An Endogenous Clustered Factor Approach to International Business Cycles*

Neville Francis
Department of Economics
University of North Carolina, Chapel Hill

Michael T. Owyang
Research Division
Federal Reserve Bank of St. Louis

Özge Savaşcan
Department of Economics
University of North Carolina, Chapel Hill

Keywords: factor model, block factor, logistic prior, cluster analysis
PRELIMINARY AND INCOMPLETE
January 29, 2011

Abstract

Factor models have become useful tools for studying international business cycles. Block factor models [e.g., Kose, Otrok, and Whiteman (2003)] can be especially useful as the block restrictions may provide some economic intuition about the nature of the factors. These block factor models, however, require the econometrician to predefine the blocks, leading to potential misspecification. We propose an alternative model in which the blocks are chosen endogenously. The model is estimated using a hierarchical prior which allows us to incorporate series-level covariates which may influence and explain how the series are grouped. Using similar international business cycle data as theirs, we find our country clusters differ in important ways to Kose, Otrok, and Whiteman’s leading to an important reduction in the explanatory power of the factor model.

[JEL: C310; E320]

*Kristie M. Engemann, Charles Gascon, and Kate Vermann provided research assistance. The authors benefitted from discussions with Jim Hamilton, Chris Otrok, and Dave Rapach and comments from seminar participants at Emory University and conference participants at the Duke University Special Workshop on Time Series and Macro and the Applied Econometrics Workshop at the Federal Reserve Bank of St. Louis. Views expressed here are the authors’ alone and do not reflect the opinions of the Federal Reserve Bank of St. Louis or the Federal Reserve System.

\(^1\)corresponding author: owyang@stls.frb.org
1 Introduction

The nature of business cycles is an issue central to macroeconomics. One way to better understand business cycles is to examine their relationships across countries. In particular, we might ask whether some characteristics (e.g., industrial similarity, proximity, language, trade) lead some countries’ business cycles to be correlated. For example, Norrbin and Schlagenhauf (1996) estimated the role of industrial similarities in international business cycles but find a limited role for industry-specific shocks in explaining the forecast error variance of output across countries. Alternatively, coordinated (systematic) policies maybe the impetus behind any synchronicity in business cycles across countries. McKinnon (1982) suggested coordinated monetary policies as a candidate (common) factor for synchronous cross-country business cycles. Finally, correlation between macroeconomic aggregates across countries could be due to unobservable innovations – e.g., common international shocks or country-specific shocks having spillover effects. Using structural vector autoregressions, Ahmed et al. (1993) concludes that spillovers from country-specific labor supply shock are more important than common shocks in generating international business cycles.

Empirical models comparing business cycles across countries generally take one of two approaches: (1) country cycles are estimated separately and then compared or (2) cycles are estimated jointly with numerous assumptions made on the correlation structure. For the most part, these approaches are motivated by the need to reduce complexity and potential parameter proliferation. The first approach leaves the country combinations unrestricted (i.e., any two countries’ cycles can be correlated). The second approach explicitly excludes this. Which approach is taken can depend both on the question to be answered and the econometric techniques used to compute the cycle. For example, the first approach might define a country’s cycle based on a Markov-switching

---

1 Relationship between countries is not restricted to simple correlation of macroeconomic series. One could imagine any one of the common features studied in Engle and Kozicki (1993) – namely, serial correlation, seasonality, deterministic trends, ARCH, heteroskedasticity, and excess kurtosis, as defining the relationship between country pairs. Engle and Kozicki applied their common features concept and find that G7 countries shared similar serial correlation properties.

2 See Baxter and Kouparitsas (2005) for a list of other potential determinants of business cycle comovements across countries.

3 This conclusion was reached after finding data consistent with the substitutability of national monies. In particular, McKinnon found that domestic money demand functions are unstable when no controls are made for foreign exchange rates. Additionally, in an empirical test of the role of borders in the synchronization of business cycles between US Census regions and across European countries Clark and Wincoop (2001) found limited roles for both monetary and fiscal policies.

4 Norrbin and Schlagenhauf (1996) reached a similar conclusion about the relative importance of nation-specific shocks in international business cycles.
or a trend-cycle decomposition, methods typically reserved for smaller systems of equations.\textsuperscript{5} The second approach might define a common cycle via a factor model, where the factor loadings reflect the degree of correlation between country cycles.\textsuperscript{6}

In a series of recent papers, Kose, Otrok, and Whiteman (2003, 2008) posit a factor model with a block structure for the factor loadings.\textsuperscript{7} This block structure provides a straightforward interpretation that may be lacking in standard factor models. Countries within a block have cycles which are correlated through a regional factor. Countries in different blocks are correlated only through a (set of) global factors. The standard factor model can emulate a block factor model if the loadings on the regional factors are close to zero. Even in that case, however, the factors will produce some cross country correlation for countries outside its block. The significant advantage of the block factor model is that it allows a larger number of less pervasive (regional) factors, only a few of which affect any particular country. The disadvantage of the block factor structure is that the blocks (or clusters) are predetermined, meaning we must make significant \textit{ex ante} assumptions about which countries’ cycles are correlated.

In this paper, we take the block factor approach but relax the assumption that the blocks are known \textit{ex ante}.\textsuperscript{8} By being agnostic about block membership we are basically allowing international data to form groups based any one, or combination of potential factors; for example, countries could form clusters due to proximity, coordinated policies, and/or structural innovations. In this sense we are not, \textit{a priori}, guided by any one particular theoretical model. However, once the \textit{ex post} country groupings are determined, potential commonalities within groups could aid in determining important features that any successful model of the international business cycle would need to possess.

The structure of the model is similar to the block factor structure with the addition of a membership indicator determining which block a country belongs to. We assume block membership is mutually exclusive – that is, a country cannot belong to more than one block. This multinomial

\textsuperscript{5}Exceptions are Hamilton and Owyang (2009) and Kaufmann (2010) which use similar approaches to this paper in a Markov-switching environment.

\textsuperscript{6}See, for example, Bai (2003); Bai and Ng (2002); Forni, Hallin, Lippi and Reichlin (2000, 2005); and Stock and Watson (2002a,b).

\textsuperscript{7}See also Boivin and Ng (2006); Onatski (2007); and Hallin and Liska (2008).

\textsuperscript{8}In addition to endogenizing cluster membership, we can use Bayesian techniques to determine the number of clusters and, hence, the number of factors.
approach to the block structure lends itself to estimation with Bayesian methods. In the simplest execution of the multinomial approach, we can assume either a uniform or Dirichlet prior on the membership indicator, giving the model the appearance of a clustering algorithm. For the uniform prior, cluster membership depends solely on the business cycle characteristics of the country’s data as compared to the other members of the cluster. For the Dirichlet prior, the size of the cluster determines the ex ante probability a country is sorted to it. Another approach we explore is the use of a multinomial logistic prior on cluster membership [see also Frühwirth-Schnatter and Kaufmann (2008); Hamilton and Owyang (2009)]. The use of the logistic prior will allow us to incorporate country characteristics (e.g., geographic proximity, industrialization) and enables us to test competing hypotheses about what influences which countries comove.

We find some evidence against the prevailing belief that geographic proximity is a major determinant of cross-country comovements. Using either the flat prior or the logistic prior, we find one pervasive cluster: a set of mostly industrialized nations (the U.S., the U.K., Canada, Germany, Australia, Denmark, and Costa Rica) which share a common "regional" factor.

The balance of the paper is as follows: Section 2 presents the endogenous clustered factor model. Section 3 outlines the Bayesian techniques we use to estimate the model. In this section, we focus on estimation of the model with a uniform prior on cluster membership. Section 4 presents some Monte Carlo evidence showing how well our algorithm identifies the clusters and the consequences of exogenously misidentifying them. Section 5 extends the model and the sampler with a multinomial logistic prior. Section 4 demonstrates how the algorithm differentiates the clusters in small samples using Monte Carlo experiments. In addition, this section shows the consequences of misspecifying the composition of the clusters on the estimation of the factors and the overall model performance. Section 6 presents results from the model with international business cycle data. Section ?? summarizes and concludes.

2 Empirical Model

The model we consider is an extension of the framework used in Kose, Otrok, and Whiteman (2003, KOW) that includes endogenously determined groupings or “clusters”. Suppose that we

---

9Our model has a similar flavor to the sparse factor model of Carvalho, Lopes, and Aguilar (2010).
are interested in $N$ series, each of length $T$.$^{10}$ Each series is composed of a global factor and a less pervasive cluster factor. We assume there are $M$ clusters for which each series $n$ belongs to a single cluster $i$ exclusively.$^{11}$ That is, each series is a function of the global factor and a single cluster factor:

$$y_{nt} = \alpha_n + \beta_{n0}F_{0t} + \beta_{ni}F_{it} + \epsilon_{nt}, \quad (1)$$

where $y_{nt}$ is series $n$’s time-$t$ vector economic indicators; $\alpha_n$ is a vector of scalar intercepts; $\beta_{n0}$ and $\beta_{ni}$ are factor loadings; $F_{0t}$ is the global factor affecting each of the $N$ series; $F_{it}$ is the $i$th cluster factor; and $\epsilon_{nt}$ is a series-specific innovation. In this case, we have imposed that series $n$ belongs to cluster $i$, meaning that it is influenced by the $i$ cluster factor. In the KOW specification, a series’ cluster is predetermined.$^{12}$

We assume that the idiosyncratic error terms, $\epsilon_{nt}$, are serially correlated and follow a $p^\epsilon$ order autoregression process:

$$\epsilon_{nt} = \psi_n(L)\epsilon_{nt-1} + \epsilon_{nt},$$

where innovations $\epsilon_{nt}$, $n = 1...N$, are independent normal random variables with zero mean and variance $\sigma_n^2$, i.e., $\epsilon_{nt} \sim N(0, \sigma_n^2)$. The diagonality of the variance-covariance matrix implies that comovements between series not in the same cluster arise solely from the global factor. Series within the same cluster, on the other hand, can comove through either the global factor or the cluster factor.

In addition to the observation equation, (1), we must place some structure on the evolution of the factors. Suppose that each factor, $F_{it}$, (including the global factor) follows an autoregressive process of the form:

$$F_{it} = \phi_i(L)F_{it-1} + \epsilon_{it}, \quad (2)$$

where $\phi_i(L)$ is a $p^\epsilon_i$ order polynomial in the lag operator and $\epsilon_{it} \sim N(0, \omega_i^2)$. Thus, we explicitly

---

$^{10}$We can relax the assumption that each series has the same length.

$^{11}$Exclusivity is implicit in the KOW framework. It can be relaxed here but would require modifications to the estimation algorithms. These issues have been explored in other papers [e.g., Frühwirth-Schnatter and Lopes (2009)].

$^{12}$In KOW countries are grouped according to their geographic location, i.e., grouped into regions. Clark (1998) and Clark and Shin (2000) find region-specific shocks to be important in international economic fluctuations.
assume that each factor is orthogonal to all other factors. As in KOW, because the sign and the scale of the factors is not uniquely identified from the loadings, the system (1) and (2) requires additional restrictions. That is, the data cannot differentiate between transformations which changes the sign on the factor and its loading.

The addition of a block structure for factors is an important innovation to regional or sectoral analysis. If we believe that shocks may have a hierarchical structure – that is, some are global but some remain confined to the region from which they originate – the block factor structure provides a framework with which to perform this regionally- or industrially-differentiated analysis [see also Moench, Ng, and Potter (2009)]. But what if we are unsure which series should move together? KOW imposes that the countries on the same continent comove. Moench, Ng, and Potter impose that sectoral data comove. While geographic proximity or industrial similarity may be a reason for two countries comovement, other causes – trade, common language – may also be influential. We, therefore, augment the KOW model to account for endogenous clusters.

For endogenous clustering, we augment (1) by allowing the data to choose each grouping. We define a grouping indicator $\gamma_{ni} = \{0, 1\}$, $\sum_i \gamma_{ni} = 1$, which signifies whether series $n$ belongs to cluster $i$, retaining the restriction that a series can belong only to a single cluster. Then, we have

$$y_{nt} = \alpha_n + \beta_{n0}F_{0t} + \sum_i \gamma_{ni}\beta_{ni}F_{it} + \varepsilon_{nt}.$$  

Here, we have assumed that each series is influenced only by a single global and a single cluster factor. The model preserves the restrictions on the comovement of the series. Again, series in different clusters comove only through the global factor, while series within the same cluster can comove apart from the global factor. As opposed to (1), the model implied by (3) allows the data to determine the clusters through a membership indicator, $\gamma_{ni}$. Generalization to models with additional layers of (either endogenously or exogenously determined) factor clusters is straightforward.

---

13 Note that this is a modelling assumption and may not be true for the posterior distribution of the factors generated from small samples.

14 In principle, we could allow the order of the lag polynomial to vary with the factor.

15 For identification, KOW first normalize the sign of one of the elements in each vector of factor loadings (that is, one cross-sectional element). In particular, they normalize the sign of the (1,1) element of the $\beta_{n0}$ matrix. Second, they place an assumption on the variance of the factors. This amounts to assuming that $\omega_i^2$ is known for each factor. While the second assumption is permissible in a generalized framework, the first assumption, for reasons discussed later, will not be.

16 The restriction that a series belong to a single cluster highlights the difference between our paper and that of Frühwirth-Schnatter and Lopes (2009). The end result may reduce to the interpretability of the cluster factors.
ward. For our application, we include an idiosyncratic country factor, $f_{nt}$, which models additional correlation between the within-country variables:

$$y_{nt} = \alpha_n + \beta_{n0} F_{0t} + \sum_{i}^{M} \gamma_{ni} \beta_{ni} F_{it} + \beta_{n1} f_{nt} + \varepsilon_{nt}.$$ 

For these factors, we impose zero loadings across countries, as per KOW.

Our endogenous cluster model, (2) and (3), is subject to the same scale-sign identification issues previously mentioned. Consistent with KOW, we can impose unit variance on the factors. Unlike KOW, we cannot restrict the sign of the first factor loading in each grouping as the clusters are not a priori known. We can, however, impose a sign on the first element (period) of each factor to resolve our identification problem.

3 Estimation

The endogenous cluster factor model outlined in the preceding section can be estimated using Bayesian techniques [see Gelfand and Smith (1990); Casella and George (1992); Carter and Kohn (1994)]. In principle, one could use classical techniques by estimating each cluster combination. The final model could be determined by some model selection criteria. However, employing classical techniques would mean estimating and comparing a very large number of different possible models. Bayesian methods allow us to estimate the cluster membership parameter directly using reversible jump Metropolis-Hastings.

3.1 The Prior

The prior for the parameters of each country’s measurement equation is normal-inverse gamma, i.e., $\rho_n = [\alpha_n, \beta_{n0}]’ \sim N(\mathbf{r}, \mathbf{R})$, $\beta_{ni} \sim N(b_0, B_0)$, and $\sigma_n^2 \sim \Gamma^{-1}(\nu_0, \Sigma_0)$. Note that we have split the slope parameters into two subgroups, $\rho_n$ and $\beta_{ni}$. The first group includes the intercept term and the loading on the global factor. The second subgroup includes the cluster factor loading. The reasons for splitting the slopes will be apparent in our description of the estimation. The factor and measurement error AR parameters also have a normal prior, $\phi \sim N(\Phi, \mathbf{V}^{-1})$ and $\psi \sim N(\Psi, \mathbf{W}^{-1})$ respectively. As a first pass, we will assume a uniform prior over all clusters – that is, a priori, a series is equally likely to belong to any cluster. Recall also that the factor innovation variances,
\( \omega_i^2 \), are constant and predetermined. The hyperparameters for the priors are given in Table 1.

### 3.2 The Sampler

The Metropolis-Hastings–in–Gibbs sampler is an MCMC algorithm in which we draw from the conditional distributions of sets of parameter blocks conditional on the previous draws from the remaining parameters. The sequence of draws from the conditional distributions then converges to the joint posterior density of the entire system. Let \( \mathbf{Y} \) represent the data, \( \Theta \) represent the full set of model parameters, and \( \mathbf{F} \) represent the factors. Then, we can partition the set of parameters into six blocks: (1) the set of intercepts and global factor loadings, \( \mathbf{\rho} \); (2) the set of innovation variances, \( \sigma^2 \); (3) the set of innovation autoregressive parameters, \( \psi_i \); (4) the factors, \( \mathbf{F} \); (5) the set of factor autoregressive parameters, \( \phi_i \); and (6) the group membership indicators, \( \gamma \), along with the cluster factor loadings \( \beta \). After initializing the sampler, we execute 20,000 iterations discarding the first 10,000 to allow for convergence.

#### 3.2.1 Preliminaries

Before discussing conditional distributions for each block, it will be useful to write out a few key quantities. First, the variance-covariance matrix for each observation error term, \( n \), and for each factor, \( i \), are \( \sigma_n^2 \Sigma_n \) and \( \omega_i^2 \Omega_i \), respectively. For convenience, we only specify the notation for each factor here. The observation error terms have similar matrices for each \( y_{nt} \) with a subscript \( n \).

\[
vec(\Omega_i) = (I - \Phi_i \otimes \Phi_i)^{-1} vec\left(u_{pF}' u_{pF}\right),
\]

\[
\Phi_i = \begin{bmatrix}
\phi_i \\
I_{p_F-1} & 0_{p_F-1 \times 1}
\end{bmatrix}
\]

is the companion matrix associated with the \( i \)th factor, and \( u_{pF} \) is a \((p_F \times 1)\) vector with a 1 as the first element and zeros as the rest. Define \( Q_i \) as the Cholesky factor of \( \Omega_i \) and
\[ \Lambda_i = \begin{bmatrix} -\phi_{i1} & \cdots & -\phi_{iP_F} & 1 & 0 & \cdots & \cdots & 0 \\ 0 & -\phi_{i1} & \cdots & -\phi_{iP_F} & 1 & 0 & \cdots & \vdots \\ \vdots & 0 & -\phi_{i1} & \cdots & -\phi_{iP_F} & 1 & 0 & \vdots \\ \vdots & \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & \cdots & -\phi_{i1} & \cdots & -\phi_{iP_F} & 1 \end{bmatrix} , \]

and

\[ S_i^{-1} = \begin{bmatrix} Q_i^{-1} & 0 \\ \Lambda_i \end{bmatrix} . \]

We would have \( S_n^{-1} \) for each observation error term using the Cholesky factor of \( \Sigma_n \) and the matrix \( \Lambda_n \) that are formed with the AR parameters for the error terms, \( \psi_i \). Then, we can use \( S_n^{-1} \) and \( S_i^{-1} \) to quasi-difference the data and the factors.

The factor measurement error is

\[ e_{it} = F_{it} - \phi_i (L) F_{it-1} , \]

by definition. Then, stack the factor measurement error as a vector \( \hat{e}_{ik} = (e_{i,p_{F}+1-k}, \ldots, e_{i,T-k})' \) and define

\[ e_i = [\hat{e}_{i1}, \ldots, \hat{e}_{iP_F}] . \]

Likewise, observation errors can be written as

\[ \varepsilon_{nt} = y_{nt} - \alpha_n - \beta_{n0} F_{0t} - \sum_i \gamma_{ni} \beta_{ni} F_{it} - \beta_{nt} f_{nt} \]

and \( \hat{\varepsilon}_{nk} = (\varepsilon_{n,p_k+1-k}, \ldots, \varepsilon_{n,T-k})' \) and \( e_n = [\hat{\varepsilon}_{n1}, \ldots, \hat{\varepsilon}_{np_k}] \) are defined accordingly.

### 3.2.2 Generating \( \rho|\Theta_{-\rho}, F, Y \)

We draw the intercept and factor loadings conditional on \( Y, F, \) and \( \Theta_{-\rho} \). Because the innovation variance-covariance matrix is assumed diagonal, we can conduct these draws independently for each
Following Chib and Greenberg (1994) and Otrok and Whiteman (1998) we build the likelihood for the first \( p_e \) observations and continue building the posterior distribution for the rest of the likelihood to draw the parameters. To begin define, \( X_n^* = [1_T F_0] \) where \( 1_T \) is a \((T \times 1)\) vector of ones and \( F_0 \) is the last draw of the global factor, then define \( Y_n^* = Y_n - \sum_i \gamma_{ni} \beta_{ni} F_i - \beta_0^i f_n \).

Let;

(i) \( \bar{X}_{n,1}^* = \begin{bmatrix} 1 & F_{0,1} \\ \vdots & \vdots \\ 1 & F_{0,p_e} \end{bmatrix} \) denote the first \( p_e \) rows of \( X_n^* \);

(ii) \( \bar{Y}_{n,1}^* = (Y_{n,1}^*, Y_{n,2}^*, \ldots, Y_{n,p_e}^*) \) denote the first \( p_e \) observations of \( Y_n^* \);

(iii) \( \bar{X}_{n,1} = Q_n^{-1} \bar{X}_{n,1}^* \) and \( \bar{Y}_{n,1} = Q_n^{-1} \bar{Y}_{n,1}^* \);

(iv) \( \bar{X}_{n,2} \) be a \((T - p_e) \times 2\) matrix with \( t^{th} \) row given by \( \psi_n(L)(X_{n,t}^*)' \);

(v) \( \bar{Y}_{n,2} \) be a vector of length \((T - p_e)\) with \( t^{th} \) row given by \( \psi_n(L)(Y_{n,t}^*)' \);

Finally, in stack form define \( \bar{X}_n = \begin{bmatrix} \bar{X}_{n,1} \\ \bar{X}_{n,2} \end{bmatrix} \) and \( \bar{Y}_n = \begin{bmatrix} \bar{Y}_{n,1} \\ \bar{Y}_{n,2} \end{bmatrix} \). Then, \( \rho_{n} = [\alpha_n, \beta_0]^' \) is drawn from

\[
\rho_{n|\Theta-\rho_{n}, F, Y} \sim N(r_n, R_n),
\]

where \( R_n = \left( R_0^{-1} + \sigma_n^{-2} \bar{X}' \bar{X} \right)^{-1} \) and \( r_n = R_n \left( R_0^{-1} r_0 + \sigma_n^{-2} \bar{X}' \bar{Y} \right) \).

### 3.2.3 Generating \( \sigma^2|\Theta-\sigma^2, F, Y \)

Next, we draw \( \sigma_n^2 \) conditional on \( Y \), and \( \Theta-\sigma^2_n \). The innovation variance can be drawn from the inverse gamma posterior

\[
\sigma_n^2|Y, X, \Theta-\sigma_n^2 \sim IG \left( \frac{\nu_0 + T}{2}, \frac{\bar{Y}_0 + \varsigma_n/\varsigma_n}{2} \right),
\]

where each element \( \varsigma_n = \bar{Y}_n - \bar{X}_n \rho_n \).

### 3.2.4 Generating \( \psi|\Theta-\psi, F, Y \)

The draw of \( \psi_n = [\psi_{n1}, \ldots, \psi_{np_e}] \) conditional on the factors, data, and remaining parameters is a straightforward application of Chib and Greenberg (1994). We can sample each set of factor AR parameters in an MH step. For each iteration, the MH algorithm draws a candidate vector from
a proposal distribution. The candidate is then accepted with a probability determined by the likelihood.\footnote{If the candidate is rejected, the draw for that iteration remains the draw from the previous iteration.} For each set of factor parameters, we draw a candidate from the following proposal density:

\[
\psi_n^* \sim N \left( \hat{\psi}_n, V_n^{-1} \right),
\]

where

\[
V_n = \underline{V}_n + \sigma_n^{-2} \varepsilon_n' \varepsilon_n
\]

and

\[
\hat{\psi}_n = V_n^{-1} \left( V_n \psi_n + \sigma_n^{-2} \varepsilon_n' \hat{\varepsilon}_n \right).
\]

Once we have obtained a candidate, it is accepted with probability \( A_{n,\psi} = \min \left\{ \hat{A}_{n,\psi}, 1 \right\} \), where

\[
\hat{A}_{n,\psi} = \frac{\Psi (\psi_n^{*})}{\Psi (\psi_n^{(g-1)})},
\]

the superscript \( g - 1 \) reflects the previous iteration, and

\[
\Psi (\psi_n) = |\Sigma_n (\psi_n)|^{-1/2} \exp \left[ -\frac{1}{2\sigma_n^2} (\hat{Y}_{n,1}^* - \hat{X}_{n,1}^* \rho_n)' \Sigma_n^{-1} (\psi_n) (\hat{Y}_{n,1}^* - \hat{X}_{n,1}^* \rho_n) \right].
\]

Thus, the acceptance probability is the ratio of the pseudolikelihoods for the candidate and the past accepted draw. The draw is accomplished by first generating the candidate from the proposal density and then drawing from a standard uniform. If the draw is less than the acceptance probability, the candidate is accepted. Otherwise, the past draw is retained.

### 3.2.5 Generating \( \phi|\Theta_{-\phi}, F, Y \)

The draw for \( \phi \) mimics the draw from the previous subsection. The autoregression coefficients for the factors, \( \phi_t = [\phi_{t0}, \phi_{t1}, ..., \phi_{tF_t}] \), conditional on the factors, data, and remaining parameters are...
drawn from
\[ \phi_i^* \sim N \left( \hat{\phi}_i, V_i^{-1} \right), \]
where \( \hat{\phi}_i \), \( V_i \) and the pseudolikelihood, \( \Psi (\phi_i) \), follows from above with the necessary change in notation.\(^{18}\)

### 3.2.6 Generating \( \gamma, \beta | \Theta, \gamma, F, Y \)

The draw of the cluster indicator, \( \gamma \), is key to the innovation of this paper. Other papers have explored endogenous clustering in a Bayesian environment but with different models – e.g., clustered Markov switching [Hamilton and Owyang (2009)] and in a general framework [Frühwirth-Schnatter and Kaufmann (2008)]. In those cases, cluster membership can be drawn from

\[
\Pr [\gamma_{ni} = 1 | \Theta, F, Y] \propto p (Y | \Theta, F, \gamma_{ni} = 1) \times \Pr [\gamma_{ni} = 1 | \rho_{-ni}],
\]

where \( p (Y | \Theta, F, \gamma_{ni} = 1) \) is the likelihood associated with the drawn parameters and \( \gamma_{ni} = 1 \).

In our model, the likelihood cannot be computed for cluster \( i \) if \( \gamma_{ni}^{[g-1]} = 0 \), where the superscript \( [g-1] \) represents the past iteration of the sampler. This is because \( \beta_{ni} \) would not have been drawn in the previous iteration.

Our alternative is to draw \( \beta \) and \( \gamma \) jointly. The joint draw of \( \beta \) and \( \gamma \) requires a Metropolis-in-Gibbs draw:

\[
\pi (\beta, \gamma | \Theta, F) = \pi (\gamma | \Theta, F) \pi (\beta | \Theta, \gamma, F),
\]

where we note that the draw of \( \gamma \) does not condition on the draw of \( \beta \). To execute this joint draw, we employ an algorithm similar to that of sections 2.5 and 2.6 Holmes and Held (2006) for a different purpose. Define the joint proposal density, \( q (\beta_n^*, \gamma_n^*) \), as

\[
q (\beta_n^*, \gamma_n^*) = \pi (\beta_n^* | \gamma_n^*, \Theta, F) q (\gamma_n^* | \gamma_n),
\]

where \( \beta_n^* \) and \( \gamma_n^* \) are the candidates and \( \beta_n \) and \( \gamma_n \) are held over from the last draw. We first

\(^{18}\)Here we have \( V_i = V_i + \omega_i^{-2} e_i e_i \), \( \hat{\phi}_i = V_i^{-1} (V_i \phi_i + \omega_i^{-2} e_i e_i) \), and \( \phi_i = |\Omega_i (\phi_i)|^{-1/2} \exp \left[ -\frac{1}{2} \hat{\varphi}_i \Omega_i^{-1} (\phi_i) \hat{\varphi}_i \right] \).
draw a candidate $\gamma^*$ from the proposal density discussed below. Given $\gamma^*$, the first term draws the candidate for $\beta_n$ from its full conditional distribution. Similar to the draw for $\rho_n$, first define,

$$\mathbf{X}_n = \sum_i \gamma_{ni}^* \mathbf{F}_i;$$

and

$$\mathbf{Y}_n = (\mathbf{Y}_n - \alpha \mathbf{1}_n - \beta_{0n} \mathbf{F}_0 - \beta_{1n}^1 \mathbf{f}_n),$$

and let;

1. $\mathbf{X}_{n,1}^* = \left[ \begin{array}{c} \sum_i \gamma_{ni}^* \mathbf{F}_{i,1} \\ \vdots \\ \sum_i \gamma_{ni}^* \mathbf{F}_{i,p_F} \end{array} \right]$ denote the first $p_F$ rows of $\mathbf{X}_n$;

2. $\mathbf{Y}_{n,1}^* = (\mathbf{Y}_{n,1}, \mathbf{Y}_{n,2}, \ldots, \mathbf{Y}_{n,p_F})$ denote the first $p_F$ observations of $\mathbf{Y}_n$;

3. $\mathbf{X}_{n,1}^* = Q_n^{-1} \mathbf{X}_{n,1}^*$ and $\mathbf{Y}_{n,1}^* = Q_n^{-1} \mathbf{Y}_{n,1}^*$;

4. $\mathbf{X}_{n,2}^*$ be a vector of length $(T - p_F)$ matrix with $t^{th}$ row given by $\psi_n(L)(\mathbf{X}_{n,t})'$;

5. $\mathbf{Y}_{n,2}^*$ be a vector of length $(T - p_F)$ with $t^{th}$ row given by $\psi_n(L)\mathbf{Y}_{n,t}$;

Finally, in stack form define

$$\begin{bmatrix} \mathbf{X}_{n,1}^* \\ \mathbf{X}_{n,2}^* \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} \mathbf{Y}_{n,1}^* \\ \mathbf{Y}_{n,2}^* \end{bmatrix}.$$ Then, $\beta_n$ is drawn from

$$\beta_n | \mathbf{\Theta}_{-\beta,\gamma}, \gamma_n^*, \mathbf{F}, \mathbf{Y} \sim N \left( \mathbf{b}_{\gamma_n^*}, \mathbf{B}_{\gamma_n^*} \right), \quad (4)$$

where $\mathbf{B}_{\gamma_n^*} = \left( \mathbf{B}_0 + \sigma_n^{-2} \mathbf{X}_n ' \mathbf{X}_n \right)^{-1}$, $\mathbf{b}_{\gamma_n^*} = \mathbf{B}_n^* \left( \mathbf{B}_n^{-1} \mathbf{b}_0 + \sigma_n^{-2} \mathbf{X}_n ' \mathbf{Y}_n \right)$.

Since we are drawing the $\beta$’s from their full conditional densities – i.e., $\pi (\beta^*|\gamma^*, \mathbf{\Theta}_{-\beta,\gamma})$, the value of the $\beta$ candidate does not appear in the acceptance probability.$^{19}$ In this case, for each $n$, acceptance probability is

$$A_{n,\gamma} = \min \left\{ 1, \frac{|\mathbf{B}_n^*|^{1/2} \exp \left( \frac{1}{2} \mathbf{b}_n^* \mathbf{B}_n^*^{-1} \mathbf{b}_n^* \right) \pi (\gamma_n^*) q (\gamma_n|\gamma_n^*)}{|\mathbf{B}_n|^{1/2} \exp \left( \frac{1}{2} \mathbf{b}_n \mathbf{B}_n^{-1} \mathbf{b}_n \right) \pi (\gamma_n) q (\gamma_n|\gamma_n)} \right\}, \quad (5)$$

where $\mathbf{b}_n$ and $\mathbf{B}_n$ are defined as above and $\mathbf{b}_n$ and $\mathbf{B}_n$ are defined for $\gamma_n$, the value held over from the past draw. Because we have chosen a uniform prior, the second term is identically one.$^{20}$

To close this portion of the algorithm, we need to supply a proposal density for $\gamma_n$. We choose a symmetric density in which we draw a random element of $\gamma_n$, $\gamma_{ni}$, and set this equal to one (setting all other elements equal to zero). The choice of the symmetric proposal makes the last term in (5) identically one.$^{21}$

$^{19}$For a formal proof of this assertion, see Appendix 1 in Troughton and Godsill (1997).

$^{20}$In future sections, we will adapt the model to account for the effect of covariates on the prior of $\gamma$. In that case, the ratio of the model priors, $\pi (\gamma^*) / \pi (\gamma)$, will no longer be unity.

$^{21}$Troughton and Godsill (1997) point out that the $\gamma$ proposal density must allow some nonzero probability of revisiting the same model. That is, the probability that the candidate $\gamma$ is equal to the last iteration's $\gamma$ must be nonzero. In this case, we still redraw $\beta$. 

12
3.2.7 Generating $F|\Theta, Y$

We apply Kose, Otrok, and Whiteman’s multiple factor extension of the methods described in Otrok and Whiteman (1998) to draw $F$. We can iteratively draw each factor conditional on the parameters and draws of the other factors. For example, the global factor is drawn from

$$
\pi (F_0|Y, \Theta, F_{-F_0}) \sim N (h_0, H_0^{-1}) ,
$$

where

$$
H_0 = \frac{1}{\omega} S_0^{-\frac{1}{2}} S_0^{-1} + \sum_{n=1}^{N} \left( \frac{\beta_n^2}{\sigma_n^2} S_n^{-\frac{1}{2}} S_n^{-1} \right),
$$

and

$$
h_0 = H_0^{-1} \left[ \sum_{n=1}^{N} \frac{\beta_n^2}{\sigma_n^2} S_n^{-\frac{1}{2}} S_n^{-1} \left( Y_n - \alpha_n 1_T - \sum_{i} \gamma_n \beta_n^i F_i - \beta_n^i f_n \right) \right].
$$

Then, each group factor is drawn conditional on the global factor and the series for which the latest draw of $\gamma$ are one:

$$
\pi (F_i|Y, \Theta, F_{-F_i}) \sim N (h_i, H_i^{-1}) ,
$$

where

$$
H_i = \frac{1}{\omega} S_i^{-\frac{1}{2}} S_i^{-1} + \sum_{n=1}^{N} \gamma_n \left( \frac{\beta_n^2}{\sigma_n^2} S_n^{-\frac{1}{2}} S_n^{-1} \right),
$$

and

$$
h_i = H_i^{-1} \left[ \sum_{n=1}^{N} \gamma_n \frac{\beta_n^2}{\sigma_n^2} S_n^{-\frac{1}{2}} S_n^{-1} \left( Y_n - \alpha_n 1_T - \beta_n^0 F_0 - \beta_n^i f_n \right) \right].
$$

The draw of the indicator $\gamma_{ni}$ determines which series enter into the distribution of each group factor.

The idiosyncratic factor is drawn similarly.

3.3 Determining the Number of Clusters

As suggested by Frühwirth-Schnatter and Kaufmann (2008), the optimal number of clusters (and, thus, the number of factors) can be ascertained by computing the Bayes factors for models with
different numbers of clusters. We can compute Bayes factors by computing the marginal likelihoods via the subsampling procedures proposed in Chib (1995) and Chib and Jeliazkov (2001). The marginal likelihoods are generated by sequentially subsampling the model parameters. Chib (1995) uses the basic marginal likelihood identity to approximate the marginal likelihood using the output from the Gibbs sampler:

\[
\ln \hat{m}(Y) = \ln f(Y|\Theta^*) + \ln \pi(\Theta^*) - \ln \hat{\pi}(\Theta^*|Y),
\]

where \(\Theta\) is the vector of model parameters, \(\ln \hat{m}(Y)\) is the log marginal likelihood, \(\ln f(Y|\Theta^*)\) is the log likelihood evaluated at a given \(\Theta = \Theta^*\), \(\ln \pi(\Theta^*)\) is the log of the prior evaluated at \(\Theta^*\), and \(\ln \hat{\pi}(\Theta^*|Y)\) is an approximation of the posterior ordinate. For Chib's method, \(\Theta^*\) need only be a high density value of \(\Theta\) (e.g., a modal point).

Computation of \(\ln f(Y|\Theta^*)\) is relatively straightforward and, conditional on a set of priors, computation of \(\ln \pi(\Theta^*)\) is typically straightforward (see the implementation notes on these values below). Chib's method centers on the computation of the posterior ordinate \(\ln \hat{\pi}(\Theta^*|Y)\). His technique relies on expanding the expression \(\hat{\pi}(\Theta^*|Y)\) as

\[
\hat{\pi}(\Theta^*|Y) = \pi(\Theta_1^*|Y) \times \pi(\Theta_2^*|Y, \Theta_1^*) \times \ldots \times \pi(\Theta_N^*|Y, \Theta_1^*, \ldots, \Theta_{N-1}^*),
\]

where here \(N\) represents the number of blocks in the sampler. The first term can be obtained from the initial run of the Gibbs sampler. A typical term for block \(n\) in the above representation is

\[
\pi(\Theta_n^*|Y, \Theta_1^*, \ldots, \Theta_{n-1}^*),
\]

which can be estimated by additional sampling of \(\{\Theta_{n+1}, \ldots, \Theta_N, F\}\) holding constant \(\{\Theta_1^*, \ldots, \Theta_n^*\}\). The estimate \(\hat{\pi}(\Theta_n^*|Y)\) is then

\[
\hat{\pi}(\Theta_n^*|Y) = \frac{1}{G} \sum_{g=1}^{G} \pi(\Theta_n^*|Y, \Theta_1^*, \ldots, \Theta_{n-1}^*, \Theta_n^{(g)}_{n+1}, \ldots, \Theta_N^{(g)}, F_0^{(g)}, F_1^{(g)}, \ldots, F_M^{(g)}, f_1^{(g)}, \ldots, f_N^{(g)}).
\]

\(^{22}\) See Ghosh and Dunson (2008) who estimate posterior probabilities of models with different number of factors. The approach relies on development of parameter expanded Gibbs sampler for all models. The paper computes Bayes factors for models differing only by one factor. This results are then used in model selection.
Details for the resampling methods are given in the appendix.

4 Monte Carlo Experiments

[Results to be inserted]

5 Incorporating Prior Beliefs for Cluster Membership

In the previous section, we assumed a flat prior over cluster membership. There are cases, however, for which prior information could be useful in characterizing the clusters. For example, similar industrial composition or geographic proximity could lead countries to respond to the same common factor. In this section, we consider an alternative logistic prior for the cluster membership indicator, \( \gamma_{ni} \). For this multinomial prior, we include additional blocks consisting of the hyperparameters \( \delta \) and \( \lambda \) and the latent vector \( \xi \).

5.1 Adding a prior for cluster membership

Suppose there exists a \((N_i \times 1)\) vector, \( z_{ni} \), that influences whether a series \( n \) belongs to cluster \( i \)

\[
p(\gamma_i) = \begin{cases} 
\exp \left( z'_{ni} \delta_i \right) / \left[ 1 + \exp \left( z'_{ni} \delta_i \right) \right] & \text{if } \gamma_{ni} = 1 \quad i = 1, ..., M - 1 \\
1 / \left[ 1 + \sum \exp \left( z'_{ni} \delta_i \right) \right] & \text{if } \gamma_{ni} = 1 \quad i = M
\end{cases},
\]

for \( n = 1, ..., N \) and where we have normalized \( \delta_M = 0 \) for identification. In this multinomial framework, series \( n \) cannot be affiliated with more than one idiosyncratic cluster. Note also that the vector, \( z_{ni} \), need not be composed of the same variables for each cluster \( i \). The standard approach to estimating the multinomial logistic is to augment the system with a latent vector in the spirit of Tanner and Wong (1987). The latent vector has the characteristic that the index of the only the nonnegative element also reflects the cluster to which series \( n \) belongs. Formally, let \( \xi_i = (\xi_{1i}, ..., \xi_{Ni})' \) denote a set of latent vector such that

\[
\xi_{ni} \geq 0 \quad \text{if } \gamma_{ni} = 1 \\
\xi_{ni} < 0 \quad \text{otherwise}
\]
Further, each $\xi_{ni}$ follows from a truncated logistic distribution. Sampling from the logistic distribution is nontrivial and typically requires a MH step. Instead, we follow Holmes and Held (2006) by defining a new latent variable, $\chi_{ni}$, that will allow us to sample the hyperparameters of the priors along with the latent variables as additional Gibbs steps in the algorithm above.

Suppose that $\chi_{ni}$ has the limiting distribution of the Kolmogorov-Smirnov test statistic:\(^{23}\)

$$p(\chi_{ni}) = 8 \sum_{j=1}^{\infty} (-1)^{j+1} j^2 \chi_{ni} \exp \left(-2j^2 \chi_{ni}^2\right). \quad (9)$$

If $\chi_{ni} \sim KS$ and $o_{ni} \sim N(0,1)$, then $\xi_{ni} = z_{ni}' \delta_i + 2 \chi_{ni} o_{ni}$ has a logistic distribution with mean $z_{ni}' \delta_i$ and unit scale parameter. The cluster probabilities can be rewritten in terms of the new latent variables:

$$\Pr(\xi_{nk} > 0) = \frac{\exp (z_{nk}' \delta_k)}{1 + \sum_{i=1}^{M-1} \exp (z_{ni}' \delta_i)}. \quad (10)$$

The following subsections demonstrate how to draw the parameters governing the cluster prior.

### 5.2 Augmenting the Sampler

The sampler outlined in Section 3 can be augmented to account for the logistic prior described above. Conditional on the cluster affiliations, $\gamma_i$, draws of most of the model parameters remain unchanged. The change to the logistic prior does alter the acceptance probability in the joint draw of $\gamma_{ni}$ and $\beta_{ni}$. The only other modification is in the form of two additional blocks sampling the three prior parameters: covariate effects, $\delta$; the logistic variances, $\lambda$; and the vector of latent variables, $\xi$. Each of these blocks is drawn by iterating (jointly) over the $M-1$ unnormalized clusters.

#### 5.2.1 Generating $\delta|\Theta, \xi, \lambda, F, Y$

Conditional on $\xi$ and $\lambda$, $\delta_i$ are the slope coefficients from a standard Normal regression model for each of the form:

$$\xi_i = Z_i \delta_i + v_i,$$

\(^{23}\)See Devroye (1986).
where

\[
Z_i = \begin{bmatrix}
z_{i1}' \\
\vdots \\
z_{Ni}'
\end{bmatrix},
\]

\[
v_i \sim N(0, W_i),
\]

and

\[
W_i = \text{diag} [\lambda_1, ..., \lambda_N].
\]

We assume a conjugate normal prior for the logistic slope parameters, \( \delta_i \sim N(d_i, D_i) \). Thus, the covariate effects can be drawn from the posterior

\[
\delta_i | Y, \Theta, F \sim N(d_i^*, D_i^*),
\]

where

\[
d_i^* = \left( D_i^{-1} + Z_i' W_i^{-1} Z_i \right)^{-1} \left( D_i^{-1} d_i + Z_i' W_i^{-1} \xi_i \right)
\]

and

\[
D_i^* = \left( D_i^{-1} + Z_i' W_i^{-1} Z_i \right)^{-1}.
\]

### 5.3 Generating \( \xi_i \) and \( \lambda_i | \Theta, \delta, F, Y \)

If we condition on \( \lambda_{ni} \), then \( \xi_{ni} \) would be Normal, \( \xi_{ni} | \delta_i, \lambda_{ni} \sim N(m_{ni}, \lambda_{ni}) \), for the \( i = 1, ..., M - 1 \) unnormalized clusters. The mean of the Normal distribution reflects this normalization:

\[
m_{ni} = z_{ni}' \delta_i - \log \sum_{j \neq i} \exp \left( z_{nj}' \delta_j \right)
\]

Without that conditioning but given \( \gamma_{ni} \), \( \xi_{ni} \) is a truncated logistic with mean \( m_{ni} \). The truncation point is at zero, where \( \gamma_{ni} \) determines the direction of the truncation: \( \xi_{ni} \geq 0 \) if \( \gamma_{ni} = 1 \) and \( \xi_{ni} < 0 \) if \( \gamma_{ni} = 0 \).

Then, to sample \( \lambda_{ni} \), Holmes and Held (2006) suggest that we can draw a candidate \( \hat{\lambda}_{ni} \) from a Generalized Inverse Gaussian distribution. The candidate, \( \hat{\lambda}_{ni} \), is accepted or redrawn based on the algorithm described by Holmes and Held (2006). Additional details for these draws are given
6 Estimation Results

Table 1: Priors for Estimation

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Prior Distribution</th>
<th>Hyperparameters</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\rho_n = [\alpha_n, \beta_n, \delta_n]$</td>
<td>$N(0,0)$</td>
<td>$\alpha_0 = 0 ; \beta_0 = 1$</td>
<td>$\forall n$</td>
</tr>
<tr>
<td>$\beta_{ni}$</td>
<td>$N(0,0)$</td>
<td>$\beta_0 = 0 ; \beta_0 = 1$</td>
<td>$\forall n, i = 1, 2, ..., M$</td>
</tr>
<tr>
<td>$\sigma_n^2$</td>
<td>$\Gamma(\nu_0, \lambda_0)$</td>
<td>$\nu_0 = 6 ; \lambda_0 = 0.01$</td>
<td>$\forall n$</td>
</tr>
<tr>
<td>$\gamma_n$</td>
<td>Uniform ($\kappa_0, \lambda_0$)</td>
<td>$\kappa_0 = \frac{1}{M}$</td>
<td>$\forall n$</td>
</tr>
<tr>
<td>$\phi$</td>
<td>$N(\phi, \psi^{-1})$</td>
<td>$\phi = 0_p, \psi = 2I_p$</td>
<td>$\forall i$</td>
</tr>
<tr>
<td>$\psi$</td>
<td>$N(\phi, \psi^{-1})$</td>
<td>$\phi = 0_q, \psi = 2I_q$</td>
<td>$\forall i$</td>
</tr>
</tbody>
</table>

Table 1: Priors. Notes: $n$ denotes the series. $i$ indicates the cluster, where $M$ is the total number of cluster factors. $p$ and $q$ are the maximum number of lags in the error and factor lag polynomials, respectively.
References


[27] Heston, Alan; Summers, Robert; and Aten, Bettina. Penn World Table Version 6.3, Center for International Comparisons of Production, Income, and Prices at the University of Pennsylvania, August 2009.


