Dynamic Hierarchical Factor Models*

Serena Ng† Emanuel Moench‡ Simon Potter§

August 22, 2008
Preliminary Draft

Abstract

This paper presents an approach to dynamic factor modeling in which variations can be idiosyncratic, block-specific, or common across blocks and units. Existing two level factor models do not usually account for variations at the block level which implies that these can be confounded with genuine common co-movements in the data. Specifying the block structure also facilitates interpretation of shocks to economic activity. Our approach is aimed at three types of applications: (i) decomposition of variance to assess the relative importance of different shocks, (ii) understanding how block-specific shocks impact overall economic activity, and (iii) monitoring and economic forecasting. We estimate a three level model for housing and find evidence of aggregate housing shocks, but these are small relative to common regional shocks and shocks to individual series within regions. We also estimate a model for real economic activity excluding housing. The 315 time series are organized into six blocks according to the timing of data releases. The results suggest that block-specific variations are important. Data released in the household survey component of the monthly employment situation report turn out to be mostly idiosyncratic and bear little information about the state of real economic activity. Finally, we estimate a four level model consisting of 402 series organized into an output, an employment, and a demand block that each are broken down into sub-blocks. According to this model, the level of real economic activity at the end of our sample in February 2008 was about 1.5 standard deviations below average, but still well above the level of 3 standard deviations below average reached at the trough of the recession in 2001.

*We would like to thank Evan LeFlore for excellent research assistance on this project. The first author would like to acknowledge 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.

†Columbia University, 420 W. 118 St., MC 3308, New York, NY 10025. serena.ng@columbia.edu

‡Federal Reserve Bank of New York, 33 Liberty St., New York, NY 10045. emanuel.moench@ny.frb.org

§Federal Reserve Bank of New York, 33 Liberty St., New York, NY 10045. simon.potter@ny.frb.org

1 Introduction

Macroeconomists define the business cycle as common fluctuations across a wide range of economic variables. Starting with the seminal work by Geweke (1977), dynamic factor models have been used to extract the common fluctuations in economic variables. As predicted by economic theory and verified by numerous empirical papers with Sargent and Sims (1977) being the first, the business cycle fluctuations can be captured by a relatively low dimensional factor compared to the large number of economic time series available. This has lead to intensive research in the last decade on principal component type estimates that are able to extract the low dimensional factor under very weak conditions on the underlying time series and factors. Work by Stock and Watson (2002) and Forni et al. (2005) stimulated much interest in this area of research. Thus, whereas early work extracts the common factors from a relatively small number of series, the more modern practice is to extract them from large panels of data. These are usually constructed by taking a number of individual series from different data releases produced by statistical agencies often with the intent to give a balanced representation to the various sectors of the economy. In this paper we propose a new dynamic factor model that exploits the block structure of data releases by statistical agencies, information on the sectoral structure of the economy, and prior views about how economy activity might be related across market, region, industry etc. to improve the estimation and interpretation of the low dimensional business cycle factor.[1]

We develop a state space estimation methodology for a highly flexible hierarchical (multi-level) dynamic factor model. The hierarchy is produced by splitting a large panel of data collected into a much smaller number of blocks, each of which consists of a reasonably large number of series. Each block is driven by its own block-specific factors that has a component driven by factors that are common across all of the blocks. Higher level models can be obtained by further splitting some blocks into finer sub-blocks. This hierarchical structure implies that the transition equation for the common factors at a given level has a time varying intercept that depends on the common factors at the next level and must be taken into account in the filtering algorithm. The state space methodology applied to this hierarchical structure naturally facilitates the monitoring and interpretation of changes in the state of the economy as new data arrives in real time. It also allows us to easily deal with missing data, data collected at different frequencies, and can handle a much larger number of time series than the standard state space approach to dynamic factor models.

[1] The widely used macroeconomic data provided by Stock and Watson (2006) is already loosely organized around blocks of data on output, consumption, prices, etc. Although it is not always clear which block some series belong to, this ambiguity does not matter as the block structure is not exploited in the analysis. In contrast, in one of our main applications data are placed into blocks by data source. For example, we might have a retail sales block based on the underlying detail of the the Census Bureau's monthly retail sales release.

We work with large dimensional panels, meaning that the number of cross-section units (N) and time series observations (T) are both large, but that the number of blocks, B, is much smaller than N. Let $i = 1, \ldots, N$ be an index for units, $t = 1, \ldots, T$ be an index for time, and $b = 1, \ldots, B$ be the index for blocks. Abstracting from dynamics to focus on the hierarchical structure for now, an observation on economic variable/unit i belonging to block b observed at time t, denoted x_{bit}, is modeled as

$$x_{bit} = \lambda'_{G.bi} G_{bt} + e_{Xbit}, i = 1, ..., n_b \tag{1}$$

where G_{bt} is a set of k_{Gb} latent, block-specific factors, N_b is the total number of variables/units in block b and e_{Xbit} is a purely idiosyncratic error (i.e., independent of all other errors and factors but not necessarily an IID process). We posit that the latent block-specific factors are modeled as depending on some common factors:

$$G_{kbt} = \lambda'_{F.bk} F_t + e_{Gkbt} \tag{2}$$

where F_t is a vector of k_F factors that are common across blocks and $e_{G_{kbt}}$ is a block-specific idiosyncratic error (i.e., independent of all other block specific errors and all factors but not necessarily an IID process). Thus, variables/units within a block can be correlated through F_t or e_{Gkbt}, but variables/units between blocks can be correlated only through F_t. In the terminology of multilevel models, (1) is the level-1 equation, (2) is the level-2 equation. A stochastic process for F_t would constitute a level-3 equation. If the process for F_t did not depend on another higher level of factors this would constitute a level-3 model.

The difference between a multilevel and a standard factor model is best understood when there is a single common and a single block-specific factor ($k_G = k_F = 1$),

$$\begin{aligned} x_{bit} &= \lambda_{G.bi}(\lambda_{F.b1} F_t + e_{Gb1t}) + e_{Xbit} \\ &= \lambda_{bi} F_t + v_{bit} \end{aligned} \tag{3}$$

where $\lambda_{bi} = \lambda_{G.bi}\lambda_{F.b1}$ and

$$v_{bit} = \lambda_{G.bi} e_{Gb1t} + e_{Xbit}.$$

A standard factor model ignores the block structure and stacks all observations up irrespective of which block an observation belongs to. The data are thus analyzed using the level-2 representation

$$x_{it} = \lambda_i F_t + v_{it}.$$

We would obtain an exact factor model if $\{e_{Gb1t} : b = 1, ..., B\}$ was a zero stochastic process. We would obtain an 'approximate factor model' if v_{it} is 'weakly correlated' across i and t. In

practice, this means that the number of idiosyncratic errors that are serially and/or cross-sectionally correlated cannot be too large. When this assumption is satisfied, the largest eigenvalue of the population covariance matrix of the $N \times 1$ vector x_t will increase with N, but the 2nd eigenvalue will be bounded. To the extent that the sample principal components are consistent for the population principal components, and the population principal components span the space of the underlying factor, the sample principal components will consistently estimate the space spanned by the latent factor as $N, T \to \infty$.

The approximate factor model does not rule out correlation within blocks per se. It merely restricts extensive correlation between v_{bit} and v_{bjt}, a condition that will be satisfied if block specific effects are absent. Alternatively it could be satisfied if the number of series in each block was relatively small and as $N \to \infty$, $B \to \infty$ with $N_b < \infty$ for all blocks. However, the standard approach to increasing the number of series is to increase the level of disaggregation within each block. For example, one might start at the two digit industry level and then expand the number of series by going to the three digit level and so on. In this case the largest eigenvalues of the covariance matrix of X may not be sufficiently separated, and as noted in Boivin and Ng (2006) and Onatski (2006) for a range of cases, the consequence is that the principal components estimator may not estimate the factor space precisely. A multi-level model directly tackles this problem by explicitly modeling the block structure. To illustrate, suppose we are interested in having a factor for prices. Rather than estimating a single price factor from the collection of producer prices, consumer prices, commodity prices, wages, and import prices, we can first find factors specific to the sub-blocks separately. This would yield producer price factors, consumer price factors, and so forth. Common price factors can then be found from the block factors and perhaps passed up to a higher level with factors related to real activity. As the block-level factors are estimated using more homogeneous data, we reduce the possibility that block-level variations are mis-identified as factors common to a pooled panel of data. Further, we can expand the number of series in any of the blocks holding the number of block specific factors fixed without running the risk that it will affect the estimation of factors from the other blocks.

Explicitly modeling the dynamics at the individual, block, and the aggregate levels is useful in a variety of economic models. Asset returns are commonly assumed to depend on a set of risk factors where the proxy for the factor affecting all assets is the market factor. But in addition to the sources of risk that all assets are exposed to, there might also be return variation that is specific to assets in a given industry. To construct portfolios that are well diversified or shall be used as a hedge against a certain type of risk, it is important to properly identify the different sources of comovement among asset returns. A natural approach to do so is to start by studying the comovement of returns at the

industry level and then to analyze to what extent the industry-factors are driven by a market-wide source of comovement. As another example, one might argue that business cycle variations are driven by different types of shocks related to global, regional or country-specific events as in Kose et al. (2003). Here again, it seems natural to start by studying the comovement among different business cycle indicators at the country-level, then to analyze to what extent countries in the same region of the world are affected by the same source of common variation, and finally to identify the global component behind the regional variation. Our hierarchical dynamic factor model naturally lends itself to analyzing such empirical questions.

The proposed dynamic hierarchical factor model has two other advantages. First, it gives more structure to the latent objects that are being estimated at each level. For example, factors extracted from the retail sales block can directly be interpreted as a retail sales factor. Variations that are common amongst the block-specific factors are naturally the common factors. Common, block-specific, and purely idiosyncratic variations are clearly distinguished, making a decomposition of variance analysis possible. Further, in many cases we can recreate aggregates produced by statistical agencies within a block of data allowing us to decompose the shocks to well-known aggregates such as private payroll employment. This goes part way in alleviating a common critique of factor models, namely, that the factors do not have a meaningful economic interpretation. In our setup, a sharp drop in the factor corresponding to the durable goods block without a concurrent drop in the factors of the remaining blocks is a shock to the durable goods sector. In contrast, a simultaneous drop in factors across blocks would be indicative of a general decline in the state of the real economy.

Second, different counter-factual experiments can be specified in our framework. Suppose we want to assess the state of the economy following a hypothetical one percent drop in housing activity. This can be accomplished by reducing each series in the housing block by one percent, re-estimate the factor in the block, and then construct the counter-factual F_t, holding other block-specific factors fixed. Alternatively, if we want to consider the state of economic activity following a one percent reduction in home prices in the northeast, we would reduce the data of the appropriate series by one percent, reconstruct the factor for the housing block, and subsequently the counter-factual F_t. If on the other hand, we want to consider a one percent drop in all economic activity, we would reduce all series in the panel by one percent, re-construct G_{bt} for all b, and subsequently F_t. In each case, we can trace out the chain of effects that account for changes in F_t.

This paper sets up the model and the framework for estimation. How to use the model for monitoring and counterfactual experiments will be discussed in a companion paper. Our goal here is to show that block-level variations are significant, and that going from a two to a three level model not only has statistical advantages, it also provides a better understanding of economic fluctuations.

We consider three examples. The first uses housing data where the blocks are determined by geography. The second uses six blocks of data on output unrelated to housing where the blocks are defined according to the timing of data release. Building on this six block model for output, data from the demand side are then added, leading to a four-level model for real economic activity. The decomposition of variance in each case shows that block-level variations tend to be stronger than the common variations, though both are small relative to the purely idiosyncratic component in the series.

2 A Three Level Dynamic Factor Model

Denote the observed data by x_{it}^D (possibly after logarithmic transformation), where $i = 1, \ldots N$ index units and $t = 1, \ldots T$ index time. We have B blocks of data, each of size N_b. Thus, $N = \sum_{b=1}^{B} N_b$ is the total number of cross-section units in the panel. A three level 'static' factor model can be specified as

$$\begin{aligned} x_{bit}^D &= c_b + c_i + \lambda'_{G.bi} g_{bt} + e_{Xbit} \\ g_{bkt} &= c_{g.k} + \lambda'_{F.bk} f_t + e_{Gbkt} \\ f_{rt} &= c_{f.r} + \epsilon_{Frt}. \end{aligned}$$

We use the notation that g_{bkt} is the k-th factor specific to block b, while f_{rt} is the r-th common factor. As c_b and c_i are not separately identifiable, we work with deviations from means as in Stock and Watson (1989) and Kim and Nelson (2000). For $i = 1, \ldots N$, let

$$\Phi_i(L) x_{it}^D = X_{it}^D$$

and

$$\Phi_i(L) = \Big(1 - \phi_{i1} L - \ldots \phi_{ip} L^p\Big)\Big(1 - \phi_{i0} L\Big).$$

By suitable choice of $\Phi_i(L)$, X_{it}^D is covariance stationary. The first difference filter obtains if $\phi_{i1} = \ldots = \phi_{ip} = 0$ and $\phi_{i0} = 1$. If a series does not require transformation to be stationary, $\phi_{i0} = 0$. A factor structure is sometimes imposed on variations not explained by the series' own past. We therefore also allow $\phi_{ij}(s), j = 1, \ldots, p$ to be non-zero, so that X_{it}^D would be the residuals from autoregressions in x_{it}^D (or Δx_{it}^D). Let $\text{mean}(X_i^D) = T^{-1} \sum_{t=1}^{T} X_{it}^D$ and $\text{std}(X_i^D)$ be the corresponding sample variance. Our multi-level factor model is defined in terms of the demeaned and standardized data,

$$X_{it} = \frac{X_{it}^D - \text{mean}(X_i^D)}{\text{std}(X_i^D)}.$$

The mean zero block-specific factors are $G_{bkt} = g_{bkt} - \mu_{g_k}$, and the mean zero common factors are $F_{rt} = f_{rt} - \mu_{f.r}$ for $k = 1, \ldots k_b$, $r = 1, \ldots K_F$.[2] We now present a three-level dynamic factor model for X_{it}.

There are a total number B of blocks of data. We assume that there are k_b common factors G_b in each block $b = 1, \ldots, B$. Hence, there is a total number $K_G = (k_1 + \ldots + k_B)$ of block-specific factors. We assume that these K_G block-specific factors share a total of K_F common factors F. Moreover, we denote N_b the number of variables in block b. This implies a total number $N = (N_1 + \ldots + N_B)$ of variables in the economy.

Each time series in a given block b is decomposed into a serially correlated idiosyncratic component, e_{Xbit}, and a component $\Lambda_{G.bi}(L)G_{bt}$ which it shares with other variables in the same block. The dynamics of each block-specific factor G_{bt} are decomposed into a serially correlated block-specific component e_{Gbt} and a component $\Lambda_{F.b}(L)F_t$ which it shares with all other blocks. Finally, the economy-wide factors F_t are assumed to be serially correlated. Let

$$\begin{aligned} X_{bt} &= \begin{pmatrix} X_{bt.1} & X_{bt.2} & \ldots X_{bt.N_b} \end{pmatrix}' \\ G_{bt} &= \begin{pmatrix} G_{bt.1} & G_{bt.2} & \ldots G_{bt.k_b} \end{pmatrix}'. \end{aligned}$$

The model can be summarized by the following equations:

$$X_{bt} = \Lambda_{G.b0} G_{bt} + \ldots + \Lambda_{G.bs_{Gb}} G_{b,t-s_{Gb}} + e_{Xbt}, \tag{4}$$

$$G_{bt} = \Lambda_{F.b0} F_t + \ldots + \Lambda_{F.bs_F} F_{t-s_F} + e_{Gbt} \tag{5}$$

$$F_t = \Psi_{F.1} F_{t-1} + \ldots + \Psi_{F.q_F} F_{t-q_F} + \epsilon_{Ft} \tag{6}$$

$$e_{Gbt} = \Psi_{G.b1} e_{Gb,t-1} + \ldots + \Psi_{G.bq_{Gb}} e_{Gb,t-q_{Gb}} + \epsilon_{Gbt} \tag{7}$$

$$e_{Xbit} = \Psi_{X.bi1} e_{Xbi,t-1} + \ldots + \Psi_{X.biq_{Xb}} e_{Xbi,t-q_{Xb}} + \epsilon_{Xbit}. \tag{8}$$

The idiosyncratic components e_{Xbi} are AR processes of order q_{Xbi} with innovations,

$$\epsilon_{Xbi} \sim N(0, \sigma^2_{Xbi}).$$

Similarly, the block-specific component is an AR process of order q_{G_b} and the block-specific innovations are normally distributed. For $b = 1, \ldots, B$:

$$\epsilon_{Gb} \sim N(0, \sigma^2_{Gb}).$$

Finally, the economy-wide factors F_{kt} are AR processes of order q_{F_k} and we also assume a normal distribution for the innovations to the economy-wide factors:

$$\epsilon_{F_k} \sim N(0, \sigma^2_{F_k}).$$

[2] Estimates of parameters μ_g and μ_f can be uncovered once the model parameters are estimated. For graphing purposes, it is actually more useful to standardize estimates G_{bt} so they can be compared across blocks.

The dynamics of the model can be enriched by allowing for stochastic volatility and markov switching effects. The current specification allows the lag order of the factor loading matrix and the factor specific errors to differ across blocks as well as within blocks. Similarly, the lag order of the idiosyncratic errors can also vary across blocks and units. Thus, $s_{Gb} = (s_{Gb.1}, \ldots s_{Gb.N_b})$ is a vector. Similarly, $q_{Xb} = (q_{Xb.1}, \ldots q_{Xb.N_b})$ and $q_{Gb} = (q_{Gb.1}, \ldots q_{Gb.k_b})$ are also vectors with possibly non-identical entries. Stacking up the data by blocks and letting

$$\begin{aligned} X_t &= \begin{pmatrix} X_{1t} & X_{2t} & \ldots X_{Bt} \end{pmatrix}' \\ G_t &= \begin{pmatrix} G_{1t} & G_{2t} & \ldots G_{Bt} \end{pmatrix}', \end{aligned}$$

we have

$$\begin{aligned} X_t &= \Lambda_G(L) G_t + e_{Xt} \\ G_t &= \Lambda_F(L) F_t + e_{Gt} \\ \Psi_X(L) e_{Xt} &= \epsilon_{Xt} \\ \Psi_G(L) e_{Gt} &= \epsilon_{Gt} \\ \Psi_F(L) F_t &= \epsilon_{Ft}. \end{aligned}$$

Let $s_G = \max_b(\max_{i \in N_b} s_{Gb.i})$, $q_X = \max_b(\max_{i \in N_b} q_{Xbi})$, and $q_G = \max_b(\max_{r \in k_b} q_{Gb.r})$. The model can be compactly written in matrix form as

$$\begin{aligned} \Lambda_G(L) &= \Lambda_{G0} + \Lambda_{G1} L + \ldots + \Lambda_{G_{s_G}} L^{s_G} \\ \Lambda_F(L) &= \Lambda_{F0} + \Lambda_{F1} L + \ldots + \Lambda_{F_{s_F}} L^{s_F} \\ \Psi_X(L) &= I_N - \Psi_{X1} L - \ldots - \Psi_{X_{q_X}} L^{q_X} \\ \Psi_G(L) &= I_{K_G} - \Psi_{G1} L - \ldots - \Psi_{G_{q_G}} L^{q_G} \\ \Psi_F(L) &= I_{K_F} - \Psi_{F1} L - \ldots - \Psi_{F_{q_F}} L^{q_F}. \end{aligned}$$

The polynomial matrices in L above have the following dimensions:

$\Lambda_G(L)$ is a $N \times K_G$ matrix polynomial of order s_G

$\Lambda_F(L)$ is a $K_G \times K_F$ matrix polynomial of order s_F

$\Psi_X(L)$ is a $N \times N$ matrix polynomial of order q_X

$\Psi_G(L)$ is a $K_G \times K_G$ matrix polynomial of order q_G

$\Psi_F(L)$ is a $K_F \times K_F$ matrix polynomial of order q_F

To ensure identification of the block-specific factors G, we assume that for $s = 0, \ldots, s_G$

$$\Lambda_{G.s} = \begin{bmatrix} \Lambda_{G.1s} & 0 & \cdots & 0 \\ 0 & \Lambda_{G.2s} & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & \Lambda_{G.Bs} \end{bmatrix}$$

The block-diagonal structure of $\Lambda_{G.s}$ implies that each block X_b of variables exclusively loads on the block-specific factors G_b. Moreover, we assume that at lag 0, and for each $b = 1, \ldots, B$, each $\Lambda_{G.b0}$, is a $N_b \times k_b$ matrix whose upper-left $k_b \times k_b$ block is lower-triangular with ones on the diagonal. More precisely, for $k_b = 3$, we would have

$$\Lambda_{G.b0} = \begin{bmatrix} 1 & 0 & 0 \\ \Lambda_{G.b0_{2,1}} & 1 & 0 \\ \Lambda_{G.b0_{3,1}} & \Lambda_{G.b0_{3,2}} & 1 \\ \Lambda_{G.b0_{4,1}} & \Lambda_{G.b0_{4,2}} & \Lambda_{G.b0_{4,3}} \\ . & . & . \\ \Lambda_{G.b0_{K_G,1}} & \Lambda_{G.b0_{K_G,2}} & \Lambda_{G.b0_{K_G,3}} \end{bmatrix}$$

This implies that the first variable within each block exclusively loads on the contemporaneous observation of the first block-specific factor, the second exclusively on the contemporaneous observations of the two first block-specific factors and so on. We order the variables in each block such that the first k_b series are variables of economic interest and yet have independent information. The loadings of the remaining variables in each block are unrestricted.

To ensure identification of the economy-wide factors F, we assume that the upper $K_F \times K_F$ submatrix of $\Lambda_{F.0}$ is lower triangular, i.e. for $K_F = 3$, we have

$$\Lambda_{F.0} = \begin{bmatrix} 1 & 0 & 0 \\ \Lambda_{F.0_{2,1}} & 1 & 0 \\ \Lambda_{F.0_{3,1}} & \Lambda_{F.0_{3,2}} & 1 \\ \Lambda_{F.0_{4,1}} & \Lambda_{F.0_{4,2}} & \Lambda_{F.0_{4,3}} \\ . & . & . \\ \Lambda_{F.0_{K_G,1}} & \Lambda_{F.0_{K_G,2}} & \Lambda_{F.0_{K_G,3}} \end{bmatrix}$$

This normalization assumes that the blocks are ordered so that the common factors load heavily on the first block-specific factors. An alternative is to normalize the variance of ϵ_F to unity. However, the diagonal elements of Λ_{F0} will still have to be constrained to be positive, even though they will not be constrained to be one.

2.1 The State Space Representation

Let $\Theta = (\Theta_F; \Theta_G; \Theta_X)$ where Θ_F, Θ_G are parameters that characterize F_t, G_t respectively, and Θ_X are the remaining parameters. By assumption,

$$(i)\ x_t \perp\!\!\!\perp \Theta | G_t, \Theta_X \quad (ii)\ G_t \perp\!\!\!\perp \Theta | F_t, \Theta_G, \quad (iii)\ F_t \perp\!\!\!\perp \Theta | \Theta_F$$

where $\perp\!\!\!\perp$ stands for stochastic independence. Stepwise specification of the sub-models lead to the statistical model

$$f(x_t, F_t, G_t; \Theta) = f(x_t | G_t; \Theta_X) f(G_t | F_t; \Theta_G) f(F_t; \Theta_F).$$

The data density is

$$f(x_t; \Theta) = \int \int f(x_t | G_t; \Theta_X) f(G_t | F_t; \Theta_G) f(F_t | \Theta_F) dG_t dF_t.$$

Because of the assumed hierarchical structure, the data density can be constructed recursively from the pair of equations:

$$\begin{aligned} f(G_t | \Theta_F, \Theta_G) &= \int f(G_t | F_t; \Theta_G) f(F_t | \Theta_F) dF_t \\ f(x_t | \Theta) &= \int f(x_t | G_t; \Theta_X) f(G_t | \Theta_F, \Theta_G) dG_t. \end{aligned}$$

Here, $f(x_t; G_t; \Theta_G)$ is the measurement equation and $f(G_t | F_t; \Theta_F)$ is the structural model for the latent factor F_t. As discussed in Mouchart and Martin (2003), strong identification of the measurement model is required to obtain weak identification of the statistical model. Our assumptions ensure that $\Theta_X = (\Psi_X, \Sigma_X, \Lambda_G)$ are identified from the measurement model, $\Theta_G = (\Psi_G, \Sigma_G, \Lambda_F)$ are identified from the structural model for G_t, and $\Theta_F = (\Psi_F, \Sigma_F)$ are identified from the transition equation for F_t. These equations are now made precise.

Common Factor Dynamics The common factors evolve according to

$$\Psi_F(L) F_t = \epsilon_{Ft},$$

where $\Psi_F(L) = I_{K_F} - \Psi_{F.1} L - \ldots \Psi_{F.q_F} L^{q_F}$. This can be rewritten in companion form as

$$\begin{pmatrix} F_t \\ F_{t-1} \\ \vdots \\ F_{t-q_F+1} \end{pmatrix} = \begin{bmatrix} \Psi_{F.1} & \Psi_{F.2} & \cdots & \Psi_{F.q_F} \\ I & 0 & & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & I & 0 \end{bmatrix} \begin{pmatrix} F_{t-1} \\ F_{t-2} \\ \vdots \\ F_{t-q_F} \end{pmatrix} + \begin{pmatrix} \epsilon_{Ft} \\ 0 \\ \vdots \\ 0 \end{pmatrix} \tag{9}$$

or

$$\vec{F}_t = \vec{\Psi}_F \vec{F}_{t-1} + \vec{\epsilon}_{Ft}$$

where $\epsilon_{Ft} \sim N(0, \Sigma_F)$ and $||\vec{\Psi}_F|| < 1$.

Block-Level Dynamics We have the structural (instead of measurement) equation

$$\begin{aligned} G_t &= \Lambda_{F.0}F_t + \Lambda_{F.1}F_{t-1} + \ldots \Lambda_{F.s_F}F_{t-s_F} + e_{Gt} \\ &= \Lambda_F(L)F_t + e_{Gt} \end{aligned} \tag{10}$$

with $G_t' = \begin{pmatrix} G_{1t} & \vdots & G_{Bt} \end{pmatrix}$, G_{bt} being a $k_b \times 1$ vector,

$$\Psi_G(L)e_{Gt} = \varepsilon_{Gt} \tag{11}$$

and where $\Lambda_{F.s}$ is a $K_G \times K_F$ matrix of factor loadings. We call this a structural instead of a measurement equation because G_t is not observed.

Denote $\Psi_G(L)$ to be a diagonal matrix with diagonal elements $\Psi_{Gb}(L)$, where for $b = 1, \ldots B$, $\Psi_{G.b}(L)$ is itself a diagonal matrix with elements $\psi_{G.bi}(L)$,

$$\psi_{G.bi}(L) = 1 - \psi_{G.bi1}L - \ldots - \psi_{G.biq_G}L^{q_G}$$

We restrict $||\psi_{G.bi}(L)|| < 1$ for stationarity and assume $\epsilon_{Gt} \sim N(0, \Sigma_G)$. Together, (10) and (11) imply that

$$\Psi_G(L)G_t = \Psi_G(L)\Lambda_F(L)F_t + \epsilon_{Gt}.$$

This leads to the block-level transition equation

$$\begin{pmatrix} G_t \\ G_{t-1} \\ \vdots \\ G_{t-q_G+1} \end{pmatrix} = \begin{pmatrix} \alpha_{Ft} \\ 0 \\ \vdots \\ 0 \end{pmatrix} + \begin{bmatrix} \Psi_{G.1} & \Psi_{G.2} & \cdots & \Psi_{G.q_G} \\ I & 0 & & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & I & 0 \end{bmatrix} \begin{pmatrix} G_{t-1} \\ G_{t-2} \\ \vdots \\ G_{t-q_G} \end{pmatrix} + \begin{pmatrix} \epsilon_{Gt} \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

or

$$\vec{G}_t = \vec{\alpha}_{Ft} + \vec{\Psi}_G\vec{G}_{t-1} + \vec{\varepsilon}_{Gt}$$

where

$$\alpha_{Ft} = \Psi_G(L)\Lambda_F(L)F_t. \tag{12}$$

Within-Block Dynamics For each $b = 1, \ldots, B$, we have

$$\begin{aligned} X_{bt} &= \Lambda_{G.b0}G_{bt} + \ldots + \Lambda_{G.bs_G}G_{bt-s_G} + e_{Xbt}, \\ \Psi_{X.b}(L)e_{Xbt} &= \varepsilon_{Xbt} \end{aligned}$$

where we recall that the top $k_G \times k_G$ block of $\Lambda_{G.b0}$ is lower triangular, and with X_{bt} stacked so that G_{bt} have non-trivial loadings on the first k_b. For $b = 1, \ldots B$, the measurement equation for each block can be rewritten as

$$\Psi_{X.b}(L)X_{bt} = \Psi_{X.b}(L)\Lambda_{G.b}(L)G_{bt} + \varepsilon_{Xbt},$$

or

$$\tilde{X}_{bt} = \tilde{\Lambda}_{G.b}(L)G_{bt} + \varepsilon_{Xbt} \tag{13}$$

where

$$\begin{aligned} \tilde{X}_{bt} &= \Psi_{X.b}(L)X_{bt} \\ \text{and} \qquad \tilde{\Lambda}_{G.b}(L) &= \Psi_{X.b}(L)\Lambda_{G.b}(L) \end{aligned}$$

These can be stacked to produce

$$\tilde{X}_t = \tilde{\Lambda}_G(L)G_t + \varepsilon_{Xt}$$

Decomposition of Variance Given the state space representation of the model, it is not hard to see that for each individual variable X_{bi},

$$vec(Var(X_{bi})) = \gamma'_{F.bi}var(F_t) + \gamma'_{G.bi}var(G_{bt}) + var(e_{Xbi}) \tag{14}$$

where

$$\begin{aligned} \gamma'_{F.bi} &= \left(\sum_{s=0}^{s_G} \lambda'_{G.bis} \otimes \lambda'_{G.bis}\right) \cdot \left(\sum_{s=0}^{s_F} \Lambda_{F.bs} \otimes \Lambda_{F.bs}\right) \\ \gamma'_{G.bi} &= \left(\sum_{s=0}^{s_G} \lambda'_{G.bis} \otimes \lambda'_{G.bis}\right) \\ var(F_t) &= \left[I - \sum_{q=1}^{q_F} (\Psi_{F.q} \otimes \Psi_{F.q})\right]^{-1} vec(\Sigma_F) \\ var(G_{bt}) &= \left[I - \sum_{q=1}^{q_G} (\Psi_{G.bq} \otimes \Psi_{G.bq})\right]^{-1} \cdot vec(\Sigma_{G_b}) \\ var(e_{Xbi}) &= \left[1 - \sum_{q=1}^{q_X} \psi^2_{X.biq}\right]^{-1} \cdot \sigma^2_{Xbi}. \end{aligned}$$

The total variance is the sum of the unconditional variance of the components multiplied by the effective loadings on the component. Dividing the three components on the right hand side of (14) gives the fraction of the variance in X explained by the common innovations ϵ_F, block-specific innovations ϵ_{Gb}, and idiosyncratic errors ϵ_X, respectively. We denote these by share$_F$, share$_G$, and share$_X$. These shares depend on the factor loadings, the dynamic parameters, as well as the size of the shocks. A two level factor model does not distinguish between F_t and G_t; in these models, 1-share$_X$ is the size of the common component.

2.2 Markov Chain Monte Carlo

We use the method of Markov Chain Monte Carlo (MCMC) to estimate the posterior distribution of the parameters of interest. The method samples a Markov chain that has the posterior density of the parameters as its stationary distribution. MCMC has been used by Kim and Nelson (2000), Aguilar and West (2000), Geweke and Zhou (1996) and Lopes and West (2004), among others, to estimate two level factor models. These algorithms are variations and extensions of the method developed in Carter and Kohn (1994). Although in theory, the algorithm allows for multiple factors, most previous studies have limited attention to estimation of a single factor. We allow both F_t and G_{bt} to be vector valued.

Our setup is a hierarchical dynamic factor model where each level admits a state-space representation that has a measurement and a transition equation. The MCMC algorithm thus needs to be extended to handle this hierarchical structure. Let $\mathbf{\Lambda} = (\Lambda_G, \Lambda_F)$, $\mathbf{\Psi} = (\Psi_F, \Psi_G, \Psi_X)$, $\mathbf{\Sigma} = (\Sigma_F, \Sigma_G, \Sigma_X)$. The main steps are as follows:

1. Initialize by using principal components to obtain initial estimates of $\{G_t\}$ and $\{F_t\}$ and the ordering of each X_{bt}.Use these to produce initial values for $\mathbf{\Lambda}$, $\mathbf{\Psi}$, $\mathbf{\Sigma}$.

2. Conditional on $\mathbf{\Lambda}, \mathbf{\Psi}$, $\mathbf{\Sigma}$ and $\{F_t\}$ draw $\{G_t\}$ using standard methods for state space models with time varying intercepts.

3. Conditional on $\mathbf{\Lambda}, \mathbf{\Psi}$, $\mathbf{\Sigma}$ and $\{G_t\}$ draw $\{F_t\}$ using standard methods for state space models.

4. Conditional on $\{F_t\}$ and $\{G_t\}$ draw $\mathbf{\Lambda}$, $\mathbf{\Psi}$, and $\mathbf{\Sigma}$ using standard methods assuming conjugate priors.

5. Return to 2.

The main complication going from a two to a three level model lies in the way $\{G_t\}$ is sampled in Step (2). Recall that the transition equation for G_{bt} is of the form

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

This involves the term $\alpha_{F.bt} = \Psi_{G.b}(L)\Lambda_F(L)F_t$, which given a draw of F_t can be interpreted as a time-varying intercept that is known for all t. By conditioning on F_t, our updating and smoothing equations for G_t explicitly take into account the information that F_t has to bear. Details of Steps 2 and 3 are given in the Appendix. The complete algorithm is given in an appendix which is available upon request.

3 Related Work

Multilevel factor models have been considered extensively in the psychology literature, see Goldstein and Browne (2002). With the size of the panel being large in only one dimension and assuming a strict factor structure, estimation of these models tend to be likelihood based. As well, these models have no dynamics. Our three-level factor model shares common features with a few approaches that have previously been suggested in the macroeconomic literature. Kose et al. (2003) and Kose et al. (2008) used multi-level factor models to study international business cycle comovements. In their model, economic fluctuations in each country are attributed to three types of shocks: a world, a regional and a country-specific business cycle component. For each observable variable i in country b, they have

$$x_{bit} = b_i F_t + c_{bi} G_{bt} + e_{bit}$$

where F_t is a world factor, G_{bt} is a factor specific to region b, and where e_{bit} is a component specific to variable i in country b.[3]

Our model differs from their model in a number of ways. While their F_t and G_{bt} are scalars, we allow for multiple common and multiple block-specific factors. Comparing this setup to (3), their loading on the world factor b_i plays the role of our $\lambda_{G.bi}\lambda_{F.b1}$ and their loading c_{bi} on the regional factor is our $\lambda_{G.bi}$. Since we impose the structure that G_{bt} is linear in F_t, the responses of shocks to F_t for all variables in block b can only differ to the extent that their exposure to the block-specific factors differs, whereas b_i is unconstrained in Kose et al. (2008). However, by imposing this 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. (2008) have $N \times K_F$ and $N \times K_G$ parameters, respectively. As K_G is much smaller than N, our framework is considerably more parsimonious, an issue that has computational consequences as we discuss below.

Another related approach is the Structural Factor-Augmented VAR setup of Milani and Belviso (2006). Similar to our model, they organize a large panel of macroeconomic time series for the US into blocks of data and use the block structure to identify factors such as a real activity factor, an inflation factor, a financial market factor etc. They do not assume the existence of comovement beyond the block structure, but instead model the dynamic evolution of the different block factors jointly within a VAR. Clearly, this approach imposes a constraint on the number of block factors that one can allow for in their model. In contrast, we attribute variation that is common across blocks to a small number of common factors F whose dynamics we describe as a VAR. This allows us to have a potentially much larger set of block-specific factors and also enables us to study to

[3] A similar framework was recently used by Stock and Watson (2008a) to analyze national and regional factors in housing construction.

what extent common variation is block-specific or due to economy-wide shocks.

In a recent paper, Hallin and Liska (2008) also study dynamic factor models with a block structure. While we assume the block structure of the data as known, they develop methods to estimate the common, block-specific, and idiosyncratic factors from the data. To separate the different sources of comovement, Hallin and Liska (2008) assume a factor space that contains a total of 2^K factors where K denotes the number of blocks. In setups with many blocks, this might give rise to computational limitations of their approach. Moreover, we think that in many empirical applications of factor models both theoretical considerations and background information on how the data are constructed provide the researcher with valuable a priori knowledge of the block structure that can be readily exploited using our hierarchical model.

In terms of estimation, Otrok and Whiteman (1998) estimate latent dynamic factors by considering their conditional joint distribution. The main practical limitation is that they have to invert a variance-covariance matrix of rank T where T denotes the number of time series observations of the data. Hence, estimation becomes computationally demanding for problems when N and T are both large. In Kose et al. (2003), $N = 180$ and $T = 30$, and in Kose et al. (2008), $T = 176$ and $N = 21$. Our experimental models here have up to $N > 400$ series and T close to 200, and we anticipate using as many as 1,000 series in a full fledged analysis. To circumvent the dimensionality problem, we put more structure on the block factors G_t. This enables us to cast the conditional model for F_t and G_t in state space representation, which can be estimated using standard methods.

An alternative to Gibbs sampling is to estimate G_t by principal components, and then estimate F_t from the principal components estimates of G_t. This method was implemented in Beltratti and Morana (2008). However, sequential estimation by principal components would not take into account the dependence of G_t on F_t through α_{Ft}. These 'unrestricted' estimates of G_t should thus be less efficient than our one step estimates. Another advantage of our approach is that the posterior distributions allow us to assess sampling uncertainty about the estimated factors. While the large sample theory for $\tilde{G}_t$ and $\tilde{F}_t$ is given in Bai and Ng (2006), the sampling distribution of $\tilde{F}_t(\tilde{G}_t)$ is not known. It remains unclear how to obtain theoretical prediction intervals or assess the sampling uncertainty of counter-factual analysis within the two-step principal components framework.

Since the true F_t are latent variables it will be difficult to compare the precision of the estimates. However, as a cross-check on our results we deem it useful to compare the estimates produced by our three level model with those obtained from principal components analysis. Hereafter, we use a 'tilde' to denote estimates obtained by the method of principal components, and a 'hat' to denote estimates obtained from our MCMC algorithm. That is, $\hat{G}_t$ denote the posterior means of the block-specific dynamic factors while $\hat{F}_t$ are the posterior means of the factors common to G_t. In

contrast, we refer to $\tilde{F}_t$ as the principal component estimates obtained using all data at once and let $\tilde{F}_t(\tilde{G}_t)$ denote the two step principal components estimates, obtained from extracting principal components from the block-specific principal components estimates. When comparing the results, It should be kept in mind that the method of principal components estimates the static factors, whereas we estimate the dynamic factors, which should generally be smoother than the static factors.

4 Empirical Analysis

We assume diffuse priors throughout. Specifically, all factor loadings Λ and all autocorrelation coefficients ψ are Gaussian with mean zero and variance of one. 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.[4] After discarding the first 2,000 draws as a burn-in, we take another 25,000 draws, storing every 50-th draw. The reported statistics for posterior distributions are based on these 500 draws. Results obtained from storing every one of the first 8,000 draws after burn-in are similar.

We use the principal components estimated for each block, denoted $\tilde{G}_{b,PC}$, as starting values for G_t. The principal components extracted from the data pooled across blocks are then used as starting values for F. Note that the principal components only identify the factor space using the normalization that $\tilde{\Lambda}'_{G,PC}\tilde{\Lambda}_{G,PC}/N = I_r$ and the matrix $\tilde{G}'_{PC}\tilde{G}_{PC}$ is diagonal. We use alternative identification assumptions. Therefore, our starting values may be far from the true values. As a cross-check on our choice of initialization, we also run the MCMC algorithm using randomly generated numbers for the factors as starting values and find that the sampler converges to the same posterior means.

4.1 Data Releases

Non-financial data on the state of the economy are being released at almost a daily frequency. Many economic analysts closely follow the release calendar and prior to any new announcement assemble information about the expectations for key indicators. While the media coverage of economic data typically focuses on a few individual series as e.g. the number of non-farm payroll employees, the majority of data releases actually contain the most recent number for a block of related time series. This cross-sectional information is likely taken into account by analysts when they update their beliefs about the state of the economy. In doing so, however, they need to filter out the

[4] If θ is distributed as inverse χ^2 with ν degrees of freedom and a sale of d, written $\theta \sim I\chi^2(v, d^2)$, then θ is distributed as an inverse gamma with parameters $\alpha/2$ and $\beta/2$, where $\alpha = \nu$ and $\beta = d^2\nu$. We use this equivalence in our procedure and sample variance parameters based on the χ^2 distribution.

aggregate from the idiosyncratic or noise components. In our model, we try to optimally exploit the cross-sectional information by assuming a hierarchical factor structure for the data that takes into account the timing of the releases.

Figure 1 shows the release of non-financial data for the month of June 2008. For the non-financial data, we use the highest level of aggregation that will give a reasonable number of cross-section units. In this paper, we use data from nine blocks of non-housing real economic data. These are industrial production (IP) and capacity utilization (CU), the establishment survey (ES), the household survey (HS), and manufacturers' surveys (MS), durable good (DS), retail sales (RS), wholesale trade (WT), and auto sales (AUTO). The establishment survey and the household survey are released together in the so-called employment report in the first week of the month. Retail sales data are published in the second week. This is followed by IP and CU released in the third week of the month and so on. We also consider three blocks of housing data: home price, housing starts, home sales. These data are reorganized into three new blocks by geographical regions, the West, the Northeast, and others. These will be denoted West, NE, and CTL.

We use a balanced panel of monthly data from 1992:01-2008:02. After the data transformation, our sample effectively starts in 1992:4, giving $T = 191$ observations for all blocks. We only have one recession in this sample, but the analysis can be extended to allow for a non-balanced panel in future work. The data are transformed using Stock and Watson (2008b) as a guide. For quantity variables, we compute month-on-month growth rates. For price variables, we take quarterly differences of the logged series. Indices (such as capacity utilization) are not transformed.

An important aspect of our model is that we use prior information to identify the factors. This is achieved by ordering the variables thought most likely to be representative of comovement in a given blocks in positions one through k_{Gb}. Table 1 list the first two variables in each block, along with some summary statistics based upon the factors estimated by principal components. Properties of the principal component estimates for each block are given in Table 2.

4.2 A Housing Model Using Actual and Simulated Data

Our first model uses three blocks of U.S. regional housing data with 7, 8, and 18 series in each block, respectively. Note that the principal component estimates will not be precise because of the small number of units in each block. We assume that there are two factors in each block, and all block-specific factors are in turn driven by one common factor. The common and block-specific factors are modeled to have AR(1) dynamics. We allow for two lags of the respective factors in the measurement equations. In addition to being of interest in its own right, we also use this simple example to assess the precision of the estimator. Treating the posterior means of the

parameters estimates, $\hat{F}_t$ and $\hat{G}_t$ as 'true' values, we reconstruct and then re-sample ϵ_X to obtain a set of simulated data. The simulated data are then used to estimate the parameters. Comparing the estimates obtained from the simulated data with the 'true' values gives an assessment of the precision of the estimates. Table 3 reports the results. Since the 'actual' parameters are themselves estimates, we report standard errors for them as well.

The ψ_F estimated from the data indicates that the common factor is highly persistent, and the estimate based on the simulated data is very similar. The ψ_G's estimated from the data are generally not significant from zero, and the estimates based on the simulated data also have this near white noise feature. The $\hat{F}_t$ has a correlation of 0.98 with the 'true' F_t. Since there are multiple factors within each block, we assess the precision of the estimates by regressing $\hat{G}_{bk}$ on the entire set of block factors, G_b. The correlation between $\hat{G}_{b1}$ and the true factors in the block are quite high (.877, .965, and .707), but considerably lower for the respective second block-specific factors $\hat{G}_{b2}$. This suggests that the second block-specific factors are not well determined in this dataset, and the model could have been estimated with one block-specific factor, respectively.

The bottom panel of Table 3 provides the decomposition of variance into economy-wide, region-specific, and idiosyncratic shocks. At each draw of the Gibbs sampler, we obtain the shares for each series and then average over units within a block. We then report the mean and standard deviation of these block-specific decompositions across all draws. Since the data are standardized to have unit variance, the estimated variance of X also gives a sense of the fit of the model. According to these numbers, the common housing factor accounts for 10 to 17% of the regional fluctuations in the housing market while region specific factors account for anther 12 to 24%. However, by far the largest source of variations in the housing market are idiosyncratic. The decomposition of variances based on the simulated data match these estimates very well. Our methodology thus seems capable of disentangling the different levels of cross-sectional variation.

The top left panel of Figure 2 graphs the F against $\hat{F}$. Clearly, $\hat{F}$ is significantly below zero in recent years. At the end of our estimation sample (2008:2), $\hat{F}$ is -0.843, with sample standard deviation of 0.320. In comparison, the maximum value of $\hat{F}$ over the sample is 0.427. The top right panel of Figure 2 graphs G_{b1} against $\hat{G}_{b1}$ for the West, which is far more volatile than the factors in the Northeast and the other regions. The $\hat{G}_b$ are quite close to the true factor processes. It is evident that shocks to all three regions have been mostly been negative since 2006. While the state of housing in the CTL region appears to be still in a downward trend, there are signs of rebounding in both the West and Northeast in 2008:2. In sum, we find that regional and to a smaller extent economy-wide shocks have contributed to the recent housing slump.

4.3 A Six Block Model for Output

We estimate a dynamic hierarchical factor model for six blocks of data related to output in the US economy: industrial production (IP), capacity utilization (CU), the establishment survey (ES), the household survey (HS), manufacturers' surveys (MS), and durable goods (DG). According to the IC_2 criterion of Bai and Ng (2002), four of the six blocks have either one or two factors. However, the criterion suggests that the HS and MS blocks may have as many as eight factors. Although our Bayesian estimation approach generally allows for different numbers of factors across blocks, we let all blocks be driven by two block-specific components so as to enhance comparability of the results. Moreover, we assume one common factor at the aggregate level. Initial estimation assuming two common factors suggests that the second factor has a very small variance, and dropping it did not lead to any noticeable change in the decomposition of variance. As before, we let the factor loadings at both levels to be polynomials of order two, i.e. two lags of the factors enter the observation equations, respectively. Finally, we assume that the common factors as well as the idiosyncratic and block-specific disturbances are AR(1) processes. Altogether, our model is thus described by the following set of parameters : $K_F = 1$, $k_{Gb} = 2$ for all b, $s_{Fr} = 2$, $s_{Gb} = 2$, $q_{Fr} = q_{Gb.k} = q_{Xb.i} = 1$ for all $b = 1, \ldots B$, $r = 1, \ldots K_F$, $k = 1, \ldots K_{Gb}$, and $i = 1, \ldots N_b$. We note that the estimated factors and idiosyncratic errors are generally mildly persistent, suggesting that the transformed data used in the analysis are stationary.

The top panel of Table 4 reports the posterior means and standard errors of the dynamic parameters. The common factor has an autoregressive coefficient of .795. The block-specific factors have varying degrees of persistence, and many of the block-specific factors are close to white noise. Moreover, the block-specific shocks tend to have larger variance than the shocks to the common factors.

In this model, there are $2 \times N$ loadings on G_t, and $K_G \times 1$ loadings on F_t, where $K_G = 12$ and N=315. Instead of reporting all the loadings, we summarize the properties of the model by evaluating the relative importance of the common, block-specific, and idiosyncratic variation. According to the model, the DG block has the best fit. The bottom panel of Table 4 shows that there is substantial heterogeneity across blocks. Of the six blocks considered, the CU, the IP, and the ES blocks have the largest common component, explaining 20% or more of the variation in the data of the block. The block-specific shocks roughly explain another 15% of the variation in these three blocks. Thus, the common and block-specific factors in our sample of economic variables explain close to 40% of the variation in the blocks. This is similar to what one finds in principal components analysis applied to the much analyzed Stock and Watson dataset with 132 series, where the first five factors are found to explain about 40% of the data.

While aggregate shocks to the CU, IP, and ES blocks are more important as shocks specific to these blocks, the block-specific component is larger than the common component in all remaining blocks. Shocks specific to MS account for around 24% of the variations, compared to the common component of about 4%. The result that stands out in Table 4 is that the idiosyncratic component always explains the largest share of variation. In particular, 80% of the variation in the Household Survey block is idiosyncratic, and only 2% of the variation in that block is explained by the common factor F. Although the monthly employment report (which contains the HS data) is well-watched by financial markets, our findings thus suggest that the HS data contain little information about the level of non-housing real economic activity. The results generally highlight the difficulty in distilling information relevant for aggregate policy from observed data, as block-specific information can be disguised as common variations, and a large idiosyncratic component can make precise estimation of the common factor space difficult.

The relative importance of the common factors based on principal components estimation is also reported in Table 5. The IC_2 criterion of in Bai and Ng (2002) indicated a total of 2 factors in the panel of 315 series. Both one and two step principal component estimation of F_t suggests that the first two factors explain about 40% of the variation in the data. As noted earlier, if block-specific variations are important, some of the principal components extracted from the entire panel of data might correspond to block-specific factors. The correlation between our first factor $\hat{F}_1$ and the first principal component is 0.80.

To investigate if the estimated principal components capture block-specific variations, we regress the principal components $\tilde{F}_{rt}$ extracted from the entire panel on $\hat{F}$ to obtain residuals $\tilde{e}_{rt}$. These are variations deemed common by the method of principal components but not by our hierarchical dynamic factor model estimated using MCMC. We then check if these residuals can be explained by our estimated block-specific factors by regressing $\tilde{e}_{rt}$ on $\tilde{G}_{bkt}$. To conserve space, Table 5 reports the R^2s that exceed 0.1. Evidently, many of the block-specific variations are correlated with the factors estimated by the method of principal components. The first and second principal components are correlated with variations in the Establishment Survey block ($b = 3$) with a correlation as high as 0.716, while the third principal component is highly correlated with the Household Survey block ($b = 4$). This could be a consequence of the fact that the employment block constitutes one third of the data, and common variations in the Household Survey block are deemed more important in principal component analysis than in our framework. The factors of the Durable Goods block ($b = 6$) are correlated with the second and fourth principal component. Overall, we interpret these results as suggesting that variations identified as common by principal component analysis may in fact be block-specific.

Figure 3 graphs the factors estimated using the three different methods. The top panel plots $\hat{F}_t$ estimated using our hierarchical model against the first principal component $\tilde{F}_t(\tilde{G}_t)$ extracted from the block-specific principal components. The lower panel graphs $\hat{F}_t$ against the first principal component $\tilde{F}_t$ extracted from the entire data panel. All estimates indicate that the trough of the last recession occurred towards the end of 2001. This is in agreement with the NBER business cycle dates which report November 2001 as the trough of the last recession. All estimates also indicate a slowdown in the level of real activity since the middle of 2005, with the common factor $\hat{F}_t$ estimated using our hierarchical model suggesting a weaker economy than the principal component estimates.

Notice that our $\hat{F}_t$ tends to be smoother than $\tilde{F}_t$ and $\tilde{F}_t(\tilde{G}_t)$. For example, around the end of 1995, both principal components estimates show large spikes which in our model are picked up by the Establishment Survey factor G_{3t}. One potential explanation for the large spikes in the principal components factor relates to the government shutdown over the budget in January 1996. Due to the large number of employment related series in the dataset, the first principal component extracted from the entire panel reflects this block-specific event. In contrast, our hierarchical model allows to distinguish this block-specific from genuinely common shocks.

5 A Four Level Model

Some blocks of data are naturally organized by sub-blocks. For example, the producer and consumer price blocks together form a price block. Data on the establishment and household surveys are released together in the employment report, while data for industrial production and capacity utilization are also released at the same time. It is therefore natural to allow some blocks to have a sub-block structure.

We continue to let X_{bit} denote variables associated with block-specific factors G_{bt}. To distinguish data associated with blocks that have sub-blocks from those that do not, let Z_{bsit} be the observed data for block b where s is an index for the sub-blocks. Let H_{bst} be the factors of sub-block b. Then a four-level model can be represented by

$$Z_{bsit} = \lambda_{H.bsi}(L)H_{bst} + \epsilon_{Xbsit} \tag{15}$$

$$H_{bst} = \lambda_{G.bs}(L)G_{bt} + \epsilon_{Hbst} \tag{16}$$

$$G_{bt} = \lambda_{F.b}(L)F_t + \epsilon_{Gbt} \tag{17}$$

$$e_{F_k t} = \Psi_{F.k} e_{F_k t-1} + \ldots \Psi_{F.q_{Fk}} e_{F_k t-q_F} + \epsilon_{Fkt} \tag{18}$$

$$e_{Gbt} = \Psi_{G.b1} e_{Gb,t-1} + \ldots + \Psi_{G.bq_{Gb}} e_{Gb,t-q_{Gb}} + \epsilon_{Gbt} \tag{19}$$

$$e_{Hbst} = \Psi_{H.bs1} e_{Hbs,t-1} + \ldots + \Psi_{H.bsq_{Hbs}} e_{Hbs,t-q_{Hbs}} + \epsilon_{Hbst} \tag{20}$$

$$e_{Xbsit} = \Psi_{X.bsi1} e_{Xbsi,t-1} + \ldots + \Psi_{X.bsiq_{Xb}} e_{Xbsi,t-q_{Xb}} + \epsilon_{Xbsit} \tag{21}$$

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 $\alpha_{G.bst} = \Psi_{H.bs}(L)\Lambda_{G.b}(L)G_{bt}$. As in the level three model, the dependence of G_{bt} on F_t in turn implies

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

Conditional on G_{bt}, F_t, and Θ, we can draw H_{bst} for each s and b, and conditional of F_t, we can draw G_{bt} for each b. Blocks with a sub-block structure can be combined with blocks that do not. A model with more levels can always be decomposed into a sequence of two-level models. Of course, we will need to have a reasonable number of series at the sub-block level, and a multi-level model would be more time intensive to estimate. But conceptually, a model with 'branches' in some but not all blocks is possible.

We extend our six block model considered in the previous section to include three more blocks of data: retail sales (RS), wholesale trade (WT), and autosales (AUTO). We then reorganize the 402 series into 3 blocks. The first is an output block with sub-blocks IP, CU, MS, and DG representing goods production. The second is a labor market block consisting of sub-blocks ES and HS. The third is a demand block consisting of sub-blocks RS, WT, and AUTO. This covers demand by both firms and consumers. The common factor extracted by this model can thus be interpreted as a factor for real economic activity. We estimate one common factor and let $K_G = (1, 2, 1)$.

To conserve space, Table 6 only reports the autoregressive parameters for G_t and F_t. The common factor is slightly more persistent than that estimated for the three level model. The ψ_G for the output specific factor is close to that found for CU and IP, while that for the employment block is higher than that found for ES or HS. The demand-specific factor is the least persistent of all block factors. Table 7 reports the decomposition of variance, which is now performed at the sub-block level. The CU and IP blocks continue to have the largest common component. The sub-blocks of the demand block have relatively large variations due to factors common to series in the sub-blocks, but the overall picture remains that series specific idiosyncratic shocks dominate.

Perhaps of most interest is an analysis of the state of real economic activity estimated with the model. This is presented in Figure 4. The solid line is the $\hat{F}_t$ based on our model and the dotted line is the principal component estimate, $\tilde{F}_t$, both standardized to have a mean of zero and unit variance. Our $\hat{F}_t$ is noticeably smoother than $\tilde{F}_t$. The large spikes in employment in 1996 due to the budget crisis once again weighs heavily in $\tilde{F}_t$ but is appropriately treated as variations specific to a sub-block of employment using our four-level hierarchical model. Our (non-standardized) estimates suggest that the state of real economic activity at the end of our sample in 2008:02 stood at -.328.

With the sample standard deviation of $\hat{F}_t$ being .202, the level of real activity was thus considerably below average. However, according to our estimate activity in 2008:02 was still stronger than at the trough of the 2001 recession for which we record a value of -0.598.

6 Conclusion

This paper lays out a framework for analyzing dynamic hierarchical factor models. The approach has three advantages. First, by extracting common components from blocks, the estimated factors have more precise interpretation. Explicitly modeling the block-specific variation also resolves an important drawback of standard (two-level) factor models in which block-specific shocks can be confounded with common shocks. Second, the blocks can be defined to take advantage of the timing of data releases, which makes the framework suitable for real time monitoring of economic activity. Third, the framework allows for a more disaggregated analysis of economic fluctuations while still achieving a reasonable level of dimension reduction. While a two-level model only enables counter-factual analyses of aggregate or idiosyncratic shocks, the effects of aggregate, block-specific, and idiosyncratic shocks can be coherently analyzed in our framework.

References

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

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

——— (2006): "Confidence Intervals for Diffusion Index Forecasts and Inference with Factor-Augmented Regressions," *Econometrica*, 74:4, 1133–1150.

Beltratti, A. and C. Morana (2008): "International Shocks and National House Prices," Bocconi University.

Boivin, J. and S. Ng (2006): "Are More Data Always Better for Factor Analysis," *Journal of Econometrics*, 132, 169–194.

Carter, C. K. and R. Kohn (1994): "On Gibbs Sampling for State Space Models," *Biometrika*, 81:3, 541–533.

Forni, M., M. Hallin, M. Lippi, and L. Reichlin (2005): "The Generalized Dynamic Factor Model, One Sided Estimation and Forecasting," *Journal of the American Statistical Association*, 100, 830–840.

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

Geweke, J. and G. Zhou (1996): "Measuring the Pricing Error of the Arbitrage Pricing Theory," *Review of Financial Studies*, 9:2, 557–87.

Goldstein, H. and W. Browne (2002): "Miltilevel Factor Analysis Using Markov Chain Monte Carlo Estimation," in *Latent variable and Latent Structure Models*, 225–243.

Hallin, M. and R. Liska (2008): "Dynamic Factors in the Presence of Block Structure," European University Institute WP 2008/22.

Kim, C. and C. Nelson (2000): *State Space Models with Regime Switching*, MIT Press.

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

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

Lopes, H. and M. West (2004): "Bayesian Model Assessment in Factor Analysis," *Statistical Sinica*, 14, 41–87.

Milani, F. and F. Belviso (2006): "Structural Factor-Augmented VARs and the Effects of Moentary Policy," *Topics in Macroeconomics*, 6:3, Article 2.

Mouchart, M. and E. S. Martin (2003): "Specification and Identification Issues in Models Involving a Latent Hierarhical Structure," *Journal of Statistical Planning and Inference*, 111, 143–163.

Onatski, A. (2006): "Asymptotic Distribution of the Principal Components Estimator of Large Factor Models when Factors are Relatively Weak," Mimeo, Columbia University.

Otrok, C. and C. Whiteman (1998): "Bayesian Leading Indicators: Measuring and Predicting Economic Conditions in Iowa," *International Economic Review*, 39:4, 997–1014.

Sargent, T. and C. Sims (1977): "Business Cycle Modelling without Pretending to have too much a Priori Economic Theory," in *New Methods in Business Cycle Research*, ed. by C. Sims, Minneapolis: Federal Reserve Bank of Minneapolis.

Stock, J. H. and M. Watson (1989): "New Indexes of Coincident and Leading Economic Indications," in *NBER Macroeconomics Annual 1989*, ed. by O. J. Blanchard and S. Fischer, Cambridge: M. I. T. Press.

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

——— (2006): "Forecasting with Many Predictors," in *Handbook of Forecasting*, North Holland.

——— (2008a): "The Evolution of National and Regional Factors in U.S. Housing Construction," Princeton University.

——— (2008b): "Forecasting in Dynamic Factor Models Subjet to Structural Instability," Princeton University.

Table 1: Data

Block		Variables Ordered 1 and 2
1	CU	Capacity Utilization: Machinery (SA, Percent of Capacity)
		Capacity Utilization: Motor Vehicles and Parts (SA, Percent of Capacity)
2	IP	IP: Durable Consumer Goods (SA, 2002=100)
		IP: Nondurable Consumer Goods (SA, 2002=100)
3	ES	All Employees: Wholesale Trade (SA, Thous)
		Avg Wkly Earnings: Construction (SA, $/wk)
4	HS	Civilian Labor Force: Men: 25-54 Years (SA, Thous)
		Unemployment Rate: Full-Time Workers: Men (SA, %)
5	MS	ISM Mfg: PMI Composite Index (SA, 50+ = Econ Expand)
		Phila FRB Bus Outlook: General Activity, Current, Diffusion Index (SA,%)
6	DG	Mfrs' Inventories: Machinery (EOP, SA, Mil.$)
		Mfrs' Unfilled Orders: Machinery (EOP, SA, Mil.$)
7	RS	Retail Sales: General Merchandise Stores (SA, Mil.$)
		Retail Sales: Motor Vehicle Dealers (SA, Mil$)
8	WT	Merchant Wholesalers: Sales: Automotive (SA, Mil.$)
		Merchant Wholesalers: Sales: Apparel (SA, Mil.$)
9	AUTO	Domestic Car Retail Sales (SAAR, Mil. Units)
		Domestic Light Truck Retail Sales (SAAR, Mil. Units)
10	H- NE	Housing Starts: 1-Unit: Northeast (SAAR, Thous.Units)
		Housing Completions: 1-Unit: Northeast (SAAR, Thous.Units)
11	H- WEST	Housing Starts: 1-Unit: West (SAAR, Thous.Units)
		Housing Completions: 1-Unit: West (SAAR, Thous.Units)
12	H- CTL	Housing Starts: 1-Unit: Midwest (SAAR, Thous.Units)
		Housing Starts: 1-Unit: South (SAAR, Thous.Units)

Summary Statistics

b	T	N	IC_2	$R^2_{\tilde{F}_1}$	$R^2_{\tilde{F}_2}$	$AR_{\tilde{F}_1}$	$AR_{\tilde{F}_2}$
CU	191	25	1	0.210	0.092	0.112	0.031
IP	191	38	1	0.208	0.086	0.123	0.033
ES	191	72	2	0.189	0.132	-0.190	0.675
HS	191	85	8	0.113	0.081	-0.122	-0.446
MS	191	35	4	0.143	0.108	-0.062	-0.104
DG	191	60	2	0.115	0.088	0.273	0.302
RS	191	30	1	0.187	0.086	-0.333	-0.301
WT	191	53	1	0.093	0.064	-0.312	-0.115
AS	191	4	4	0.483	0.242	-0.360	-0.365
H- NE	191	8	8	0.190	0.181	-0.352	-0.531
H- WEST	191	7	7	0.240	0.196	-0.310	-0.199
H- CTL	191	18	0	0.119	0.091	-0.112	-0.246

Note: IC_2 is the Bai-Ng (2002) criteria for determining the number of factors. R^2_j is the j-th eigenvalue of $x'x$ divided by the sum of the eigenvalues. $AR_{\tilde{F}_j}$ is the first order autocorrelation of j-th principal component of F.

Table 2: Univariate Analysis of Principal Component Estimates

Two Step Model:

$$
\begin{aligned}
\tilde{G}_{bjt} &= \tilde{\Lambda}_{F.bj}\tilde{F}_t(\tilde{G}_t) + \tilde{e}_{Gbjt} \\
\tilde{G}_{bjt} &= \tilde{\Psi}_{G.bj}\tilde{G}_{bjt-1} + \tilde{\epsilon}_{Gbjt} \\
\tilde{e}_{Gbjt} &= \tilde{\Psi}_{e_{Gbj}}\tilde{e}_{Gbjt-1} + \tilde{\epsilon}_{Gbjt} \\
\tilde{F}_{kt}(\tilde{G}_t) &= \tilde{\Psi}_{F.k}\tilde{F}_{kt-1}(\tilde{G}_t) + \tilde{\epsilon}_{Fkt}.
\end{aligned}
$$

Block	T	N_b	IC_2	$R^2_{G_{b.1}}$	$R^2_{Gb.2}$	$\tilde{\Psi}_{\tilde{G}.b1}$	$\tilde{\Psi}_{\tilde{G}.b2}$
CU	191	25	1	0.210	0.092	0.112	0.031
IP	191	38	1	0.208	0.086	0.123	0.033
ES	191	72	2	0.189	0.132	-0.190	0.675
HS	191	85	8	0.113	0.081	-0.122	-0.446
MS	191	35	4	0.143	0.108	-0.062	-0.104
DG	191	60	2	0.115	0.088	0.273	0.302
RS	191	30	1	0.187	0.086	-0.333	-0.301
WT	191	53	1	0.093	0.064	-0.312	-0.115
AS	191	4	4	0.483	0.242	-0.360	-0.365
H- NE	191	8	8	0.190	0.181	-0.352	-0.531
H- WEST	191	7	7	0.240	0.196	-0.310	-0.199
H- CTL	191	18	0	0.119	0.091	-0.112	-0.246

Note: Let $\tilde{G}_{bjt}$ be the j-th factor obtained by the method of principal components using data from block b. Then $R^2_{Gb_j}$ is the explanatory power of the j factor, obtained as the ratio of j-th largest eigenvalue $X'X$ to the sum of the eigenvalues. $\hat{\Psi}_{\tilde{G}.bj}$ is the estimated first order autocorrelation coefficient of $\tilde{G}_{bj}$.

Table 3: A Three Block Housing Model

True and Posterior Means: $\hat{\psi}_G$ and $\hat{\psi}_F$

		Actual Data		Standard Errors		Simulated Data		Standard Errors	
b	j	$\psi_{G.bj}$	$\hat{\sigma}^2_{G.bj}$	$\psi_{G.bj}$	$\hat{\sigma}^2_{G.bj}$	$\psi_{G.bj}$	$\hat{\sigma}^2_{G.bj}$	$\psi_{G.bj}$	$\hat{\sigma}^2_{G.bj}$
1	1	-0.007	0.366	0.177	0.098	-0.008	0.409	0.157	0.090
1	2	-0.105	0.082	0.114	0.031	-0.049	0.070	0.117	0.024
2	1	0.009	0.310	0.126	0.085	-0.083	0.244	0.150	0.066
2	2	-0.251	0.091	0.128	0.038	-0.043	0.054	0.112	0.020
3	1	0.110	0.116	0.179	0.066	0.088	0.121	0.118	0.049
3	2	-0.114	0.040	0.114	0.015	-0.006	0.067	0.118	0.032
		ψ_F	σ^2_F	ψ_F	$\hat{\sigma}^2_F$	ψ_F	$\hat{\sigma}^2_F$	ψ_F	$\hat{\sigma}^2_F$
	1	0.945	0.016	0.033	0.009	0.943	0.012	0.035	0.006

Decomposition of Variance

	Estimates				standard Errors			
block	σ^2_X	share$_F$	share$_G$	share$_X$	σ^2_X	share$_F$	share$_G$	share$_X$
	Data							
1	1.490	0.173	0.237	0.590	2.299	0.133	0.059	0.087
2	1.742	0.106	0.230	0.664	5.924	0.103	0.051	0.064
3	1.368	0.140	0.118	0.742	1.269	0.101	0.032	0.077
	Simulated Data							
1	1.338	0.162	0.233	0.606	0.798	0.112	0.057	0.064
2	1.234	0.106	0.223	0.670	0.457	0.104	0.046	0.072
3	1.191	0.121	0.119	0.760	0.506	0.099	0.024	0.080

Table 4: A Six Block three Level Model for Production:

$$
\begin{aligned}
G_{bt} &= \Lambda_{F.b}(L)F_t + e_{Gbt} \\
\Psi_{F.r}(L)F_{rt} &= \epsilon_{Frt}, \quad r = 1, \ldots K_F \\
\Psi_{G.bj}(L)e_{Gbjt} &= \epsilon_{Gbjt}, \quad j = 1, \ldots K_{Gb}.
\end{aligned}
$$

Block	j	$\hat{\Psi}_{G.bj}$	$\hat{\sigma}^2_{\epsilon bj}$	S.E	
CU: 1	1	0.272	0.031	0.114	0.026
CU: 1	2	0.288	0.072	0.166	0.033
IP: 2	1	0.322	0.043	0.117	0.020
IP: 2	2	0.098	0.064	0.107	0.018
ES: 3	1	0.013	0.027	0.185	0.010
ES: 3	2	-0.034	0.027	0.185	0.010
HS: 4	1	-0.067	0.103	0.100	0.021
HS: 4	2	-0.021	0.059	0.101	0.013
MS: 5	1	0.381	0.860	0.130	0.103
MS: 5	2	0.087	0.106	0.099	0.023
DG: 6	1	-0.101	0.039	0.173	0.009
DG: 6	2	0.041	0.041	0.175	0.009
Factor		$\Psi_{F.k}$	$\hat{\sigma}^2_{F.k}$	S.E.	
1		0.795	0.020	0.068	0.008

Principal Component Estimates

	N	$R^2_{F.1}$	$R^2_{F.2}$	$\tilde{\Psi}_{F.1}$	$\tilde{\Psi}_{F.2}$
$\tilde{F}_t(\tilde{G}_t)$	14	.205	.133	.074	.272
$\tilde{F}_t$	315	.210	.208	.187	.339

Decomposition of Variance

	Estimates				Standard Errors			
block	σ^2_X	share$_F$	share$_G$	share$_X$	σ^2_X	share$_F$	share$_G$	share$_X$
1 CU:	1.235	0.197	0.144	0.659	0.120	0.048	0.021	0.042
2 IP:	1.318	0.245	0.151	0.604	0.153	0.054	0.019	0.043
3 ES:	1.116	0.201	0.138	0.661	0.112	0.053	0.020	0.037
4 HS:	1.061	0.021	0.170	0.809	0.029	0.015	0.011	0.013
5 MS:	1.130	0.044	0.244	0.712	0.053	0.025	0.020	0.018
6 DG:	1.033	0.062	0.141	0.798	0.040	0.028	0.013	0.024

Table 5: Correlation Between $\hat{G}_{bkt}$ and $\tilde{e}_{rt|\hat{F}}$

r	b	k	ρ	r	b	k	ρ
1	3	2	0.235	4	1	2	0.109
2	3	2	0.716	4	6	1	0.233
2	5	1	0.124	5	4	2	0.139
2	6	2	0.196	6	2	2	0.138
3	4	1	0.616	7	5	2	0.224

Table 6: A Nine Block Four Level Model for Real Activity

$$
\begin{aligned}
H_{bst} &= \Lambda_{H.bs}(L)G_{bt} + e_{Hbst} \\
G_{bt} &= \Lambda_{G.b}(L)F_t + e_{Gbt} \\
\Psi_{F.r}(L)F_{rt} &= \epsilon_{Frt} \\
\Psi_{G.bj}(L)e_{Gbjt} &= \epsilon_{Gbjt}, \quad j = 1, \ldots k_{Gb}. \\
\Psi_{H.bsk}(L)e_{Hbskt} &= \epsilon_{Hbskt}, \quad j = 1, \ldots k_{Hbs}.
\end{aligned}
$$

b	j	$\hat{\Psi}_{G.bj}$	$\hat{\sigma}^2_{\epsilon bj}$	S.E	
1	1	0.302	0.101	0.140	0.024
2	1	0.351	0.037	0.193	0.015
2	2	0.342	0.019	0.328	0.007
3	1	0.182	0.063	0.173	0.024
Factor		$\Psi_{F.k}$	$\hat{\sigma}^2_{F.k}$	S.E.	
1		0.895	0.013	0.049	0.007

Table 7: Decomposition of Variance

block	sub-block	σ^2_X	share$_F$	share$_G$	share$_H$	share$_X$
		Estimates				
1	CU	1.239	0.104	0.137	0.169	0.590
1	IP	1.157	0.099	0.129	0.153	0.619
1	MS	1.050	0.027	0.033	0.237	0.703
1	DG	0.975	0.022	0.027	0.172	0.779
2	ES	1.032	0.051	0.103	0.205	0.641
2	HS	1.026	0.018	0.037	0.177	0.767
3	RS	1.091	0.045	0.072	0.213	0.669
3	WT	1.007	0.015	0.023	0.164	0.798
3	AU	1.050	0.040	0.066	0.552	0.342
		Standard Errors				
1	CU	0.525	0.063	0.033	0.027	0.040
1	IP	0.404	0.061	0.029	0.020	0.041
1	MS	0.076	0.027	0.012	0.018	0.023
1	DG	0.072	0.023	0.008	0.013	0.021
2	ES	0.167	0.044	0.026	0.027	0.033
2	HS	0.058	0.023	0.013	0.014	0.022
3	RS	0.509	0.044	0.020	0.022	0.032
3	WT	0.147	0.022	0.010	0.015	0.022
3	AU	0.332	0.046	0.034	0.051	0.027

1 Appendix

Sampling $\{F_t\}$

To obtain estimates of the global factors F given the block factors G, we have to perform the following steps. First, pre-Whiten the observation equation

$$G_t = \Lambda_F(L)F_t + e_{Gt}$$

so that its errors are i.i.d. This gives

$$\begin{array}{rcl} \Psi_G(L)G_t & = & \Psi_G(L)\Lambda_F(L)F_t + \varepsilon_{Gt}, \\ \text{or} \quad \tilde{G}_t & = & \tilde{\Lambda}_F(L)F_t + \varepsilon_{Gt} \end{array}$$

where $\tilde{G}_t = \Psi_G(L)G_t$, and where $\tilde{\Lambda}_F(L) = \Psi_G(L)\Lambda_F(L) = \tilde{\Lambda}_{F0} + \tilde{\Lambda}_{F1}L + ... + \tilde{\Lambda}_{Fs_F^*}L^{s_F^*}$ is a $K_G \times K_F$ matrix polynomial of order $s_F^* = q_G + s_F$. Stacking the lags of F, this gives the companion form:

$$\tilde{G}_t = \begin{bmatrix} \tilde{\Lambda}_{F0} & \tilde{\Lambda}_{F1} & \cdots & \tilde{\Lambda}_{Fs_F^*} \end{bmatrix} \begin{pmatrix} F_t \\ F_{t-1} \\ \vdots \\ F_{t-s_F^*} \end{pmatrix} + \epsilon_{Gt}$$

$$\begin{pmatrix} F_t \\ F_{t-1} \\ \vdots \\ F_{t-s_{F+1}^*} \end{pmatrix} = \begin{bmatrix} \Psi_{F1} & \cdots & \Psi_{Fq_F} & 0 & \cdots & 0 \\ I & 0 & \vdots & \vdots & & \vdots \\ \vdots & \ddots & \ddots & \vdots & & \vdots \\ 0 & \cdots & I & 0 & \cdots & 0 \end{bmatrix} \begin{pmatrix} F_{t-1} \\ F_{t-2} \\ \vdots \\ F_{t-s_F^*} \end{pmatrix} + \begin{pmatrix} \epsilon_{Ft} \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

or

$$\begin{array}{rcl} \tilde{G}_t & = & \vec{\tilde{\Lambda}}_F \vec{F}_t + \epsilon_{Gt} \\ \vec{\Psi}_F(L)\vec{F}_t & = & \vec{\varepsilon}_{Ft} \end{array}$$

where $\Omega_F = Var(\vec{\varepsilon}_{Ft}) = \begin{pmatrix} \Sigma_F & 0 \\ 0 & 0 \end{pmatrix}$.

Denote Ξ_F the set of parameters $\{\Lambda_F, \Psi_F, \Sigma_G, \Sigma_F\}$. Then, following Carter and Kohn (1994), the conditional distribution of the factors $\vec{F}$ given the pre-Whitened block factors $\{\tilde{G}_t\}$ and the parameters Ξ_F can be obtained by performing the following steps. First run the Kalman filter forward to obtain estimates $\vec{F}_{T|T}$ of the (stacked) factors and their variance covariance matrix $\vec{P}_{T|T}$ in period T based on all available sample information:

$$
\begin{aligned}
\vec{F}_{t+1|t} &= \vec{\Psi}_F \; \vec{F}_{t|t} \\
\vec{P}_{Ft+1|t} &= \vec{\Psi}_F \vec{P}_{Ft|t} \vec{\Psi}'_F + \vec{\Sigma}_F \\
\vec{F}_{t|t} &= \vec{F}_{t|t-1} + \vec{P}_{Ft|t-1} \vec{\tilde{\Lambda}}'_F \left(\vec{\tilde{\Lambda}}_F \vec{P}_{Ft|t-1} \vec{\tilde{\Lambda}}'_F + \vec{\Sigma}_F \right)^{-1} \left(\tilde{G}_t - \vec{\tilde{\Lambda}}_F \vec{F}_{t|t-1} \right) \\
\vec{P}_{Ft|t} &= \vec{P}_{Ft|t-1} - \vec{P}_{Ft|t-1} \vec{\tilde{\Lambda}}'_F \left(\vec{\tilde{\Lambda}}_F \vec{P}_{Ft|t-1} \vec{\tilde{\Lambda}}'_F + \vec{\Sigma}_F \right)^{-1} \vec{\tilde{\Lambda}}_F \vec{P}_{Ft|t-1}
\end{aligned}
$$

Next, draw $\vec{F}_T$ from its conditional distribution given Ξ_F and the data through period T:

$$
\vec{F}_T | \{\tilde{G}_t\}, \Xi_F \sim N(\vec{F}_{T|T}, \vec{P}_{FT|T})
$$

Then, for $t = T-1, \ldots, 1$ proceed backwards to generate draws $\vec{F}_{t|T}$ from

$$
\begin{aligned}
\vec{F}_{t|T} | \vec{F}^*_{t+1}, \{\tilde{G}_T\}, \Xi_F &\sim N(\vec{F}_{t|t,\vec{F}^*_{t+1}}, \vec{P}_{t|t,\vec{F}^*_{t+1}}) \qquad (1) \\
\text{where} \quad \vec{F}_{t|t,\, \vec{F}^*_{t+1}} &= \vec{F}_{t|t} + \vec{P}_{t|t} \vec{\Psi}^{*\prime}_F \left(\vec{\Psi}^*_F \; \vec{P}_{t|t} \; \vec{\Psi}^{*\prime}_F + \Omega^*_F \right)^{-1} \left(\vec{F}^*_{t+1} - \vec{\Psi}^*_F \; \vec{F}_{t|t} \right) \\
\text{and} \quad \vec{P}_{t|t,\vec{F}^*_{t+1}} &= \vec{P}_{t|t} - \vec{P}_{t|t} \; \vec{\Psi}^{*\prime}_F \left(\vec{\Psi}^*_F \; \vec{P}_{t|t} \; \vec{\Psi}^{*\prime}_F + \Omega^*_F \right)^{-1} \vec{\Psi}^*_F \; \vec{P}_{t|t}.
\end{aligned}
$$

where Ω^*_F is the upper left $K_F \times K_F$ block of Ω_F which is positive-definite and where $\vec{F}^*_t$ and $\vec{\Psi}^*_F$ are the first K_F rows of $\vec{F}_t$ and $\vec{\Psi}_F$, respectively. Note also that we initialize the Kalman filter with the unconditional mean and variance of the states $\vec{F}$, i.e. $\vec{F}_{1|0} = E[\vec{F}] = 0$ and $vec(\vec{P}_{1|0}) = \left[I_{s_F} - \left(\vec{\Psi}_F \otimes \vec{\Psi}_F \right) \right]^{-1} vec(\Omega_F)$.

Sampling $\{G_t\}$

A similar algorithm can be used to sample the block factors G. Since the block-dynamics are assumed to be independent, this can be done block by block. Recall that

$$
\tilde{X}_{bt} = \tilde{\Lambda}_{Gb}(L) G_{bt} + \varepsilon_{Xbt}, \qquad \forall \; b = 1, \ldots, B,
$$

where $\tilde{X}_{bt} = \Psi_{Xb}(L) X_{bt}$ and where $\tilde{\Lambda}_{Gb}(L) = \Psi_{Xb}(L) \Lambda_{Gb}(L)$ is a $N_b \times K_b$ matrix polynomial of order $s^*_G = q_X + s_G$. Moreover, recall that

$$
\begin{pmatrix} G_{bt} \\ G_{bt-1} \\ \vdots \\ G_{bt-q_G+1} \end{pmatrix} = \begin{pmatrix} \alpha_{Fbt} \\ 0 \\ \vdots \\ 0 \end{pmatrix} + \begin{bmatrix} \Psi_{Gb1} & \Psi_{Gb2} & \cdots & \Psi_{Gbq_G} \\ I & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & I & 0 \end{bmatrix} \begin{pmatrix} G_{bt-1} \\ G_{bt} \\ \vdots \\ G_{bt-q_G} \end{pmatrix} + \begin{pmatrix} \epsilon_{Gbt} \\ 0 \\ \vdots \\ 0 \end{pmatrix}
$$

where

$$
\alpha_{Fbt} = \Psi_{Gb}(L) \Lambda_F(L) F_t, \qquad \forall \; b = 1, \ldots, B,
$$

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

$$\tilde{X}_{bt} = \begin{bmatrix} \tilde{\Lambda}_{Gb0} & \tilde{\Lambda}_{Gb1} & \cdots & \tilde{\Lambda}_{Gbs^*_G} \end{bmatrix} \begin{pmatrix} G_{bt} \\ G_{bt-1} \\ \vdots \\ G_{bt-s^*_G} \end{pmatrix} + \epsilon_{Xbt}$$

$$\begin{pmatrix} G_{bt} \\ G_{bt-1} \\ \vdots \\ G_{bt-s^*_G} \end{pmatrix} = \begin{pmatrix} \alpha_{Fbt} \\ 0 \\ \vdots \\ 0 \end{pmatrix} + \begin{bmatrix} \Psi_{Gb1} & \cdots & \Psi_{Gbq_G} & 0 & \cdots & 0 \\ I & 0 & \vdots & \vdots & & \vdots \\ \vdots & \ddots & \ddots & \vdots & & \vdots \\ 0 & \cdots & I & 0 & \cdots & 0 \end{bmatrix} \begin{pmatrix} G_{bt-1} \\ G_{bt-2} \\ \vdots \\ G_{bt-s^*_G-1} \end{pmatrix} + \begin{pmatrix} \epsilon_{Gbt} \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

or

$$\begin{aligned} \tilde{X}_{bt} &= \vec{\tilde{\Lambda}}_{Gb}\vec{G}_{bt} + \epsilon_{Xbt} \\ \vec{\Psi}_{Gb}(L)\vec{G}_{bt} &= \vec{\alpha}_{Fbt} + \vec{\epsilon}_{Gbt} \end{aligned}$$

Denote Ξ_{Gb} the set of parameters $\{\tilde{\Lambda}_{Gb}, \vec{\Psi}_{Gb}, \Sigma_{Gb}, \Sigma_{Xb}\}$. Conditional on Ξ_{Gb} and $\{F_t\}$, the above equations represent a state-space system with a time-varying intercept. We therefore need to slightly adjust the Carter and Kohn (1994) method laid out before. The complete set of equations is as follows.

First, run the Kalman filter forward to obtain estimates $\vec{G}_{bT|T}$ of the factors and their variance covariance matrix $\vec{P}_{bT|T}$ in period T based on all available sample information. With the time-varying intercept $\vec{\alpha}_{Fbt}$, this implies the following steps:

$$\begin{aligned} \vec{G}_{bt+1|t} &= \vec{\alpha}_{Fbt} + \vec{\Psi}_{Gb}\vec{G}_{bt|t} \\ \vec{P}_{Gbt+1|t} &= \vec{\Psi}_{Gb}\vec{P}_{Gbt|t}\vec{\Psi}'_{Gb} + \vec{\Sigma}_{Gb} \\ \vec{G}_{bt|t} &= \vec{G}_{bt|t-1} + \vec{P}_{Gbt|t-1}\vec{\tilde{\Lambda}}'_{Gb}\left(\vec{\tilde{\Lambda}}_{Gb}\vec{P}_{Gbt|t-1}\vec{\tilde{\Lambda}}'_{Gb} + \vec{\Sigma}_{Xb}\right)^{-1}\left(\tilde{X}_{bt} - \vec{\tilde{\Lambda}}_{Gb}\vec{G}_{bt|t-1}\right) \\ \vec{P}_{Gbt|t} &= \vec{P}_{Gbt|t-1} - \vec{P}_{Gbt|t-1}\vec{\tilde{\Lambda}}'_{Gb}\left(\vec{\tilde{\Lambda}}_{Gb}\vec{P}_{Gbt|t-1}\vec{\tilde{\Lambda}}'_{Gb} + \vec{\Sigma}_{Xb}\right)^{-1}\vec{\tilde{\Lambda}}_{Gb}\vec{P}_{Gbt|t-1} \end{aligned}$$

We again initialize the filter with the unconditional mean and variance of the states $\vec{G}_b$, i.e. $\vec{G}_{b1|0} = E[\vec{G}_b]$ and $vec(\vec{P}_{1|0}) = vec(var(\vec{G}_b))$. Precisely, these are given by

$$\begin{aligned} E[\vec{G}_{bt}] &= E\left[\vec{\alpha}_{Fbt} + \vec{\Psi}_{Gb}\vec{G}_{bt-1} + \vec{\epsilon}_{Gbt}\right] = 0 \\ Var(\vec{G}_{bt}) &= Var(\vec{\alpha}_{Fbt} + \vec{\Psi}_{Gb}\vec{G}_{bt-1} + \vec{\epsilon}_{Gbt}) \\ &= Var(\vec{\alpha}_{Fbt}) + \vec{\Psi}_{Gb}Var(\vec{G}_{bt-1})\vec{\Psi}'_{Gb} + \vec{\Sigma}_{Gb} + 2\vec{\Psi}_{Gb}Cov(\vec{\alpha}_{Fbt}, \vec{G}_{bt-1}) \end{aligned}$$

Altogether, we therefore have

$$vec(Var(\vec{G}_b)) = \left[I - \left(\vec{\Psi}_{Gb} \otimes \vec{\Psi}_{Gb}\right)\right]^{-1}\left(\vec{\Sigma}_{\alpha_F} + \vec{\Sigma}_{Gb} + 2\vec{\Psi}_{Gb}\vec{\Sigma}_{\alpha_{Fb}}\right)$$

where

$$\begin{aligned}
\vec{\Sigma}_{\alpha_F} &= Var(\vec{\alpha}_{Fbt}) = \begin{bmatrix} \Sigma_{\alpha_F} & 0 \\ 0 & 0 \end{bmatrix} \\
\text{and} \qquad \Sigma_{\alpha} &= Var(\alpha_{bt}) = Var(\vec{\tilde{\Lambda}}_{Fb0}\vec{F}_t) \\
&= \vec{\tilde{\Lambda}}_{Fb0} Var(\vec{F}_t)\vec{\tilde{\Lambda}}'_{Fb0}
\end{aligned}$$

Moreover,

$$\vec{\Sigma}_{Gb} = \begin{bmatrix} \Sigma_{Gb} & 0 \\ 0 & 0 \end{bmatrix}$$

and

$$\begin{aligned}
\vec{\Sigma}_{\alpha b} &= Cov(\vec{\alpha}_{bt}, \vec{G}_{bt-1}) = \begin{bmatrix} \Sigma_{ab} & 0 \\ 0 & 0 \end{bmatrix}, \\
\text{with} \qquad \Sigma_{\alpha b} &= Cov(\alpha_{bt}, G_{bt-1}) \\
&= Cov(\vec{\tilde{\Lambda}}_{Fb0}\vec{F}_t, \lambda'_{Fb}(L)F_{t-1} + \varepsilon_{Gbt-1}) \\
&= Cov(\vec{\tilde{\Lambda}}_{Fb0}\vec{F}_t, \lambda'_{Fb}(L)F_{t-1}) \\
&= \vec{\tilde{\Lambda}}_{Fb1} Var(\vec{F}_{t-1})\vec{\tilde{\Lambda}}'_{Fb0}
\end{aligned}$$

where $vec(Var(\vec{F}_t)) = vec(Var(\vec{F}_{t-1})) = \left[I_{s_F} - \left(\vec{\Psi}_F \otimes \vec{\Psi}_F\right)\right]^{-1} vec(\Omega_F)$ The Kalman filter iterations provide us with the conditional distribution of $\vec{G}_{bT|T}$ given Ξ_{Gb} and the data through period T:

$$\vec{G}_{bT}|\{\tilde{X}_{bt}\}, \Xi_{Gb} \sim N(\vec{G}_{bT|T}, \vec{P}_{GbT|T})$$

Carter and Kohn show us how to sample the entire set of factor observations conditional on the parameters Ξ_{Gb} and all the data. Given the Gaussianity and Markovian structure of the state-space model, the distribution of $\vec{G}_{bt}$ given $\vec{G}_{bt+1}$ and $\tilde{X}_{bt}$is normal:

$$\vec{G}_{bt}|\tilde{X}_{bt}, \vec{G}^*_{bt+1}, \Xi_{Gb} \sim N(\vec{G}_{bt|t,\vec{G}^*_{bt+1}}, \vec{P}_{Gbt|t,\vec{G}^*_{bt+1}}) \tag{2}$$

where

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

and where $\vec{G}^*_{bt+1}$ and $\vec{\Psi}^*_{Gb}$ denote the first k_b rows of $\vec{G}_{bt+1}$ and $\vec{\Psi}_{Gb}$, respectively. Given these conditional distributions, we can then proceed backwards to generate draws $\vec{G}^*_{bt}$ for $t{=}T{-}1, \ldots, 1$.

Figure 2: Housing

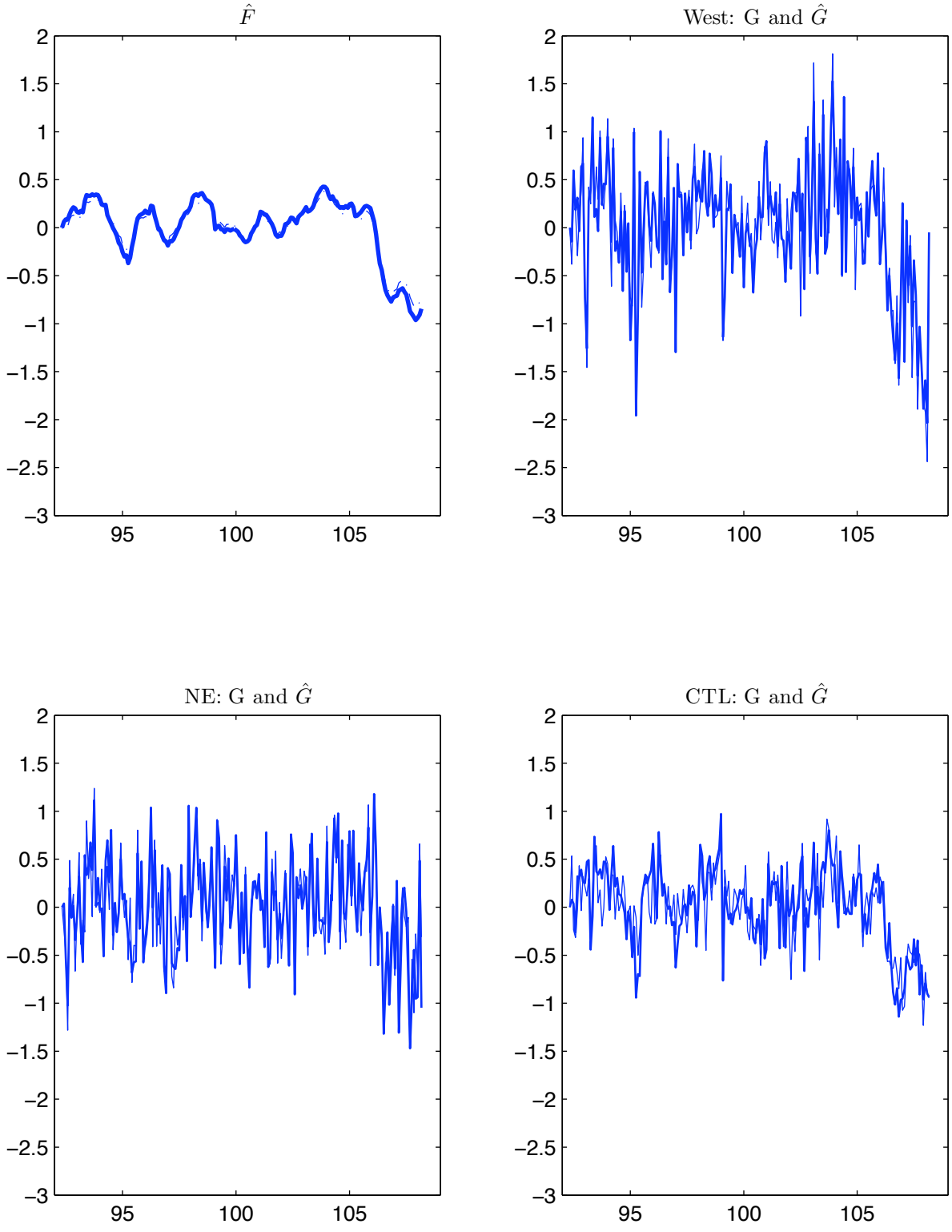
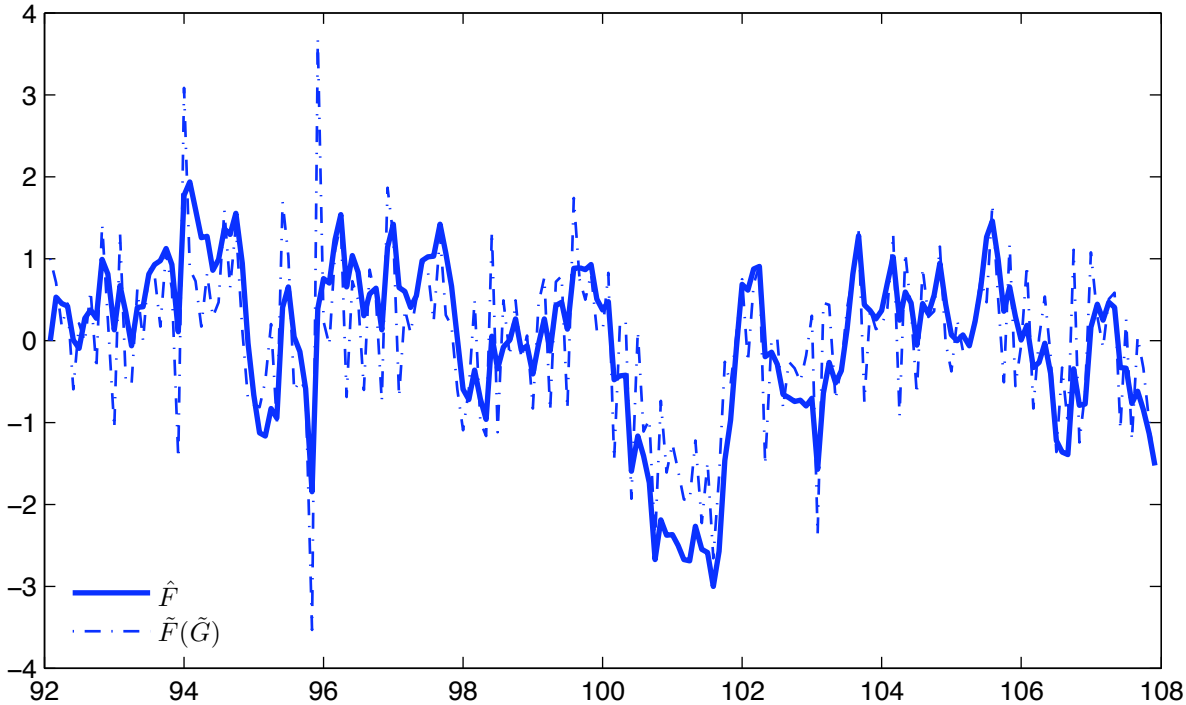


Figure 3: Six Block for Output



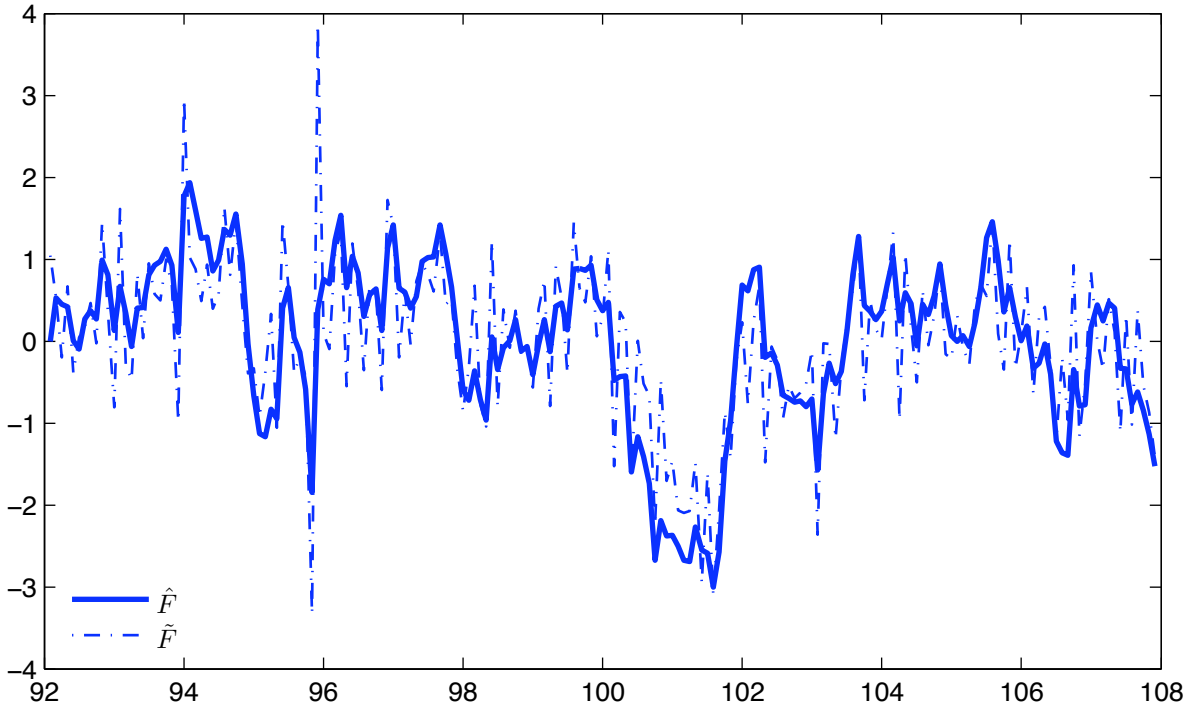
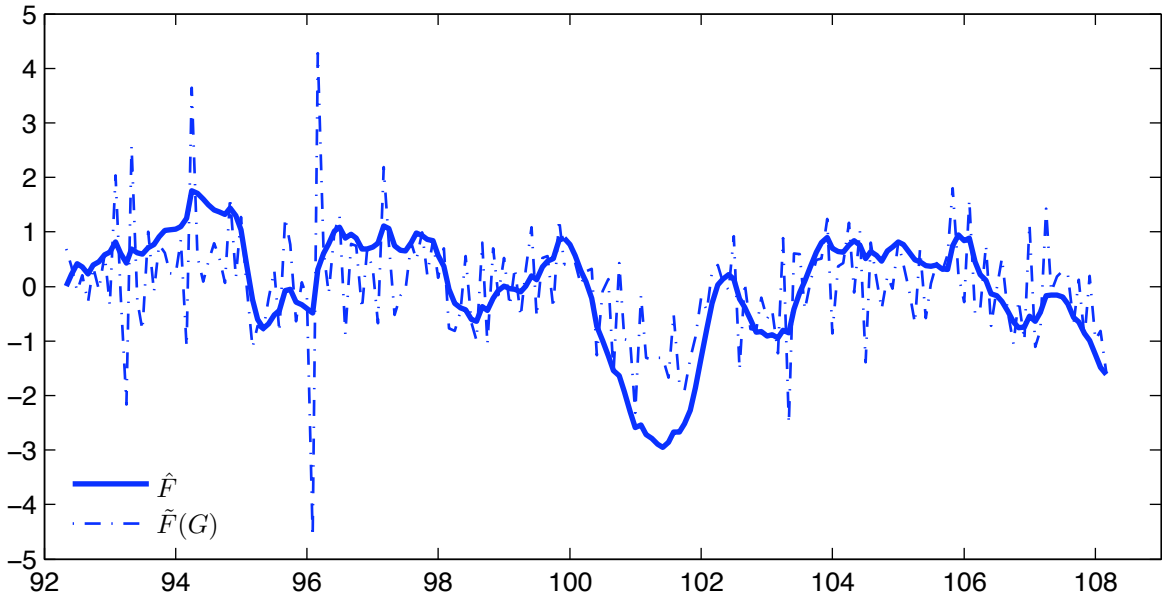
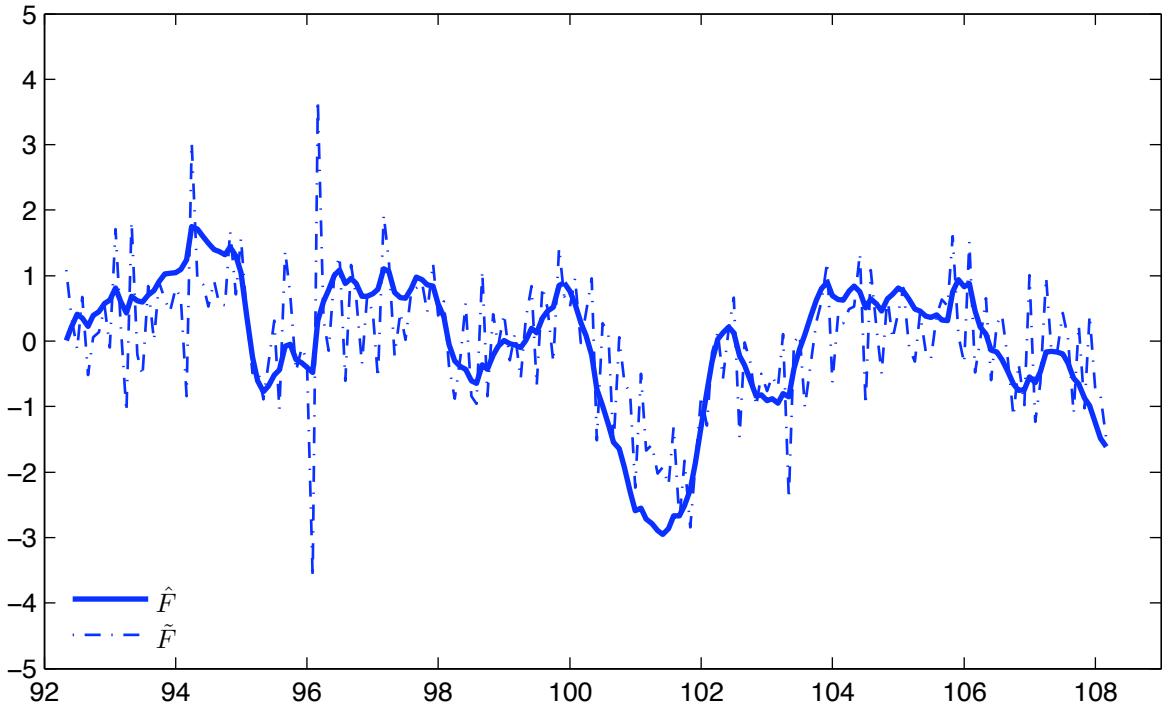



Figure 4: A Four Level Factor Estimate of Real Activity