NOTE

DYNAMIC HIERARCHICAL FACTOR MODELS

Emanuel Moench, Serena Ng, and Simon Potter*

Abstract—This paper uses multilevel factor models to characterize within- and between-block variations as well as idiosyncratic noise in large dynamic panels. Block-level shocks are distinguished from genuinely common shocks, and the estimated block-level factors are easy to interpret. The framework achieves dimension reduction and yet explicitly allows for heterogeneity between blocks. The model is estimated using an MCMC algorithm that takes into account the hierarchical structure of the factors. The importance of block-level variations is illustrated in a four-level model estimated on a panel of 445 series related to different categories of real activity in the United States.

I. Introduction

RECENT research has found that dimension reduction in the form of common factors is useful for forecasting and policy analysis in a data-rich environment. However, a fair criticism of factor models is that the estimated factors are difficult to interpret. One reason is that the factors are typically estimated from a large panel of data without taking full advantage of the data structure. This paper proposes a factor model that uses common and block-specific factors to capture the between- and within-block variations in the data. Each block can be further divided into subblocks to arrive at a hierarchical (multilevel) model. A distinctive feature of the model is that the transition equations for the factors at each level have time-varying intercepts that depend on the factors at the next higher level. We show how this can be taken into account in state-space estimation.

A natural use of the hierarchical model is real-time monitoring of economic activity, which requires filtering news from noise as data arrive on a staggered basis. This can be handled by using the timing of the data releases to organize the data into blocks. More generally, the model can be applied whenever a panel of data can be organized into blocks using a priori information or statistical procedures. The block structure provides a parsimonious way to allow for covariations that are not sufficiently pervasive to be treated as common factors. For example, in multicountry data, there could be series-specific, country (subblock), region (block), and global (common) variations. If the country and regional variations are not properly modeled, they would appear as either weak common factors or idiosyncratic errors that would be cross-correlated among series in the same region.

The remainder of this paper is organized as follows. Section II introduces the hierarchical model and its state-space representation. Estimation via Markov chain Monte Carlo methods is presented. In section III, a four-level model is used to analyze 445 economic time series on real economic activity in the United States. We find that comovement at the block level tends to be more important than comovement across all variables. Furthermore, the principal components estimator tends to treat block-level variation as common. Section IV concludes.

Received for publication April 14, 2010. Revision accepted for publication May 30, 2012.
* Moench: Federal Reserve Bank of New York; Ng: Columbia University; Potter: Federal Reserve Bank of New York.
We thank Chris Otrok and seminar participants at the Bundesbank, the Kansas City Fed, the European Economic Association Annual Meeting, and the Royal Economic Society Annual Conference, as well as two anonymous referees and Mark Watson (the editor) for helpful comments. We thank Evan LeFlore and Jenny Chan for excellent research assistance. S.N. acknowledges financial support from the National Science Foundation under grant SES 0549978. The views expressed in this paper are those of the authors and do not necessarily reflect the views of the Federal Reserve Bank of New York or the Federal Reserve System.
A supplemental appendix is available online at http://www.mitpressjournals.org/doi/suppl/10.1162/REST_a_00359.

II. A Hierarchical Dynamic Factor Model

We assume that the data are stationary, mean zero, standardized to have unit variance after possible logarithmic transformation and detrending. Let N_b denote the number of variables in block $b = 1, \ldots, B$, and let $N = (N_1 + \ldots + N_B)$ be the total number of variables, each with T time series observations. We assume that N and T are both large but that B is much smaller than N.

Consider the two-level dynamic factor model considered in Geweke (1977) and Sargent and Sims (1977). For $t = 1, \ldots, T, n = 1, \ldots, N$, and $k = 1, \ldots, K_F$, the data are assumed to be generated as

$$\begin{aligned} X_{nt} &= \lambda^n(L)F_t + v_{nt}, \\ \psi_{F.k}(L)F_{kt} &= \epsilon_{Fkt}, \end{aligned} \tag{1}$$

where $F_t = (F_{1t}, \ldots, F_{K_Ft})'$ is a $K_F \times 1$ vector of common factors, $\lambda^n(L)$ is a distributed lag of loadings on F_t, and v_{nt} is the idiosyncratic error. We generalize this two-level model by positing that at each t, series n in a given block b has three sources of variations: idiosyncratic, block specific, and common. Let the mean zero block-level factors be $G_{bt} = (G_{b1t}, \ldots, G_{bK_{Gb}t})$. For $n = 1, \ldots, N_b$, a three-level representation of the data is

$$X_{bnt} = \lambda^n_{G.b}(L)G_{bt} + e_{Xbnt}, \tag{2}$$

$$G_{bjt} = \lambda^j_{F.b}(L)F_t + e_{Gbjt}, \tag{3}$$

where $\lambda^n_{G.b}(L)$ is a distributed lag in the block-level factor loadings and $\lambda^j_{F.b}(L)$ is a distributed lag of loadings on the common factors. In the terminology of multilevel models, equation (2) is the level 1 equation and equation (3) is the level 2 equation. The stochastic process for F_t given in equation (1) would constitute the level 3 equation. In this three-level model, variables within a block are correlated because of the common factors F_t or the block-specific variations e_{Gbjt}. However, correlations between blocks are possible only through F_t.

For some blocks, it may be appropriate to break up the data into subblocks. Let Z_{bsnt} be the nth series in subblock s of block b at time t. Let H_{bst} be the K_{Hbs} factors in subblock s. The level 4 dynamics are defined by

$$\begin{aligned} Z_{bsnt} &= \lambda^n_{H.bs}(L)H_{bst} + e_{Zbsnt}, \\ H_{bst} &= \Lambda_{G.bs}(L)G_{bt} + e_{Hbst}, \\ G_{bt} &= \Lambda_{F.b}(L)F_t + e_{Gbt}. \end{aligned}$$

Because not all series need to belong to blocks and subblocks, the data used in a level 4 model are a mixture of Z_{bsnt}, X_{bnt}, and X_{nt}. In general, a model at any level can always be decomposed into a sequence of two-level models as long as there is a reasonable number of series at each level.

An example helps to illustrate the key features of the model. Suppose that we are given data for production, employment, consumption, and so on. Then X_{1it} would be one of the N_1 series collected for production, X_{2it} would be one of the N_2 series collected for employment, X_{3it} would be one of the N_3 consumption series, and so forth. The production, employment, and consumption factors G_{1t}, G_{2t}, and G_{3t} would be correlated because of economy-wide fluctuations, as captured by F_t. However, if the N_2 employment series are derived from two different surveys, specifying two employment subblocks would allow us to model the two independent signals about the state of the labor market.

To close the model, the idiosyncratic components, the subblock-specific, block-specific, and economy-wide factors are assumed to be stationary, normally distributed autoregressive processes of order $q_{Zbsn}, q_{Xbn}, q_{H_{bsi}}, q_{G_{bj}}$, and q_{F_k}, respectively. That is, for $b = 1, \ldots, B$,

$$\begin{aligned}
&\psi_{F.k}(L)F_{kt} = \epsilon_{Fkt}, && \epsilon_{F_k} \sim N(0, \sigma^2_{F_k}) \quad k = 1, \ldots, K_F,\\
&\psi_{G.bj}(L)e_{Gbjt} = \epsilon_{Gbjt}, && \epsilon_{Gbj} \sim N(0, \sigma^2_{Gbj}) \quad j = 1, \ldots, K_{Gb},\\
&\psi_{H.bsi}(L)e_{Hbsit} = \epsilon_{Hbsit}, && \epsilon_{Hbsi} \sim N(0, \sigma^2_{Hbsi}) \quad i = 1, \ldots, K_{H_{bs}},\\
&\psi_{X.bn}(L)e_{Xbnt} = \epsilon_{Xbnt}, && \epsilon_{Xnbt} \sim N(0, \sigma^2_{Xbn}) \quad n = 1, \ldots N_b,\\
&\psi_{Z.bsn}(L)e_{Zbsnt} = \epsilon_{Zbnst}, && \epsilon_{Znbst} \sim N(0, \sigma^2_{Zbsn}) \quad n = 1, \ldots N_{bs}.
\end{aligned}$$

The lag orders can differ across units, subblocks, and blocks. The model could be further enriched by allowing for stochastic volatility or Markov switching effects at different levels of the hierarchy.

The factors and the loadings are not separately identified even in a two-level dynamic factor model. To see this, let $X_t = (X_{1t}, \ldots, X_{N_t})'$ so that in vector form, the observation equation of the model is $X_t = \Lambda(L)F_t + e_t$. Obviously there could exist an invertible polynomial matrix $\Theta(L)$ of arbitrary order such that the common component $\Lambda(L)F_t$ is observationally equivalent to $\tilde{\Lambda}(L)\tilde{F}_t$, where $\tilde{\Lambda}(L) = \Lambda(L)\Theta(L)$ and $\tilde{F}_t = \Theta(L)^{-1}F_t$. To achieve identification, two-level models often assume that $\Lambda(L) = \Lambda$ is a constant lower triangular matrix of order 0 where the elements on the diagonal have a fixed sign (see Geweke & Zhou, 1996; Aguilar & West, 2000). The assumption of constant, lower triangular factor loading matrices can still be used to handle multiple factors in a hierarchical setting. Note that the lower triangular structure is necessary but not sufficient for identification when $\lambda^n(L)$ has lagged dynamics. As shown in theorem 3 of Heaton and Solo (2004), additional restrictions on the polynomial structure will be necessary even for two-level models. Since the data are standardized to have unit variance, we further assume that innovations to the factors have fixed variances. Then $\Psi_F, \Psi_{G.b}, \Psi_{H.bs}, \psi_{X.bn}, \psi_{Z.bsn}$, and the idiosyncratic variances σ_{Xbn} and σ_{Zbsn} are free parameters that adjust to satisfy the variance decomposition identity.

A unique feature of the hierarchical structure is that the transition equation at the block and subblock level has a time-varying intercept since the autoregressive dynamics of e_{Gbjt} imply that

$$\Psi_{G.b}(L)G_{bt} = \Psi_{G.b}(L)\Lambda_{F.b}(L)F_t + \epsilon_{Gbt}.$$

This leads to the block-level transition equation,

$$G_{bt} = \alpha_{F.bt} + \Psi_{G.b1}G_{bt-1} + \ldots + \Psi_{G.bq_{Gb}}G_{bt-q_{Gb}} + \epsilon_{Gbt}, \tag{4}$$

where $\alpha_{F.bt} = \Psi_{G.b}(L)\Lambda_{F.b}(L)F_t$ is correlated across blocks due to F_t. Intuitively, knowledge of the comovement across blocks is useful in estimating the block-specific dynamics. Similarly, the dependence of H_t on G_t implies that

$$H_{bst} = \alpha_{G.bst} + \Psi_{H.bs1}H_{bst-1} + \ldots + \Psi_{H.bsq_{Hbs}}H_{bst-q_{Hbs}} + \epsilon_{Hbst},$$

where $\alpha_{G.bst} = \Psi_{H.bs}(L)\Lambda_{G.bs}(L)G_{bt}$ is common across subblocks. In section IIB, we show how this additional term can be incorporated into a standard sampling method for linear state-space models. Section A.1 in the appendix summarizes the main equations of the four-level model.

A. *Related Work*

A vast number of papers in macroeconomics and finance have studied variants of the two-level dynamic factor model. The difference between our multilevel and a two-level model is best understood when there is a single factor at each level. With $K_{Gb} = K_F = 1$,

$$\begin{aligned}
X_{bnt} &= \lambda^n_{G.b}(\lambda^j_{F.b}F_t + e_{Gbt}) + e_{Xbnt}\\
&= \lambda_{bn}F_t + v_{bnt},
\end{aligned} \tag{5}$$

where $\lambda_{bn} = \lambda^n_{G.b}\lambda^j_{F.b}$ and $v_{bnt} = \lambda^n_{G.b}e_{Gbt} + e_{Xbnt}$. A standard two-level factor model ignores the block structure and simply stacks all observations up. The data would be modeled as

$$X_{nt} = \lambda_n F_t + v_{nt}.$$

This two-level representation corresponds to an exact factor model if the block-specific components $\{e_{Gbt} : b = 1, \ldots, B\}$ were 0 for all t, but is an approximate factor model if v_{nt} was weakly correlated across n and t. Weak cross-sectional correlation requires that the variation in v_{bnt} is not dominated by e_{Gbt} as $N \to \infty$ and $N_b \to \infty$. Instead of imposing this possibly invalid assumption, our hierarchical model explicitly specifies the block structure. The factors are also easier to interpret because the data blocks have a well-defined interpretation.

Multilevel factor models have been considered extensively in the psychology literature. As seen from the review in Goldstein and Browne (2002), for example, these models do not allow for dynamics and typically assume that either T or N is small. Dynamic hierarchical linear models were considered by Gammerman and Migon (1993), but there are no latent variables.

Two models closely related to ours are Diebold, Canlin, and Yue (2008) and Giannone, Reichlin, and Small (2008). Diebold et al. (2008) study a three-level hierarchical factor model for government bond yield data from four countries. They estimate their model in two steps. Country-level yield factors are first estimated by nonlinear least squares and then treated as data in the estimation of the global factors. Hence, they do not take into account the global factor dynamics in the estimation of the country-level block factors. Giannone et al. (2008) are interested in taking advantage of the different timing of data releases for the purpose of "now-casting." First, they estimate the static factors in a two-level model by principal components. Then they estimate the loadings in a second step using Kalman filtering and smoothing techniques.

Kose, Otrok, and Whiteman (2003, 2008) use multilevel factor models to study international business cycle comovements. For each observable variable n in country b, they have

$$x_{bnt} = c_n F_t + d_{bn}e_{Gbt} + e_{Xbnt},$$

where F_t is a world factor, e_{Gbt} is a common shock specific to region b (such as Europe or Asia), and e_{Xbnt} is a component specific to variable n in country b. A similar framework has been used by Stock and Watson (2009) to analyze national and regional factors in housing construction. While we take a bottom-up approach, which explicitly estimates the factors at each level, their top-down approach yields only a block-level component e_{Gbt} that is orthogonal to F_t. Thus Kose et al. (2003, 2008) can estimate only European or Asian factors that are uncorrelated with the global factors, but not factors for those regions per se. Another important difference is that c_n is unconstrained in Kose et al. (2003), which vastly increases the number of parameters to estimate. Since we impose that G_{bt} is linear in F_t, the responses of shocks to F_t for all variables in block b can differ only to the extent that their exposure to the block-level factors differs. Because of this hierarchical structure, we have a total of $K_G \times K_F$ and $N \times K_G$ parameters characterizing loadings on F_t and G_t, whereas Kose et al. (2003) have $N \times K_F$ and $N \times K_G$ parameters, respectively. As K_G is much smaller than N, our framework is more parsimonious.

B. *Estimation via Markov Chain Monte Carlo*

The simplest way to estimate the hierarchical model is to first estimate H_t subblock by subblock via principal components, then estimate G_t from the principal components estimates of H_t, and finally estimate F_t from the principal components estimates of G_t. However, sequential estimation by principal components would not take into account the dependence of H_t on G_t and G_t on F_t through α_{Gt} and α_{Ft}, respectively. Furthermore, principal components estimates the static and not the dynamic factors.

Kose et al. (2003) use Gibbs sampling to estimate latent dynamic factors. By considering their conditional joint distribution, they have to invert a variance-covariance matrix of rank T at each sweep of the sampler. The procedure is computationally costly when N and T are both large. We put more structure on the factors between levels, and we exploit the prediction error decomposition of F_t and G_t to avoid inverting large matrices.

Specifically, we use the MCMC, which samples a Markov chain that has the posterior density of the parameters as its stationary distribution. Kim and Nelson (2000) and Lopes and West (2004), among others, have used the algorithm proposed in Carter and Kohn (1994) and Frühwirth-Schnatter (1994) to estimate two-level factor models with a single factor. We generalize the algorithm to allow for a multilevel structure with multiple factors.

Let $\mathbf{\Lambda} = (\Lambda_H, \Lambda_G, \Lambda_F), \mathbf{\Psi} = (\Psi_F, \Psi_G, \Psi_H, \Psi_Z), \mathbf{\Sigma} = (\Sigma_F, \Sigma_G, \Sigma_H, \Sigma_Z)$. The main steps are:

1. Organize the data into blocks and subblocks to yield $Z_{bst}, b = 1, \dots B, s = 1, \dots B_S$. Get initial values for $\{H_t\}, \{G_t\}$, and $\{F_t\}$ using principal components. Use these to produce initial values for $\mathbf{\Lambda}, \mathbf{\Psi}, \mathbf{\Sigma}$.
2. Conditional on $\mathbf{\Lambda}, \mathbf{\Psi}, \mathbf{\Sigma}, \{G_{bt}\}$ and the data Z_{bst}, draw $\{H_{bst}\}\ \forall\ b\ \forall\ s$.
3. Conditional on $\mathbf{\Lambda}, \mathbf{\Psi}, \mathbf{\Sigma}, \{H_{bt}\}$ and $\{F_t\}$, draw $\{G_{bt}\}\ \forall\ b$.
4. Conditional on $\mathbf{\Lambda}, \mathbf{\Psi}, \mathbf{\Sigma}$, and $\{G_t\}$, draw $\{F_t\}$.
5. Conditional on $\{F_t\}, \{G_t\}$, and $\{H_t\}$, draw $\mathbf{\Lambda}, \mathbf{\Psi}$, and $\mathbf{\Sigma}$.
6. Return to 2.

To analyze series X_{bnt} without a subblock structure, H_t would be dropped from the algorithm and step 2 would be omitted. To analyze series X_{nt} without a block structure, G_t as well as steps 2 and 3 would be omitted, and the algorithm reduces to that for a two-level model. The only complication going from the two-level to a multilevel model is that the transition equations for the subblock and the block-specific factors feature time-varying intercepts that depend on the factors at the next higher level. This dependence needs to be taken into account in sampling the factors.

Denote Ξ_{Gb} the set of parameters $\{\vec{\tilde{\Lambda}}_{G.b}, \vec{\Psi}_{G.b}, \vec{\Sigma}_{G.b}, \Sigma_{\vec{X}.b}\}$. The modified algorithm consists of first running the Kalman filter forward to obtain the sequence $\{\vec{G}_{bt|t}\}$ that accounts for $\vec{\alpha}_{F.bt}$ and the corresponding covariance matrix $\vec{P}_{GbT|T}$ in period T based on all available sample information. This implies the following prediction and updating equations:

$$
\begin{aligned}
\vec{G}_{bt+1|t} &= \vec{\alpha}_{F.bt} + \vec{\Psi}_{G.b}\vec{G}_{bt|t},\\
P_{Gbt+1|t} &= \vec{\Psi}_{G.b} P_{Gbt|t} \vec{\Psi}'_{G.b} + \vec{\Sigma}_{G.b},\\
\vec{G}_{bt|t} &= \vec{G}_{bt|t-1} + P_{Gbt|t-1} \vec{\tilde{\Lambda}}'_{G.b}\\
&\quad \times \left(\vec{\tilde{\Lambda}}_{G.b} P_{Gbt|t-1} \vec{\tilde{\Lambda}}'_{G.b} + \Sigma_{\vec{X}.b}\right)^{-1} \left(\tilde{X}_{bt} - \vec{\tilde{\Lambda}}_{G.b}\vec{G}_{bt|t-1}\right),\\
P_{Gbt|t} &= P_{Gbt|t-1} - P_{Gbt|t-1} \vec{\tilde{\Lambda}}'_{G.b}\\
&\quad \times \left(\vec{\tilde{\Lambda}}_{G.b} P_{Gbt|t-1} \vec{\tilde{\Lambda}}'_{G.b} + \Sigma_{\vec{X}.b}\right)^{-1} \vec{\tilde{\Lambda}}_{G.b} P_{Gbt|t-1}.
\end{aligned}
$$

We can then sample the entire set of factor observations conditional on the parameters Ξ_{Gb} and all the data. The Gaussian and Markovian structure of the state-space model imply that the distribution of $\vec{G}_{bt}$ given $\vec{G}_{bt+1}$ and $\tilde{X}_{bt}$ is normal. Thus,

$$
\vec{G}_{bt}|\tilde{X}_{bt}, \vec{G}^*_{bt+1}, \Xi_{Gb} \sim N\left(\vec{G}_{bt|t,\vec{G}^*_{bt+1}}, \vec{P}_{Gbt|t,\vec{G}^*_{bt+1}}\right),
$$

where using $\vec{G}^*_{bt+1}$ and $\vec{\Psi}^*_{G.b}$ to denote the first K_{Gb} rows of $\vec{G}_{bt+1}$ and $\vec{\Psi}_{G.b}$, respectively:

$$
\begin{aligned}
\vec{G}_{bt|t,\vec{G}^*_{bt+1}} &= E[\vec{G}_{bt}|\tilde{X}_{bt}, \vec{G}^*_{bt+1}]\\
&= \vec{G}_{bt|t} + \vec{P}_{Gbt|t}\vec{\Psi}^{*\prime}_{G.b}\left(\vec{\Psi}^*_{G.b}\vec{P}_{Gbt|t}\vec{\Psi}^{*\prime}_{G.b} + \Sigma_{\vec{G}.b}\right)^{-1}\\
&\quad \times (\vec{G}^*_{bt+1} - \vec{\alpha}_{F.bt+1} - \vec{\Psi}^*_{G.b}\vec{G}_{bt|t}).\\
\vec{P}_{Gbt|t,\vec{G}^*_{bt+1}} &= \mathrm{Var}(\vec{G}_{bt}|\tilde{X}_{bt}, \vec{G}^*_{bt+1})\\
&= \vec{P}_{Gbt|t} - \vec{P}_{Gbt|t}\vec{\Psi}^{*\prime}_{G.b}\left(\vec{\Psi}^*_{G.b}\vec{P}_{Gbt|t}\vec{\Psi}^{*\prime}_{G.b} + \Sigma_{\vec{G}.b}\right)^{-1}\\
&\quad \times \vec{\Psi}^*_{G.b}\vec{P}_{Gbt|t}.
\end{aligned}
$$

Given these conditional distributions, we can then proceed backward to generate draws $\vec{G}^*_{bt}$ for $t = T-1, \dots, 1$. The subblock factors H_{bst} are sampled in an analogous manner, taking into account the dependence on the block-level factors via the time-varying intercept $\alpha_{G.bst}$.

III. A Four-Level Model of Real Activity

We illustrate our model with a hierarchial factor analysis of real economic activity in the United States using a balanced panel of 445 monthly time series from 1992:01 to 2011:03. The data include series on capacity utilization, industrial production, manufacturers' shipments, inventories and orders of durable goods, the labor market as perceived by firms and households, retail sales, wholesale trade, housing starts, new home sales, and manufacturing surveys.

We arrange the data into five blocks: production, employment, consumption, housing, and manufacturing surveys. The housing block comprises data on housing starts and new home sales, while the manufacturing survey block combines data from the Institute for Supply Management, the Philadelphia Fed, and the Chicago Fed. The other three blocks have subblocks that are defined as follows. Industrial production (IP), capacity utilization (CU), and durable goods (DG) constitute the production block; the establishment survey (ES) and the household survey (HS) constitute the employment block; retail sales (RS) and wholesale trade (WT) constitute the consumption block. Our blocks are thus defined using prior information about the structure of the data. An analysis could also be conducted with the blocks determined by statistical criteria such as suggested by Hallin and Liska (2008).

TABLE 1.—DATA AND MODEL STRUCTURE

Block	Subblock	Source	N	K_{Hbs}
Production	CU	Fed	25	1
	IP	Fed	38	1
	DG	Census	60	2
Employment	ES	BLS	82	2
	HS	BLS	92	1
Consumption	WT	Census	54	1
	RS	Census	30	1
Housing		Census	29	
Manufacturing surveys		ISM, Fed	35	

This table summarizes the block structure of the four-level model of real activity in the United States discussed in section III. The data sources for the variables in the block, the number of series N in the block, as well as the number of estimated subblock factors K_{Hbs} are provided. *Fed* denotes the Federal Reserve, *BLS* the Bureau of Labor Statistics, *Census* the Census Bureau, *ISM* the Institute of Supply Managers.

TABLE 2.—CORRELATION OF $\tilde{e}_{rt}$ WITH G_{bjt} AND H_{bsit}

r	Block j		Factor j	R^2
1	Employment		1	0.16
2	Employment		1	0.17
2	Manufacturing surveys		1	0.25
	Block j	**Subblock s**	**Factor i**	**R^2**
1	Production	IP	1.00	0.29
1	Production	DG	2.00	0.25
1	Employment	ES	1.00	0.22
1	Employment	ES	2.00	0.14
2	Production	CU	1.00	0.17
2	Production	DG	2.00	0.36
2	Employment	ES	2.00	0.55
2	Consumption	RS	1.00	0.15

This table summarizes the R^2's obtained from regressions of $\tilde{e}_{rt}$ onto $\hat{G}_{bt}$ and $\hat{H}_{bst}$. $\tilde{e}_{rt}$ is the residual from a regression of $\tilde{F}_{rt}$ on $\hat{F}_t$ where $\tilde{F}_{rt}$ is the rth factor estimated by principal components and $\hat{F}_t$ is the posterior mean of the aggregate factor from the four-level model of real activity estimated using MCMC.

However, the prior information facilitates interpretation of blocks and the factor estimates.

We assume that the factor loading matrix is constant and estimate one common factor, one common factor per block, and one or two factors per subblock.[1] For the two subblocks with $K_{Hbs} = 2$, we follow Aguilar and West (2000) and assume a lower triangular factor loading matrix with 1's on the diagonal:

$$\Lambda_{H.bs} = \begin{bmatrix} 1 & 0 \\ \lambda_{H.bs_{2,1}} & 1 \\ \lambda_{H.bs_{3,1}} & \lambda_{H.bs_{3,2}} \\ \vdots & \vdots \\ \lambda_{H.bs_{Nbs,1}} & \lambda_{H.bs_{Nbs,2}} \end{bmatrix}. \tag{6}$$

This normalization implies that in the presence of multiple subblock-level factors, the first factor loads only on the variable ordered first in a given subblock. It further identifies the signs of the factors. We order first the series thought most likely to represent the subblock dynamics. A summary of the data structure is provided in table 1. The data are transformed to be stationary using Stock and Watson (2008) as a guide. After the data transformation, our sample effectively starts in April 1992, giving $T = 227$ observations for all blocks. A list of all 445 series is provided in the online appendix.

We assume the prior distribution for all factor loadings Λ and autocorrelation coefficients Ψ to be Gaussian with mean zero and variance 1. The prior distribution for the variance parameters is that of an inverse chi square distribution with ν degrees of freedom and a scale of d where ν and d^2 are set to 4 and 0.01, respectively. After discarding the first 50,000 draws as a burn-in, we take another 50,000 draws, storing every fiftieth draw. The reported statistics for posterior distributions are based on these 1,000 draws. We use the principal components estimates of the factors as initial values for F_t, G_t, and H_t. As a cross-check, we run the MCMC algorithm using randomly generated numbers for the factors as starting values and find that it converges to the same posterior means. We also run the sampler on simulated data and find that it converges to the true posterior means.

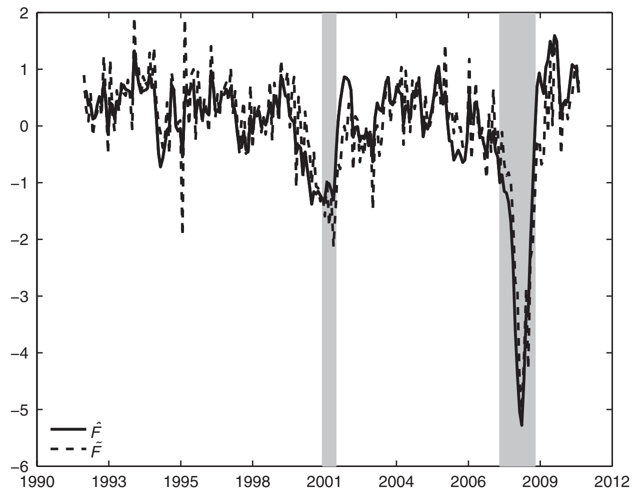
A. Comparison with Principal Components

Let a tilde denote estimates obtained by the method of principal components, and let a "hat" denote estimates obtained from our MCMC algorithm. The IC_2 criterion of Bai and Ng (2002) suggests two static factors in our panel of 445 time series. If block-level variations are important, the principal components extracted from the entire panel of data might capture block-specific rather than aggregate common dynamics. A regression of the principal components $\tilde{F}_{rt}$ on $\hat{F}_t$ yields residuals $\tilde{e}_{rt}$ for each $r = 1, \ldots, 2$. These are variations deemed common by the method of principal components but not by our hierarchical model. We use regressions of $\tilde{e}_{rt}$ on $\hat{e}_{Gbjt}$ and $\hat{e}_{Hbsit}$ to check if these residuals can be explained by our estimated block- and subblock-specific components.

Table 2 reports the R^2 of these regressions that exceed 0.1. The residuals associated with both the first and the second principal component are correlated with the employment block factor, while $\tilde{e}_{2t}$ is correlated with Manufacturing Surveys. The residuals $\tilde{e}_{1t}$ and $\tilde{e}_{2t}$ are correlated with all three subblocks of production. The largest correlation of 0.55 is between $\tilde{e}_{2t}$ and the second factor in the establishment survey subblock. Furthermore, $\tilde{e}_{2t}$ has a correlation of 0.36 with the second durable goods factor. A possible explanation as to why principal components treat these variations as pervasive is that the employment and durable goods data are overrepresented (in terms of number of series) in the panel.

Figure 1 graphs our $\hat{F}_t$ and the first principal component $\tilde{F}_{1t}$. Note that $\hat{F}_t$ is noticeably smoother than $\tilde{F}_{1t}$. For example, $\tilde{F}_{1t}$ features large

FIGURE 1.—COMPARISON TO PRINCIPAL COMPONENTS ESTIMATES



This figure plots the posterior mean of the common factor estimate $\hat{F}$ (solid line) from our four-level model of real activity along with the first principal component $\tilde{F}$ (dashed line) extracted from the entire data panel. Shaded areas indicate NBER recessions. The sample period is 1992:04 to 2011:03.

[1] We started by estimating two factors for each subblock and dropped factors whose posterior distribution was not tightly estimated.

TABLE 3.—DECOMPOSITION OF VARIANCE

Block	Subblock	$Share_F$		$Share_G$		$Share_H$		$Share_Z$	
		Posterior Mean (Standard Deviation)							
Production	CU	0.137	(0.027)	0.031	(0.006)	0.030	(0.005)	0.802	(0.035)
Production	IP	0.162	(0.028)	0.037	(0.006)	0.027	(0.005)	0.773	(0.035)
Production	DG	0.032	(0.010)	0.007	(0.002)	0.177	(0.039)	0.784	(0.046)
Employment	ES	0.012	(0.007)	0.122	(0.025)	0.213	(0.035)	0.652	(0.049)
Employment	HS	0.003	(0.002)	0.026	(0.011)	0.109	(0.030)	0.863	(0.037)
Consumption	WT	0.031	(0.011)	0.033	(0.010)	0.031	(0.010)	0.906	(0.029)
Consumption	RS	0.010	(0.005)	0.011	(0.005)	0.106	(0.029)	0.874	(0.034)
Housing		0.010	(0.006)	0.074	(0.018)			0.916	(0.018)
Manufacturing surveys		0.034	(0.015)	0.096	(0.026)			0.871	(0.035)

This table summarizes the decomposition of variance for the four-level model of real activity. For each (sub)block of data, $share_F$, $share_G$, $share_H$, and $share_Z$ denote the average variance share across all variables in the block due to aggregate, block-level, subblock-level and idiosyncratic shocks, respectively.

spikes in 1996 that are not prevalent in our common factor estimate $\hat{F}_t$. One potential explanation for this relates to the government shutdown of the budget in January 1996. Due to the large number of employment-related series in the data set, the first principal component extracted from the panel puts a lot of weight on this block-level event. In contrast, it is appropriately treated as variations associated with the employment block in our model. Notice also that our estimated common factor nicely tracks the two recessions in our sample period. According to the common factor estimate, real activity bottomed at the end of 2001, consistent with the official NBER business cycle chronology that reports November 2001 as the trough of the recession. The estimated factor also documents the striking collapse of real activity in late 2008 and its subsequent sharp rebound in early 2009.

B. *Importance of Block-Level Variations*

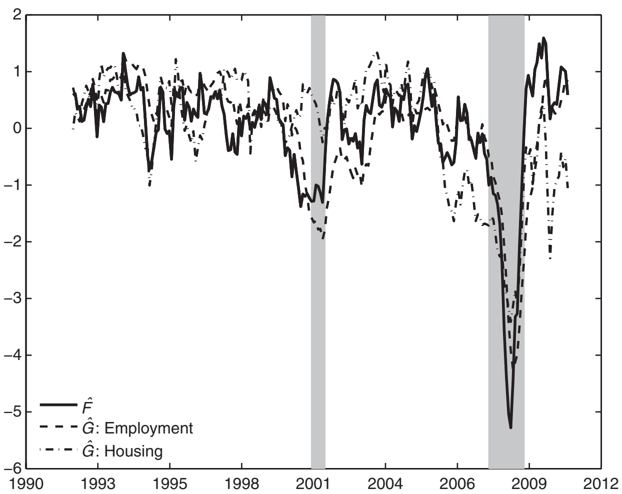
Our model can be used to compute the importance of the aggregate ($share_F$), block-specific ($share_G$), and subblock-specific ($share_H$) components as well as idiosyncratic noise ($share_Z$) relative to the total variation in the data.

A two-level factor model does not distinguish between F_t and G_t or H_t. Table 3 reports the posterior means and standard deviations of the estimated variance shares for all blocks and subblocks of our data set. The latter show that all shares are precisely estimated.

For housing and manufacturing (which have no subblock structure), the block-specific variations, dominate the aggregate variations, as $share_G$ is much larger than $share_F$. The capacity utilization and industrial production subblocks of production both have $share_F$ of around 0.15, while $share_G$ and $share_H$ are around 0.10. However, for durable goods in the same block, $share_F$ and $share_G$ are much smaller than $share_H$. Each of the two subblocks in the employment block also features larger $share_G$ and $share_H$ than $share_F$. Interestingly, in all categories, $share_Z$ exceeds 0.65 highlighting the importance of series-specific shocks.

The hierarchical model also allows us to track the developments in certain sectors of the economy. Figure 2 plots the estimated aggregate factor $\hat{F}$, along with $\hat{G}$ for the employment and the housing block. Leading into the 2001 recession, housing activity was stronger than aggregate activity $\hat{F}_t$ and only retracted briefly during the recession. The recession of 2001 was followed by a jobless recovery as the employment factor failed to keep pace with $\hat{F}_t$ after the recession. In the most recent recession of 2008–2009, housing activity declined well ahead of aggregate activity as measured by $\hat{F}_t$ and dropped back to negative growth rates after a brief recovery. Figure 2 thus shows that comovement in economic time series can coexist with heterogeneous variations between blocks. The hierarchical model allows us to jointly model these common variations at different levels.

FIGURE 2.—FOUR-LEVEL MODEL OF REAL ACTIVITY WITH FIVE BLOCKS



This figure plots the posterior mean of the common factor estimate $\hat{F}$ (solid line) along with the posterior means of the block-level factors $\hat{G}$ for the employment (dashed line) and housing (dash-dotted) block, respectively, estimated from our four-level model of real activity in the United States. Shaded areas indicate NBER recessions. The sample period is 1992:04 to 2011:03.

IV. Conclusion

This paper lays out a framework in which the effects of aggregate, block-level, and idiosyncratic shocks can be coherently analyzed while still achieving a reasonable level of dimension reduction. By extracting common components from blocks, the estimated factors have a straightforward interpretation. While multilevel models are computationally more demanding than two-level models, explicitly modeling the block level variation also makes it less likely that shocks at the block level will be confounded with genuinely common shocks. Estimation requires only a simple variation to existing MCMC methods for estimating two-level factor models.

REFERENCES

Aguilar, Omar, and Mike West, "Bayesian Dynamic Factor Models and Portfolio Allocation," *Journal of Business and Economic Statistics* 18 (2000), 338–357.

Bai, Jushan, and Serena Ng, "Determining the Number of Factors in Approximate Factor Models," *Econometrica* 70:1 (2002), 191–221.

Carter, Christopher K., and Robert Kohn, "On Gibbs Sampling for State Space Models," *Biometrika* 81:3 (1994), 541–553.

Diebold, Francis X., Li Canlin, and Vivian Yue, "Global Yield Curve Dynamics and Interations: A Dynamic Nelson-Siegel Approach," *Journal of Econometrics* 146 (2008), 315–363.

Frühwirth-Schnatter, Sylvia, "Data Augmentation and Dynamic Linear Models," *Journal of Time Series Analysis* 15 (1994), 183–202.

Gammerman, Dani, and Helio S. Migon, "Dynamic Hierarchical Models," *Journal of Royal Statistical Society Series B* 55:3 (1993), 629–642.

Geweke, John, "The Dynamic Factor Analysis of Economic Time Series," in D. J. Aigner and A. S. Goldberger (eds.), *Latent Variables in Socio Economic Models* (Amsterdam: North-Holland, 1977).

Geweke, John, and Guofu Zhou, "Measuring the Pricing Error of the Arbitrage Pricing Theory," *Review of Financial Studies* 9:2 (1996), 557–587.

Giannone, Domenico, Lucrezia, Reichlin, and David Small, "Nowcasting: The Real-Time Information Content of Macroeconomic Data," *Journal of Monetary Economics* 55:4 (2008), 665–676.

Goldstein, Harvey, and William Browne, "Multilevel Factor Analysis Using Markov Chain Monte Carlo Estimation," (pp. 225–243), in G. Marcoulides, I. Moustaki (eds.), *Latent variable and Latent Structure Models* (Mahwah, NJ: Erlbaum, 2002).

Hallin, Marc, and Roman Liska, "Dynamic Factors in the Presence of Block Structure," European University Institute working paper 2008/22 (2008).

Heaton, Chris, and Victor Solo, "Identification of Causal Factor Models of Stationary Time Series," *Econometrics Journal* 7 (2004), 618–627.

Kim, Chang-Jin, and Charles Nelson, *State Space Models with Regime Switching* (Cambridge, MA: MIT Press, 2000).

Kose, Ayhan, Christopher Otrok, and Charles H. Whiteman, "International Business Cycles: World Region and Country Specific Factors," *American Economic Review* 93:4 (2003), 1216–1239.

——— "Understanding the Evolution of World Business Cycles," *International Economic Review* 75 (2008), 110–130.

Lopes, Hedibert Freitas, and Mike West, "Bayesian Model Assessment in Factor Analysis," *Statistica Sinica* 14 (2004), 41–87.

Sargent, Thomas, and Chris Sims, "Business Cycle Modelling without Pretending to Have Too Much a Priori Economic Theory," in C. Sims (ed.), *New Methods in Business Cycle Research* (Minneapolis: Federal Reserve Bank of Minneapolis, Minneapolis, 1977).

Stock, James H., and Mark W. Watson, "Forecasting in Dynamic Factor Models Subject to Structural Instability," Princeton University working paper (2008).

——— "The Evolution of National and Regional Factors in U.S. Housing Construction," in T. Bollerslev, J. Russell, and M. Watson (eds.), *Volatility and Time Series Econometrics: Essays in Honor of Robert F. Engle* (New York: Oxford University Press, 2009).

APPENDIX

A1. Four-Level Model: State-Space Representation

Stacking all variables Z_{bsnt} in a subblock and pseudo-differencing the serially correlated idiosyncratic components e_{Zbsnt}, we can write the observation equation at the subblock level as

$$\tilde{Z}_{bst} = \tilde{\Lambda}_{H.bs}(L)H_{bst} + \epsilon_{Zbst}, \qquad \forall b = 1, \ldots, B, \forall\, s = 1, \ldots, B_S,$$

where $\tilde{Z}_{bst} = \Psi_{Z.bs}(L)Z_{bst}$ and $\tilde{\Lambda}_{H.bs}(L) = \Psi_{Z.bs}(L)\Lambda_{H.bs}(L)$ is a $N_b \times K_{Hbs}$ matrix polynomial of order $l_H^* = q_Z + l_H$. Moreover, the state equation at the subblock level is

$$H_{bst} = \alpha_{G.bst} + \Psi_{H.bs1}H_{bst-1} + \cdots + \Psi_{H.bsq_H}H_{bst-q_H} + \epsilon_{Hbst}$$

where

$$\alpha_{G.bst} = \Psi_{H.bs}(L)\Lambda_{G.bs}(L)G_{bt}, \qquad \forall b = 1, \ldots, B, \forall\, s = 1, \ldots, S.$$

Together, these two equations imply the following state-space form:

$$\tilde{Z}_{bst} = \begin{bmatrix} \tilde{\Lambda}_{H.bs0} & \tilde{\Lambda}_{H.bs1} & \cdots & \tilde{\Lambda}_{H.bsl_H^*} \end{bmatrix} \begin{pmatrix} H_{bst} \\ H_{bst-1} \\ \vdots \\ H_{bst-l_H^*} \end{pmatrix} + \epsilon_{Zbst}$$

$$\begin{pmatrix} H_{bst} \\ H_{bst-1} \\ \vdots \\ H_{bst-l_H^*} \end{pmatrix} = \begin{pmatrix} \alpha_{G.bst} \\ 0 \\ \vdots \\ 0 \end{pmatrix} + \begin{bmatrix} \Psi_{H.bs1} & \cdots & \Psi_{H.bsq_H} & 0 \cdots 0 \\ I & 0 & \vdots & \vdots & \vdots \\ \vdots & \ddots & \ddots & \vdots & \vdots \\ 0 & \cdots & I & 0 \cdots 0 \end{bmatrix} \begin{pmatrix} H_{bst-1} \\ H_{bst-2} \\ \vdots \\ H_{bst-l_H^*-1} \end{pmatrix} + \begin{pmatrix} \epsilon_{Hbst} \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

or

$$\tilde{Z}_{bst} = \vec{\tilde{\Lambda}}_{H.bs}\vec{H}_{bst} + \vec{\epsilon}_{Zbst} \tag{A1}$$
$$\vec{H}_{bst} = \vec{\alpha}_{G.bst} + \vec{\Psi}_{Hbs}\vec{H}_{bst-1} + \vec{\epsilon}_{Hbst} \tag{A2}$$

For blocks that do have a subblock structure, the observation and state equation at the block level become

$$\tilde{H}_{bt} = \begin{bmatrix} \tilde{\Lambda}_{G.b0} & \tilde{\Lambda}_{G.b1} & \cdots & \tilde{\Lambda}_{G.bl_G^*} \end{bmatrix} \begin{pmatrix} G_{bt} \\ G_{bt-1} \\ \vdots \\ G_{bt-l_G^*} \end{pmatrix} + \epsilon_{Hbt}$$

$$\begin{pmatrix} G_{bt} \\ G_{bt-1} \\ \vdots \\ G_{bt-l_G^*} \end{pmatrix} = \begin{pmatrix} \alpha_{F.bt} \\ 0 \\ \vdots \\ 0 \end{pmatrix} + \begin{bmatrix} \Psi_{G.b1} & \cdots & \Psi_{G.bq_G} & 0 \cdots 0 \\ I & 0 & \vdots & \vdots & \vdots \\ \vdots & \ddots & \ddots & \vdots & \vdots \\ 0 & \cdots & I & 0 \cdots 0 \end{bmatrix} \begin{pmatrix} G_{bt-1} \\ G_{bt-2} \\ \vdots \\ G_{bt-l_G^*-1} \end{pmatrix} + \begin{pmatrix} \epsilon_{Gbt} \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

or

$$\tilde{H}_{bt} = \vec{\tilde{\Lambda}}_{G.b}\vec{G}_{bt} + \vec{\epsilon}_{Hbt}, \tag{A3}$$
$$\vec{G}_{bt} = \vec{\alpha}_{F.bt} + \vec{\Psi}_{G.b}\vec{G}_{bt-1} + \vec{\epsilon}_{Gbt}. \tag{A4}$$

For blocks that do not have a subblock structure, the observation and state equation at the block level are

$$\tilde{X}_{bt} = \tilde{\Lambda}_{G.b}(L)G_{bt} + \epsilon_{Xbt}$$
$$\text{and} \qquad G_{bt} = \alpha_{F.bt} + \Psi_{G.b1}G_{bt-1} + \ldots + \Psi_{G.bq_{Gb}}G_{bt-q_{Gb}} + \epsilon_{Gbt}$$

where $\tilde{X}_{bt} = \Psi_{X.b}(L)X_{bt}$ and $\tilde{\Lambda}_{G.b}(L) = \Psi_{X.b}(L)\Lambda_{G.b}(L)$ is a $N_b \times K_{Gb}$ matrix polynomial of order $l_G^* = q_X + l_G$. Furthermore, $\alpha_{F.bt} = \Psi_{G.b}(L)\Lambda_{F.b}(L)F_t, \forall\, b = 1, \ldots, B$.

Finally, we have the following observation and state equations at the aggregate level:

$$\tilde{G}_t = \tilde{\Lambda}_F(L)F_t + \epsilon_{Gt}, \tag{A5}$$
$$\text{and} \qquad F_t = \Psi_{F.1}F_{t-1} + \ldots + \Psi_{F.q_F}F_{t-q_F} + \epsilon_{Ft}, \tag{A6}$$

where $G_t = \Psi_G(L)G_t$ and $\tilde{\Lambda}_F(L) = \Psi_G(L)\Lambda_F(L)$ is a $K_G \times K_F$ matrix polynomial of order $l_F^* = q_G + l_F$.

A2. Variance Decomposition

The total unconditional variance for each individual variable Z_{bsn} can be decomposed according to

$$\begin{aligned}\mathrm{Var}(Z_{bsn}) = \gamma'_{F.bsn} vec(\mathrm{Var}(F)) + \gamma'_{G.bsn} vec(\mathrm{Var}(e_{Gb})) + \cdots \\ + \gamma'_{H.bsn} vec(\mathrm{Var}(e_{Hbs})) + vec(\mathrm{Var}(e_{Zbsn}))\end{aligned} \tag{A7}$$

where

$$\begin{aligned}\gamma'_{F.bsn} &= \left(\sum_{l=0}^{l_H} \lambda'_{H.bsn}(l) \otimes \lambda'_{H.bsn}(l)\right) \times \left(\sum_{l=0}^{l_G} \lambda'_{G.bs}(l) \otimes \lambda'_{G.bs}(l)\right) \\ &\quad \times \left(\sum_{l=0}^{l_F} \lambda'_{F.b}(l) \otimes \lambda'_{F.b}(l)\right) \\ \gamma'_{G.bsn} &= \left(\sum_{l=0}^{l_H} \lambda'_{H.bsn}(l) \otimes \lambda'_{H.bsn}(l)\right) \times \left(\sum_{l=0}^{l_G} \lambda'_{G.bs}(l) \otimes \lambda'_{G.bs}(l)\right) \\ \gamma'_{H.bsn} &= \left(\sum_{l=0}^{l_H} \lambda'_{H.bsn}(l) \otimes \lambda'_{H.bsn}(l)\right) \\ vec(\mathrm{Var}(F)) &= \left[I - \sum_{q=1}^{q_F} \left(\Psi_{F.q} \otimes \Psi_{F.q}\right)\right]^{-1} vec(\Sigma_F) \\ vec(\mathrm{Var}(e_{Gb})) &= \left[I - \sum_{q=1}^{q_{Gb}} \left(\Psi_{G.bq} \otimes \Psi_{G.bq}\right)\right]^{-1} \times vec(\Sigma_{G_b}) \\ vec(\mathrm{Var}(e_{Hbs})) &= \left[I - \sum_{q=1}^{q_{Hbs}} \left(\Psi_{H.bsq} \otimes \Psi_{H.bsq}\right)\right]^{-1} \times vec(\Sigma_{H_{bs}}) \\ vec(\mathrm{Var}(e_{Zbsn})) &= \left[1 - \sum_{q=1}^{q_{Zbsn}} \psi^2_{Z.bsnq}\right]^{-1} \times \sigma^2_{Zbsn}.\end{aligned}$$