Identification and Bayesian Estimation of Dynamic Factor Models

Jushan Bai
Department of Economics, Columbia University, New York, NY 10027 (jb3064@columbia.edu)

Peng Wang
Department of Economics, The Hong Kong University of Science and Technology, Clear Water Bay, NT, Hong Kong (pwang@ust.hk)

We consider a set of minimal identification conditions for dynamic factor models. These conditions have economic interpretations and require fewer restrictions than the static factor framework. Under these restrictions, a standard structural vector autoregression (SVAR) with measurement errors can be embedded into a dynamic factor model. More generally, we also consider overidentification restrictions to achieve efficiency. We discuss general linear restrictions, either in the form of known factor loadings or cross-equation restrictions. We further consider serially correlated idiosyncratic errors with heterogeneous dynamics. A numerically stable Bayesian algorithm for the dynamic factor model with general parameter restrictions is constructed for estimation and inference. We show that a square-root form of the Kalman filter improves 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 multilevel factor models, and they allow us to study the “spill-over” effects of the shocks arising from one group to another. Supplementary materials for technical details are available online.

KEY WORDS: Cross-equation restrictions; Impulse response function; Multilevel factor model; Spill-over effects.

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, such as an oil price shock, drive the comovements of economic variables. As a result, economists are increasingly relying on these models for policy analysis. Examples include Bernanke, Boivin, and Eliasz (2005), Stock and Watson (2005), and Kose, Otrok, and White (2003), among others. This article considers identification issues for a general class of dynamic factor models and their Bayesian estimation.

Three sources of dynamics are considered. One is that the latent factor is dynamic. An example would be \( x_{it} = \lambda_{it} f_{it} + e_{it} \), where \( f_{it} \) is a vector autoregressive process. While \( f_{it} \) is dynamic, the relationship between \( x_{it} \) and \( f_{it} \) is static. The other source of dynamics is that \( f_{it} \) affects \( x_{it} \) dynamically, for instance, \( x_{it} = \gamma_{it} f_{it} + \tau_{it} f_{it-1} + e_{it} \) so that the lags of \( f_{it} \) directly impacts \( x_{it} \). We shall refer to the first example as a static factor model in spite of \( f_{it} \) being dynamic. The second example can be put into a static factor framework. Let \( F_t = (f_{it}, f_{it-1})' \) and \( \lambda_{it} = (\gamma_{it}', \tau_{it})' \), then \( x_{it} = \lambda_{it}' F_t + e_{it} \). We will also consider another source of dynamics, that is, the idiosyncratic errors, \( e_{it} \), are serially correlated. There exists a large literature for analyzing static factor models. There are drawbacks, however, to 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.

The model considered in this article has more structures than the generalized dynamic factor model of Forni et al. (2000). We incorporate restrictions on the factor loadings and study a set of simple identification schemes for the dynamic factor models. These identification restrictions have economic interpretations. For instance, 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 number of 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 article 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 show how all three sources of dynamics can be naturally handled within the Bayesian framework.

Our method is closely related to the factor-augmented vector autoregression (FAVAR) approach of Bernanke, Boivin, and Eliasz (2005), and the dynamic factor model of Stock and Watson (2005). The model of Bernanke, Boivin, and Eliasz (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 considered 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 articles consider a two-step approach that 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, Boivin, and Eliasz (2005) also considered a Bayesian approach using the Gibbs sampling method for state-space models. They imposed conjugate priors for model parameters, and derive the posterior distribution of the objective of interest, the outcome of which is the posterior joint distribution of the 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’s (1995) fractional Bayes’ factors or Berger and Pericchi’s (1996) intrinsic Bayes’ factors. The Bayesian method proposed in this article 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 approach is to conduct Bayesian analysis of dynamic factor models with time-varying parameters, such as stochastic volatilities (Aguilar and West 2000), or both time-varying factor loadings and stochastic volatilities (Del Negro and Otrok 2008). The main focus of both articles is the time-varying feature of the dynamic factor model. The identifying restrictions assume that factors are mutually independent. The measurement equations are static. 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. For example, using a time-varying parameters VAR, Cogley and Sargent (2005) have presented evidence of evolving inflation dynamics in the U.S. To limit our focus, however, we do not explore that direction in this article.

In the literature, the maximum likelihood estimation of dynamic factor models has been considered by many authors, such as Geweke (1977), Watson and Engle (1983), Quah and Sargent (1993), Kim and Nelson (1999), Junbacker and Koopman (2008), and Doz, Giannone, and Reichlin (2012). The maximization can be implemented through the expectation–maximization (EM) method as in Doz, Giannone, and Reichlin (2012). 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-square estimator for nondynamic factor models was studied by Breitung and Tenhofen (2011) and Choi (2012).

The Bayesian Markov chain Monte Carlo (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. We consider the identification of factors without rotations.

2. DYNAMIC FACTOR MODELS: IDENTIFICATION CONDITIONS

We consider a dynamic factor model featuring a dynamic factor representation of the observables 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):

\[ X_t = \Lambda_0 f_t + \Lambda_1 f_{t-1} + \cdots + \Lambda_h f_{t-h} + \varepsilon_t, \]  

(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 \( t \), and the \( N \times q \) matrix \( \Lambda_j \) is the dynamic factor loading for \( f_{t-j}, j = 0, 1, \ldots, s \). To fix this idea, we first assume that \( \varepsilon_t \) is iid \( N(0, R) \), and will relax this assumption later when we consider serially correlated error terms. The dynamic factor follows a VAR(h) process

\[ f_t = \Phi_1 f_{t-1} + \cdots + \Phi_h f_{t-h} + \varepsilon_t, \]  

(2)

where \( \varepsilon_t \) is iid \( N(0, Q) \), \( \Phi_j \) is a \( q \times q \) matrix of autoregressive coefficients, \( 1 \leq j \leq h \). We assume that \( \{\varepsilon_t\}_{t=1}^{T} \) is independent of \( \{f_t\}_{t=1}^{T} \). 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 parameterized by a VAR(h) process. This article emphasizes the dynamics in Equation (1), because it is these
dynamics that make 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 $f_t$, that is, $q$, irrespective of $s$ and $h$.

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

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

Then

$$
X_t = \Lambda F_t + \varepsilon_t. \quad (3)
$$

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 restrictions. 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 were discussed by Bai and Ng (2013). Let $F' = (F_1, \ldots, F_T')$. The restrictions in static factor analysis take 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 $\Lambda$ is lower triangular, or

(iii) $F_t$ is unrestricted, and the first $r \times r$ block of $\Lambda$ 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 $\Lambda$. In general, we only estimate rotations of $F_t$ and $\Lambda$ 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 $\varepsilon_t$, 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.

It is well known that the dynamic factor model defined in (1) and (2) is not identified without further restrictions. One can always premultiply 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

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

where $\Lambda_{01}$ 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 $\Lambda_{01}$ is a lower-triangular matrix with strictly positive diagonal elements.

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 $\Lambda_{01}$ is lower triangular imposes $q(q-1)/2$ restrictions. The latter restriction says that the first variable $X_{q1}$ is affected contemporaneously by the first dynamic factor, and the second variable $X_{q2}$ is affected contemporaneously by the first two dynamic factors, and so on. In practice, the choice of the first $q$ variables is important.

It makes more sense to assume the innovations $\varepsilon_t$ to be mutually uncorrelated than to assume $f_t$ to be mutually uncorrelated. This leaves the autoregressive coefficient matrix $\Phi_{f_j}$, $j = 1, \ldots, h$ in (2) unrestricted, allowing the dynamic factors to interact with each other. Under DFM1, we shall say $\Lambda_0$ 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 $\Lambda_j$, $j \in \{0, 1, \ldots, s\}$ are uniquely identified.

The proof is in the Appendices. Similar to Proposition 1, if $\varepsilon_t$ has an identity covariance matrix and $\Lambda_j$ is lower-triangular for some given $j \in \{0, 1, \ldots, s\}$, 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 $\var{\varepsilon_t} = I_q$, $\Lambda_j$ is a lower-triangular matrix for some given $j \in \{0, 1, \ldots, s\}$, and the diagonal elements of the first $q \times q$ block of $\Lambda_j$ are all strictly positive. Then the dynamic factors $f_t$ and the dynamic factor loadings $\Lambda_j$, $j \in \{0, 1, \ldots, s\}$ are uniquely identified.

Notice that in DFM1, other normalizations on the factor loadings also work, as long as they imply $q(q-1)/2$ nonredundant restrictions. For example, the normalization that $\Lambda_{j}'\Lambda_j$ be diagonal for some given $j \in \{0, 1, \ldots, s\}$ also works, which is similar to the normalization in principal component analysis of factor models (see Anderson 1984; Connor and Korajczyk 1986; Bai and Ng 2002; Stock and Watson 2002; Bai 2003). A variation to DFM1 is that $\var{\varepsilon_t}$ is diagonal but $\Lambda_{01}$ has 1’s on the diagonal. We next consider a different set of identification conditions.

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

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

This is also sufficient to identify the factor model. Again, there are $q^2$ restrictions, and no restrictions are imposed on $\Lambda_1, \ldots, \Lambda_s$. 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 $\varepsilon_t$ can be linked to structural shocks that are implied by economic theory, which allows a structural VAR representation of the dynamic factors; see Section 2.2. Both DFM1 and DFM2 differ from the conventional practice that 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 \( \Lambda_0 \) is an identity matrix (DFM2), the dynamic factors \( f_t \) and the dynamic factor loadings \( \Lambda_j, j \in \{0, 1, \ldots, s\} \) are uniquely identified.

Similar to Proposition 2, if the upper \( q \times q \) block of \( \Lambda_j \) is an identity matrix for some \( j \in \{0, 1, \ldots, s\} \), 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 \( \Lambda_j \) is an identity matrix for some given \( j \in \{0, 1, \ldots, s\} \), then the dynamic factors \( f_t \) and the dynamic factor loadings \( \Lambda_j, j \in \{0, 1, \ldots, s\} \) are uniquely identified.

Using a state–space representation of the factor model, Doz, Giannone, and Reichlin (2012) explicitly considered the dynamics in factors. Although they considered 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) considered a more general version of dynamic factor models. They allowed the error terms in the observation equation to be serially correlated. Like Doz, Giannone, and Reichlin (2012), their identification is also based on restricting the static factors, so that the principal component analysis is embedded into an iterative least-square estimation. Our identification scheme is directly on the dynamic factors \( f_t \), not on the static factors \( F_t = (f_t', f_{t-1}', f_{t-2}', \ldots, f_{t-s}')' \).

### 2.1 Identification for Multilevel Factor Models

The dynamic factor model with a multilevel factor structure has been increasingly applied to study the comovement of economic quantities at different levels (see Gregory and Head 1999; Kose, Otrok, and Whiteman 2003; Crucini, Kose, and Otrok 2011; Moench, Ng, and Potter 2013). When researchers confront the model with data, they usually find ways to impose economic meaning of factors. For example, Kose, Otrok, and Whiteman (2003) and a number of subsequent articles considered a dynamic factor model with a multilevel factor structure to characterize the comovement of international business cycles on the global level, regional level, and country level, respectively. For ease of presentation, we will focus on the case with only two levels of factors: a world factor and a country-specific factor. Extensions to models with more than two levels are straightforward.

Consider \( C \) countries, each having an \( n_c \times 1 \) vector of country variables \( X_{ct}, t = 1, \ldots, T, c = 1, \ldots, C \). One may model \( X_{ct} \) as being affected by a world factor \( f_{t}^W \) and a country factor \( f_{t}^c, c = 1, \ldots, C \), all factors being latent,

\[
X_{ct} = \Lambda_W f_{t}^W + \Lambda_c f_{t}^c + \varepsilon_{ct},
\]

where \( \Lambda_W, \Lambda_c \) are the matrices of factor loadings of country \( c \), \( \varepsilon_{ct} \) 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 \( F_{t}^C \) be a vector collecting all the country factors. Kose, Otrok, and Whiteman (2003) assumed both \( f_{t}^W \) and \( f_{t}^c \) are scalars and considered a VAR(1) specification of factors given below,

\[
\begin{pmatrix}
    f_{t}^W \\
    F_{t}^C
\end{pmatrix} =
\begin{pmatrix}
    \rho_W & 0 \\
    0 & \rho_C
\end{pmatrix}
\begin{pmatrix}
    f_{t-1}^W \\
    F_{t-1}^C
\end{pmatrix} +
\begin{pmatrix}
    u_{t}^W \\
    U_{t}^C
\end{pmatrix},
\]

where \( \{\rho_W, \rho_C\} \) are also diagonal matrices comparable to the dimension of corresponding factors. The innovation to factors vec\( [u_{t}^W, U_{t}^C] \) is independent of \( \varepsilon_{ct} \) at all leads and lags and is iid normal,

\[
\begin{pmatrix}
    u_{t}^W \\
    U_{t}^C
\end{pmatrix} \overset{iid}{\sim} N\left(0, \begin{pmatrix}
    \sigma_W^2 & 0 \\
    0 & \sigma_C^2
\end{pmatrix}\right),
\]

where \( \sigma_W^2, \sigma_C^2 \) 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 that the variance of vec\( [u_{t}^W, U_{t}^C] \) is an identity matrix.

Observe that it is sufficient but not necessary to assume the independence among \( \{f_{t}^W, f_{t}^c, c = 1, \ldots, C\} \) 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, that is, the impulse responses of one factor to the innovation of another factor are not zero. Take the world factor, for example. It might be affected by 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 the United States might have an impact on other economies over time. 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 an 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

\[
\begin{pmatrix}
    f_{t}^W \\
    F_{t}^C
\end{pmatrix} =
\begin{pmatrix}
    \rho_{11} & \rho_{12} \\
    \rho_{21} & \rho_{22}
\end{pmatrix}
\begin{pmatrix}
    f_{t-1}^W \\
    F_{t-1}^C
\end{pmatrix} +
\begin{pmatrix}
    u_{t}^W \\
    U_{t}^C
\end{pmatrix},
\]

where the innovation to factors vec\( [u_{t}^W, U_{t}^C] \) is independent of \( \varepsilon_{ct} \) in Equation (4) at all leads and lags and is iid normal with a diagonal covariance matrix, as described earlier. This implies that factors can be correlated with each other. Equation (7) 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 \( [u_{t}^W, U_{t}^C] \). For example, a shock to \( U_{t}^C \) not only affects the country factor \( F_{t}^C \), but also affects the world factor. This allows us to evaluate the “spill-over” effects of country-level shocks to other economies.

In addition, Equation (7) nests the conventional specification of Kose, Otrok, and Whiteman (2003) in Equation (5) as a special case if we further restrict the autoregressive coefficient
matrix in (7) to be diagonal. When confronting the model with data, we will be able to test Equation (7) against Equation (5) using a Wald-test in classical estimation, or using the Bayes’ factors in Bayesian estimation. We summarize the identification conditions in the following proposition.

**Proposition 3.** Define a dynamic factor model as in (4), (7), and (6). Assume that $\sigma_{WW}^2 = I_k, \sigma_C^2 = \Lambda_{k}^{\prime}, \Lambda_{WW}^1$ and $\{\Lambda_{C}^{c}\}_{c=1}^{C}$ are all lower-triangular matrices with strictly positive diagonal terms, and $[\Lambda_{WW}^c, \Lambda_{C}^{c}]$ is of full column rank ($c = 1, \ldots, C$), then the factor model is uniquely identified.

The proof is in the Appendices. Proposition 3 can be easily extended to the case with three or more levels. A key feature with multilevel 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 (4) is static. The presence of serially correlated measurement errors.

For efficient estimation, we will take into account the zero restrictions. Finally, note that Equation (4) is static. The presence of serially correlated measurement errors. Suppose that $Z_t$ 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 VAR is given by

$$Z_t = B_1 Z_{t-1} + \cdots + B_p Z_{t-p} + \varepsilon_t$$

or $B(L)Z_t = \varepsilon_t$, 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 not directly observable, but observable with a measurement error $\eta_t$ such that

$$Y_t = Z_t + \eta_t.$$

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

$$W_t = \Gamma_0 Z_t + \cdots + \Gamma_x Z_{t-x} + \varepsilon_{wt}.$$

Essentially, $W_t$ is also driven by the fundamental structural shocks $a_t$ through $Z_t$. Then it is possible to recover $Z_t$ so that SVAR analysis can be performed. Let

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

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

Then we have a structural dynamic factor model

$$X_t = \Lambda_0 f_t + \cdots + \Lambda_x f_{t-x} + e_t,$$

$$f_t = B_1 f_{t-1} + \cdots + B_p f_{t-p} + \varepsilon_t.$$  \hfill (8)

According to the identification scheme DFM2, Equation (8) is identified and can be analyzed using a Bayesian approach. In particular, without further assumptions, we are able to estimate $f_t = Z_t, B(L), \Lambda_t$, and $E(\varepsilon_t, \varepsilon_t') = A_0^{-1}(A_0')^{-1}$. We may also incorporate additional structural restrictions (such as long-run restrictions) as in standard SVAR analysis to identify $A_0$.

The impulse responses to structural shocks can be obtained as

$$\frac{\partial Y_{t+k}}{\partial \eta_t} = \frac{\partial Z_{t+k}}{\partial a_t} = \frac{\partial f_{t+k}}{\partial a_t}$$

and

$$\frac{\partial W_{t+k}}{\partial a_t} = \begin{bmatrix} \Gamma_0 \frac{\partial f_{t+k}}{\partial a_t} + \cdots + \Gamma_s \frac{\partial f_{t+k-s}}{\partial a_t}, & k \geq s, \\ \Gamma_0 \frac{\partial f_{t+k}}{\partial a_t} + \cdots + \Gamma_k \frac{\partial f_t}{\partial a_t}, & k < s \end{bmatrix},$$

where the partial derivative is taken for each component of $a_t$ when $a_t$ is a vector.

### 3. BAYESIAN ESTIMATION USING GIBBS SAMPLING

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 because $q \ll N$. An alternative likelihood approach is to use the expectation–maximization algorithm, which is not considered in this article. 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 in models (1) and (2). It is helpful to form a state–space representation of the dynamic factor model, and then Carter and Kohn’s (1994) mimove Gibbs-sampling algorithm can be modified for estimation. Our specific estimation procedure has several desirable features. For example, it explicitly takes into account parameter restrictions and can accommodate cases with or without serial correlation in the error terms. In addition, we focus on directly estimating the dynamic factors instead of the static factors. Last but not least, to avoid numerical singularities due to round-off errors, we 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. Here, we focus on how to formulate the parameter restrictions so that a simple Bayesian algorithm can be designed. We then discuss the case with serially correlated measurement errors.
3.1 Formulating the Restrictions on Factor Loadings

In addition to the minimal identifying restrictions as in DFM1 or DFM2, there may be other overidentifying restrictions. For example, the multilevel 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)

\[ X_t = \Lambda F_t + \varepsilon_t, \]

where \( F_t = [f_t, f_{t-1}, \ldots, f_{t-j}], \) and \( \Lambda = [\Lambda_0, \ldots, \Lambda_j]. \) Let \( X \) be the \((T - s - 1) \times N\) data matrix, \( F \) be the \((T - s - 1) \times q(s + 1)\) matrix of the static factors, then we have a matrix representation of the factor model

\[ X = FN' + E, \quad \text{or} \quad \text{vec}(X) = (I_N \otimes F) \lambda + \text{vec}(E), \]

where \( \lambda = \text{vec}(\Lambda') \) and \( \text{vec}(E) \sim N(0, R \otimes I_{T-s-1}). \) The identifying restrictions in DFM1 or DFM2 impose linear restrictions on the factor loadings. Consider the following restriction

\[ \lambda = B\delta + C, \]

where \( \delta \) 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 (10) can be obtained from DFM2.

- Example: In DFM2, the upper-left \( q \times q \) block of \( \Lambda \) 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. \( \delta \) is a \([Nq(s + 1) - q^2] \times 1 \) vector of free parameters. There are \( q \) ones and \( q^2 - q \) zeros in \( \lambda, \) 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 \( \Lambda \) equation-by-equation due to the absence of cross-equation restrictions.

In view of (10), we may rewrite the restricted factor model (9) as

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

where \( y = \text{vec}(X) - (I_N \otimes F)C \) and \( Z = [(I_N \otimes F)B]. \) Impose the Jeffreys’ prior for \( \delta \) and \( R:

\[ p(\delta, R) \propto |R|^{-(N+1)/2}, \]

which implies the following conditional posterior:

\[ \delta|R, X, F \sim N\left(B(B'B)^{-1}\left(\text{vec}(\hat{\Lambda}') - C\right), \left(B' (R^{-1} \otimes F'F) B\right)^{-1}\right), \]

with \( \hat{\Lambda} = X'F(F'F)^{-1}. \) Thus, we may draw \( \delta \) according to (11) and construct the associated loading matrix \( \Lambda(\delta). \) In the meantime, we may draw \( R \) according to an inverse-Wishart distribution

\[ R|X,F \sim \text{invWishart}(S, T - s + N + 1 - q(s + 1)), \]

where \( S = (X - F\hat{\Lambda}')(X - F\hat{\Lambda}'). \)

The Jeffreys’ prior \( p(\delta, R) \) for \( \delta \) and \( R \) is improper. An important open question is under what conditions the corresponding posterior is proper. While the simulation results below show that the Gibbs sampler always converges and the posterior is well behaved (the point estimates are close to the true factors), further theoretical investigation is needed. We leave this as a future research topic.

3.2 Restrictions on \( R \)

If the cross-section dimension \( N \) is small in (1), we may allow \( R = \text{var}(\varepsilon_i) \) to be an unrestricted covariance matrix. Given our knowledge of the model structure, the factors and model parameters can be identified through the autocorrelation function implied by the dynamic model, in contrast to the static factor model that is based on the variance only. 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. To preserve a parsimonious model structure, we may consider restricting \( R \) to be diagonal when \( N \) is large (e.g., when \( N \geq 20). \) Other plausible restrictions include making \( R \) a block-diagonal matrix. For example, in the multilevel factor model, we may allow the error term \( \varepsilon_i \) to be cross-correlated within the same group, but uncorrelated between groups.

We consider a diagonal \( R = \text{var}(\varepsilon_i) = \text{diag}[\sigma_1^2, \ldots, \sigma_N^2]. \) Impose the Jeffreys’ prior for \( \delta \) (free parameters in loadings) and \( R \) with \( p(\delta, R) \propto \Pi_{n=1}^N(1/\sigma_n^2). \) Then, we may draw \( \sigma_n^2 \) one-by-one according to the following inverse-Gamma distribution

\[ \sigma_n^2|X,F \sim \text{invGamma}(v/2 - 1, \hat{\varepsilon}_n^2/2), \]

\[ v = T - s + 2 - q(s + 1). \]

When the restrictions on the factor loadings are on an equation-by-equation basis, we may also sample the factor loadings equation-by-equation:

\[ \delta_i|R, X, F \sim N(\hat{B}_i, (\hat{B}_i\hat{B}_i)^{-1}(\hat{\lambda}_i - C_i), \sigma_i^2(\hat{B}_i\hat{B}_i)^{-1}(\hat{B}_i\hat{B}_i)^{-1}), \]

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 ordinary least-square (OLS) estimator of regressing \( X_i \) on \( F. \)

3.3 Serial Correlation in the Measurement Errors

Serial correlation in the measurement errors in (1) can be conveniently modeled using a VAR(1) process

\[ \varepsilon_t = \rho \varepsilon_{t-1} + u_t, \quad u_t \sim \text{iid} N(0, R), \]

where

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

Let \( L \) be the lag operator and apply \( (I - \rho L) \) on both sides of the measurement Equation (1), we obtain

\[ (I - \rho L)X_t = (I - \rho L)\Lambda(L)f_t + u_t, \quad u_t \sim \text{iid} N(0, R). \]
where $\Lambda(L) = [\Lambda_0, \Lambda_1, \ldots, \Lambda_s L^s]$. Let $\bar{X}_i \equiv (I - \rho L)X_i$, $\bar{\Lambda}(L) \equiv (I - \rho L)\Lambda(L)$, then

$$\bar{X}_i = \bar{\Lambda}(L) f_i + u_i.$$

Thus, conditional on parameters $[\rho, \Lambda, R, \Phi, Q]$, the latent factors can be analyzed in the same way as the case with serially uncorrelated measurement errors. Conditional on the data and the factors, the Gibbs sampler for model parameters is constructed by dividing the parameters into three blocks: $[\rho, R], [\Lambda]$, and $[\Phi, Q]$. Note that for the $i$th measurement equation,

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

where $\lambda_i(L)'$ is the $i$th row of $\Lambda(L)$. Given $[\rho, R]$, the factor loadings can be analyzed from the regression analysis with known variance. Given $\Lambda, \rho$ and $R$ can be analyzed equation-by-equation for the regression $e_{it} = \rho_i e_{i,t-1} + \sigma_i u_{it}$. Sampling $\Phi$ and $Q$ is basically the same as before. The same method applies to the case with VAR(p) measurement errors. We provide the detailed sampling algorithm in the supplementary material.

4. SIMULATION STUDY

4.1 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 are generated according to (1) and (2), where the parameters are draw from $\lambda_{ijs} \sim \text{iid } U(0, 1), e_i \sim \text{iid } N(0, R), e_t \sim \text{iid } N(0, Q)$. We choose $R = I_N, Q = I_q$, $\Phi_1 = [0.5, 0; -0.1, 0.2]$, and $\Phi_2 = [0.2, 0; 0.1, 0.1]$. Besides the 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, and the fifth row being unrestricted.

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 18 samples with different combinations of $N$ and $T$. We use the Jeffreys’ priors throughout the estimation. The length of the Gibbs sampling chain is chosen to be 10,000. (We found that our Bayesian algorithm converges very fast. The chain size of 100,000 or 5000 offers basically the same results. In addition, for the model with $N = 200, T = 200$, the CPU time for 10,000 chain size is less than 1200 sec using a standard Intel Core i7 computer with 2 GB DDR3 memory. In all the simulations, the confidence bands well cover the true factors.) We discard the first 5000 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 $\hat{f}_j$ on $\tilde{f}_j$ to obtain the adjusted $R$-squared $\tilde{R}_j^2$. We also regress $\hat{f}_j$ on both $\tilde{f}_1$ and $\tilde{f}_2$ to measure the closeness of the true 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 the adjusted $R$-squared 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 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 case with $N = 10$ or $100$ and $T = 200$, Figures 2 and 3 plot the two SE error bands of the estimated factors and compare them to the true factors. As is evident from the figures, factors are more precisely estimated when the sample has a larger $N$. This is also in line with the theory on large dimensional factor models such as Bai (2003).

Under the identification condition DFM2, we obtain similar results. We use the same DGP as described at the beginning of this section, except that we let the upper $q \times q$ block of the factor loading $\Lambda_0$ in (1) to be an identity matrix. We also impose the same extra restrictions on every five rows of the factor loadings. When estimating the model, we use DFM2 and treat the error
variance $Q$ in the factor process (2) to be unknown parameters. Figure 4 shows that with large $N$, the confidence bands are rather tight while still covering the true factors.

Given the data that produces Figure 3, we examine how the marginal likelihood varies for different choice of $q$. 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 $T = 21, \ldots, 200$. To fix this idea, we fix $s = 1$, $h = 2$ and only vary the number of factors $q$. Table 1 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 $q = 2$.

An alternative model comparison method is to use the deviance information criteria (DIC); see, for example, Berg, Meyer, and Yu (2004), Celeux et al. (2006), and Li, Zeng, and Yu (2013). The DIC does not require a proper prior and thus is applicable to our case with Jeffreys’ priors; it also works for time series observations. Out of the seven versions of DIC in Celeux et al. (2006), we adopt their DIC1 and report the results in Table 1 as well. A computationally simple approximation to DIC1 is considered by Li, Zeng, and Yu (2013). In the table, we use the MCMC approximation. Note that DIC1 is defined as

$$
\text{DIC}_1 \approx -4 \frac{1}{M} \sum_{m=1}^{M} \log p(X|\theta^{(m)}) + 2\log p(X|\hat{\theta}),
$$

where $\theta$ is the vector of model parameters and $\hat{\theta}$ is the posterior mean. We may use the MCMC posterior draws $\{\theta^{(m)}, m = 1, \ldots, M\}$ to obtain the following approximation

$$
\text{DIC}_1 \approx -4 \frac{1}{M} \sum_{m=1}^{M} \log p(X|\theta^{(m)}) + 2\log p(X|\hat{\theta}),
$$

where $p(X|\theta)$ can be evaluated at any given $\theta$ using the Kalman filter. In Table 1, the DIC achieves minimum when the number of factors equals the true value $q = 2$.

4.2 Serially Correlated Measurement Errors

This section explores the 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 a VAR(1) process given in (13). The number of dynamic factors is set at two. We fix $N = 50$, $T = 200$, $s = 1$, $h = 2$. Two methods are used to compare the estimation results, one with $\rho_i$ being 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 2 reports the results. To fix the idea, we choose $\rho_i = \rho$ for all $i = 1, \ldots, N$ and examine the adjusted $\bar{R}^2$ for different values of $\rho \in \{0.99, 0.9, 0.8, 0.5, 0.0\}$.
Table 2. Fit of dynamic factors: $\bar{R}^2_j$

For $j = 1, 2$, regress $f_j$ on

| $\rho_i$ | Fix $\rho_i = 0$ | Free $\rho_i$ | Fix $\rho_i = 0$ | Free $\rho_i$ |
|---------|-----------------|---------------|-----------------|---------------|
| 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) |

NOTES. 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.

From the table, we notice that for not-so-persistent measurement errors, say $\rho \leq 0.5$, both estimation methods provide reasonably good 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 $\rho_i$ 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. To demonstrate the performance of our Bayesian algorithm in the case of persistent measurement errors, we estimate the model for $N = 50$, $T = 200$, and $\rho_i = 0.9$ for all $i$. Figure 5 plots the two SE error bands of the estimated factors and compares them to the true factors. As is evident from the figures, the true factors are well covered by the tight confidence bands.

5. COMOVEMENT IN INTERNATIONAL BOND YIELDS

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 cover nine countries from January 1971 to May 2009. We will use a balanced panel from December 1992 to May 2009. The nine countries are the U.S. Canada, U.K., 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 $X_c(t, \tau)$ be the time $t$ bond yield of country $c$ at maturity $\tau$. For $c = 1, \ldots, C$, assume that a global factor $g_t$ and a country specific factor $f^c_t$ combine to affect the bond yield:

$$X_c(t, \tau) = g_t + f^c_t + \epsilon_{c}(t, \tau)$$

Figure 5. $N = 50$, $T = 200$. Interval estimates of the factors: two SE error bands.
\[ X^c_t(\tau) = \mu^c(\tau) + \gamma^c(\tau) g_t + \lambda^c(\tau) f^c_t + e^c_t(\tau), \]
\[ e^c_t(\tau) \sim \text{iid.} \mathcal{N}(0, \sigma^2_c(\tau)). \]  

\[ (14) \]

We further assume that the factors follow a VAR(1) process. In particular, let \( F_t = [g_t, f^1_t, \ldots, f^C_f]^\prime \), then
\[ F_t = \Phi F_{t-1} + \varepsilon_t, \varepsilon_t \sim N(0, Q). \]

\[ (15) \]

Assuming that \( Q = I_C + \), both \( \gamma^c(\tau) \) and \( \lambda^c(\tau) \) being strictly positive for \( \tau = 3 \) months, then the multilevel 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, U.S., and U.S. Our empirical exercise differs from theirs in two important ways. First, we adopt a much larger dataset with nine countries and more maturities, which expects to better represent the global bond yield. Second, we treat the AR coefficient matrix \( \Phi \) as unrestricted, which allows us to study the “spill-over” effects of one country to another, such as how a shock to the U.S. factor affects U.S. On the other hand, almost all existing literature on multilevel factor models assumes that different factors are independent of each other. 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. Models (14) and (15) 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 6. 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 report the variance decomposition results for each country in Table 3. (We calculate the variances due to the global factor, the country factors, and the idiosyncratic errors,

|                | U.S.  | Canada | U.K.   | 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

|                | 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%       |

Figure 7. Variance decomposition of country yields at maturities of 3, 12, 60, 120 months. Remark: the line with squares represents the variance explained by the global factor, and the line with crosses by the country factors.
respectively. And then use the variance ratio to measure the variance contribution due to each component.) The result is similar to Diebold, Li, and Yue (2008) in that the global factor explains a large proportion of variance in country bond yields, and compared with most other markets the United States 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 7 shows that, for all the nine 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 multilevel 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. The top left panel of Figure 8 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 \( f_{ct} \) given a shock to the U.S. factor, the U.S. factor, and the Japan factor, respectively, in the other three panels in Figure 8. (Here, only report results for U.S., U.K., and Japan. The results for other countries are given in the appendices.)

To briefly summarize, the global factor does not seem to respond much to country factors, except for some mild responses to several countries as can be seen from the top left panel of Figure 8. On the other hand, there are significant interactions between country factors based on evidence from other panels in Figure 8. For example, the shock originated from the U.S. factor seems to significantly affect most other countries, especially Canada and the European countries including U.K. The U.K. factor seems not to have a statistically significant impact on the U.S. factor, although it significantly affects Canada, Sweden, Japan, Australia, and New Zealand. On the other hand, the Japan factor affects most other countries such as the U.S. and U.K. Such rich evidence in the impulse response functions points to complex interstate relations, which could not be simply accounted for by the geographic relations. Our modeling strategy naturally allows the study of such complicated

Figure 8. Impulse responses of global factor and country factors.
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 suitable for the study of such rich interstate relations.

6. CONCLUSION

In this article, we discuss the minimal requirement of identifying 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. The empirical study of the cross-country linkage of bond yields well demonstrates the importance of allowing correlated factors. In addition, a more general class of impulse response functions can be derived under our framework, and the associated confidence bands are readily constructed from the Bayesian estimation results. The proposed identification scheme naturally combines the structural VAR analysis as well. We also show how the Bayes’ factors can help determine the number of dynamic factors. Furthermore, the Bayesian method can easily handle the Bayes’ factors can help determine the number of dynamic factors. In addition, a more general class of impulse response functions yields well demonstrates the importance of allowing correlated factors, which provides a useful framework to study the structural implications of the factor model. The empirical study of the cross-country linkage of bond yields well demonstrates the importance of allowing correlated factors. In addition, a more general class of impulse response functions can be derived under our framework, and the associated confidence bands are readily constructed from the Bayesian estimation results. The proposed identification scheme naturally combines the structural VAR analysis as well. We also show how the Bayes’ factors can help determine the number of dynamic factors. Furthermore, the Bayesian method can easily handle the number of dynamic factors. In addition, a more general class of impulse response functions yields well demonstrates the importance of allowing correlated factors, which provides a useful framework to study the structural implications of the factor model. The empirical study of the cross-country linkage of bond yields well demonstrates the importance of allowing correlated factors. In addition, a more general class of impulse response functions can be derived under our framework, and the associated confidence bands are readily constructed from the Bayesian estimation results. The proposed identification scheme naturally combines the structural VAR analysis as well. We also show how the Bayes’ factors can help determine the number of dynamic factors. Furthermore, the Bayesian method can easily handle the number of dynamic factors. In addition, a more general class of impulse response functions yields well demonstrates the importance of allowing correlated factors, which provides a useful framework to study the structural implications of the factor model. The empirical study of the cross-country linkage of bond yields well demonstrates the importance of allowing correlated factors. In addition, a more general class of impulse response functions can be derived under our framework, and the associated confidence bands are readily constructed from the Bayesian estimation results. The proposed identification scheme naturally combines the structural VAR analysis as well. We also show how the Bayes’ factors can help determine the number of dynamic factors. Furthermore, the Bayesian method can easily handle the number of dynamic factors. In addition, a more general class of impulse response functions yields well demonstrates the importance of allowing correlated factors, which provides a useful framework to study the structural implications of the factor model. The empirical study of the cross-country linkage of bond yields well demonstrates the importance of allowing correlated factors. In addition, a more general class of impulse response functions can be derived under our framework, and the associated confidence bands are readily constructed from the Bayesian estimation results. The proposed identification scheme naturally combines the structural VAR analysis as well. We also show how the Bayes’ factors can help determine the number of dynamic factors. Furthermore, the Bayesian method can easily handle the number of dynamic factors. In addition, a more general class of impulse response functions yields well demonstrates the importance of allowing correlated factors, which provides a useful framework to study the structural implications of the factor model. The empirical study of the cross-country linkage of bond yields well demonstrates the importance of allowing correlated factors. In addition, a more general class of impulse response functions can be derived under our framework, and the associated confidence bands are readily constructed from the Bayesian estimation results. The proposed identification scheme naturally combines the structural VAR analysis as well. We also show how the Bayes’ factors can help determine the number of dynamic factors. Furthermore, the Bayesian method can easily handle the number of dynamic factors. In addition, a more general class of impulse response functions yields well demonstrates the importance of allowing correlated factors, which provides a useful framework to study the structural implications of the factor model.

APPENDIX A. PROOF OF PROPOSITIONS

Proof of Proposition 1: It is well known that a static factor model is identified up to an rotation, where is the number of static factors (Lawley and Maxwell 1971). We show that under the dynamic specification in the observation equation, the only admissible rotation matrix is a block diagonal matrix with the same matrix on the main diagonal. To see this, consider the case and define the static factor as

\[ F_t \equiv \begin{bmatrix} f_t \\ f_{t-1} \end{bmatrix} . \]

From the theory of static factor models, we may identify a rotation of , given by

\[ \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} f_t \\ f_{t-1} \end{bmatrix} = \begin{bmatrix} \tilde{f}_t \\ \tilde{f}_{t-1} \end{bmatrix} . \]

This implies \( \tilde{f}_{t-1} = C f_t + D f_{t-1} = A f_{t-1} + B f_{t-2} \), which in turn implies

\[ (A - D) f_{t-1} = C f_t - B f_{t-2} , \text{ for all } t . \]

Under the VAR specification of , we have

\[ (A - D)(\Phi(L) f_t + \epsilon_t) = C [\Phi(L)(\Phi(L) f_t + \epsilon_t) + \epsilon_t] - B f_{t-2} , \]

where \( \Phi(L) \equiv \Phi_1 + \Phi_2 L + \cdots + \Phi_s L^{s-1} \). Combining terms to obtain

\[ ((A - D) \Phi(L) - C \Phi^2(L) + B) f_{t-2} \]

\[ + (A - D - C \Phi(L)) \epsilon_{t-1} - C \epsilon_t = 0 . \]

By assumption, \( \epsilon_t \) is uncorrelated with both \( (A - D - C \Phi(L)) \epsilon_{t-1} \) and \(((A - D) \Phi(L) - C \Phi^2(L) + B) f_{t-2} \). This implies \( C \epsilon_t = 0 \) or \( C = 0 \) because \( \text{var}(\epsilon_t) > 0 \). And thus \( (A - D)(\Phi(L) + B) f_{t-2} + (A - D) \epsilon_{t-1} = 0 \).

No correlation between \( \epsilon_{t-1} \) and \(((A - D) \Phi(L) + B) f_{t-2} \) implies \( A - D = 0 \) or \( A - D \), which leads to \( B f_{t-2} = 0 \) and thus \( B = 0 \) because \( \text{var}(f_{t-2}) > 0 \). To summarize, the dynamic structure implies \( A = D \), \( B = 0 \), which implies that the dynamic factor model is identified up to a rotation of the dynamic factor.

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 \( \Lambda_j \) by \( A^{-1} \), \( j \in \{0,1,\ldots,s\} \). After the rotation, the new factors \( \tilde{f}_t = A f_t \) have a VAR(1) representation given below

\[ \tilde{f}_t = A \Phi_1 A^{-1} \tilde{f}_{t-1} + \cdots + A \Phi_s A^{-1} \tilde{f}_{t-s} + A \epsilon_t . \]

The observation Equation (1) after the rotation becomes

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

If we can show that 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 \( \text{var}(\epsilon_t) = I_q \) implies that

\[ AA' = I_q \]

or \( A \) is an orthonormal matrix. Next, let

\[ \Lambda_0 = \begin{bmatrix} \lambda_{11} & 0 & \cdots & 0 \\ \lambda_{21} & \lambda_{22} & \cdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_{q1} & \cdots & \cdots & \lambda_{qq} \end{bmatrix} , \quad A^{-1} = \begin{bmatrix} a_{11} & \cdots & a_{1q} \\ \vdots & \ddots & \vdots \\ a_{q1} & \cdots & a_{qq} \end{bmatrix} , \]

Then the normalization of factor loadings requires \( \Lambda_0 A^{-1} \) to be a lower triangular matrix, that is,

\[ \begin{bmatrix} \lambda_{11} & 0 & \cdots & 0 \\ \lambda_{21} & \lambda_{22} & \cdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_{q1} & \cdots & \cdots & \lambda_{qq} \end{bmatrix} , \quad A^{-1} = \begin{bmatrix} a_{11} & \cdots & a_{1q} \\ \vdots & \ddots & \vdots \\ a_{q1} & \cdots & a_{qq} \end{bmatrix} , \]

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 \( \lambda_{ij} \neq 0, \lambda_{ij} \neq 0, i = 1, \ldots, q \). An orthogonal matrix that is lower triangular must be diagonal. In sum we have proved that

\[ A^{-1} = \text{diag}(a_{11}, \ldots, a_{qq}) . \]

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

This proves that the proposed normalization DF1 identifies both the dynamic factors and the associated 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.

Proof of Proposition 2: Similar to the 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 \( \Lambda_j \) by \( A^{-1} \), \( j \in \{0,1,\ldots,s\} \). DF2 implies that both \( \Lambda_0 \) 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 \( A = I_q \), in which turn implies that the factors and factor loadings are identified.
Proof of Proposition 3: Write (4) in matrix form as,

\[
\begin{pmatrix}
X_1^i \\
\vdots \\
X_n^i \\
\end{pmatrix} = 
\begin{bmatrix}
\Lambda_0^W & \Lambda_0^C \\
\vdots & \vdots \\
\Lambda_s^W & \Lambda_s^C \\
\end{bmatrix}
\begin{pmatrix}
X_0^W \\
X_0^C \\
\end{pmatrix} + 
\begin{pmatrix}
e_1^i \\
\vdots \\
e_n^i \\
\end{pmatrix},
\]

(A.3)

Let \(M\) be a rotation matrix. Left-multiply the factors by \(M\) and right-
multiply the loading matrix by \(M^{-1}\). If under the current normalization,
the only admissible \(M\) is an identity matrix, then both the factors and
factor loadings are identified. From Wang (2010), the multilevel factor
structure implies that the only admissible rotation matrix \(M\) takes the
following form,

\[
M = 
\begin{bmatrix}
A & 0 & \cdots & 0 \\
B_1 & A_1 & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
B_C & 0 & \cdots & A_C \\
\end{bmatrix},
\]

(A.4)

Denote \(H_i = [(f_0^W)^T, (f_1^W)^T, \ldots, (f_s^W)^T]\). Then \(H_i\) follows a VAR(1)
process \(H_i = \rho H_{i-1} + U_i\), with \(U_i \sim i.i.d.\{0, I\}\). After the rotation,
we obtain \(\tilde{H}_i = MH_i\). And \(\tilde{H}_i\) follows a VAR(1) process \(\tilde{H}_i = 
MPM^{-1}H_{i-1} + \tilde{U}_i\), with \(\tilde{U}_i = MU_i \sim N(0, MM')\). The normaliza-
tion on \(U_i\) and \(\tilde{U}_i\) implies that \(MM' = I\). Along with (A.4), this in
turn implies that \(M\) is block diagonal: \(M = \text{blkdiag}(A, A_1, \ldots, A_C)\).
Similar to the proof of Proposition 1, it follows that \(\tilde{\Lambda}_0^W\) being lower
triangular with strictly positive diagonal terms implies \(A\) is an identity
matrix. Likewise, \(\{A_c, c = 1, \ldots, C\}\) are all identity matrices. In sum,
the rotation matrix \(M\) becomes an identity matrix. This proves that the
proposed normalization separately identifies both the world factors and
the country factors and the corresponding loadings. \(\square\)

APPENDIX B. IDENTIFYING DYNAMIC FACTOR MODELS WITH A MULTILEVEL FACTOR STRUCTURE

This section extends the result of Section 2.1 to the case with dy-
namic factors. Consider a dynamic version of Equation (4)

\[
X_t^i = \Lambda_t^W f_t^W + \Lambda_t^C f_t^C + \cdots + \Lambda_s^W f_{s-t}^W + \Lambda_s^C f_{s-t}^C + \epsilon_t^i,
\]

(B.1)

Stacking the observations from all countries, we can write (B.1) in
matrix form as,

\[
\begin{pmatrix}
X_1^i \\
\vdots \\
X_n^i \\
\end{pmatrix} = \Lambda_0^W \begin{pmatrix} f_0^W \\ f_1^W \\ \vdots \\ f_s^W \end{pmatrix} + \cdots + \Lambda_s^W \begin{pmatrix} f_0^W \\ f_1^W \\ \vdots \\ f_s^W \end{pmatrix} + \begin{pmatrix} e_1^i \\ \vdots \\ e_n^i \end{pmatrix},
\]

(B.2)

where

\[
\Lambda_j = 
\begin{bmatrix}
\Lambda_j^W & \Lambda_j^C \\
\vdots & \vdots \\
\Lambda_s^W & \Lambda_s^C \\
\end{bmatrix}, \quad j = 0, 1, \ldots, s.
\]

The following corollary is a natural extension of Proposition 3.

Corollary B.1 Define a dynamic factor model as in (B.1), (7), and
(6). Assume that \(\sigma_{0}^2 = I_0\), \(\sigma_1^2 = I_{0+c}\), \(\Lambda_0^W\) and \(\Lambda_0^C\) are
all lower-triangular matrices with strictly positive diagonal terms, and
\(\Lambda_0^W, \Lambda_0^C\) is of full column rank \((c = 1, \ldots, C)\), then the factor model
is uniquely identified.

The proof of Corollary 3 is basically the same as that of Proposition 3.

APPENDIX C. MODEL COMPARISON WITH IMPROPER 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 \(\Lambda_0\) being iden-
tity) throughout. Let \(M_1\) and \(M_2\) be two possible competing models for
the time series \(X \equiv \{X_1, \ldots, X_T\}\), 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

\[
p(X|M_j) = \int p(X, \theta|M_j) d\theta
\]

\[
= \int p(X, \theta|M_j)\pi(\theta) d\theta,
\]

where \(p(X, \theta|M_j)\) is the likelihood under model \(M_j\) and \(\pi(\theta)\) is
the prior. Note that if we use the diffuse prior, \(p(X|M_j)\) in general is not
well defined unless \(\theta\) has a compact support. Possible alternatives to
Bayes’ factors in case of diffuse priors include fractional Bayes’ fac-
tors (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 \(X^0 \equiv \{X_1, \ldots, X_{TB}\}\), calculate the posterior
based on the training sample

\[
p(\theta|M_j, X_1, \ldots, X_{TB}) \equiv p(\theta|M_j, X^0).
\]

Then calculate the marginal likelihood for \(X^1 \equiv \{X_{TB+1}, \ldots, X_T\}\)
conditional on the training sample

\[
p(X^1|M_j, X^0) = \int p(X^1, \theta|M_j, X^0) d\theta
\]

\[
= \int p(X^1|\theta, M_j, X^0)p(\theta|M_j, X^0) d\theta.
\]

The fractional Bayes’ factor is then given by

\[
B_{21}^f \equiv \frac{p(X^1|M_2, X^0)}{p(X^1|M_1, X^0)}.
\]

Note that

\[
p(\theta|X^1, X^0) \propto p(X^1|\theta, X^0) p(\theta|X^0)
\]

\[
\propto p(X^1, X^0|\theta) \pi(\theta),
\]

where \(\pi(\theta)\) is the diffuse prior. This suggests two equivalent sampling
schemes to sample from the posterior \(p(\theta|X^1, X^0)\). One is to treat
\(p(\theta|X^0)\) as the prior and \(p(X^1|\theta, X^0)\) as the likelihood. The other is
to treat \(p(\theta)\) as the prior and \(p(X^1, X^0|\theta)\) 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 \(p(\theta|X^1, X^0)\).
The algorithm consists the following three steps.

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

\[
p(\theta|M_j, X^0) = \int p(\theta, F|M_j, X^0) dF
\]

\[
= \int p(\theta|M_j, F, X^0) p(F|M_j, X^0) dF
\]

\[
\approx \frac{1}{T} \sum_{l=1}^{T} p(\theta|M_j, F^{(l)}, X^0).
\]

Note that \(p(\theta|M_j, F^{(l)}, X^0)\) 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 $\{\theta^{(i)}, F^{(i)}\}_{i=1}^{L}$. Obtain the posterior density given $X$

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

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

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

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

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

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)

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

Thus, we may obtain the logarithm of the fractional marginal likelihood

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

where $\theta^*$ is the posterior mean for $\theta$.

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

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

which is multivariate normal:

$$p(X^1 | \theta, M_j, X^0) = \Pi_{t=0}^{T} p(X_t | X_{t-1}, \ldots, X_1, \theta, M_j)$$

$$= \Pi_{t=0}^{T} N(E(X_t | X_{t-1}, \ldots, X_1, \theta, M_j), \text{var}(X_t | X_{t-1}, \ldots, X_1, \theta, M_j))$$

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) proposed three alternatives: $b = \frac{m}{n}$, $b = \frac{m \log(m)}{\log(n)}$, or $b = \sqrt{\frac{n}{m}}$, ranked by increasing robustness, where $n_0$ is the minimal training sample size such that a posterior is well defined.

**APPENDIX D. NUMERICALLY POSITIVE-DEFINITE CONDITIONAL COVARIANCE 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. (The algorithm is implemented in the Matlab function schur_pos.m in the accompanying computer program files.) 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 $Y | X$

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

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

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

$$P^{-1} = \begin{bmatrix} * & (P_{22} - P_{21} P_{11}^{-1} P_{21})^{-1} \end{bmatrix}.$$
So \( \text{var}(Y|X) = P_{22} - P_{21}P_{11}^{-1}P_{12} = ((P^{-1})_{22})^{-1} \). 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

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

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

\[
M = \begin{bmatrix} U_{21}S_{11}^{-1} & U_{22}S_{22}^{-1} \end{bmatrix}.
\]

Figure 10. Impulse response functions with 2 SE error bands given one std change of \( f_2 \)'s innovation.

Figure 11. Impulse responses of country factors given one unit shock to Canada factor.
then
\[(P^{-1})_{22} = MM'.\]

Perform singular value decomposition for \(M'\) to obtain
\[M' = U_mS_mV'_m,\]
where \(U_m\) is \((p + q) \times (p + q)\), \(S_m\) is \((p + q) \times q\), and \(V_m\) is \(q \times q\).

This admits a \(V_mD V'_m\) representation of \((P^{-1})_{22}\) given below
\[(P^{-1})_{22} = V_mS'_mS_mV'_m = V_mDV'_m,\]
where \(D = S'_mS_m\) is \(q \times q\). This shows that
\[(P^{-1})_{22}^{-1} = V_mD^{-1}V'_m\]
whose numerical outcome is always positive definite. In computationally efficient Matlab implementation, $D = \text{diag}(\text{diag}(S_m)^2)$.

APPENDIX E. SQUARE-ROOT FORM OF THE KALMAN FILTER

We write the square-root form of Kalman filter Matlab program, \textit{kfilter.srf.m}, based on Bierman (1977), Evensen (2009), and Tippett et al. (2003). Consider a state–space representation of the dynamic factor model

\begin{align*}
X_t &= \Lambda F_t + \epsilon_t, \quad \epsilon_t \sim \text{iid}.N(0, R), \\
F_t &= \Phi F_{t-1} + \epsilon_{t-1}, \quad \epsilon_{t-1} \sim \text{iid}.N(0, Q).
\end{align*}

Let $P_{t|t} = \text{var}(F_t|X_t, \ldots, X_1)$. The Kalman filter covariance evolution equations are given by

\begin{align*}
P_{t|t-1} &= \Phi P_{t-1|t-1} \Phi' + Q, \quad P_{t|t} = (I - K_t \Lambda) P_{t|t-1}, \\
K_t &= P_{t|t-1} \Lambda' (\Lambda P_{t|t-1} \Lambda' + R)^{-1} = P_{t|t} \Lambda' R^{-1}. \quad (E.1)
\end{align*}

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

Figure 15. Impulse responses of country factors given one unit shock to Australia factor.
Let the matrix square-root representation of $P_{\beta-1}$ and $P_{\beta}$ be (not unique)

$$P_{\beta-1} = Z_t' (Z_t')', \quad P_{\beta} = Z_t Z_t'. \quad (E.2)$$

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

$$P_{\beta} = Z_t' (I - P_{\beta-1} \Lambda (\Lambda P_{\beta-1} \Lambda + R)^{-1} \Lambda) P_{\beta-1}^{-1}$$

$$= Z_t' (I - (Z_t')' \Lambda (\Lambda Z_t (Z_t')' \Lambda + R)^{-1} \Lambda Z_t') (Z_t')'.$$

where $V_t = \Lambda Z_t, D_t = V_t V_t + R$. Then update $Z_t'$ as

$$Z_t^n = Z_t^1 M_t,$$

where $M_t$ is a square-root of $I - V_t D_t^{-1} V_t'$, that is, $M_t M_t' = I - V_t D_t^{-1} V_t'$ (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

$$I - V_t D_t^{-1} V_t' = (I + (Z_t')' \Lambda R^{-1} \Lambda Z_t')^{-1}.$$

In fact, by the Sherman–Morrison–Woodbury identity

$$(A + UCV)^{-1} = A^{-1} - A^{-1} U (C^{-1} + VA^{-1} U) V A^{-1},$$

it follows that

$$(I + (Z_t')' \Lambda R^{-1} \Lambda Z_t')^{-1} = I - (Z_t')' \Lambda (R + \Lambda Z_t (Z_t')' \Lambda)^{-1} \Lambda Z_t' = I - V_t D_t^{-1} V_t'.$$

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

$$Z_t^n = Z_t^1 H_t (\Gamma_t + I)^{-1/2}, \quad (E.3)$$

where $H_t \Gamma_t H_t'$ is the eigenvalue decomposition of $(Z_t')' \Lambda R^{-1} \Lambda Z_t'$. Note that the eigenvalue decomposition of $I - V_t D_t^{-1} V_t'$ is $H_t (\Gamma_t + I)^{-1} H_t'$ and thus $M_t = H_t (\Gamma_t + I)^{-1/2}$ 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 sec to obtain a Gibbs-sampling chain size of 10,000 using a standard Intel Core i7 2GB RAM computer.

### APPENDIX F. MORE SIMULATION AND EMPIRICAL RESULTS

Using the same data that produced Figures 2 and 3 in Section 4, we provide more simulation results regarding the impulse response functions. We 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 9 and 10 show that with a larger $N$ the error bands tend to be narrower.

Following Section 5, we report the impulse responses for other countries’ shocks in Figure 11 to 16. All figures demonstrate rich evidence regarding interstate relations and point to cross-correlated country factors.

### SUPPLEMENTARY MATERIALS

The supplementary materials provide technical details for Bayesian estimation of the dynamic factor models in Section 2. Gibbs sampling to implement the estimation is discussed.

### ACKNOWLEDGMENTS

We thank the editor and two anonymous referees for helpful comments. Jushan Bai acknowledges partial financial support from the NSF (SES0962410) and Peng Wang acknowledges partial financial support from the Hong Kong Research Grant Council (Project 643910) and the Direct Allocation Grant.
of Hong Kong University of Science and Technology (Project DAG09/10.BM06).

[Received April 2013. Revised May 2014.]

REFERENCES

Aguilar, O., and West, M. (2000), “Bayesian Dynamic Factor Models and Portfolio Allocation,” *Journal of Business and Economic Statistics*, 18, 338–357. [222]

Amengual, D., and Watson, M. (2007), “Consistent Estimation of the Number of Dynamic Factors in Large $N$ and $T$ Panel,” *Journal of Business and Economic Statistics*, 25, 91–96. [224]

Anderson, T. W. (1984), *An Introduction to Multivariate Statistical Analysis*, New York: Wiley. [223]

Anderson, T. W., and Rubin, H. (1956), “Statistical Inference in Factor Analysis,” in *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability* (Vol. 5), ed. J. Neyman, Berkeley, CA: University of California Press, pp. 111–150. [223]

Bai, J. (2003), “Inferential Theory for Factor Models of Large Dimensions,” *Econometrica*, 71, 135–172. [223,227,228]

Bai, J., and Ng, S. (2002), “Determining the Number of Factors in Approximate Factor Models,” *Econometrica*, 70, 191–221. [221,223]

——— (2007), “Determining the Number of Primitive Shocks in Factor Models,” *Journal of Business and Economic Statistics*, 25, 52–60. [222]

——— (2013), “Principal Components Estimation and Identification of Static Factors,” *Journal of Econometrics*, 176, 18–29. [223]

Bai, J., and Wang, P. (2014), “Identification Theory for High Dimensional Static and Dynamic Factor Models,” *Journal of Econometrics*, 178, 794–804. [222]

Berg, A., Meyer, R., and Yu, J. (2004), “Deviance Information Criterion for Comparing Stochastic Volatility Models,” *Journal of Business and Economic Statistics*, 22, 107–120. [229]

Berger, J. O., and Pericchi, L. R. (1996), “The Intrinsic Bayes Factor for Model Selection and Prediction,” *Journal of the American Statistical Association*, 91, 109–122. [222,234]

Bernanke, B., Boivin, J., and Eliasz, P. (2005), “Measuring Monetary Policy: A Factor Augmented Vector Autoregressive (FAVAR) Approach,” *Quarterly Journal of Economics*, 120, 387–422. [221,222]

Bierman, G. J. (1977), *Factorization Methods for Discrete Sequential Estimation*, *Dover Books on Mathematics*, Mineola, NY: Dover Publications. [225,238,239]

Breitung, J., and Tenhofen, J. (2011), “GLS Estimation of Dynamic Factor Models,” *Journal of the American Statistical Association*, 106, 1150–1166. [222]

Carter, C. K., and Kohn, R. (1994), “On Gibbs Sampling for State Space Models,” *Biometrika*, 81, 541–553. [225]

Celeux, G., Forbes, F., Robert, C., and Titterington, D. (2000), “Deviance Information Criterion for Missing Data Models,” *Bayesian Analysis*, 1, 651–673. [229]

Chib, S. (1995), “Marginal Likelihood From the Gibbs Output,” *Journal of the American Statistical Association*, 90, 1313–1321. [235]

Choi, I. (2012), “Efficient Estimation of Factor Models,” *Econometric Theory*, 28, 274–308. [222]

Cogley, T., and Sargent, T. J. (2005), “Drifts and Volatilities: Monetary Policies and Outcomes in the Post WWII US,” *Review of Economic Dynamics*, 8, 262–302. [222]

Connor, G., and Korajczyk, R. (1986), “Performance Measurement With the Arbitrage Pricing Theory: A New Framework for Analysis,” *Journal of Financial Economics*, 15, 373–394. [223]

Crucini, M. J., Kose, M. A., and Otrok, C. (2011), “What are the Driving Forces of International Business Cycles?” *Review of Economic Dynamics*, 14, 156–175. [224]

Del Negro, M., and Otrok, C. (2008), “Dynamic Factor Models With Time-Varying Parameters: Measuring Changes in International Business Cycles,” working paper, Federal Reserve Bank of New York Staff Reports No. 326, New York, NY. [222]

Diebold, F. X., Li, C., and Yue, V. (2008), “Global Yield Curve Dynamics and Interactions: A Generalized Nelson-Siegel Approach,” *Journal of Econometrics*, 146, 351–363. [231]

Doz, C., Giannone, D., and Reichlin, L. (2012), “A Quasi-Maximum Likelihood Approach for Large Approximate Dynamic Factor Models,” *Review of Economic and Statistics*, 94, 1014–1024. [222,224]

Evensen, G. (2009), *Data Assimilation: The Ensemble Kalman Filter* (2nd ed.), Berlin: Springer. [238]

Forni, M., Hallin, M., Lippi, M., and Reichlin, L. (2000), “The Generalized Dynamic Factor Model: Identification and Estimation,” *Review of Economics and Statistics*, 82, 540–554. [221]

Geweke, J. F. (1977), “The Dynamic Factor Analysis of Economic Time Series,” in *Latent Variables in Socio-Economic Models*, eds. D. J. Aigner and A. S. Goldberger, Amsterdam: North-Holland, pp. 365–383. [222]

Gregory, A. W., and Head, A. C. (1999), “Common and Country-Specific Fluctuations in Productivity, Investment, and the Current Account,” *Journal of Monetary Economics*, 44, 423–451. [224]

Junacker, B., and Koopman, S. J. (2008), “Likelihood-Based Analysis for Dynamic Factor Models,” working paper, Tinbergen Institute Discussion Paper 08-007/4. [222]

Kim, C., and Nelson, C. (1999), *State Space Models With Regime Switching: Classical and Gibbs Sampling Approaches With Applications*, Cambridge, MA: MIT Press. [222]

Kose, M. A., Otrok, C., and Whiteman, C. H. (2003), “International Business Cycles: World, Region, and Country-Specific Factors,” *American Economic Review*, 93, 1216–1239. [221,224]

Lawley, D. N., and Maxwell, A. E. (1971), *Factor Analysis in a Statistical Method*, London: Butterworth. [223,233]

Li, Y., Zeng, T., and Yu, J. (2013), “Robust Deviance Information Criterion for Latent Variable Models,” working paper, Singapore Management University, Research Collection School of Economics, Paper 1403. [229]

Moench, E., Ng, S., and Potter, S. (2013), “Dynamic Hierarchical Factor Models,” *Review of Economics and Statistics*, 95, 1811–1817. [224]

O’Hagan, A. (1995), “Fractional Bayes Factors for Model Comparison,” *Journal of the Royal Statistical Society*, Series B, 57, 99–138. [222,234,235]

Quah, D., and Sargent, T. (1993), “A Dynamic Index Model for Large Cross Sections,” in *Business Cycles, Indicators and Forecasting*, NBER Chapters, Cambridge, MA: National Bureau of Economic Research, Inc., pp. 285–310. [222]

Sims, C. A., and Zha, T. (1999), “Error Bands for Impulse Responses,” *Econometrica*, 67, 1113–1156. [239]

Stock, J. H., and Watson, M. W. (1998), “Diffusion Indexes,” Cambridge, MA: National Bureau of Economic Research, NBER working paper 6702. [221]

——— (2005), “Forecasting Using Principal Components From a Large Number of Predictors,” *Journal of the American Statistical Association*, 97, 1167–1179. [223]

——— (2005), “Implications of Dynamic Factor Models for VAR Analysis,” Cambridge, MA: National Bureau of Economic Research, NBER working paper 11467. [221,222,224]

Tippett, M. K., Anderson, J. L., Bishop, C. H., Hamill, T. M., and Whitaker, J. S. (2003), “Ensemble Square Root Filters,” *Monthly Weather Review*, 131, 1485–1490. [238,239]

Wang, P. (2010), “Large Dimensional Factor Models With a Multi-Level Factor Structure: Identification, Estimation, and Inference,” working paper, Hong Kong University of Science and Technology, Hong Kong. [225,234]

Watson, M. W., and Engle, R. F. (1983), “Alternative Algorithms for the Estimation of Dynamic Factor, mimic and Varying Coefficient Regression Models,” *Journal of Econometrics*, 23, 385–400. [222]

Wright, J. H. (2011), “Term Premia and Inflation Uncertainty: empirical Evidence From an International Panel Dataset,” *American Economic Review*, 101, 1514–1534. [230]