1, \theta, M_j\right)$$

$$= \Pi_{t=[Tb]+1}^{T} N\left(E\left(X_t | X_{t-1}, ..., X_1, \theta, M_j\right), Var\left(X_t | X_{t-1}, ..., X_1, \theta, M_j\right)\right)$$

in which the mean and variance of the normal distribution are given by the Kalman filter.

Regarding the choice of the fraction, O'Hagan (1995) proposes three alternatives: $b = \frac{n_0}{n}$, $b = \frac{n_0 log(n)}{n log(n_0)}$, or $b = \sqrt{\frac{n_0}{n}}$, ranked by increasing robustness, where n_0 is the minimal training sample size such that a posterior is well defined.

6   Serial correlation in the measurement errors

The above analysis is based on i.i.d. measurement errors. The methodology can be readily extended to the case with serially correlated measurement errors. In particular, consider:

$$X_t = \Lambda(L)f_t + e_t,$$

where $\Lambda\left(L\right) = [\Lambda_0, \Lambda_1 L, ..., \Lambda_s L^s]$, where L is the lag operator. The measurement error follows an VAR(1) process

$$e_t = \rho e_{t-1} + u_t, \quad u_t \sim i.i.d.\, N(0, R)$$

where

$$\rho = \begin{bmatrix} \rho_1 & & \\ & \ddots & \\ & & \rho_N \end{bmatrix}, \quad R = \begin{bmatrix} \sigma_1^2 & & \\ & \ddots & \\ & & \sigma_N^2 \end{bmatrix}.$$

The factors follow a $VAR\left(h\right)$ process

$$f_t = \Phi_1 f_{t-1} + \cdots + \Phi_h f_{t-h} + \varepsilon_t.$$

Applying $(I - \rho L)$ on both sides of the measurement equation, we obtain

$$(I - \rho L) X_t = (I - \rho L) \Lambda(L)f_t + u_t, \quad u_t \sim i.i.d.\, N(0, R).$$

Let $\tilde{X}_t \equiv (I - \rho L) X_t$, $\tilde{\Lambda}\left(L\right) \equiv (I - \rho L) \Lambda\left(L\right)$, then

$$\tilde{X}_t = \tilde{\Lambda}\left(L\right) f_t + u_t.$$

Thus conditional on parameters $\{\rho, \Lambda, R, \Phi, Q\}$, the latent factors can be analyzed using the same backward sampling algorithm as before. Conditional on the data and the factors, the Gibbs sampler for model parameters is constructed as follows. Note that for the ith measurement equation

$$x_{it} - \rho_i x_{i,t-1} = \lambda_i(L)' \left(f_t - \rho_i f_{t-1}\right) + u_{it},$$

where $\lambda_i(L)'$ is the ith row of $\Lambda(L)$.

Conditional on ρ and R, the factor loadings can be analyzed from simple regression analysis with known variance. Conditional on $\Lambda(L)$, the ρ and R can also be analyzed equation-by-equation for the regression $e_{it} = \rho_i e_{i,t-1} + \sigma_i u_{it}$.

The details are given below.

6.1 Sampling model parameters conditional on the factors

Let θ be the collection of all the model parameters

$$\theta = \{\Lambda_0, \cdots, \Lambda_s, \Phi_1, \cdots, \Phi_h, \rho, R, Q\}.$$

We will again use Jeffreys diffuse priors (Zellner, 1971) for θ. Note that sampling $\{\Phi_j, j = 1, ..., h\}$ and Q conditional on the factors is basically the same as before. After adding serially correlated error terms, we need to revise the Gibbs sampler of $\{\Lambda_0, \cdots, \Lambda_s, \rho, R\}$ conditional on the factors and the identifying restrictions.

Similarly, impose the diffuse independent prior for Λ, ρ, and R:

$$p(\Lambda) = const, \; p(\rho) = const, \; p(\sigma_n^2) \propto 1/\sigma_n^2$$

$$p(\Lambda, \rho, R) = p(\Lambda)\, p(\rho) \left[\Pi_{n=1}^N p(\sigma_n^2) \right].$$

Given the diffuse prior, the posterior for Λ conditional on ρ and R is

$$p(\Lambda | X, F, \rho, R) \propto p(X | F, \Lambda, \rho, R)$$

$$\propto \prod_{i=1}^N exp \left\{ -\frac{1}{2} \frac{\sum_{t=s+2}^T \left(x_{it} - \rho_i x_{i,t-1} - \lambda_i(L)'(f_t - \rho_i f_{t-1}) \right)^2}{\sigma_i^2} \right\}.$$

Let $y_{it} = x_{it} - \rho_i x_{i,t-1}$, $g_{it} = f_t - \rho_i f_{t-1}$. The measurement equation is given by

$$y_{it} = \lambda_i(L)' g_{it} + u_{it}$$

$$= \lambda_i' \begin{bmatrix} g_{it} \\ \vdots \\ g_{i,t-s} \end{bmatrix} + u_{it}$$

$$= \begin{bmatrix} g_{it}' & \cdots & g_{i,t-s}' \end{bmatrix} \lambda_i + u_{it}.$$

Or

$$y_i = G_i \lambda_i + u_i,$$

where

$$G_i = \begin{bmatrix} g_{i,s+2}' & \cdots & g_{i,2}' \\ & \ddots & \\ g_{iT}' & \cdots & g_{i,T-s}' \end{bmatrix} \quad \text{is } (T-s-1) \times (s+1)q.$$

Also note that

$$G_i = \begin{bmatrix} f_{s+2}' & \cdots & f_2' \\ & \ddots & \\ f_T' & \cdots & f_{T-s}' \end{bmatrix} - \rho_i \begin{bmatrix} f_{s+1}' & \cdots & f_1' \\ & \ddots & \\ f_{T-1}' & \cdots & f_{T-s-1}' \end{bmatrix},$$

$$Y = \begin{bmatrix} y_1' \\ \vdots \\ y_N' \end{bmatrix} = X - \begin{bmatrix} \rho_1 & & \\ & \ddots & \\ & & \rho_N \end{bmatrix} X_{-1}.$$

We are able to sample $\lambda'_i = [\lambda'_{i0}, ..., \lambda'_{is}]$ equation by equation:

$$\lambda_i | \lambda_{-i}, R, X, F \sim N\left(\hat{\lambda}_i, \sigma_i^2 \left(G'_i G_i\right)^{-1}\right),$$

where λ_{-i} is λ with λ_i removed, and

$$\hat{\lambda}_i = \left(G'_i G_i\right)^{-1} G'_i y_i.$$

Let $e_{it} = x_{it} - \lambda_i(L)' f_t$, then

$$e_{it} = \rho_i e_{i,t-1} + u_{it}.$$

We obtain the following joint posterior for ρ and R conditional on Λ and F:

$$p\left(\rho, R | X, F, \Lambda\right) \propto \prod_{i=1}^{N} \frac{1}{\left(\sigma_i^2\right)^{\frac{T-s-1}{2}+1}} exp\left\{-\frac{1}{2} \frac{\sum_{t=s+2}^{T}\left(e_{it} - \rho_i e_{i,t-1}\right)^2}{\sigma_i^2}\right\}$$

$$= \prod_{i=1}^{N} \frac{1}{\left(\sigma_i^2\right)^{\frac{T-s}{2}}} \frac{1}{\left(\sigma_i^2\right)^{\frac{1}{2}}} exp\left\{-\frac{1}{2} \frac{\sum_{t=s+2}^{T}\left(e_{it} - \rho_i e_{i,t-1}\right)^2}{\sigma_i^2}\right\}$$

Thus (ρ_i, σ_i^2) can be analyzed on an equation-by-equation basis. Let

$$\hat{u}_{it} = e_{it} - \hat{\rho}_i e_{i,t-1},$$

$$\hat{\rho}_i = \frac{e'_{i,-1} e_i}{e'_{i,-1} e_{i,-1}},$$

$$\hat{u}_i = \left[\hat{u}_{i,s+1}, ..., \hat{u}_{iT}\right]'.$$

In sum

$$p(\sigma_i^2 | X, F, \Lambda) \propto \left(\sigma_i^2\right)^{-(v/2-1)-1} exp\left(-\frac{\hat{u}'_i \hat{u}_i/2}{\sigma_i^2}\right), \quad v = T - s.$$

Thus we may obtain a draw for σ_i^2 one-by-one according to the following inverse-Gamma distribution

$$\sigma_i^2 | X, F, \Lambda \sim invGamma\left(v/2 - 1, \hat{u}'_i \hat{u}_i/2\right), v = T - s.$$

Alternatively, we may drawn the precision $1/\sigma_i^2$ according the Gamma distribution

$$\frac{1}{\sigma_i^2} | X, F, \Lambda \sim Gamma\left(v/2 - 1, (\hat{u}'_i \hat{u}_i/2)^{-1}\right).$$

The posterior distribution of ρ_i is given by

$$\rho_i | \rho_{-i}, X, F, \Lambda, R \sim N\left(\hat{\rho}_i, \sigma_i^2 / \left(e'_{i,-1} e_{i,-1}\right)\right).$$

7 Implications for structural VAR analysis

The dynamic factor model is useful for analyzing the structural VAR (SVAR) models, especially when the variables are measured with errors. Consider a standard SVAR given by

$$A(L) Z_t = a_t$$

where Z_t is a $q \times 1$ vector of economic variables, and a_t is the vector of structural shocks with $E(a_t a_t') = I_q$. Let

$$A(L) = A_0 - A_1 L - \cdots - A_p L^p,$$

with $A_0 \neq I_q$. The reduced form is given by

$$Z_t = B(L) \varepsilon_t,$$

where $a_t = A_0 \varepsilon_t$, or $E(\varepsilon_t \varepsilon_t') = A_0^{-1} (A_0')^{-1}$.

SVAR analysis aims to identify A_0 under structural restrictions. Suppose that Z_t is only observable with measurement error η_t

$$Y_t = Z_t + \eta_t.$$

In this case, it is difficult to analyze the SVAR model. Now suppose that a large number of other observable variables (W_t) are determined by

$$W_t = \Gamma_0 Z_t + \cdots + \Gamma_s Z_{t-s} + e_{wt}.$$

Essentially, W_t is also driven by the fundamental structural shocks a_t through Z_t. This is the essence of comovement for economic variables.

Let

$$X_t = \begin{bmatrix} Y_t \\ W_t \end{bmatrix}, e_t = \begin{bmatrix} \eta_t \\ e_{wt} \end{bmatrix}, f_t = Z_t,$$

$$\Lambda_0 = \begin{bmatrix} I_q \\ \Gamma_0 \end{bmatrix}, \Lambda_j = \begin{bmatrix} 0 \\ \Gamma_j \end{bmatrix}, j \neq 0.$$

Then we have a structural dynamic factor model

$$X_t = \Lambda_0 f_t + \cdots + \Lambda_s f_{t-s} + e_t, \tag{23}$$
$$B^{-1}(L) f_t = \varepsilon_t.$$

According to our identification scheme DFM2, equation (23) is identified and can be analyzed using a Bayesian approach. In particular, without further assumptions, we are able to estimate $f_t = Z_t$, $B(L)$, Λ_i, and $E(\varepsilon_t \varepsilon_t') = A_0^{-1} (A_0')^{-1}$. We may also incorporate additional structural restrictions as in standard SVAR analysis, so as to estimate A_0.

8 Simulation study

8.1 Serially uncorrelated measurement errors

This section uses some simulation studies to show the effectiveness of the Bayesian estimation method given our identifying restrictions. We first consider a dynamic factor model with $q = 2, s = 1, h = 2$. In particular, the data is generated according to (1) and (2):

$$X_t = \Lambda_0 f_t + \Lambda_1 f_{t-1} + \cdots + \Lambda_s f_{t-s} + e_t,$$
$$f_t = \Phi_1 f_{t-1} + \cdots + \Phi_h f_{t-h} + \varepsilon_t,$$

where the parameters are draw from

$$\lambda_{ijs} \sim i.i.d. \quad U(0, 1),$$
$$e_t \sim i.i.d. \quad N(0, R),$$
$$\varepsilon_t \sim i.i.d. \quad N(0, Q).$$

We choose $R = I_N, Q = I_q$, and

$$[\Phi_1, \Phi_2] = \begin{bmatrix} 0.5 & 0 & 0.2 & 0 \\ -0.1 & 0.2 & 0.1 & 0.1 \end{bmatrix}.$$

Besides our identifying restrictions DFM1, we also impose extra restrictions on the factor loadings. For every five rows, we assume that the first row is only affected by the first dynamic factors, the second row being affected by the second dynamic factors only, the third row being affected by lagged factors only, the fourth row being affected by current factors only, the fifth row being unrestricted. Thus the matrix of factor loadings is restricted as follows

$$\Lambda = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ * & * & * & * \\ * & 0 & * & 0 \\ 0 & * & 0 & * \\ 0 & 0 & * & * \\ * & * & 0 & 0 \\ * & * & * & * \\ \vdots & \vdots & \vdots & \vdots \end{bmatrix}$$

where "$*$" denotes the position of free parameters.

With the sample size chosen to be $N = 10, 20, 50, 100, 150, 200$ and $T = 50, 100, 200$, we first generate a sample of size $N = 200, T = 200$. Then we estimate the model according to eighteen samples with different combinations of N and T. We use the

Jeffreys diffuse prior throughout the estimation. The length of the Gibbs sampling chain is chosen to be ten thousand.[5] We discard the first five thousand draws before calculating the posterior mean and standard deviations. We calculate the posterior mean as our point estimates for the dynamic factors and compare them to the true factors. To evaluate the estimation outcome, we project the true factors on the estimates to obtain the adjusted R-squared of a regression through the origin. For example, for the j-th factor, we regress f_j on $\hat{f}_j$ to obtain the adjusted R-squared $\bar{R}_j^2$. We also regress f_j on both $\hat{f}_1$ and $\hat{f}_2$ to obtain $\bar{R}_{j,space}^2$ to measure the closeness of factors to the estimated factor space. Figure 1 shows the result. In each one of the four graphs, there are three lines, each connecting the adjusted R-squared of different N for a given T. For example, the solid line shows the pattern how $\bar{R}_j^2$ or $\bar{R}_{j,space}^2$ increases over the cross-section dimension N given $T = 200$. Such a pattern is in line with the theory on large dimensional factor models (such as Bai and Ng, 2002, Bai, 2003). The other two lines show the same pattern given a time dimension of either $T = 50$ or 100. In each graph, all three lines are very close to each other, especially when N is large. This implies that for the given data generating process, the estimated factors reasonably well approximate the true factors and the overall estimation precision mainly comes from the cross-section dimension N.

For the sample of size $N = 10, T = 200$, Figure 2 and Figure 3 compare the Bayesian posterior mean of the factors with the true factors, as well as providing the two SE error bands of the posterior mean. Figure 4 and Figure 5 conduct the same exercise for $N = 100, T = 200$. The factors are more precisely estimated when the sample has a larger N. This is best seen as the much tighter two SE bands around the true factors. This is also in line with the theory on large dimensional factor models. We then compare the two SE error bands of the impulse responses of each dynamic factor given one standard deviation change of each shock in the state equation. The SE is calculated as the sample standard deviation of the posterior draws of the impulse response function. It is worth mentioning that other methods such as Sims and Zha (1999) can be applied as well, which we do not explore here. Figures 6 and 7 show that with a larger N the error bands tend to be narrower.

8.2   Serially correlated measurement errors

This section explores implications of serially correlated error terms for the analysis of the dynamic factors. We use the same data generating process as the previous section, except that the error terms in (1) follow an AR(1) process:

$$e_{it} = \rho_i e_{i,t-1} + u_{it}, \ u_{it} \sim i.i.d.N(0, \sigma_i^2).$$

The number of dynamic factors is still set at two. We fix $N = 50, T = 200, s = 1, h = 2$. We compare the estimation results from two methods, one with ρ_i being

---

[5]We found that our Bayesian algorithm converges very fast. The chain size of 100 thousand offers basically the same results. In addition, for the model with $N = 200, T = 200$, the CPU time for ten thousand chain size is less than 1200 seconds using an Intel Core i7 laptop with 2GB DDR3 memory.

Figure 1: Adjusted R-squared from regressing the true factors on the estimates
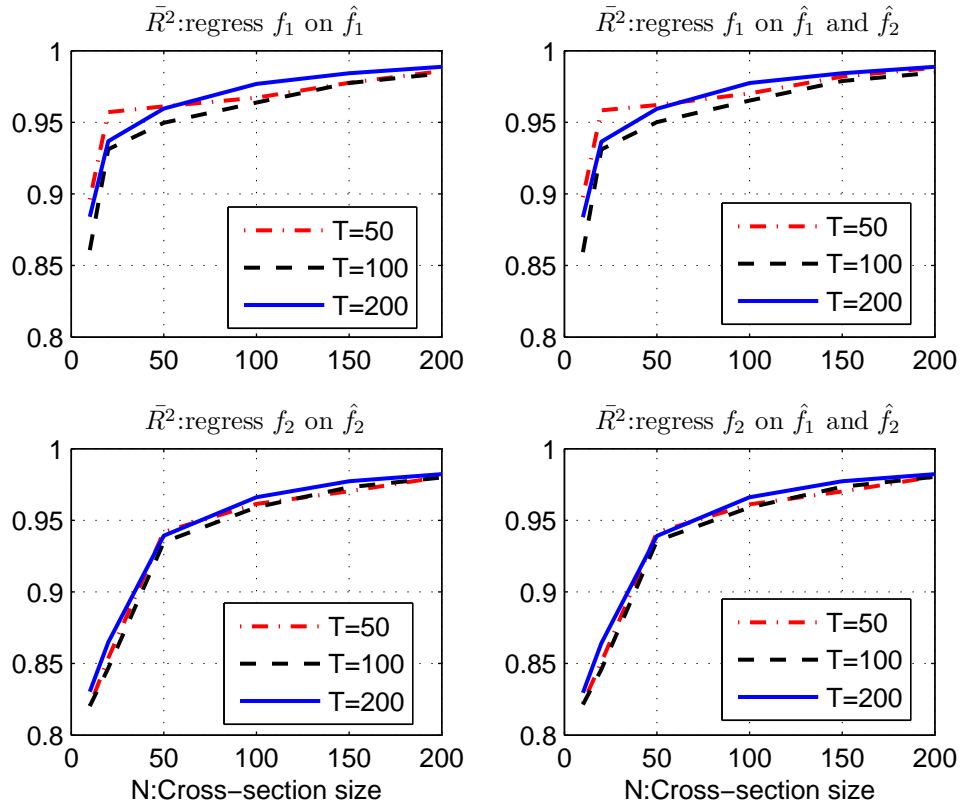
Table 1: Fit of dynamic factors: $\bar{R}_j^2$

	For $j = 1, 2$, regress f_j on				
	$\hat{f}_j$			$\left\{\hat{f}_1, \hat{f}_2\right\}$	
ρ_i	Fix $\rho_i = 0$	Free ρ_i		Fix $\rho_i = 0$	Free ρ_i
0.99	N.A.	(0.8329, 0.7929)		N.A.	(0.8345, 0.7950)
0.9	(0.6211, 0.0010)	(0.8998, 0.8659)		(0.6489, 0.3718)	(0.9007, 0.8714)
0.8	(0.7160, 0.1693)	(0.8948, 0.8127)		(0.7353, 0.4806)	(0.8958, 0.8172)
0.5	(0.9220, 0.9106)	(0.9350, 0.9277)		(0.9218, 0.9132)	(0.9353, 0.9274)
0.0	(0.9557, 0.9319)	(0.9542, 0.9293)		(0.9555, 0.9316)	(0.9540, 0.9291)

Remark: The two numbers within each pair of parentheses are the adjusted R^2 of regressing the true f_1 and f_2, respectively, on the corresponding estimated factors (the left pane) or the entire factor space (the right pane). N.A. means the MCMC chain fails to converge.

free parameters to be estimated, the other being under the assumption that $\rho_i = 0$ for all i. Thus the first estimation is conducted under the true model, while the second one is under the assumption of serially uncorrelated error terms and thus is using a misspecified model. This exercise helps us to understand how serial correlation in the error terms affects the estimation of the dynamic factors.

We use the adjusted R-squared to measure the overall fit of estimated factors. We project the true factors on the corresponding estimated factors to obtain one set of $\bar{R}^2$. We also project the true factors on the entire estimated factor space to obtain another set of $\bar{R}^2$. Table 1 reports the results. To fix idea, we choose $\rho_i = \rho$ for all $i = 1, ..., N$ and examine the adjusted $\bar{R}^2$ for different values of $\rho \in \{0.99, 0.9, 0.8, 0.5, 0.0\}$.

From the table, we notice that for not so persistent measurement errors, say $\rho \leq 0.5$, both estimation methods provide reasonably well fit. That is, ignoring the serial correlation still gives good fit. However, as the persistence increases, the fit of factors deteriorates when ignoring the serial correlation. For example, when $\rho \geq 0.8$, the fit becomes rather poor when ρ_i is set to zero. On the other hand, across all cases, the estimation taking into account serial correlation of error terms performs uniformly well in terms of fit, even in the near unit root case.

8.3 Model comparison

This section considers the fractional Bayes factors to calculate the marginal likelihood of data under different model specifications. We use the same data generating process as in Section 8.1. The true model is specified as $N = 100, T = 200, q = 2, s = 1, h = 2$. We use $T_0 = 20$ to obtain the prior from the training sample, and then use the prior to obtain the marginal likelihood for the sample $T = 21, ..., 200$. To fix idea, we fix $s = 1, h = 2$ and only vary the number of factors. Table 2 reports the associated marginal likelihood of the data. The marginal likelihood achieves maximum when the number of factors is the same as the true value $q = 2$.

Table 2: Model comparison: logarithm of marginal likelihood.

q	1	2	3
log(marginal likelihood)	-13659	-13473	-13708

9 Comovement in international bond yields: an application

In this section, we apply the dynamic factor models to the analysis of international bond yields. We adopt the monthly nominal zero-coupon bond yields data constructed by Wright (2011), which covers nine countries from 1971 January to 2009 May. We will use a balanced panel from December 1992 to May 2009. The nine countries are US, Canada, UK, Germany, Sweden, Switzerland, Japan, Australia, and New Zealand. For each country, the data are constructed at 60 maturities from 3 months to 180 months, except for Sweden with 40 maturities. Let $X_t^c(\tau)$ be the time t bond yield of country c at maturity τ. For $c = 1, ..., C$, assume that a global factor g_t and a country specific factor f_t^c combine to affect the bond yield:

$$X_t^c(\tau) = \mu^c(\tau) + \gamma^c(\tau) g_t + \lambda^c(\tau) f_t^c + e_t^c(\tau), \ e_t^c(\tau) \sim iid.N\left(0, \sigma_c^2(\tau)\right). \quad (24)$$

We further assume that the factors follow a VAR(1) process. In particular, let $F_t = \left[g_t, f_t^1, ..., f_t^C\right]'$, then

$$F_t = \Phi F_{t-1} + \varepsilon_t, \ \varepsilon_t \sim N(0, Q). \quad (25)$$

Assuming that $Q = I_{C+1}$, both $\gamma^c(\tau)$ and $\lambda^c(\tau)$ being strictly positive for $\tau = 3$ months, then the multi-level factor model is identified. Using a similar but different model, Diebold, Li, and Yue (2008) found an economically significant global yield factor for Germany, Japan, UK and US. Our empirical exercise differs from theirs in two important ways. Firstly, we adopt a much larger data set with nine countries and more maturities, which expects to better represent the global bond yield. Secondly, we treat the AR coefficient matrix Φ as unrestricted, which allows us to study the spill-over effects of one country to another. On the other hand, almost all existing literature on multi-level factor models assumes that different factors are independent of each other. We think this is restrictive, because the country factors could be correlated with each other due to regional interactions or other economic linkages such as cross-border capital flows. Model (24) and (25) are estimated using the Bayesian method described in the previous sections. The posterior mean of the global factor g_t along with two standard deviation error bands is reported in Figure 8. The large cross-section size N leads to very narrow error bands for both the global factor and the country factors. To save space, we do not report results for country factors. We also reports the variance decomposition results for each country in Table 3.[6] The result is similar to Diebold, Li, and Yue (2008) in that the global factor

---

[6]We calculate the variances due to the global factor, the country factors, and the idiosyncratic errors respectively. And then use the variance ratio to measure the variance contribution due to each component.

Table 3: Variance decomposition for nine countries.

	US	Canada	UK	Germany	Sweden
Global factor	52.07%	74.08%	77.11%	75.25%	83.40%
Country factor	44.16%	23.97%	20.96%	23.76%	15.84%
Idiosyncratic error	3.77%	1.94%	1.93%	0.99%	0.76%
	Switzerland	Japan	Australia	New Zealand	
Global factor	63.17%	62.61%	69.69%	50.97%	
Country factor	35.07%	36.48%	28.72%	44.04%	
Idiosyncratic error	1.76%	0.91%	1.59%	4.99%	

explains a large proportion of variance in country bond yields, and compared with most other markets the US has a stronger country factor.

We conduct a further variance decomposition exercise of country yields at four representative maturities of 3, 12, 60, and 120 months. Figure 9 shows the results for all nine countries in our sample. For all countries considered, the global factor tends to explain more variation at longer maturities, while the country factor shows the opposite pattern. For Sweden, the majority of variation is explained by the global factor.

Given the multi-level factor structure, we are particularly interested in how different factors interact with each other. To estimate the impulse response functions, we use the median to denote the point estimate, and lower and upper five percentile to denote the error bands around the point estimate. Figure 10 shows the impulse responses of the global factor g_t up to 24 months to one standard deviation shock of each one of the nine country factors. We report the impulse responses of the country factors f_t^c in Figures 11 to 19.

To briefly summarize, the global factor does not seems to response much to country factors, except for some mild responses to several countries as can be seen from Figure 10. On the other hand, there are significant interactions between country factors based on evidence from Figures 11 to 19. For example, the shock originated from the US factor seems to significantly affect most other countries, especially Canada and the European countries. The Canada factor, however, only has notable impact on the Sweden factor, which might be due to their close economic relations. In the meantime, the Sweden factor affects the Canada factor among several other countries. Such rich evidence in the impulse response functions points to complicated interstate relations, which could not be simply accounted for by the geographic relations. Canada, usually treated as a member of the North America region only, are closely related to both US and Sweden. To see another example, there is strong relation between the New Zealand factor and the Australia factor. Both factors also strongly affect the Sweden factor but are only mildly affected by the latter. Our modelling strategy naturally allows the study of such complicated pairwise relations between countries, which are not strong enough to be summarized by global factors or by geographically defined regional factors. Most literature on dynamic factor models assumes that factors are independent of each other, and thus is not able to

Figure 2: N=10, T=200. Point estimates of the factors: Posterior mean
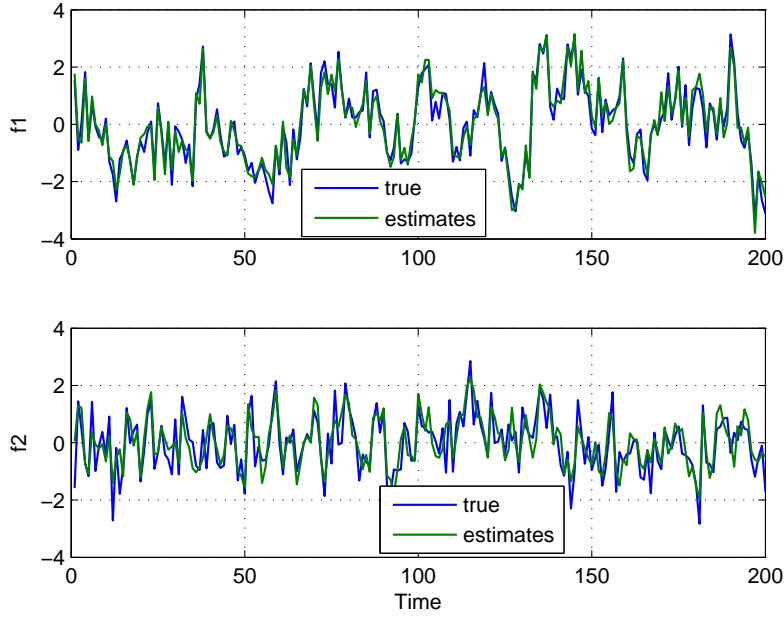
study such rich interstate relations.

10   Conclusion

In this paper, we discuss the minimal requirement of identifying the dynamic factor models. Our identification scheme allows the interaction among different factors, which provides a useful framework to study the structural implications of the factor model. In addition, a more general class of impulse response functions can be derived under our framework, the confidence bands of which are readily constructed from the Bayesian estimation results. The proposed identification scheme naturally combines the structural VAR analysis as well. We also derive the Bayes factors that help determine the number of dynamic factors as well as the order of lags in the dynamic factor model.

Figure 3: N=10, T=200. Interval estimates of the factors: two SE error bands

Two sigma bands (shaded area) V.S. true factors, (N=10, s=1, h=2)
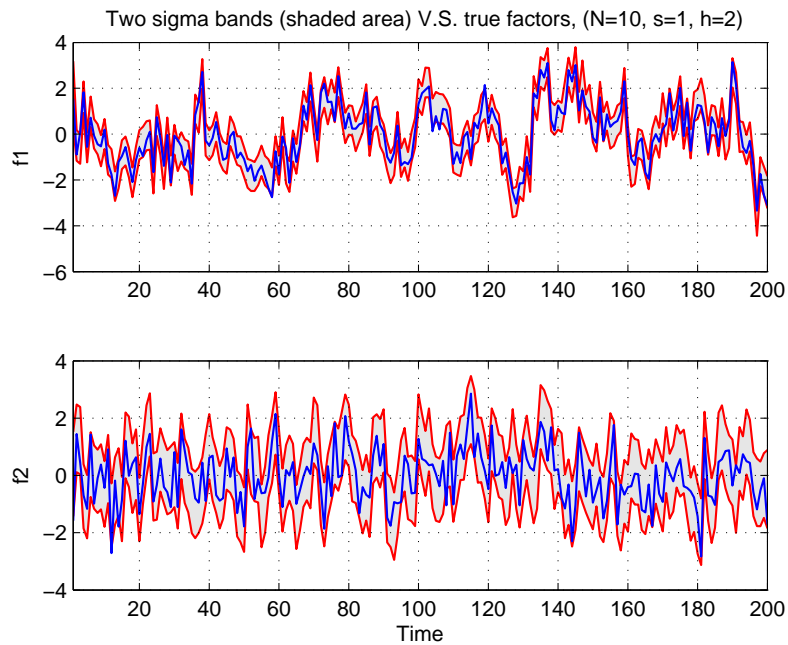
Figure 4: N=100,T=200. Point estimates of the factors: Posterior mean

34
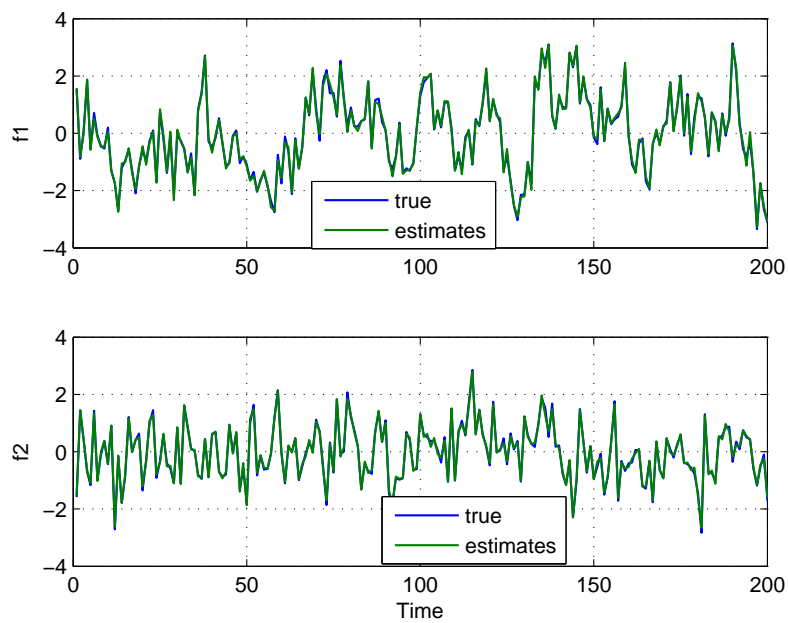
Figure 5: N=100,T=200. Interval estimates of the factors: two SE error bands

Two sigma bands (shaded area) V.S. true factors, (N=100, s=1, h=2)
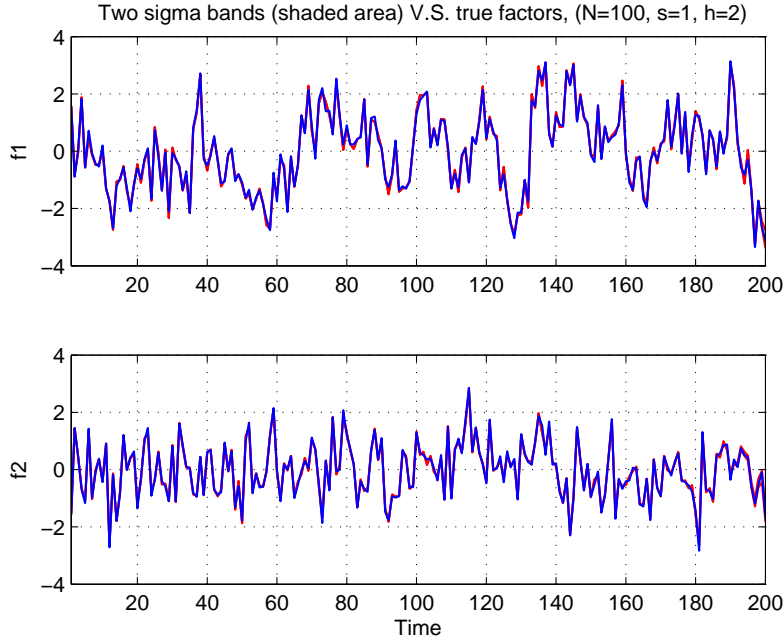
Figure 6: Interval estimates of the impulse response function given one std change of f_1's innovation.
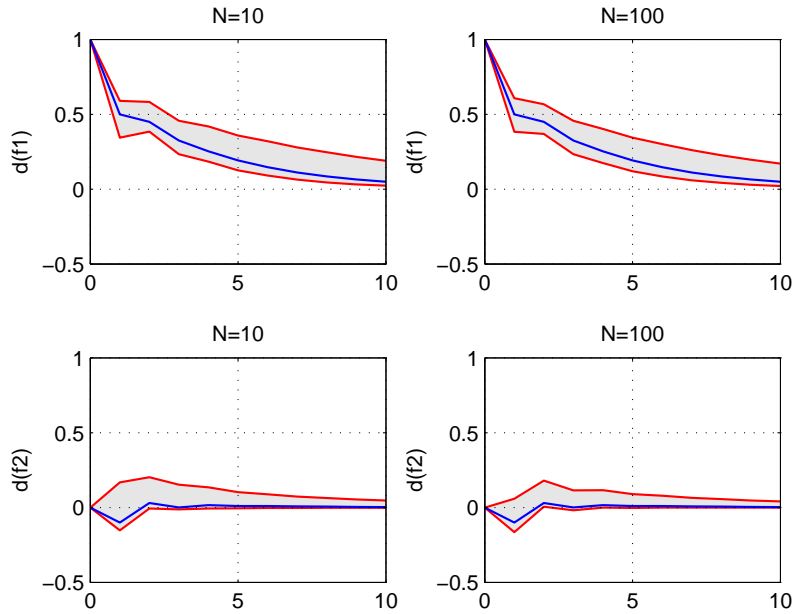
Remark: the blue line represents the true impulse response function.

Figure 7: Interval estimates of the impulse response function given one std change of f_2's innovation.
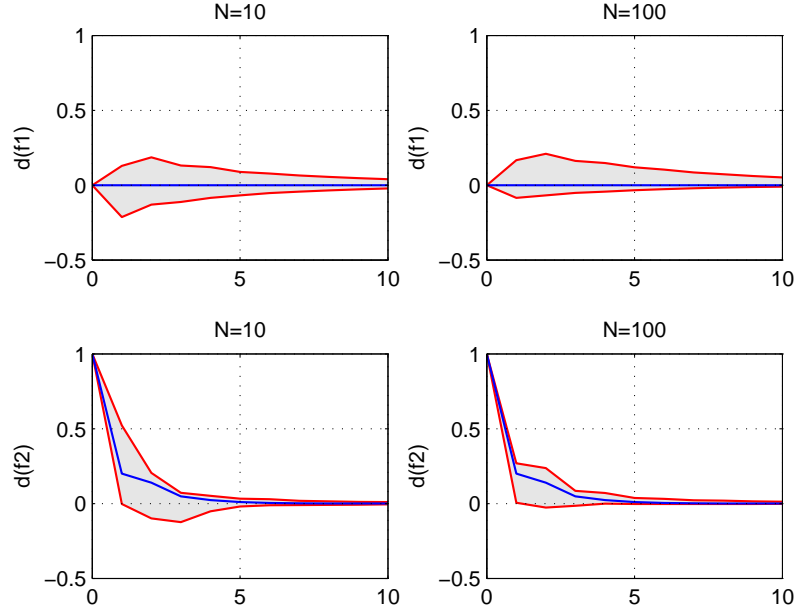
Figure 8: The global yield factor: posterior median and 90% confidence bands.
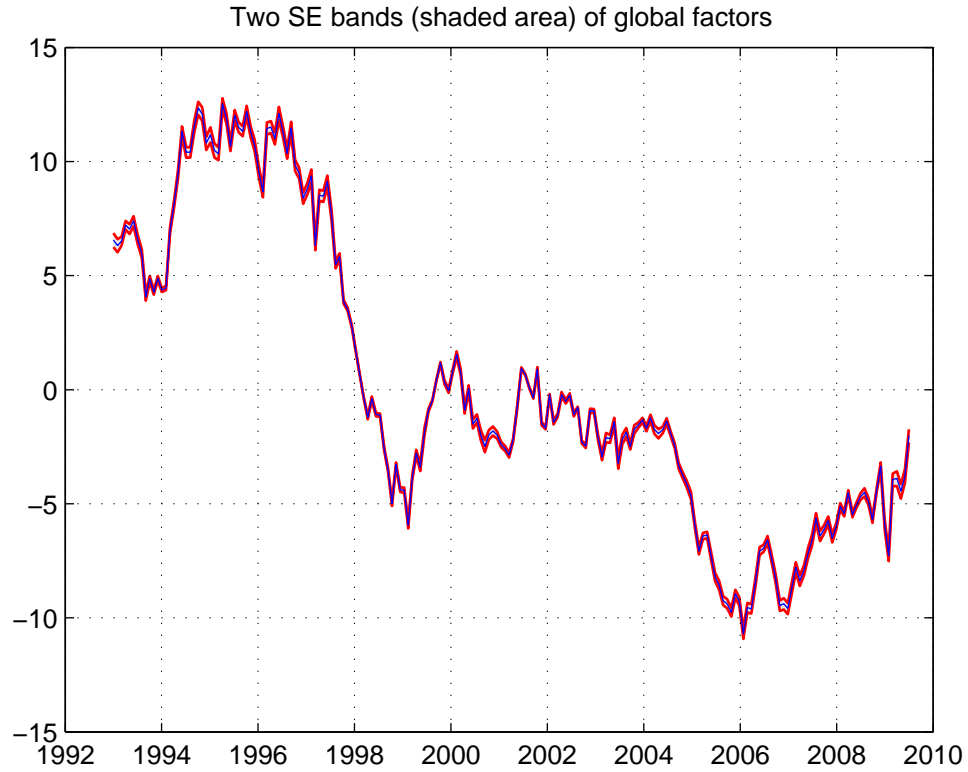
Figure 9: Variance decomposition of country yields at maturities of 3, 12, 60, 120 months.
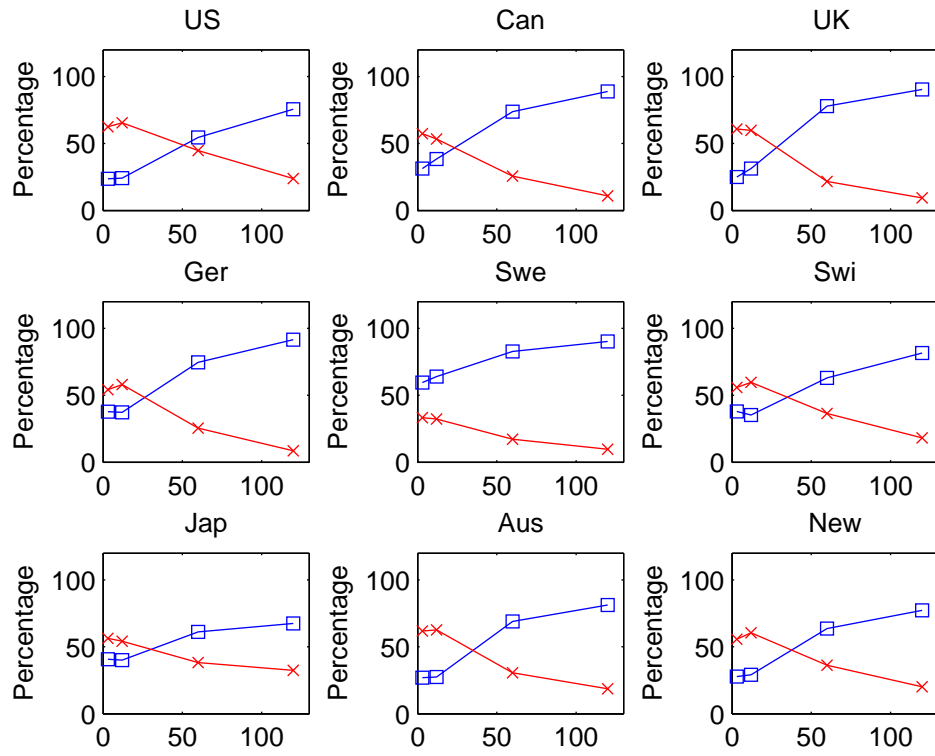
Remark: the blue line with squares represents the variance explained by the global factor, and the red line with crosses by the country factors.

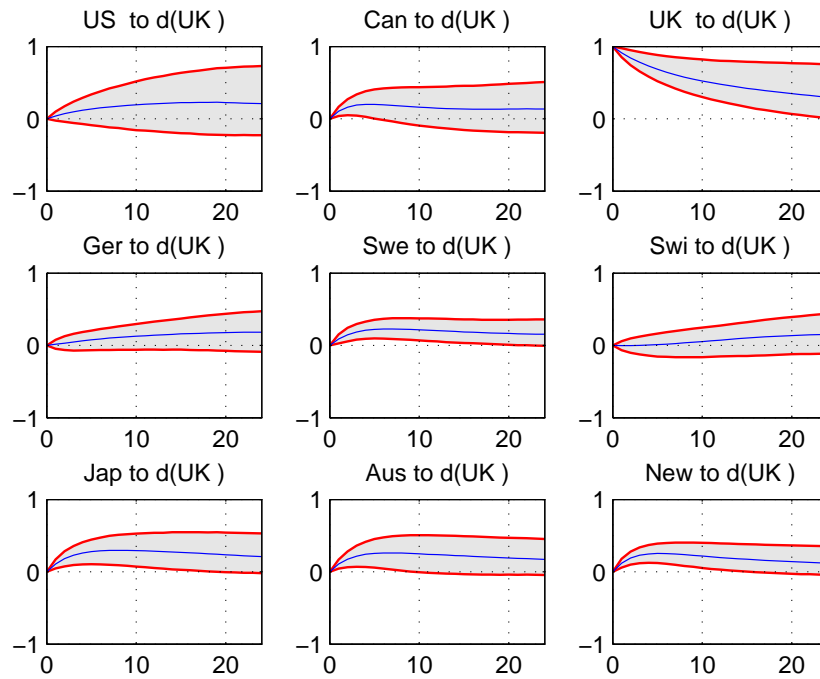Figure 10: Impulse responses of global factor to shocks of country factors.
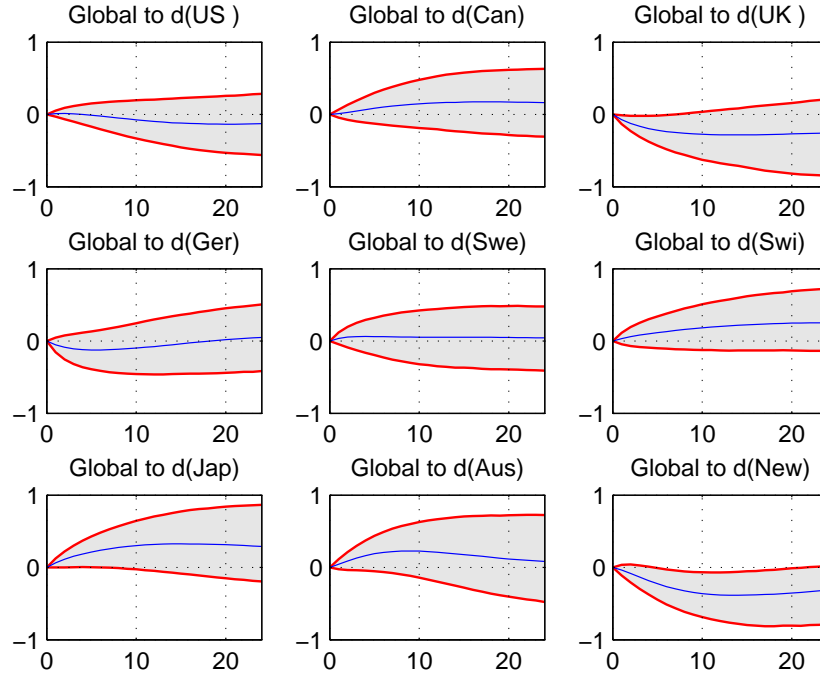
Figure 11: Impulse responses of country factors given one unit shock to US factor.
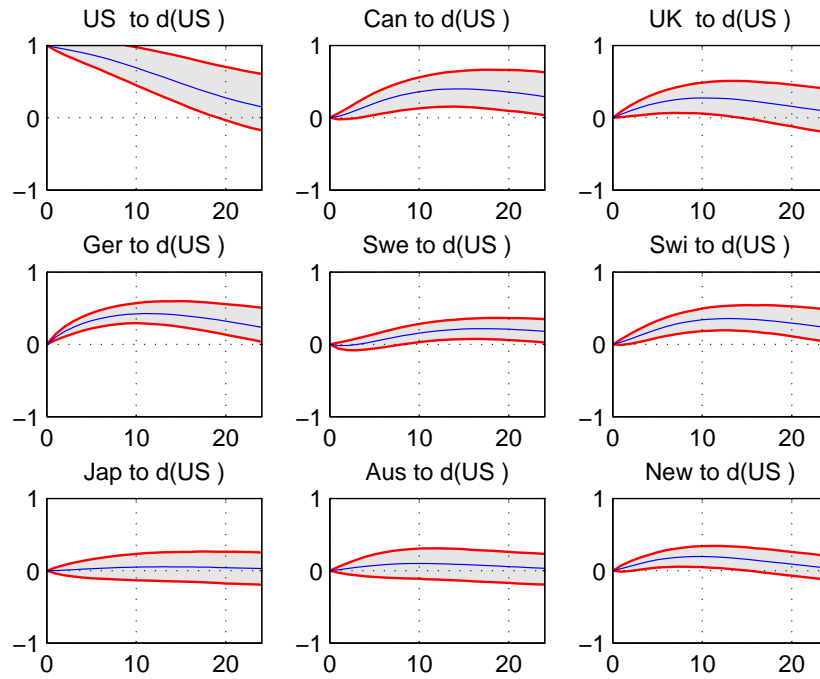
Figure 12: Impulse responses of country factors given one unit shock to Canada factor.
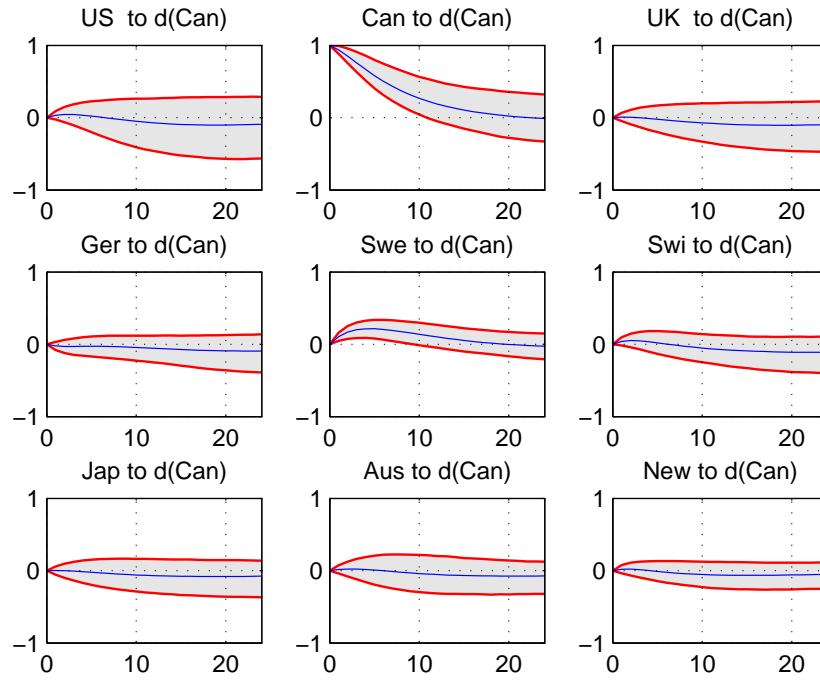
Figure 13: Impulse responses of country factors given one unit shock to UK factor.

Figure 14: Impulse responses of country factors given one unit shock to Germany factor.
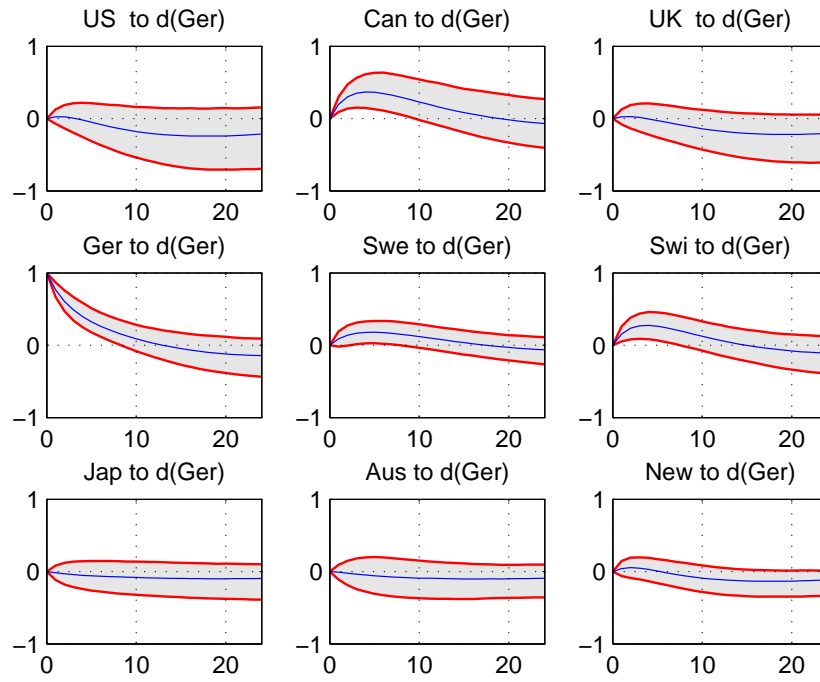

Figure 15: Impulse responses of country factors given one unit shock to Sweden factor.
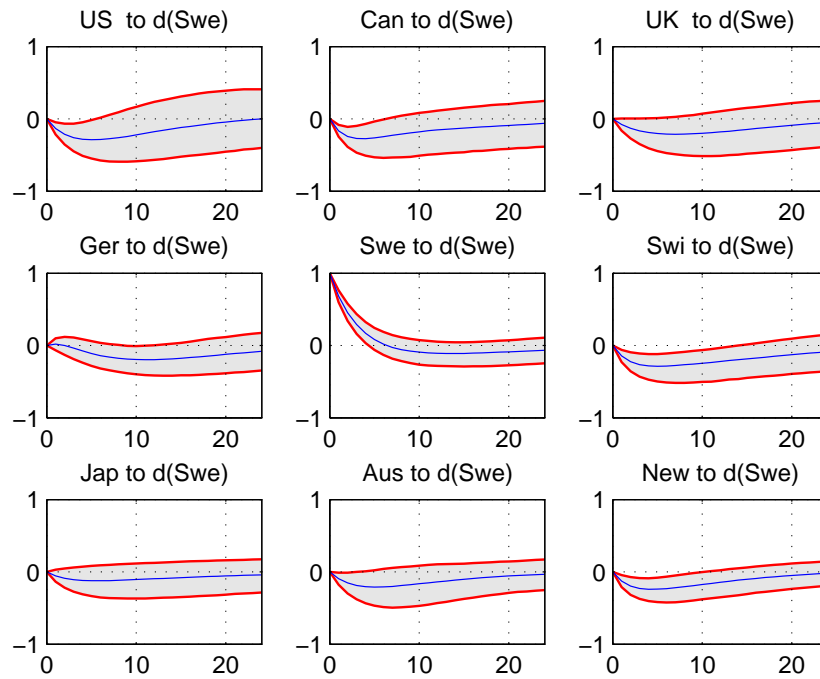
Figure 16: Impulse responses of country factors given one unit shock to Switzerland factor.
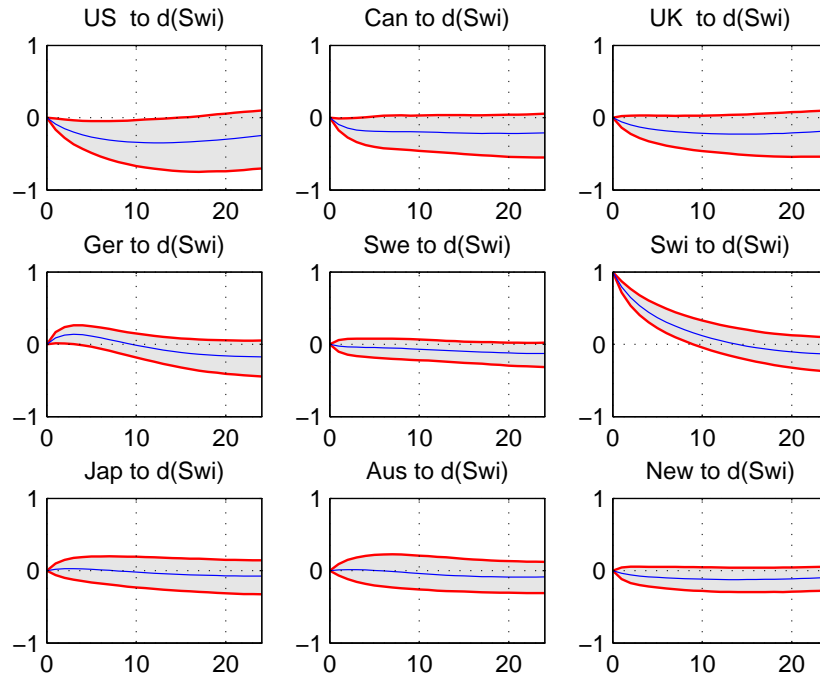

Figure 17: Impulse responses of country factors given one unit shock to Japan factor.
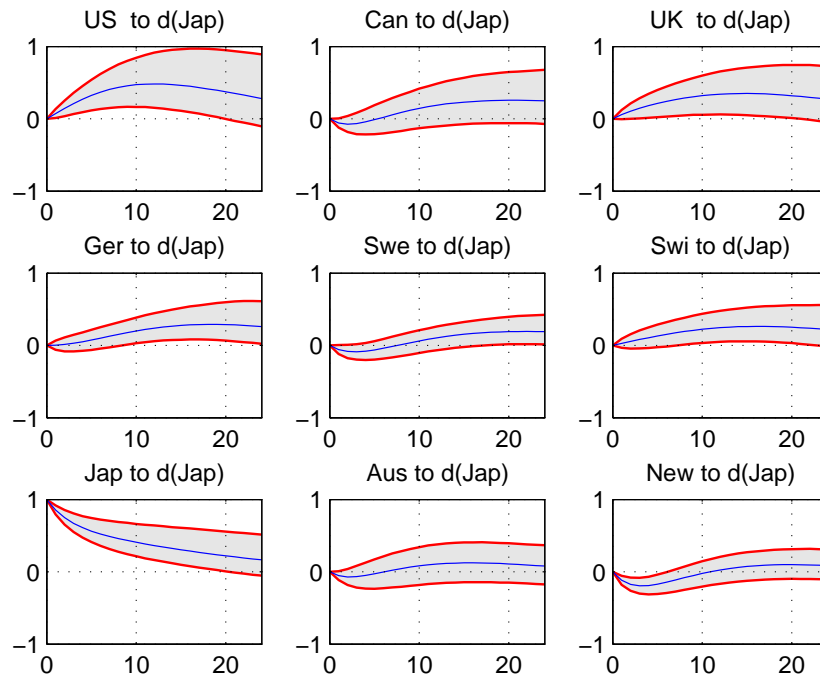
Figure 18: Impulse responses of country factors given one unit shock to Australia factor.
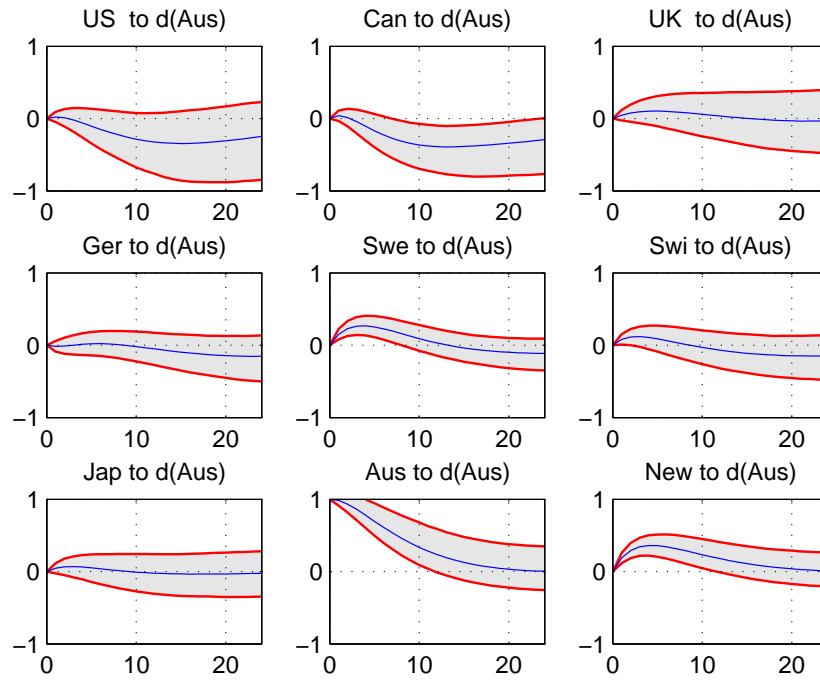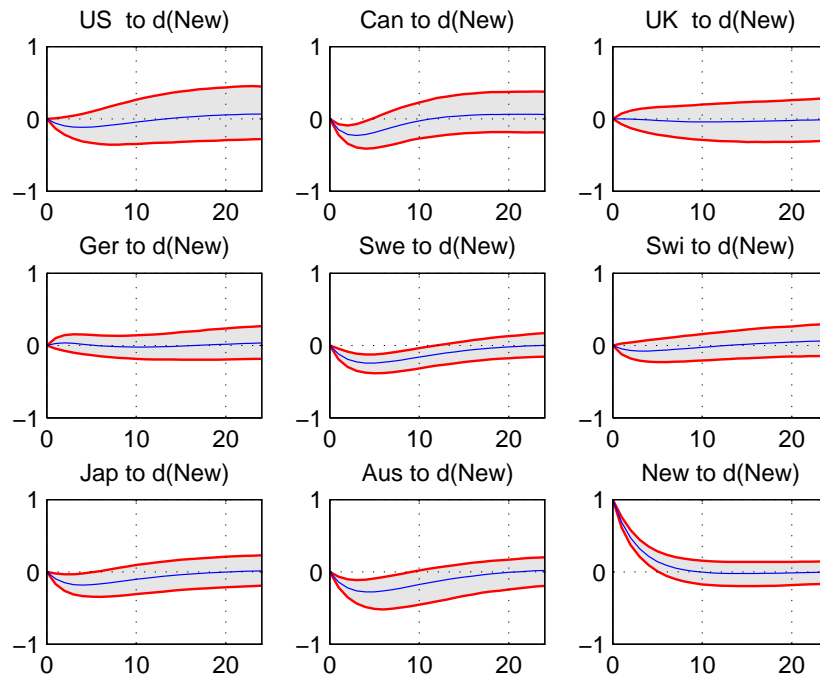

Figure 19: Impulse responses of country factors given one unit shock to New Zealand factor.

11 Appendix A: Proof of Propositions

Proof of Proposition 1: Let A be a full rank $q \times q$ rotation matrix. Left-multiply the dynamic factors f_t by A and right-multiply the loading matrix Λ_j by A^{-1}, $j \in \{0, 1, ..., s\}$. After the rotation, the new factors $\tilde{f}_t = Af_t$ have a $VAR(h)$ representation given below

$$\tilde{f}_t = A\Phi_1 A^{-1}\tilde{f}_{t-1} + \cdots + A\Phi_h A^{-1}\tilde{f}_{t-h} + A\varepsilon_t.$$

The observation equation (1) after the rotation becomes

$$X_t = \Lambda_0 A^{-1}\tilde{f}_t + \Lambda_1 A^{-1}\tilde{f}_{t-1} + \cdots + \Lambda_s A^{-1}\tilde{f}_{t-s} + e_t.$$

If under the current normalization, the only admissible A is a diagonal matrix with either 1 or -1 on the diagonal, then both the factors and factor loadings are identified up to a sign change. The normalization $var(A\varepsilon_t) = I_q$ implies that

$$AA' = I_q \tag{26}$$

or A is an orthonormal matrix. Next, let

$$\Lambda_0 = \begin{bmatrix} \lambda_{11} & 0 & \cdots & 0 \\ \lambda_{21} & \lambda_{22} & \ddots & \vdots \\ \vdots & \vdots & \ddots & 0 \\ \lambda_{q1} & \cdots & \cdots & \lambda_{qq} \\ \vdots & \cdots & \cdots & \vdots \\ \lambda_{N1} & \cdots & \cdots & \lambda_{Nq} \end{bmatrix}, A^{-1} = \begin{bmatrix} a_{11} & \cdots & a_{1q} \\ \vdots & \ddots & \vdots \\ a_{q1} & \cdots & a_{qq} \end{bmatrix}.$$

Then the normalization of factor loadings requires $\Lambda_0 A^{-1}$ to be a lower triangular matrix, i.e.,

$$\begin{bmatrix} \lambda_{11} & 0 & \cdots & 0 \\ \lambda_{21} & \lambda_{22} & \ddots & \vdots \\ \vdots & \vdots & \ddots & 0 \\ \lambda_{q1} & \cdots & \cdots & \lambda_{qq} \\ \vdots & \cdots & \cdots & \vdots \\ \lambda_{N1} & \cdots & \cdots & \lambda_{Nq} \end{bmatrix} \begin{bmatrix} a_{11} & \cdots & a_{1q} \\ \vdots & \ddots & \vdots \\ a_{q1} & \cdots & a_{qq} \end{bmatrix} = \begin{bmatrix} \lambda_{11}^* & 0 & \cdots & 0 \\ \lambda_{2}^* & \lambda_{22}^* & \ddots & \vdots \\ \vdots & \vdots & \ddots & 0 \\ \lambda_{q1}^* & \cdots & \cdots & \lambda_{qq}^* \\ \vdots & \cdots & \cdots & \vdots \\ \lambda_{N1}^* & \cdots & \cdots & \lambda_{Nq}^* \end{bmatrix}, \tag{27}$$

from which we obtain $a_{ij} = 0$ for any i, j such that $i < j$, or A^{-1} is lower triangular given the assumption that $\lambda_{ii} \neq 0, \lambda_{ii}^* \neq 0, i = 1, ..., q$. Also note that equation (26) implies $A^{-1} \cdot (A^{-1})' = I_q$. Use $(A^{-1})'$ to right multiply both sides of equation (27) to obtain

$$\begin{bmatrix} \lambda_{11} & 0 & \cdots & 0 \\ \lambda_{21} & \lambda_{22} & \ddots & \vdots \\ \vdots & \vdots & \ddots & 0 \\ \lambda_{q1} & \cdots & \cdots & \lambda_{qq} \\ \vdots & \cdots & \cdots & \vdots \\ \lambda_{N1} & \cdots & \cdots & \lambda_{Nq} \end{bmatrix} = \begin{bmatrix} \lambda_{11}^* & 0 & \cdots & 0 \\ \lambda_{2}^* & \lambda_{22}^* & \ddots & \vdots \\ \vdots & \vdots & \ddots & 0 \\ \lambda_{q1}^* & \cdots & \cdots & \lambda_{qq}^* \\ \vdots & \cdots & \cdots & \vdots \\ \lambda_{N1}^* & \cdots & \cdots & \lambda_{Nq}^* \end{bmatrix} \begin{bmatrix} a_{11} & \cdots & a_{q1} \\ \vdots & \ddots & \vdots \\ a_{1q} & \cdots & a_{qq} \end{bmatrix},$$

from which we obtain $a_{ij} = 0$ for any i, j such that $i > j$. In sum we have proved that

$$A^{-1} = diag\{a_{11}, ..., a_{qq}\}.$$

Given that $A^{-1} \cdot (A^{-1})' = I_q$, we obtain $a_{ii} = 1$ or -1 for all $i = 1, ..., q$. So the rotation matrix A is also a diagonal matrix with either 1 or -1 on the diagonal.

This proves that the proposed normalization DFM1 identifies both the dynamic factors and the corresponding dynamic factor loadings up to a sign change.

As a normalization, we may assume $\lambda_{ii} > 0$ so that both the dynamic factors and the associated dynamic factor loadings are fully identified. **Q.E.D.**

Proof of Proposition 2: Let A be a full rank $q \times q$ rotation matrix A. Left-multiply the dynamic factors f_t by A and right-multiply the loading matrix Λ_j by A^{-1}, $j \in \{0, 1, ..., s\}$. If under DFM2, the only admissible A is I_q, then both the factors and factor loadings are uniquely identified. DFM2 implies that both Λ_0 and the new dynamic factor loading $\tilde{\Lambda}_0$ must have I_q as its upper $q \times q$ block. This implies $I_q A^{-1} = I_q$, or $A^{-1} = I_q$, or $A = I_q$. **Q.E.D.**

Proof of Proposition 3: By equation (8),

$$
\begin{bmatrix} X_t^1 \\ \vdots \\ X_t^C \end{bmatrix} =
\begin{bmatrix} \Lambda_W^1 & \Lambda_C^1 & & \\ \vdots & & \ddots & \\ \Lambda_W^C & & & \Lambda_C^C \end{bmatrix}
\begin{bmatrix} f_t^W \\ f_t^1 \\ \vdots \\ f_t^C \end{bmatrix} +
\begin{bmatrix} e_t^1 \\ \vdots \\ e_t^C \end{bmatrix}. \tag{28}
$$

Let R be a rotation matrix. Left-multiply the factors by R and right-multiply the loading matrix by R^{-1}. If under the current normalization, the only admissible R is a diagonal matrix with either 1 or -1 on the diagonal, then both the factors and factor loadings are identified up to a sign change. From Wang (2010), the multi-level factor structure implies that the only admissible rotation matrix R takes the following form,

$$
R = \begin{bmatrix} A & 0 & \cdots & 0 \\ B_1 & A_1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ B_C & 0 & \cdots & A_C \end{bmatrix}.
$$

Denote

$$
H_t = \begin{bmatrix} f_t^W \\ f_t^1 \\ \vdots \\ f_t^C \end{bmatrix}.
$$

Then the DGP for the factors implies that H_t follows a VAR(1) process with the innovation being i.i.d. $N(0, I)$,

$$H_t = \rho H_{t-1} + U_t, \quad U_t \sim N(0, I).$$

After the rotation, we obtain $\tilde{H}_t = RH_t$. And $\tilde{H}_t$ follows a VAR(1) process given below,

$$\tilde{H}_t = R\rho R^{-1}\tilde{H}_{t-1} + \tilde{U}_t, \quad \tilde{U}_t = RU_t \sim N(0, RR').$$

where the normalization on the innovation of factors implies that

$$RR' = \begin{bmatrix} A & 0 & \cdots & 0 \\ B_1 & A_1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ B_C & 0 & \cdots & A_C \end{bmatrix} \begin{bmatrix} A' & B_1' & \cdots & B_C' \\ 0 & A_1' & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & A_C' \end{bmatrix} = I.$$

This in term implies that

$$B_1 = \cdots = B_C = 0, \text{ and}$$
$$AA' = I, A_c A_c' = I, c = 1, ..., C. \tag{29}$$

In sum we have

$$R = \begin{bmatrix} A & 0 & \cdots & 0 \\ 0 & A_1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & A_C \end{bmatrix}.$$

We then show that the assumption that Λ_W^1 is lower triangular implies A is a diagonal matrix with either 1 or -1 on the diagonal. Notice that the inverse of R has the following form

$$R^{-1} = \begin{bmatrix} A^* & 0 & \cdots & 0 \\ 0 & A_1^* & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & A_C^* \end{bmatrix},$$

where $A^* = A^{-1}, A_c^* = A_c^{-1}$. Then after the rotation the factor loading matrix in (28) becomes

$$\begin{bmatrix} \Lambda_W^1 & \Lambda_C^1 & & \\ \vdots & & \ddots & \\ \Lambda_W^C & & & \Lambda_C^C \end{bmatrix} \begin{bmatrix} A^* & 0 & \cdots & 0 \\ 0 & A_1^* & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & A_C^* \end{bmatrix} = \begin{bmatrix} \Lambda_W^1 A^* & \Lambda_C^1 A_1^* & & \\ \vdots & & \ddots & \\ \Lambda_W^C A^* & & & \Lambda_C^C A_C^* \end{bmatrix}$$

Let

$$\Lambda_W^1 = \begin{bmatrix} \lambda_{11} & 0 & \cdots & 0 \\ \lambda_{21} & \lambda_{22} & \ddots & \vdots \\ \vdots & \vdots & \ddots & 0 \\ \lambda_{k1} & \cdots & \cdots & \lambda_{kk} \\ \vdots & \cdots & \cdots & \vdots \\ \lambda_{n_1 1} & \cdots & \cdots & \lambda_{n_1 k} \end{bmatrix}, A^* = \begin{bmatrix} a_{11} & \cdots & a_{1k} \\ \vdots & \ddots & \vdots \\ a_{k1} & \cdots & a_{kk} \end{bmatrix}.$$

Then the normalization of factor loadings implies $\Lambda_W^1 A^*$ is a lower triangular matrix, i.e.,

$$
\begin{bmatrix}
\lambda_{11} & 0 & \cdots & 0 \\
\lambda_{21} & \lambda_{22} & \ddots & \vdots \\
\vdots & \vdots & \ddots & 0 \\
\lambda_{k1} & \cdots & \cdots & \lambda_{kk} \\
\vdots & \cdots & \cdots & \vdots \\
\lambda_{n_1 1} & \cdots & \cdots & \lambda_{n_1 k}
\end{bmatrix}
\begin{bmatrix}
a_{11} & \cdots & a_{1k} \\
\vdots & \ddots & \vdots \\
a_{k1} & \cdots & a_{kk}
\end{bmatrix}
=
\begin{bmatrix}
\lambda_{11}^* & 0 & \cdots & 0 \\
\lambda_2^* & \lambda_{22}^* & \ddots & \vdots \\
\vdots & \vdots & \ddots & 0 \\
\lambda_{k1}^* & \cdots & \cdots & \lambda_{kk}^* \\
\vdots & \cdots & \cdots & \vdots \\
\lambda_{n_1 1}^* & \cdots & \cdots & \lambda_{n_1 k}^*
\end{bmatrix},
\qquad (30)
$$

from which we obtain $a_{ij} = 0$ for any i, j such that $i < j$, or A^* is lower triangular. Also note that equation (29) implies $A^* \cdot (A^*)' = I$. Use $(A^*)'$ to right multiply both sides of equation (30) to obtain

$$
\begin{bmatrix}
\lambda_{11} & 0 & \cdots & 0 \\
\lambda_{21} & \lambda_{22} & \ddots & \vdots \\
\vdots & \vdots & \ddots & 0 \\
\lambda_{k1} & \cdots & \cdots & \lambda_{kk} \\
\vdots & \cdots & \cdots & \vdots \\
\lambda_{n_1 1} & \cdots & \cdots & \lambda_{n_1 k}
\end{bmatrix}
=
\begin{bmatrix}
\lambda_{11}^* & 0 & \cdots & 0 \\
\lambda_2^* & \lambda_{22}^* & \ddots & \vdots \\
\vdots & \vdots & \ddots & 0 \\
\lambda_{k1}^* & \cdots & \cdots & \lambda_{kk}^* \\
\vdots & \cdots & \cdots & \vdots \\
\lambda_{n_1 1}^* & \cdots & \cdots & \lambda_{n_1 k}^*
\end{bmatrix}
\begin{bmatrix}
a_{11} & \cdots & a_{k1} \\
\vdots & \ddots & \vdots \\
a_{1k} & \cdots & a_{kk}
\end{bmatrix},
$$

from which we obtain $a_{ij} = 0$ for any i, j such that $i > j$. In sum we have proved that

$$
A^* = diag\left\{a_{11}, ..., a_{kk}\right\}.
$$

Again, from $A^* \cdot (A^*)' = I$ we obtain $a_{ii} = 1$ or -1 for all $i = 1, ..., k$. Since A is the inverse of A^*, A is also a diagonal matrix with either 1 or -1 on diagonal terms. Similarly, we may prove that the assumption Λ_C^c being lower triangular implies A_c is a diagonal matrix with either 1 or -1 on the diagonal. In sum, the rotation matrix R is a diagonal matrix with either 1 or -1 on the diagonal. With the extra condition that the diagonal terms of factor loadings are strictly positive, the rotation matrix R becomes an identify matrix. This proves that the proposed normalization separately identifies both the world factor and the country factors and the corresponding loadings. **Q.E.D.**

12 Appendix B: Positive-definite conditional co-variance matrix

In real applications, numerical round-off errors could result in an indefinite covariance matrix although the matrix should be positive definite. This situation arises more frequently for high dimensional covariance matrices. We provide an algorithm to calculate the variance of a conditional normal distribution such that it is always

positive-definite[7]. Let

$$\begin{bmatrix} X_{p \times 1} \\ Y_{q \times 1} \end{bmatrix} \sim N \left(\begin{bmatrix} \mu_X \\ \mu_Y \end{bmatrix}, \begin{bmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{bmatrix} \right).$$

Then for the conditional distribution of $Y|X$

$$var\,(Y|X) = P_{22} - P_{21}P_{11}^{-1}P_{12}$$

which is assumed to be positive definite. However, numerical round-off errors might produce an indefinite $var\,(Y|X)$. To handle this, we propose a numerically stable algorithm to calculate $var\,(Y|X)$ which is always positive definite as long as $var\,([X;Y])$ is positive definite.

Let $P = [P_{11}, P_{12}; P_{21}, P_{22}]$. Then

$$P^{-1} = \begin{bmatrix} * & * \\ * & \left(P_{22} - P_{21}P_{11}^{-1}P_{12} \right)^{-1} \end{bmatrix}.$$

So $var\,(Y|X) = P_{22} - P_{21}P_{11}^{-1}P_{12} = ((P^{-1})_{22})^{-1}$. We first conduct singular-value-decomposition (SVD) of P to obtain

$$P = USV'$$

then

$$P^{-1} = US^{-1}V'.$$

Because P is positive definite, S is diagonal with positive diagonal elements, and $U = V$, $UU' = I_{p+q}$. Write U and S as block matrices conformable to the blocks of P

$$U = \begin{bmatrix} U_{11} & U_{12} \\ U_{21} & U_{22} \end{bmatrix}, S = \begin{bmatrix} S_{11} & 0 \\ 0 & S_{22} \end{bmatrix}.$$

Define a $q \times (p+q)$ matrix

$$M = \left[U_{21}S_{11}^{-\frac{1}{2}}, U_{22}S_{22}^{-\frac{1}{2}} \right]$$

then

$$\left(P^{-1} \right)_{22} = MM'.$$

Perform singular value decomposition for M' to obtain

$$M' = U_m S_m V_m'$$

where U_m is $(p+q) \times (p+q)$, S_m is $(p+q) \times q$, V_m is $q \times q$. This admits a $V_m DV_m'$ representation of $(P^{-1})_{22}$ given below

$$\left(P^{-1} \right)_{22} = V_m S_m' S_m V_m' = V_m DV_m'$$

---

[7]The algorithm is implemented in the Matlab function "*schur_pos.m*" in the accompanying computer program files.

where $D = S'_m S_m$ is $q \times q$. This shows that

$$\left(\left(P^{-1}\right)_{22}\right)^{-1} = V_m D^{-1} V'_m$$

whose numerical outcome is always positive definite. In computationally efficient Matlab implementation, $D = diag(diag(S_m)^2)$.

13   Appendix C: Square-root form of the Kalman filter

We write the square-root form of Kalman filter Matlab program, "*k_filter_srf.m*", based on Bierman (1977), Evenson (2009), and Tippett et al. (2003).

Consider a state space parametrized by $\{A, C, Q, R\}$. The Kalman filter covariance evolution equations are given by

$$P_{t|t-1} = A P_{t-1|t-1} A' + Q$$
$$P_{t|t} = \left(I - K_t C\right) P_{t|t-1}$$
$$K_t = P_{t|t-1} C' \left(C P_{t|t-1} C' + R\right)^{-1} = P_{t|t} C' R^{-1} \tag{31}$$

Let the matrix square root representation of $P_{t|t-1}$ and $P_{t|t}$ be (not unique)

$$P_{t|t-1} = Z_t^f \left(Z_t^f\right)', \quad P_{t|t} = Z_t^a \left(Z_t^a\right)'. \tag{32}$$

The Potter method for the square-root form Kalman update (Bierman, 1977) is

$$P_{t|t} = Z_t^a \left(Z_t^a\right)' = \left(I - P_{t|t-1} C' \left(C P_{t|t-1} C' + R\right)^{-1} C\right) P_{t|t-1}$$
$$= Z_t^f \left(I - \left(Z_t^f\right)' C' \left(C Z_t^f \left(Z_t^f\right)' C' + R\right)^{-1} C Z_t^f\right) \left(Z_t^f\right)'$$
$$= Z_t^f \left(I - V_t D_t^{-1} V_t'\right) \left(Z_t^f\right)'$$

where $V_t = \left(C Z_t^f\right)'$, $D_t = V_t' V_t + R$. Then update Z_t^a as

$$Z_t^a = Z_t^f M_t$$

where M_t is a square-root of $I - V_t D_t^{-1} V_t'$, i.e., $M_t M_t' = I - V_t D_t^{-1} V_t'$ (notice that this decomposition is not unique, subject to any full rank orthogonal transformation). The technical issue would be the numerical stable calculation of M_t.

To obtain M_t, we use

$$I - V_t D_t^{-1} V_t' = \left(I + \left(Z_t^f\right)' C' R^{-1} C Z_t^f\right)^{-1}.$$

In fact, by the Sherman-Morrison-Woodbury identity

$$(A + UCV)^{-1} = A^{-1} - A^{-1} U \left(C^{-1} + V A^{-1} U\right) V A^{-1}.$$

It follows that

$$\left(I + \left(Z_t^f\right)' C'R^{-1}CZ_t^f\right)^{-1} = I - \left(Z_t^f\right)' C' \left(R + CZ_t^f \left(Z_t^f\right)' C'\right)^{-1} CZ_t^f$$
$$= I - V_t D_t^{-1} V_t'$$

Then we may proceed to use the ensemble transform Kalman filter (ETKF) (Tippett et al. (2003))

$$Z_t^a = Z_t^f H_t \left(\Gamma_t + I\right)^{-1/2} \tag{33}$$

where $H_t \Gamma_t H_t'$ is the eigenvalue decomposition of $\left(Z_t^f\right)' C'R^{-1}CZ_t^f$. Note that the eigenvalue decomposition of $I - V_t D_t^{-1} V_t'$ is $H_t \left(\Gamma_t + I\right)^{-1} H_t'$ and thus $M_t = H_t \left(\Gamma_t + I\right)^{-1/2}$ is a square root of $I - V_t D_t^{-1} V_t'$.

For a model with $N = T = 100, q = 2, h = 2, s = 1$, it takes less than 400 seconds to obtain a Gibbs-sampling chain size of ten thousand using a standard Intel Core i7 2GB RAM computer.

References

[1] D. Amengual and M. Watson. Consistent estimation of the number of dynamic factors in large N and T panel. *Journal of Business and Economic Statistics*, 25(1):91–96, 2007.

[2] T.W. Anderson. *An Introduction to Multivariate Statistical Analysis*. New York: Wiley, 1984.

[3] T.W. Anderson and H. Rubin. Statistical inference in factor analysis. In *In J. Neyman (Ed.): Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability.*, volume 5, pages 111–150. Univ. of California Press, 1956.

[4] J. Bai. Inferential theory for factor models of large dimensions. *Econometrica*, 71(1):135–172, 2003.

[5] J. Bai and S. Ng. Determining the number of factors in approximate factor models. *Econometrica*, 70(1):191–221, 2002.

[6] J. Bai and S. Ng. Determining the number of primitive shocks in factor models. *Journal of Business and Economic Statistics*, 25(1):52–60, 2007.

[7] J. Bai and S. Ng. Principal components estimation and identification of the factors. Working paper, 2010.

[8] J.O. Berger and L.R. Pericchi. The intrinsic Bayes factor for model selection and prediction. *Journal of the American Statistical Association*, 91(433):109–122, 1996.

[9] B. Bernanke, J. Boivin, and P. Eliasz. Measuring monetary policy: a factor augmented vector autoregressive (FAVAR) approach. *Quarterly Journal of Economics*, 120(1):387–422, 2005.

[10] G.J. Bierman. *Factorization methods for discrete sequential estimation.* Dover Books on Mathematics, Dover Publications, 1977.

[11] J. Breitung and J. Tenhofen. GLS estimation of dynamic factor models. *Journal of the American Statistical Association*, 106(495):1150–1166, 2011.

[12] C.K. Carter and R. Kohn. On Gibbs sampling for state space models. *Biometrika*, 81:541–553, 1994.

[13] S. Chib. Marginal likelihood from the Gibbs output. *Journal of the American Statistical Association*, 90 (432):1313–1321, 1995.

[14] I. Choi. Efficient estimation of factor models. *Econometric Theory*, 28(2):274–308, 2012.

[15] T. Cogley and T.J. Sargent. Drifts and volatilities: monetary policies and outcomes in the post WWII US. *Review of Economic Dynamics*, 8 (2):262–302, 2005.

[16] G. Connor and R. Korajzcyk. Performance measurement with the arbitrage pricing theory: a new framework for analysis. *Journal of Financial Economics*, 15:373–394, 1986.

[17] M.J. Crucini, M.A. Kose, and C. Otrok. What are the driving forces of international business cycles? *Review of Economic Dynamics*, 14:156–175, 2011.

[18] M. Del Negro and C. Otrok. Dynamic factor models with time-varying parameters: measuring changes in international business cycles. Working paper, 2008.

[19] F.X. Diebold, C. Li, and V. Yue. Global yield curve dynamics and interactions: a generalized Nelson-Siegel approach. *Journal of Econometrics*, 146(2):351–363, 2008.

[20] C. Doz, D. Giannone, and L. Reichlin. A quasi-maximum likelihood approach for large approximate dynamic factor models. *forthcoming, Review of Economics and Statistics*, 2011.

[21] G. Evensen. *Data assimilation: the ensemble Kalman filter.* 2nd ed. Springer, 2009.

[22] M. Forni, M. Hallin, M. Lippi, and L. Reichlin. The generaized dynamic factor model: identification and estimation. *Review of Economics and Statistics*, 82(4):540–554, 2000.

[23] J.F. Geweke. The dynamic factor analysis of economic time series. In *In D. J. Aigner and A. S. Goldberger (Eds.): Latent Variables in Socio-economic Models.* Amsterdam: North-Holland, 1977.

[24] A.W. Gregory and A.C. Head. Common and country-specific fluctuations in productivity, investment, and the current account. *Journal of Monetary Economics*, 44(3):423–451, 1999.

[25] B. Junbacker and S.J. Koopman. Likelihood-based analysis for dynamic factor models. Working paper, 2008.

[26] C. Kim and C. Nelson. *State Space Models With Regime Switching: Classical and Gibbs Sampling Approaches With Applications.* Massachusetts: MIT Press, 1999.

[27] M.A. Kose, C. Otrok, and C.H. Whiteman. International business cycles: World, region, and country-specific factors. *American Economic Review*, 93(4):1216–1239, 2003.

[28] D.N. Lawley and A.E. Maxwell. *Factor Analysis in a Statistical Method.* Butterworth, London, 1971.

[29] E. Moench, S. Ng, and S. Potter. Dynamic hierarchical factor models. Working paper, 2011.

[30] A. O'Hagan. Fractional Bayes factors for model comparison. *Journal of the Royal Statistical Society. Series B (Methodological)*, 57(1):99–138, 1995.

[31] D. Quah and T. Sargent. A dynamic index model for large cross sections. In *Business Cycles, Indicators and Forecasting*, NBER Chapters, pages 285–310. National Bureau of Economic Research, Inc, December 1993.

[32] C.A. Sims and T. Zha. Error bands for impulse responses. *Econometrica*, 67(5):1113–1156, 1999.

[33] J.H. Stock and M.W. Watson. Diffusion indexes. NBER Working Paper 6702, 1998.

[34] J.H. Stock and M.W. Watson. Forecasting using principal components from a large number of predictors. *Journal of the American Statistical Association*, 97:1167–1179, 2002.

[35] J.H. Stock and M.W. Watson. Implications of dynamic factor models for VAR analysis. NBER Working Paper 11467, 2005.

[36] M.K. Tippett, J.L. Anderson, C.H. Bishop, T.M. Hamill, and J.S. Whitaker. Ensemble square root filters. *Monthly Weather Review*, 131:1485–1490, 2003.

[37] P. Wang. Large dimensional factor models with a multi-level factor structure: identification, estimation, and inference. Working paper, 2010.

[38] M.W. Watson and R.F. Engle. Alternative algorithms for the estimation of dynamic factor, mimic and varying coefficient regression models. *Journal of Econometrics*, 23(3):385–400, 1983.

[39] M. West and J. Harrison. *Bayesian Forecasting and Dynamic Models.* Springer Series in Statistics, 1997.

[40] J.H. Wright. Term premia and inflation uncertainty: empirical evidence from an international panel dataset. *American Economic Review*, 101(4):1514–1534, 2011.

[41] A. Zellner. *An introduction to Bayesian inference in econometrics.* John Wiley and Sons, Inc, 1971.