The dynamic structure of high-dimensional factor models

Mario Forni, Università di Modena and RECent
Marc Hallin, ECARES, Bruxelles
Marco Lippi, Università di Roma “La Sapienza” and EIEF
Paolo Zaffaroni, Imperial College, London and La Sapienza, Roma

CREATES, Aarhus University

December 2, 2016

Outline

Outline

Outline

Outline

Outline

Forecasting.

We would like to use a large macroeconomic dataset x_{it}, $i = 1, \ldots, n$, to forecast, say, inflation. However, the number of observations available for each series is too small to regress inflation on all available series (curse of dimensionality).
In some cases, with macroeconomic datasets in particular, it is reasonable to assume that the variables x_{it} are driven by a small number of common factors plus variable-specific components:

$$\begin{aligned} x_{it} &= \chi_{it} + \xi_{it} = \mathbf{b}_i(L)\mathbf{u}_t + \xi_{it} \\ &= b_{i1}(L)u_{1t} + b_{i2}(L)u_{2t} + \cdots + b_{iq}(L)u_{qt} + \xi_{it} \end{aligned}$$

for $i = 1, \ldots, n$, where q is very small as compared to n. If we are able to estimate the shocks u_{jt}, the filters $b_{ij}(L)$, and therefore ξ_{it}, then we can separately forecast the common components χ_{it} and the idiosyncratic components ξ_{it}. See Stock and Watson (2002a and b), Forni, Hallin, Lippi and Reichlin (2005).

EUROCOIN

Let me just recall here an important application of GDFM, namely the construction of Eurocoin, an indicator of the medium- long-run component of the European GNP, published monthly by the Banca d'Italia and CEPR. Eurocoin is based on smooth factors, obtained by maximizing the variance of linear combinations of the data within a high-frequency band. See Altissimo, Cristadoro, Forni, Lippi and Veronese (2009).

Aggregation, an example

Let y_{1t} and y_{2t} be sectoral output rates of growth, aggregate output rate of growth is $y_t = y_{1t} + y_{2t}$.
Let U_t be the unemployment rate.
Let s_{1t} and s_{2t} be two supply shocks (permanent) and d_t a (transitory) monetary shock. Suppose that

$$\begin{pmatrix} y_{1t} \\ y_{2t} \\ U_t \end{pmatrix} = A(L) \begin{pmatrix} s_{1t} \\ s_{2t} \\ d_t \end{pmatrix}$$

is the structural microrelashionship. We observe

$$\begin{pmatrix} y_t \\ U_t \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} y_{1t} \\ y_{2t} \\ U_t \end{pmatrix} = \mathcal{G} A(L) \begin{pmatrix} s_{1t} \\ s_{2t} \\ d_t \end{pmatrix} = B(L) \begin{pmatrix} w_t \\ z_t \end{pmatrix} = B^*(L) \begin{pmatrix} s_t^* \\ d_t^* \end{pmatrix}$$

where $B^*(L)$ is correctly identified à la Blanchard and Quah.

Aggregation, an example

[From previous slide]

$$\begin{pmatrix} y_t \\ U_t \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} y_{1t} \\ y_{2t} \\ U_t \end{pmatrix} = \mathcal{G} A(L) \begin{pmatrix} s_{1t} \\ s_{2t} \\ d_t \end{pmatrix} = B(L) \begin{pmatrix} w_t \\ z_t \end{pmatrix} = B^*(L) \begin{pmatrix} s_t^* \\ d_t^* \end{pmatrix}$$

We have

$$\begin{pmatrix} s_t^* \\ d_t^* \end{pmatrix} = B^*(L)^{-1} \mathcal{G} A(L) \begin{pmatrix} s_{1t} \\ s_{2t} \\ d_t \end{pmatrix},$$

which is a pretty complicated relationship. Each of the “observable” shocks is a linear combination of current and past supply and demand microshocks. In general, s_t^* depends on both supply and demand microshocks, and so does d_t^*.

Aggregation, an example

However, we actually do not observe the aggregates without measurement errror, i.e.

$$\begin{pmatrix} y_t \\ U_t \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} y_{1t} \\ y_{2t} \\ U_t \end{pmatrix} + \begin{pmatrix} \xi_{1t} \\ \xi_{2t} \end{pmatrix},$$

so that the “observable” shocks are

$$\begin{pmatrix} s_t^* \\ d_t^* \end{pmatrix} = B^*(L)^{-1} \left[\mathcal{G} A(L) \begin{pmatrix} s_{1t} \\ s_{2t} \\ d_t \end{pmatrix} + \begin{pmatrix} \xi_{1t} \\ \xi_{2t} \end{pmatrix} \right],$$

that is, linear combinations of current and past values of the structural shocks and errors.

Aggregation, an example

Suppose now that y_t and U_t belong to a high-dimensional vector

$$\begin{pmatrix} x_{1t} & x_{2t} & \cdots & x_{Nt} \end{pmatrix}$$

(y_t and U_t are, say, the first and second) and that

$$x_{it} = b_{i1}(L)s_{1t} + b_{i2}(L)s_{2t} + b_{i3}(L)d_t + \xi_{it},$$

then dynamic factor techniques allow estimation of the shocks $s_{1t}, s_{2t},\ d_t$.
Thus, **under assumptions that must be specified and are not innocuous**, both aggregation and measurement error problems are mitigated.
(On aggregation see Forni and Lippi, Aggregation and the microfoundations of dynamic macroeconomics, Oxford, 1997).

Structural analysis.

In VAR and SVAR analysis the number of macro shocks is equal to the number of variables in the vector, a rather artificial assumption. Dynamic factor models allow to determine the number of macro shocks that determine all the variables in the dataset, irrespective of their number (much like in a DSGE).

$$\begin{aligned} x_{it} &= \chi_{it} + \xi_{it} = \mathbf{b}_i(L)\mathbf{u}_t + \xi_{it} \\ &= b_{i1}(L)u_{1t} + b_{i2}(L)u_{2t} + \cdots + b_{iq}(L)u_{qt} + \xi_{it} \end{aligned}$$

Structural analysis can be applied to the common components and the common shocks. An interesting property of structural GDFM is that the common shocks are generically fundamental for the common components χ_{it} (a solution to a big problem in VAR analysis). See Stock and Watson (2005), Forni, Giannone, Lippi and Reichlin (2009), FAVAR literature, Forni and Gambetti (2010a,b).

Structural analysis: Fundamentalness.

The simplest example I know. x_t is the rate of growth of productivity, u_t is a shock to technology:

$$x_t = au_t + bu_{t-1} = v_t + (b/a)v_{t-1}, \qquad a + b = 1$$

x_t is observed whereas u_t is not. Standard econometrics estimate

$$x_t = w_t + \beta w_{t-1}, \quad |\beta| < 1$$

thus the structural shock v_t is estimated only if $a > b$. Otherwise $\beta = a/b$ and $w_t \neq v_t$.

Structural analysis: Fundamentalness.

From a time-series theoretical viewpoint, an MA(1) has two representations:

$$x_t = w_t + \beta w_{t-1} = v_t + \beta^{-1} v_{t-1}, \quad |\beta| < 1$$

and there is no way you can tell which one is the structural one. What you can say is that w_t is the one-step ahead prediction error and that βw_{t-1} is the best linear predictor of x_t at $t-1$, but no more.

Eigenvalues of the spectral density of the observables.

We assume the following:

Let x_{it}, $i \in \mathbb{N}$, $t \in \mathbb{Z}$ be a family of stochastic variables. We assume:
(a) $\mathbf{x}_{nt} = (x_{1t}\ x_{2t}\ \cdots\ x_{nt})$ is stationary for all n.
(b) There exists an integer q such that the q-th eigenvalue of the spectral density matrix of the x's diverges as $n \to \infty$, whereas the $(q+1)$-th eigenvalue is bounded.

The GDFM. The factor structure

The above assumptions are equivalent to the following factor structure:

$$\begin{aligned} x_{it} &= \chi_{it} + \xi_{it} = \mathbf{b}_i(L)\mathbf{u}_t + \xi_{it} \\ &= b_{i1}(L)u_{1t} + b_{i2}(L)u_{2t} + \cdots + b_{iq}(L)u_{qt} + \xi_{it} \end{aligned}$$

where

1. The filters b_{ij} are square summable.
2. Common and idiosyncratic components are orthogonal.
3. The first eigenvalue of the spectral density of the idiosyncratic components is bounded as $n \to \infty$.
4. The q-th eigenvalue of the spectral density matrix of the common components diverges as $n \to \infty$. See Forni and Lippi (2001), Forni, Hallin, Lippi and Reichlin, (2000), Chamberlain and Rothschild, (1983a and b).

In a recent paper, Hallin and Lippi (2013) provide an alternative presentation in the time domain.

Finite Dimension

Consider again:

$$x_{it} = \chi_{it} + \xi_{it}.$$

Define

$$C_t = \overline{\text{span}}\{\chi_{it}\ ,\ i \in \mathbb{N}\}.$$

Of course $\dim(C_t)$ is independent of t because of stationarity. If

$$\dim(C_t) = r < \infty \qquad (\ddagger)$$

then

$$\begin{aligned} \chi_{it} &= g_{i1}F_{1t} + g_{i2}F_{2t} + \cdots + g_{ir}F_{rt} \\ \mathbf{F}_t &= N(L)\mathbf{u}_t \end{aligned}$$

and conversely.

Finite Dimension

$$\begin{aligned} \chi_{it} &= g_{i1}F_{1t} + g_{i2}F_{2t} + \cdots + g_{ir}F_{rt} \\ \mathbf{F}_t &= N(L)\mathbf{u}_t \end{aligned}$$

Usually $\mathbf{F}_t$ is modeled as a singular VAR

$$(I - A_1 L - \cdots - A_k L^k)\mathbf{F}_t = R\mathbf{u}_t,$$

where A_j is $r \times r$ and R is $r \times q$. Assumption (‡) is almost universal in recent GDFM literature. See Stock and Watson (2002a and b), Bai and Ng (2002, 2007), Forni, Hallin, Lippi and Reichlin (2005), Forni, Giannone, Lippi and Reichlin (2009).

Finite Dimension

Under (‡), the model for the common components is

$$\begin{aligned} \chi_{it} &= g_{i1}F_{1t} + g_{i2}F_{2t} + \cdots + g_{ir}F_{rt} \\ \mathbf{F}_t &= N(L)\mathbf{u}_t \end{aligned}$$

and we say that there are r static and q dynamic factors. For example, if

$$\chi_{it} = d_{i0}u_t + d_{i1}u_{t-1}$$

we have $q = 1$ and $r = 2$.

Finite Dimension

The reason for the popularity of (‡), i.e. the model

$$\begin{aligned} \chi_{it} &= g_{i1}F_{1t} + g_{i2}F_{2t} + \cdots + g_{ir}F_{rt} \\ \mathbf{F}_t &= N(L)\mathbf{u}_t \end{aligned}$$

for the common components, is simple to explain. Consider again the example

$$x_{it} = d_{i0}u_t + d_{i1}u_{t-1} + \xi_{it}.$$

In this case the first two principal components of the variables x_{it} estimate the space spanned by the common components.
Under (‡), estimation by Maximum Likelihood has been studied in Doz, Giannone and Reichlin (2012).

The number of factors

Bai and Ng (2002): the number of static factors r.
Bai and Ng (2007): both r and q.
Hallin and Liška (2007): q without the finite-dimension assumption.
Onatski (2009): q without the finite-dimension assumption.

Simple Factor Models Ruled Out

It is crucial here to note that this finite-dimension assumption is ruling out (with probability one) a model as simple as

$$x_{it} = \frac{1}{1 - \alpha_i L} u_t + \xi_{it},$$

with α_i drawn from a uniform distribution between, say, $-.8$ and .8. Indeed the stochastic variables

$$\chi_{it} = u_t + \alpha_i u_{t-1} + \alpha_i^2 u_{t-2} + \cdots$$

for a given t and $i \in N$, span an infinite-dimensional space unless α_i takes on a finite number of values.

Simple Factor Models Ruled Out

More in general, schemes like:

$$x_{it} = \frac{e_{i1}(L)}{f_{i1}(L)}u_{1t} + \frac{e_{i2}(L)}{f_{i2}(L)}u_{2t} + \cdots + \frac{e_{iq}(L)}{f_{iq}(L)}u_{qt} + \xi_{it}$$

are ruled out, unless strict conditions on the coefficients of the polynomials e and f hold.

This is a theoretical motivation to keep studying the general version of the model.

From an empirical point of view, we are proposing an alternative specification, which allows for an infinite-dimensional factor space.

Frequency-domain principal components

Back to the the paper Forni, Hallin, Lippi and Reichlin (2000).

1. Start with the spectral density matrix of the vector $\mathbf{x}_{nt}$, call it $\Sigma_n^x(\theta)$.
2. Take the first q principal components of $\Sigma_n^x(\theta)$, call $G(\theta)$ the $n \times q$ matrix of such p.c.'s
3. The spectral density matrix of χ_{nt}, call it $\Sigma_n^\chi(\theta)$, is $G(\theta)G^*(\theta)$.
4. Moreover:

$$\chi_{nt} = \underline{G}(L)\mathbf{u}_t, \quad \underline{G}(L) = \sum_{k=-\infty}^{\infty} \left[\int_{-\pi}^{\pi} G(\theta) e^{ik\theta} d\theta \right] L^k.$$

Two-sidedness of the filters

However, the estimation of $\Sigma_n^\chi(\theta)$ and $G(\theta)$ is obtained using Brillinger's dynamic principal components, with the consequence that $\underline{G}(L)$ is two-sided and therefore unreliable at the end of the sample and useless for forecasting purposes.
Here we solve the problem, i.e. we obtain one-sided filters, under some additional assumptions.

Rational spectral density of the common components

We do not assume finite dimension for $C_t = \overline{\text{span}}(\chi_{it},\ i \in \mathbb{N})$ but:
(i) (First additional assumption)The spectral density of the common components, Σ_n^{χ}, is rational.
This is equivalent to assuming that the functions $b_{ij}(L)$ in

$$\begin{aligned} x_{it} &= \chi_{it} + \xi_{it} = \mathbf{b}_i(L)\mathbf{u}_t + \xi_{it} \\ &= b_{i1}(L)u_{1t} + b_{i2}(L)u_{2t} + \cdots + b_{iq}(L)u_{qt} + \xi_{it} \end{aligned}$$

are rational.
The model is semiparametric, in that we do not make parametric assumptions on the idiosyncratic components. In particular, on their covariance.

Anderson and Deistler's results

We use recent results by Anderson and Deistler (2008a,b) on tall stochastic vectors.
The n-dimensional, rational spectral density, vector $\mathbf{y}_t$ is **tall** if its rank q is smaller than n. (The spectral density matrix of $\mathbf{y}_t$ is constant almost everywhere in $[-\pi \ \pi]$.)
For example, if we consider the vector

$$(\chi_{1t} \ \chi_{2t} \ \cdots \ \chi_{q+1,t})$$

its dimension is $q+1$ while its rank cannot exceed q.

Anderson and Deistler's results

Anderson and Deistler prove the following:
Consider a tall vector

$$\mathbf{y}_t = D(L)\mathbf{w}_t$$

whose dimension is n and rank $q < n$. Let

$$D(L) = \left(\frac{e_{ij}(L)}{f_{ij}(L)} \right),$$

where $f_{ij}(0) = 1$, degree$(e_{ij}) = m$, degree$(f_{ij}) = p$. Thus we have $P = nq(m + p + 1)$ parameters. For generic values of the parameters in $\mathbb{R}^P$, $\mathbf{y}_t$ has a finite autoregressive representation

$$A(L)\mathbf{y}_t = R\mathbf{w}_t$$

where $A(L)$ is $n \times n$ and R is $n \times q$.

An example

An intuition of this result. Let $n = 2$ and $q = 1$:

$$\begin{aligned} y_{1t} &= \alpha_1 w_t + \beta_1 w_{t-1} \\ y_{2t} &= \alpha_2 w_t + \beta_2 w_{t-1} \end{aligned}$$

We see that

$$w_t = \frac{1}{\alpha_1\beta_2 - \alpha_2\beta_1}(\beta_2 y_{1t} - \beta_1 y_{2t}),$$

so that

$$\begin{pmatrix} 1 - \delta\beta_1\beta_2 L & \delta\beta_1^2 L \\ -\delta\beta_2^2 & 1 + \delta\beta_1\beta_2 L \end{pmatrix} \begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} = \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix} w_t$$

where $\delta = 1/(\alpha_1\beta_2 - \alpha_2\beta_1)$. Note that the autoregressive representation exists if and only if $\alpha_1\beta_2 - \alpha_2\beta_1 \neq 0$, thus generically. Note also that, generically, w_t is **fundamental** even if it is not fundamental in the first or the second equation.

Generic fundamentalness in the singular case

Anderson and Deistler's result implies that for generic values of the parameters, the representation

$$\mathbf{y}_t = D(L)\mathbf{w}_t$$

is fundamental, namely $\mathbf{w}_t$ belongs to the space spanned by current and past values of $\mathbf{y}_t$. This is even more easy to understand. If $D(L)$ is square then one of its zeros lies within the unit disk with, so to speak, probability 1/2. In the tall case $D(L)$ has generically no zeros at all. Thus, unlike in SVAR analysis, fundamentalness is not a problem within structural analysis of the common components of GDFM.

The example again

Now consider again our example

$$\begin{aligned} y_{1t} &= \alpha_1 w_t + \beta_1 w_{t-1} \\ y_{2t} &= \alpha_2 w_t + \beta_2 w_{t-1} \end{aligned}$$

We see that

$$w_t = \frac{1}{\alpha_1\beta_2 - \alpha_2\beta_1}(\beta_2 y_{1t} - \beta_1 y_{2t}),$$

so that

$$\begin{pmatrix} 1 - \delta\beta_1\beta_2 L & \delta\beta_1^2 L \\ -\delta\beta_2^2 & 1 + \delta\beta_1\beta_2 L \end{pmatrix} \begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} = \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix} w_t$$

where $\delta = 1/(\alpha_1\beta_2 - \alpha_2\beta_1)$. Since y_{1t} and y_{2t} are generically linearly independent, the autoregressive representation of order one is generically unique.

Another example

On the other hand, consider

$$\begin{aligned} y_{1t} &= \alpha_1 w_t + \beta_1 w_{t-1} \\ y_{2t} &= \alpha_2 w_t + \beta_2 w_{t-1} \\ y_{3t} &= \alpha_3 w_t + \beta_3 w_{t-1} \end{aligned}$$

In this case there are an infinite number of order-one autoregressive representations.

A special case of singularity

We can prove the following result:
Same assumption as in Anderson and Deistler

$$\mathbf{y}_t = D(L)\mathbf{w}_t$$

Moreover, assume that $n = q + 1$. Then generically the autoregressive representation of minimum degree, as a matrix polynomial, is unique, i.e. if

$$A(L)\mathbf{y}_t = R\mathbf{w}_t, \quad \tilde{A}(L)\mathbf{y}_t = \tilde{R}\tilde{\mathbf{w}}_t$$

with $A(0) = \tilde{A}(0) = I_n$, then

$$\tilde{A}(L) = A(L), \quad \tilde{R} = RQ, \quad \tilde{\mathbf{w}}_t = Q'\mathbf{w}_t,$$

where Q is $q \times q$ and orthogonal.

A special case of singularity

$$\mathbf{y}_t = D(L)\mathbf{w}_t$$

with, generically,

$$d_{ij(L)} = \frac{e_{ij}(L)}{f_{ij}(L)},$$

degree$(e_{ij}) \leq s_1$, degree$(f_{ij}) \leq s_2$.
Then the degree of $A(L)$ as a matrix polynomial is $qs_1 + q^2 s_2$, generically.

Second assumption

We also assume that
(ii) (Second additional assumption) Current and past values of any $(q+1)$-tuple

$$(\chi_{i_1,t}\ \chi_{i_2,t}\ \cdots\ \chi_{i_{q+1},t})$$

span the same space spanned by current and past values of all the χ_{it}.
Let me illustrate this by an example. Assume $q = 1$ and

$$\begin{aligned}\chi_{1t} &= u_{t-1}\\ \chi_{2t} &= u_{t-1}\\ \chi_{jt} &= u_t, \quad \text{for } j \geq 3\end{aligned}$$

Then Assumption (ii) does not hold. Thus (ii) rules out cases like this. On the other hand, if

$$\chi_{it} = \alpha_i u_t + \beta_i u_{t-1},$$

then generically (ii) holds.

Consequences

Under (i) and (ii), if χ_t^h and χ_t^k are two $(q+1)$-dimensional selections of the χ's, then

$$A^h(L)\chi_t^h = R^h\mathbf{u}_t^h, \quad A^k(L)\chi_t^k = R^k\mathbf{u}_t^k$$

where $\mathbf{u}_t^h = Q\mathbf{u}_t^k$, with Q orthogonal.

Consequences

Assume for convenience that $n = (q+1)m$. Under Assumptions (i) and (ii) we obtain the representation

$$\begin{pmatrix} A^1(L) & 0 & \cdots & 0 \\ 0 & A^2(L) & \cdots & 0 \\ & & \ddots & \\ 0 & 0 & \cdots & A^m(L) \end{pmatrix} \begin{pmatrix} \chi_t^1 \\ \chi_t^2 \\ \vdots \\ \chi_t^m \end{pmatrix} = \begin{pmatrix} R^1 \\ R^2 \\ \vdots \\ R^m \end{pmatrix} \mathbf{v}_t$$

$$A_n(L)\mathbf{x}_{nt} = R_n \mathbf{v}_t$$

where the vectors χ_t^k are non overlapping $(q+1)$-dimensional selections, the blocks $A^k(L)$ are $(q+1) \times (q+1)$, the matrix R_n is $n \times q$ and $\mathbf{v}_t$ is a q dimensional white noise. The matrices $A^k(L)$ are minimum order.

A static version of the model

The autoregressive representation of χ_{nt} can be rewritten as

$$A_n(L)\chi_{nt} = R_n\mathbf{v}_t$$

Lastly, using $\mathbf{x}_{nt} = \chi_{nt} + \xi_{nt}$,

$$A_n(L)\mathbf{x}_{nt} = \mathbf{z}_{nt} = R_n\mathbf{v}_t + A(L)\xi_{nt}.$$

Under mild assumptions we can prove that $A_n(L)\xi_{nt}$ is idiosyncratic. Thus the original dynamic model has been transformed into a **static** one. We use the spectral density of the χ's to obtain the matrices $A_n(L)$, thus $\mathbf{z}_{nt}$. Then use the static model to estimate $\mathbf{v}_t$.

Summary of the construction in population

A. Estimation of the spectral density of the common components: FHLR (2000). Using the spectral density of the χ's we estimate the covariances of the χ's.

$$\Gamma^{\chi}_{jj,k} = \int_{-\pi}^{\pi} \Sigma^{\chi}_{jj}(\theta) e^{ik\theta} d\theta$$

Such covariances are then used to obtain

$$A^j(L)\chi^j_t = \mathbf{w}^j_t,$$

B. The "static" model

$$A_n(L)\mathbf{x}_{nt} = R_n \mathbf{v}_t + A_n(L)\boldsymbol{\xi}_{nt}$$

is then used to estimate R_n and $\mathbf{v}_t$.

Summary of the construction in population

Under our assumptions, the particular grouping into $(q+1)$-dimensional subvectors of χ_{nt} does not matter in population. However, in practice the grouping makes a difference. This problem can be solved by randomizing the groupings and averaging over the results.

Summary of the estimation procedure

(i) We start with an estimated lag-window spectral density $\hat{\Sigma}_n^x(\theta)$. We assume that

$$\hat{\Sigma}_{ij}^x(\theta) \to \Sigma_{ij}^x(\theta)$$

in probability for $T \to \infty$, uniformly in θ, i and j, with rate $1/\sqrt{\rho_T}$, where $\rho_T = T^{2/3}/\log T$. This is the price paid to non parametric estimation of the spectrum of the x's.
(ii) Then we obtain an estimator for the spectral density of the χ's, $\hat{\Sigma}_n^\chi(\theta)$ (principal components in the frequency domain), and prove that

$$\hat{\Sigma}_{ij,n}^\chi(\theta) \to \Sigma_{ij}^\chi(\theta)$$

as T and n tend to infinity, in probability at the rate $\max(n^{-1/2}, \rho_T^{-1/2})$.

Summary of the estimation procedure

(iii) Then we prove that

$$\hat{\Gamma}^{\chi}_{ij,k,n} = \int_{-\pi}^{\pi} \hat{\Sigma}^{\chi}_{ij,n}(\theta) e^{ik\theta} d\theta \to \int_{-\pi}^{\pi} \Sigma^{\chi}_{ij,n}(\theta) e^{ik\theta} d\theta = \Gamma^{\chi}_{ij,k}$$

in probability as T and n tend to infinity, at the same rate as above.
(iv) The same rate is obtained for the autoregressive matrices

$$\hat{A}^k(L) \to A^k(L)$$

Summary of the estimation procedure

(v) Lastly, we estimate the static model

$$\mathbf{z}_{nt} = R_n \mathbf{v}_t + A_n(L)\xi_{nt} \qquad (*)$$

and obtain the same rate $\max(n^{-1/2}, \rho_T^{-1/2})$ for R^k and $\mathbf{v}_t$. Though $(*)$ is estimated by standard principal components, obtaining consistency and rates has been hard because $\mathbf{z}_{nt}$ and $A_n(L)$ are estimated, not observed.

Alternative estimators

Under further assumptions on the idiosyncratic components, $\mathbf{A}^k(L)$ and $\mathbf{R}^k$ can be estimated consistently, as $T \to \infty$, by fully parametric methods. Let us return to the k-th $(q+1)$-dimensional block $\mathbf{A}^k(L)\chi_t^k = \mathbf{R}^k\mathbf{u}_t$ and the corresponding equation for $\mathbf{x}_t^k$:

$$\mathbf{A}^k(L)\mathbf{x}_t^k = \mathbf{R}^k\mathbf{u}_t + \mathbf{A}^k(L)\xi_t^k. \tag{1}$$

Under the assumption that ξ_t^k is white noise, (1) is a VARMA with equal AR and MA orders, allowing direct estimation of $\mathbf{A}^k(L)$. Direct estimation of $\mathbf{A}^k(L)$ is also possible if ξ_t^k is a vector moving average. Assuming that ξ_t^k has a VARMA structure, $\mathbf{A}^l(L)$ and $\mathbf{R}^k$ in (1) can be estimated using unobserved components model's techniques. However, VARMA and unobserved components models consistently estimate $\mathbf{A}^k(L)$, $\mathbf{A}^k(L)$ and $\mathbf{R}^k$, respectively, as $T \to \infty$, but not $\mathbf{u}_t$. Consistent estimation of $\mathbf{u}_t$ requires that both T and n diverge.

Alternative estimators

Altissimo et al. (2009) estimate α_i in model

$$x_{it} = \frac{1}{1 - \alpha L} u_t + \xi_{it}$$

equation by equation, using

$$(1 - \alpha_i L) x_{it} = a_i u_t + (1 - \alpha_i L) \xi_{it}.$$

In this particular case estimating q-dimensional instead of $(q+1)$-dimensional blocks is correct because u_t is fundamental for χ_{it} for all i. However, if $\chi_{it} = [c_i(L)/d_i(L)] u_t$, u_t is generically fundamental for 2-dimensional but not 1-dimensional blocks, nor has $c_i(L)/d_i(L)$ a finite inverse.

The dataset

We use two macroeconomic panels. The first is the one used in Forni and Gambetti (2010a), with 101 US quarterly series, covering the period 1959 I - 2007 IV. The second is the one used in Forni and Gambetti (2010b), which is essentially an updating of the panel used in Stock and Watson (2002a, 2002b) and Bernanke, Boivin and Eliasz (2005). It includes 112 US monthly series between March 1973 and November 2007. Details on both panels and their treatment are reported in Appendix F (see the paper).

Design

(i) We run the FGLR and FHLZ methods on both the quartely and the monthly panels.
(ii) Denote by $(\text{IRF, IC})_{\text{FGLR}}$ and $(\text{IRF, IC})_{\text{FHLZ}}$ the impulse-response functions and idiosyncratic components estimated via FGLR and FHLZ, respectively. The IRF's are obtained using a recursive identifications scheme.
(iii) Based on $(\text{IRF, IC})_{\text{FGLR}}$, we generate 500 artificial quarterly panels and 200 artificial monthly panels as follows. Firstly, we produce 4 random independent standard normal shocks, filter them with the impulse-response functions, and add the resulting series to get the common components.
(iv) Then we add artificial idiosyncratic components obtained by block bootstrapping (without overlapping) the idiosyncratic components estimated as in (4) above. We take blocks of 19 periods for quarterly data and 51 periods for monthly data.

Design

(v) The same procedure is applied to obtain the 500 quarterly and 200 monthly panels based on $(\text{IRF}, \text{IC})_{\text{FHLZ}}$.
(vi) Lastly, impulse-response functions and shocks are estimated for each artificial panel using the two competing methods, with the recursive identification scheme used in the estimation of the IRF's.
(vii) estimation error of the impulse response functions is measured by

$$\text{MSE}(\text{irf}) = \frac{\sum_{i=1}^{n} \sum_{f=1}^{q} \sum_{h=0}^{H} \left(\hat{b}_{if,h} - b_{if,h}\right)^2}{\sum_{i=1}^{n} \sum_{f=1}^{q} \sum_{h=0}^{H} b_{if,h}^2}.$$

The truncation lag H was set to 20 for quarterly data, to 48 for monthly data.

Results

Table 1: Average and standard deviation (slanted) of MSE across 500 artificial data set. DGP: FHLZ. Data generating Process (DGP): FGLR. Quarterly data.

irf	$p = 1$	$p = 2$	$p = 3$	$p = 6$	$p = BIC$	$p = HQC$
FHLZ	0.2494	0.1915	**0.1857**	0.2447	0.2040	0.1928
	0.0256	*0.0274*	*0.0281*	*0.0388*	*0.0266*	*0.0281*
FGLR r = 6	0.2468	0.2030	0.2276	0.2937	0.2288	0.2070
	0.0490	*0.0628*	*0.0714*	*0.0699*	*0.0604*	*0.0604*
FGLR r = 12	0.2137	**0.1862**	0.2163	0.3302	0.2137	0.1959
	0.0298	*0.0321*	*0.0349*	*0.0445*	*0.0298*	*0.0360*
FGLR r = IC2	0.2305	0.1931	0.2190	0.3095	0.2226	0.2160
	0.0369	*0.0476*	*0.0518*	*0.0764*	*0.0663*	*0.0824*

Results

Table 2: Average and standard deviation (slanted) of MSE across 500 artificial data set. DGP: FHLZ. Quarterly data.

irf	$p = 1$	$p = 2$	$p = 3$	$p = 6$	$p = BIC$	$p = HQC$
FHLZ	0.1401	0.1186	0.1287	0.1740	**0.1184**	0.1280
	0.0179	*0.0182*	*0.0184*	*0.0232*	*0.0178*	*0.0193*
FGLR r = 6	0.1651	0.1665	0.1894	0.2659	0.1651	0.1660
	0.0204	*0.0232*	*0.0261*	*0.0325*	*0.0204*	*0.0210*
FGLR r = 12	0.1494	0.1631	0.1951	0.3149	**0.1494**	0.1546
	0.0205	*0.0239*	*0.0271*	*0.0344*	*0.0205*	*0.0350*
FGLR r = IC2	0.1585	0.1657	0.1932	0.2914	0.1764	0.1862
	0.0200	*0.0235*	*0.0300*	*0.0624*	*0.0732*	*0.0858*

Results

Table 3: Average and standard deviation (slanted) of MSE across 200 artificial data set. DGP: FGLR. Monthly data.

irf	$p = 2$	$p = 4$	$p = 6$	$p = 12$	$p = BIC$	$p = HQC$
FHLZ	0.3003	0.2768	0.2797	0.3133	0.3012	**0.2760**
	0.0435	*0.0383*	*0.0386*	*0.0338*	*0.0400*	*0.0364*
FGLR r = 8	0.2435	0.2417	0.2606	0.3274	0.2603	0.2408
	0.0919	*0.0882*	*0.0914*	*0.0916*	*0.0955*	*0.0954*
FGLR r = 16	**0.2156**	0.2325	0.2649	0.3962	0.2562	0.2320
	0.0799	*0.0837*	*0.0833*	*0.0795*	*0.0905*	*0.0893*
FGLR r = IC2	0.2273	0.2286	0.2523	0.3412	0.2632	0.2417
	0.0820	*0.0723*	*0.0765*	*0.0893*	*0.0947*	*0.1006*

Results

Table 4: Average and standard deviation (slanted) of MSE across 200 artificial data set. DGP: FHLZ. Monthly data.

irf	$p = 2$	$p = 4$	$p = 6$	$p = 12$	$p = BIC$	$p = HQC$
FHLZ	0.1226	0.1292	0.1394	0.1832	<u>0.1228</u>	**<u>0.1220</u>**
	0.0250	*0.0243*	*0.0237*	*0.0219*	*0.0236*	*0.0214*
FGLR r = 8	0.2890	0.3131	0.3423	0.4040	0.3018	0.2949
	0.1064	*0.0980*	*0.0978*	*0.0887*	*0.1129*	*0.1104*
FGLR r = 16	**0.2564**	0.2780	0.3148	0.4448	0.2619	0.2679
	0.0951	*0.0842*	*0.0823*	*0.0718*	*0.1067*	*0.1134*
FGLR r = IC2	0.2652	0.2872	0.3208	0.4134	<u>0.2826</u>	<u>0.2928</u>
	0.0906	*0.0912*	*0.0923*	*0.0849*	*0.0981*	*0.1178*