Dynamic specification tests for dynamic factor models

Gabriele Fiorentini
<fiorentini@disia.unifi.it>
Università di Firenze
RCEA

Enrique Sentana
<sentana@cemfi.es>
CEMFI

Large-Scale Factor Models in Finance Conference (SoFiE)
Universitá della Svizzera Italiana, Lugano
October 11-12, 2013

Outline of the presentation

Introduction and motivation

- Dynamic factor models have been used in macroeconomics and finance for over 35 years as a way of capturing the cross-sectional and dynamic correlations between multiple series in a parsimonious way.
- The parameters are estimated by maximising the likelihood function obtained as a by-product of the Kalman filter prediction equations or from Whittle's (1962) frequency domain asymptotic approximation.
- The latent factors are filtered with the Kalman smoother or its Wiener-Kolmogorov counterpart.
- Many important modelling issues may arise in practice, such as the number of factors or the identification of their effects.
- Another non-trivial issue is the specification of the dynamics of common and idiosyncratic factors.
- We propose LM-based specification tests for neglected serial correlation in those factors that are simple to implement and interpret.
- Once a preferred model has been specified and estimated, our score tests can be computed from simple statistics of the estimated factors.

Exact dynamic factor model with a single common factor

$$\begin{pmatrix} y_{1,t} \\ \vdots \\ y_{N,t} \end{pmatrix} = \begin{pmatrix} \pi_1 \\ \vdots \\ \pi_N \end{pmatrix} + \begin{pmatrix} c_{1,0} \\ \vdots \\ c_{N,0} \end{pmatrix} x_t + \begin{pmatrix} c_{1,1} \\ \vdots \\ c_{N,1} \end{pmatrix} x_{t-1} + \begin{pmatrix} u_{1,t} \\ \vdots \\ u_{N,t} \end{pmatrix},$$

$$\alpha_x(L)x_t = \beta_x(L)f_t, \quad \alpha_{u_i}(L)u_{i,t} = \beta_{u_i}(L)v_{i,t},$$

$$(f_t, v_{1,t}, \ldots, v_{N,t})|I_{t-1}; \boldsymbol{\theta} \sim N[0, diag(1, \gamma_1, \ldots, \gamma_N)],$$

$\alpha_x(L)$ and $\alpha_{u_i}(L)$ are polynomials of orders p_x and p_{u_i}, respectively, while $\beta_x(L)$ and $\beta_{u_i}(L)$ are polynomials of orders q_x and q_{u_i}.

- There are three different dynamic features: common factors, specific factors and loadings, which if eliminated yield static factor analysis.
- We consider hypothesis tests for $p_x = \bar{p}_x$ vs $p_x = \bar{p}_x + d_x$ or $p_{u_i} = \bar{p}_{u_i}$ vs $p_{u_i} = \bar{p}_{u_i} + d_{u_i}$, or the analogous hypotheses for q_x and q_{u_i}.
- We assume the cross-sectional dimension, N, is fixed and let the time series dimension, T, increase without bound.

Other models

- Structural Time Series and UCARIMA models

$$y_t = T_t + C_t + S_t + I_t$$

Test for neglected serial correlation in the unobserved components.

- Linear, time-invariant state space models

$$\begin{aligned} \mathbf{y}_t &= \boldsymbol{\pi} + \mathbf{C}(\boldsymbol{\theta})\mathbf{x}_t, \\ \mathbf{x}_t &= \mathbf{A}(\boldsymbol{\theta})\mathbf{x}_{t-1} + \mathbf{B}(\boldsymbol{\theta})\mathbf{u}_t, \\ \mathbf{u}_t | I_{t-1}; \boldsymbol{\pi}, \boldsymbol{\theta} &\sim N[\mathbf{0}, \boldsymbol{\Omega}(\boldsymbol{\theta})]. \end{aligned}$$

Test for neglected serial correlation in $\mathbf{u}_t$ or some of its elements.

Related literature on specification testing in these models

- Engle and Watson (1980, *Cahiers du Séminaire d'Économétrie*): LM, time-domain.
- Geweke and Singleton (1981, *International Economic Review*): LR and Wald, frequency-domain.
- Harvey (1989, *Cambridge University Press*): LM, LR and Wald, time- and frequency- domains.
- Fernández (1990, *Journal of Time Series Analysis*): LM, frequency-domain.
- Fiorentini and Sentana (2012, *CEMFI WP 1211*): LM, time-domain.

Example from FS (2012): White noise vs. AR(1)

- Our baseline model is the static factor model

$$\mathbf{y}_t = \boldsymbol{\pi} + \mathbf{c}x_t + \mathbf{u}_t, \qquad x_t = f_t, \qquad \mathbf{u}_t = \mathbf{v}_t,$$

$$\left(\begin{array}{c} f_t \\ \mathbf{v}_t \end{array}\right) | I_{t-1}, \boldsymbol{\theta}_s \sim N\left[\left(\begin{array}{c} 0 \\ \mathbf{0} \end{array}\right), \left(\begin{array}{cc} 1 & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Gamma} \end{array}\right)\right],$$

  which remains popular in empirical finance (except in term structure applications).

- The Kalman smoother yields the same factor estimates as the Kalman filter updating equations, which have simple closed form expressions:

$$\begin{aligned} f_{t|T}^K &= f_{t|t}^K = \mathbf{c}'\boldsymbol{\Sigma}^{-1}(\mathbf{y}_t - \boldsymbol{\pi}) = \frac{\mathbf{c}'\boldsymbol{\Gamma}^{-1}}{1 + \mathbf{c}'\boldsymbol{\Gamma}^{-1}\mathbf{c}}(\mathbf{y}_t - \boldsymbol{\pi}), \\ \mathbf{v}_{t|T}^K &= \mathbf{v}_{t|t}^K = \boldsymbol{\Gamma}\boldsymbol{\Sigma}^{-1}(\mathbf{y}_t - \boldsymbol{\pi}) = \mathbf{y}_t - \boldsymbol{\pi} - \mathbf{c}f_{t|t}^K. \end{aligned}$$

Example from FS (2012): White noise vs. AR(1)

- A potentially interesting alternative would be:

$$\mathbf{y}_t = \boldsymbol{\pi} + \mathbf{c}x_t + \mathbf{u}_t, \quad \mathbf{u}_t = \mathbf{v}_t,$$
$$x_t = \psi x_{t-1} + f_t.$$

- It reduces to the static specification under the null $H_0 : \psi = 0$.
- Otherwise, it has the autocorrelation structure of a $\text{V}\small{\text{ARMA}}$(1,1).
- The average score w.r.t. ψ under H_0 is

$$\bar{\mathbf{s}}_{\psi T} = \frac{1}{T}\sum_{t=2}^{T} f_{t|T}^{K} f_{t-1|T}^{K},$$

which is analogous to the score that we would use to test for first order serial correlation in f_t if we could observe the latent factors.
- However its asymptotic variance $[\mathbf{c}'(\mathbf{cc}'+\boldsymbol{\Gamma})^{-1}\mathbf{c}]^2 < 1$ reflects the unobservability of the factors.

Testing ARMA(p,q) vs ARMA(p+d,q) (or ARMA(p,q+d))

- When we move to testing say AR(1) vs AR(2) in the unobservable factors, the model is already dynamic under the null and the Kalman filter and smoother equations no longer coincide.
- More importantly, those equations are recursive and therefore difficult to characterise without solving a multivariate Riccati equation.
- Although a Lagrange Multiplier test of the new null hypothesis in the time domain is conceptually straightforward, the algebra is incredibly tedious and the recursive scores difficult to interpret.
- An alternative way to characterise a dynamic factor model is in the frequency domain.
- As we shall see, the frequency domain scores remain remarkably simple, since they closely resemble the scores of the static model.

Maximum likelihood in the frequency domain

- We assume that the observed series are covariance stationary, at least after a suitable transformation.
- Under stationarity, the spectral density matrix is proportional to

$$\mathbf{G_{yy}}(\lambda) = \mathbf{c}(e^{-i\lambda})G_{xx}(\lambda)\mathbf{c}'(e^{i\lambda}) + \mathbf{G_{uu}}(\lambda),$$

$$G_{xx}(\lambda) = \frac{\beta_x(e^{-i\lambda})\beta_x(e^{i\lambda})}{\alpha_x(e^{-i\lambda})\alpha_x(e^{i\lambda})},$$

$$\mathbf{G_{uu}}(\lambda) = diag[G_{u_1u_1}(\lambda), \ldots, G_{u_Nu_N}(\lambda)],$$

$$G_{u_iu_i}(\lambda) = \gamma_i \frac{\beta_{u_i}(e^{-i\lambda})\beta_{u_i}(e^{i\lambda})}{\alpha_{u_i}(e^{-i\lambda})\alpha_{u_i}(e^{i\lambda})},$$

which inherits the exact single factor structure of the unconditional covariance matrix of a static factor model.

Maximum likelihood in the frequency domain

- Let
$$\mathbf{I_{yy}}(\lambda) = \frac{1}{2\pi T}\sum_{t=1}^{T}\sum_{s=1}^{T}(\mathbf{y}_t - \boldsymbol{\pi})(\mathbf{y}_s - \boldsymbol{\pi})' e^{-i(t-s)\lambda}$$
denote the periodogram matrix and $\lambda_j = 2\pi j/T$ $(j = 0, \dots T-1)$ the usual Fourier frequencies.
- The discrete version of the (spectral) log-likelihood function is
$$-\frac{NT}{2}\ln(2\pi) - \frac{1}{2}\sum_{j=0}^{T-1}\ln|\mathbf{G_{yy}}(\lambda_j)| - \frac{1}{2}\sum_{j=0}^{T-1}\mathrm{tr}\left\{\mathbf{G_{yy}^{-1}}(\lambda_j)[2\pi\mathbf{I_{yy}}(\lambda_j)]\right\}.$$
- The continuous version replaces sums by integrals.
- Computations can be considerably speeded up by exploiting that
$$\mathbf{G_{yy}^{-1}}(\lambda) = \mathbf{G_{uu}^{-1}}(\lambda) - \omega(\lambda)\mathbf{G_{uu}^{-1}}(\lambda)\mathbf{c}(e^{-i\lambda t})\mathbf{c}'(e^{i\lambda t})\mathbf{G_{uu}^{-1}}(\lambda),$$
$$\omega(\lambda) = [G_{xx}^{-1}(\lambda) + \mathbf{c}'(e^{i\lambda t})\mathbf{G_{uu}^{-1}}(\lambda)\mathbf{c}(e^{-i\lambda t})]^{-1}.$$
- The MLE of $\boldsymbol{\pi}$, which only enters through $\mathbf{I_{yy}}(\lambda)$, is the sample mean.

Maximum likelihood in the frequency domain

- The score w.r.t. all the remaining parameters is

$$\mathbf{d}(\boldsymbol{\theta}) = \frac{1}{2}\sum_{j=0}^{T-1} \frac{\partial vec'[\mathbf{G}_{\mathbf{yy}}(\lambda_j)]}{\partial \boldsymbol{\theta}} \mathbf{M}(\lambda_j)\mathbf{m}(\lambda_j),$$

$$\mathbf{m}(\lambda) = vec\left[2\pi \mathbf{I}'_{\mathbf{yy}}(\lambda) - \mathbf{G}'_{\mathbf{yy}}(\lambda)\right],$$

$$\mathbf{M}(\lambda) = \left[\mathbf{G}^{-1}_{\mathbf{yy}}(\lambda) \otimes \mathbf{G}'^{-1}_{\mathbf{yy}}(\lambda)\right].$$

- Using * to denote conjugate transpose, the information matrix is

$$\mathbf{Q} = \frac{1}{4\pi}\int_{-\pi}^{\pi} \frac{\partial vec'[\mathbf{G}_{\mathbf{yy}}(\lambda)]}{\partial \boldsymbol{\theta}} \mathbf{M}(\lambda) \left[\frac{\partial vec[\mathbf{G}_{\mathbf{yy}}(\lambda)]}{\partial \boldsymbol{\theta}'} d\lambda\right]^*.$$

- Consistent estimators will be provided either by the outer product of the score or by

$$\boldsymbol{\Phi}(\boldsymbol{\theta}) = \frac{1}{2}\sum_{j=0}^{T-1} \frac{\partial vec'[\mathbf{G}_{\mathbf{yy}}(\lambda_j)]}{\partial \boldsymbol{\theta}} \mathbf{M}(\lambda_j) \left[\frac{\partial vec[\mathbf{G}_{\mathbf{yy}}(\lambda_j)]}{\partial \boldsymbol{\theta}'}\right]^*.$$

The (Kalman-)Wiener-Kolmogorov filter

- By working in the frequency domain we can easily obtain smoothed estimators of the latent variables too.
- Specifically, let

$$\begin{aligned} \mathbf{y}_t - \boldsymbol{\pi} &= \int_{-\pi}^{\pi} e^{i\lambda t} d\mathbf{Z}_{\mathbf{y}}(\lambda), \\ V[d\mathbf{Z}_{\mathbf{y}}(\lambda)] &= \mathbf{G}_{\mathbf{yy}}(\lambda) d\lambda \end{aligned}$$

  denote Cramer's spectral decomposition of the observed process (Wold's decomposition frequency-domain analogue).
- The Wiener-Kolmogorov two-sided filter for the common factor x_t at each frequency is given by

$$\mathbf{c}'(e^{i\lambda}) G_{xx}(\lambda) \mathbf{G}_{\mathbf{yy}}^{-1}(\lambda) d\mathbf{Z}_{\mathbf{y}}(\lambda)$$

  so that the spectral density of the smoother $x_{t|T}^{K}$ as $T \to \infty$ is

$$G_{xx}^2(\lambda)\mathbf{c}'(e^{i\lambda})\mathbf{G}_{\mathbf{yy}}^{-1}(\lambda)\mathbf{c}(e^{-i\lambda}) = \omega(\lambda) G_{xx}(\lambda)\mathbf{c}'(e^{i\lambda})\mathbf{G}_{\mathbf{uu}}^{-1}(\lambda)\mathbf{c}(e^{-i\lambda}).$$

The (Kalman-)Wiener-Kolmogorov filter

- Hence, the spectral density of the final estimation error $x_t - x^K_{t|T}$ will be given by

$$G_{xx}(\lambda) - \mathbf{c}'(e^{i\lambda})\mathbf{G}^{-1}_{\mathbf{yy}}(\lambda_j)\mathbf{c}(e^{i\lambda}) = \omega(\lambda).$$

- Having obtained these, we can easily obtain the smoother for f_t, $f^K_{t|T}$, by applying to $x^K_{t|T}$ the one-sided filter

$$\alpha_x(e^{-i\lambda})/\beta_x(e^{-i\lambda})$$

- Likewise, we can derive its spectral density, as well as the spectral density of its final estimaton error $f_t - f^K_{t|T}$.
- Finally, we can obtain the autocovariances of $x^K_{t|T}$, $f^K_{t|T}$ and their final estimation errors by applying the usual Fourier transformation

$$cov(z_t, z_{t-k}) = \int_{-\pi}^{\pi} e^{i\lambda k} G_{zz}(\lambda) d\lambda.$$

The minimal sufficient statistics for $\{x_t\}$ ▸ skip

- Define $x^G_{t|T}$ as the spectral GLS estimator of x_t through the transformation

$$[\mathbf{c}'(e^{i\lambda})\mathbf{G}^{-1}_{\mathbf{uu}}(\lambda)\mathbf{c}(e^{-i\lambda})]^{-1}\mathbf{c}'(e^{i\lambda})\mathbf{G}^{-1}_{\mathbf{uu}}(\lambda)d\mathbf{Z}_{\mathbf{y}}(\lambda).$$

- Similarly, define $\mathbf{u}^G_{t|T}$ though

$$\{\mathbf{I}_N - [\mathbf{c}'(e^{i\lambda})\mathbf{G}^{-1}_{\mathbf{uu}}(\lambda)\mathbf{c}(e^{-i\lambda})]^{-1}\mathbf{c}'(e^{i\lambda})\mathbf{G}^{-1}_{\mathbf{uu}}(\lambda)\}d\mathbf{Z}_{\mathbf{y}}(\lambda).$$

- It is then easy to see that the joint spectral density of $x^G_{t|T}$ and $\mathbf{u}^G_{t|T}$ will be block-diagonal, with the (1,1) element being

$$G_{xx}(\lambda) + [\mathbf{c}'(e^{i\lambda})\mathbf{G}^{-1}_{\mathbf{uu}}(\lambda)\mathbf{c}(e^{-i\lambda})]^{-1}$$

and the (2,2) block

$$\mathbf{G}_{\mathbf{yy}}(\lambda) - \mathbf{c}(e^{-i\lambda})[\mathbf{c}'(e^{i\lambda})\mathbf{G}^{-1}_{\mathbf{uu}}(\lambda)\mathbf{c}(e^{-i\lambda})]^{-1}\mathbf{c}'(e^{i\lambda}),$$

whose rank is $N-1$.

The minimal sufficient statistics for $\{x_t\}$

- This block-diagonality allows us to factorise the spectral log-likelihood function of $\mathbf{y}_t$ as the sum of the log-likelihood function of $x^G_{t|T}$, which is univariate, and the log-likelihood function of $\mathbf{u}^G_{t|T}$.
- Importantly, the parameters characterising $G_{xx}(\lambda)$ only enter through the first component.
- In contrast, the remaining parameters affect both components.
- Moreover, we can easily show that
  1. $x^G_{t|T} = x_t + \zeta^G_{t|T}$, with x_t and $\zeta^G_{t|T}$ orthogonal at all leads and lags.
  2. The smoothed estimator of x_t obtained by applying the Wiener-Kolmogorov filter to $x^G_{t|T}$ coincides with $x^K_{t|T}$.
- This confirms that $x^G_{t|T}$ constitute minimal sufficient statistics for x_t.
- In addition, the degree of unobservability of x_t depends exclusively on the size of $[\mathbf{c}'(e^{i\lambda})\mathbf{G}^{-1}_{\mathbf{uu}}(\lambda)\mathbf{c}(e^{-i\lambda})]^{-1}$ relative to $G_{xx}(\lambda)$.

AR(1) vs AR(2) for observable x_t

- Although all the previous calculations are straightforward, they might seem daunting unless one is familiar with spectral methods.
- Fortunately, they have remarkably simple time domain counterparts.
- For pedagogical purposes, let us initially assume that x_t is observable.
- The model under the alternative is

$$(1-\psi_{x1}L)(1-\alpha_{x1}L)x_t = f_t.$$

- Therefore, the null is $H_0 : \psi_{x1} = 0$.
- This is a multiplicative alternative.
- Instead, we could test $H_0 : \alpha_{x2} = 0$ in the additive alternative

$$(1-\alpha_{x1}L-\alpha_{x2}L^2)x_t = f_t.$$

- In that case, it would be more convenient to reparametrise the model in terms of partial autocorrelations as $\alpha_{12} = \alpha_1/\left(1-\alpha_2\right), \alpha_{22} = \alpha_2$.
- We stick to multiplicative alternatives, which cover MA terms too.

AR(1) vs AR(2) for observable x_t

- Under the alternative, the spectral density of x_t is

$$\frac{\sigma_f^2}{(1-\alpha_{x1}e^{-i\lambda})(1-\alpha_{x1}e^{i\lambda})}\frac{1}{(1-\psi_{x1}e^{-i\lambda})(1-\psi_{x1}e^{i\lambda})}.$$

- The derivative of $G_{xx}(\lambda)$ w.r.t. ψ_{x1} evaluated at the null is

$$\frac{\partial G_{xx}(\lambda)}{\partial \psi_{x1}} = 2(e^{-i\lambda}+e^{i\lambda})\frac{\sigma_f^2}{(1-\alpha_{x1}e^{-i\lambda})(1-\alpha_{x1}e^{i\lambda})} = 2\cos\lambda G_{xx}(\lambda).$$

- Hence the spectral version of the score w.r.t. ψ_{x1} under H_0 is

$$\sum_{j=0}^{T-1}\cos\lambda_j G_{xx}^{-1}(\lambda_j)[2\pi I_{xx}(\lambda_j)-G_{xx}(\lambda_j)] = \sum_{j=0}^{T-1}\cos\lambda_j[2\pi I_{ff}(\lambda_j)],$$

  where we have exploited the fact that

$$\sum_{j=0}^{T-1}\frac{\partial G_{xx}(\lambda_j)}{\partial \psi_{x1}}G_{xx}^{-1}(\lambda_j) = \sum_{j=0}^{T-1}\cos\lambda_j = 0.$$

AR(1) vs AR(2) for observable x_t

- Given that

$$I_{ff}(\lambda_j) = \hat{\gamma}_{ff}(0) + 2\sum_{k=1}^{T-1} \hat{\gamma}_{ff}(k)\cos(k\lambda_j),$$

the spectral version of the score becomes

$$\sum_{j=0}^{T-1} \cos\lambda_j[2\pi I_{ff}(\lambda_j)] = T[\hat{\gamma}_{ff}(1) + \hat{\gamma}_{ff}(T-1)].$$

- In turn, the time domain version of the score will be

$$\sum_t (x_t - \alpha_{x1}x_{t-1})(x_{t-1} - \alpha_{x1}x_{t-2}) = \sum_t f_t f_{t-1},$$

which is (almost) identical as $\hat{\gamma}_{ff}(T-1) = T^{-1}x_T x_1$.

- Therefore, the spectral test is simply checking that the first sample autocorrelation of f_t coincides with its theoretical value under H_0.

AR(1) vs AR(2) in **common factor**

- After some straightforward algebraic manipulations, we can show that under the null of $H_0 : \psi_{x1} = 0$ this score can be written as

$$\sum_{j=0}^{T-1} cos(\lambda_j) G_{xx}^{-1}(\lambda_j)[2\pi I_{x^K x^K}(\lambda_j) - G_{x^K x^K}(\lambda_j)]$$
$$= \sum_{j=0}^{T-1} cos(\lambda_j)[2\pi I_{f^K f^K}(\lambda_j) - G_{f^K f^K}(\lambda_j)].$$

- Once again, the time domain counterpart to the spectral score w.r.t. ψ_{x1} is (asymptotically) proportional to the difference between the first sample autocovariance of $f_{t|T}^K$ and its theoretical counterpart under H_0.
- Therefore, the only difference with the observable case is that the autocovariance of $f_{t|T}^K$, which is a forward filter of the Wold innovations of $\mathbf{y}_t$, is no longer 0 when $\psi_{x1} = 0$, although it approaches 0 as the signal to noise ratio increases.

AR(1) vs AR(2) in common factor: A simple example

- Imagine that the model under the alternative is:

$$\mathbf{y}_t = \boldsymbol{\pi} + \mathbf{c}x_t + \mathbf{u}_t, \quad \mathbf{u}_t = \mathbf{v}_t,$$
$$(1-\psi_{x1}L)(1-\alpha_{x1}L)x_t = f_t,$$
$$\left(\begin{array}{c} f_t \\ \mathbf{v}_t \end{array}\right)|I_{t-1}, \boldsymbol{\theta} \sim N\left[\left(\begin{array}{c} 0 \\ \mathbf{0} \end{array}\right), \left(\begin{array}{cc} 1 & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Gamma} \end{array}\right)\right].$$

- Straighforward algebra shows that if $\psi_{x1} = 0$ then $x_{t|T}^K$ will have the autocorrelation structure of an $\textsc{Ar}(2)$, while $f_{t|T}^K$ will follow an $\textsc{Ar}(1)$ with first order autocovariance $(\mathbf{c}'\boldsymbol{\Gamma}^{-1}\mathbf{c})\alpha_{x1}/(1-\alpha_{f^K}^2)$, where α_{f^K} is

$$\frac{1+\alpha_{x1}^2+(\mathbf{c}'\boldsymbol{\Gamma}^{-1}\mathbf{c}) - \sqrt{[(1+\alpha_{x1})^2+(\mathbf{c}'\boldsymbol{\Gamma}^{-1}\mathbf{c})]\,[(1-\alpha_{x1})^2+(\mathbf{c}'\boldsymbol{\Gamma}^{-1}\mathbf{c})]}}{2\alpha_{x1}}$$

- The larger $(\mathbf{c}'\boldsymbol{\Gamma}^{-1}\mathbf{c})$ is, the closer this autocovariance will be to 0.
- The LM test of $H_0 : \psi_{x1} = 0$ will simply compare the first sample autocovariance of $f_{t|T}^K$ with its theoretical value above.

AR(1) vs AR(2) in specific factors

- Let $\boldsymbol{\psi}'_{\mathbf{u}1} = (\psi_{u_1 1}, \ldots, \psi_{u_N 1})$.
- In this case we have that

$$\frac{\partial vec[\mathbf{G_{yy}}(\lambda)]}{\partial \boldsymbol{\psi}'_{\mathbf{u}_1}} = \mathbf{E}_N \frac{\partial vecd[\mathbf{G_{uu}}(\lambda)]}{\partial \boldsymbol{\psi}'_{\mathbf{u}_1}},$$

  where $\mathbf{E}_N$ is the "diagonalisation" matrix that maps $vecd$ into vec.
- Straightforward algebraic manipulations allow us to write the score w.r.t. $\psi_{u_i 1}$ under the null of $H_0 : \boldsymbol{\psi}_{\mathbf{u}_1} = \mathbf{0}$ as

$$\sum_{j=0}^{T-1} cos(\lambda_j) G^{-1}_{u_i u_i}(\lambda_j)[2\pi I_{u_i^K u_i^K}(\lambda_j) - G_{u_i^K u_i^K}(\lambda_j)]$$
$$= \sum_{j=0}^{T-1} cos(\lambda_j)[2\pi I_{v_i^K v_i^K}(\lambda_j) - G_{v_i^K v_i^K}(\lambda_j)].$$

- Thus, the time domain counterpart to the spectral score w.r.t. $\psi_{u_i 1}$ will again be proportional for large T to the difference between the first sample autocovariance of $v^K_{it|T}$ and its theoretical value under H_0.
- Joint tests can be easily obtained by combining the scores involved.

Parameter uncertainty

- So far we have implicitly assumed known model parameters, but in practice some of them will have to be estimated under the null.
- Maximum likelihood estimation of the dynamic factor model parameters can be done either in the time domain using the Kalman filter or in the frequency domain.
- The sampling uncertainty surrounding $\boldsymbol{\pi}$ is asymptotically inconsequential because the information matrix is block diagonal.
- The sampling uncertainty surrounding the other parameters is not necessarily so.
- The solution is the standard one: replace the inverse of the $(\boldsymbol{\psi}, \boldsymbol{\psi})$ block of the information matrix by the $(\boldsymbol{\psi}, \boldsymbol{\psi})$ block of the inverse information matrix in the quadratic form that defines the LM test.
- We provide computationally efficient expressions for the entire information matrix.
- We also discuss some special cases in which it is block diagonal.

Monte Carlo simulation: Size experiment ▸ skip

- To evaluate possible finite sample size distorsions, we generate 10,000 samples of length 500 from the following heterogeneous dynamic factor model:

$$\begin{bmatrix} y_{1,t} \\ y_{2,t} \\ y_{3,t} \end{bmatrix} = \begin{bmatrix} .1 \\ .1 \\ .1 \end{bmatrix} + \begin{bmatrix} 0.7 \\ 0.5 \\ 0.4 \end{bmatrix} x_t + \begin{bmatrix} u_{1,t} \\ u_{2,t} \\ u_{3,t} \end{bmatrix},$$

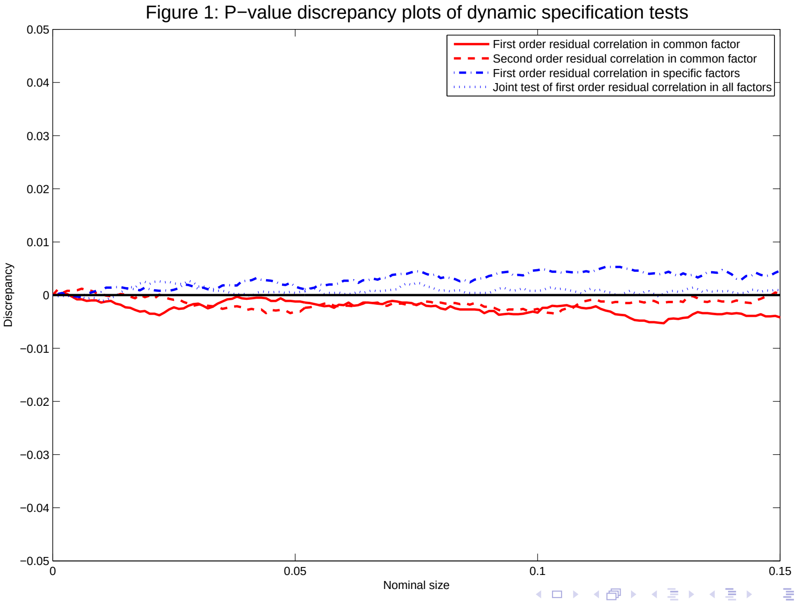
$$(1 - .4L - .2L^2)x_t = f_t,$$

$$(1 + .4L)u_{1,t} = v_{1,t}, \quad (1 - .6L)u_{2,t} = v_{2,t}, \quad (1 - .2L)u_{3,t} = v_{3,t},$$

$$V(f_t) = 1, \qquad vecd'[V(\mathbf{v}_t)] = (0.4, 0.3, 0.8).$$

- We then compute LM tests against
  1. First order residual serial correlation in the common factor (χ^2_1)
  2. Second order residual serial correlation in the common factor (χ^2_2)
  3. First order residual serial correlation in all the specific factors (χ^2_3)
  4. First order residual serial correlation in both common and specific factors (χ^2_4)

Size experiment (T=500)



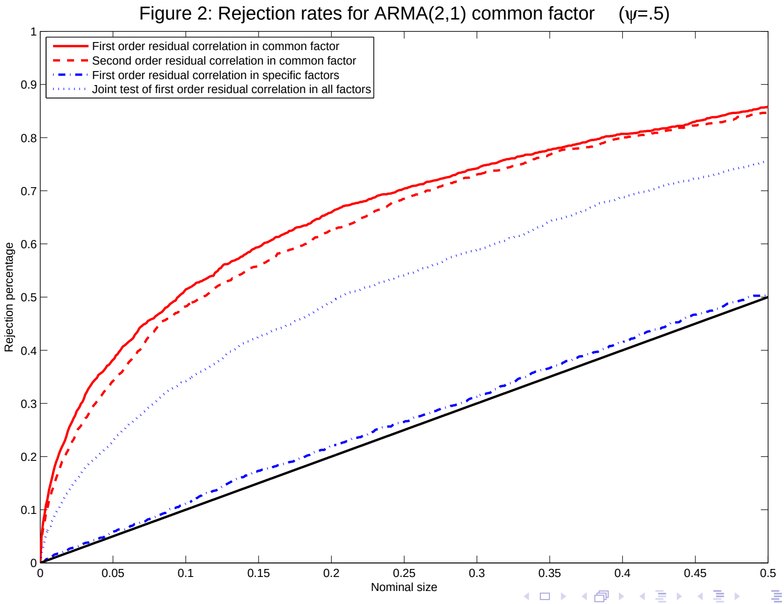
Monte Carlo simulation: Power experiment I

We first simulate and estimate 2,000 samples of length 500 of a DGP in which the MA polynomial of the common factor is $\psi_x(L) = (1 - .5L)$ while we simultaneously adjust its AR polynomial to keep the same first and second-order autocorrelations, so that

$$x_t = 0.874x_{t-1} + 0.037x_{t-2} + f_t - 0.5f_{t-1}.$$

We also rescale the loadings so as to mantain the same unconditional signal to noise ratio as under the null in order to make sure that our power results are unaffected by the degree of observability of the factors. Everything else is unchanged.

Power when common factor is misspecified



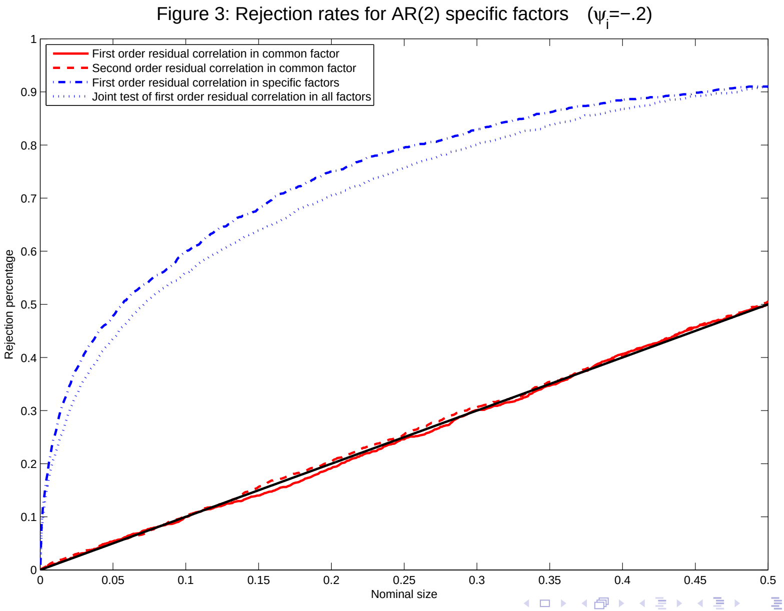
Monte Carlo simulation: Power experiment II

We also simulate and estimate 2,000 samples of length 500 of a DGP in which the AR polynomials of the specific factors are multiplied by $\psi_{u_i}(L) = (1 + .2L)$ for $i = 1, 2, 3$, while simultanously adjusted to maintain the same first-order autocorrelation, so that

$$
\begin{aligned}
u_{1,t} &= -0.418 u_{1,t-1} - 0.044 u_{1,t-2} + v_{1,t}, \\
u_{2,t} &= 0.514 u_{2,t-1} + 0.143 u_{2,t-2} + v_{2,t}, \\
u_{3,t} &= 0.185 u_{3,t-1} + 0.077 u_{3,t-2} + v_{3,t}.
\end{aligned}
$$

In this case we rescale $V(\mathbf{v}_t)$ in order to match the same unconditional signal to noise ratio as under the null but we leave everything else unchanged.

Power when specific factors are misspecified

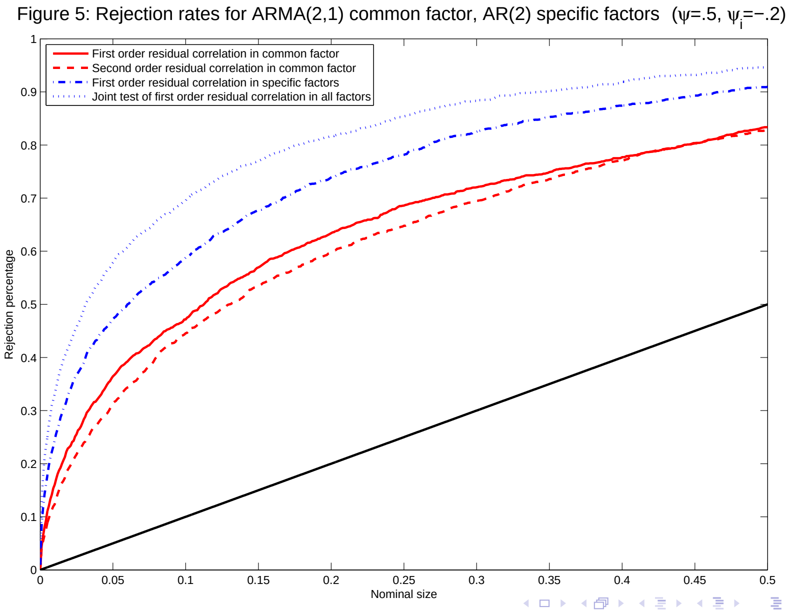


Monte Carlo simulation: Power experiment III

Finally, we simulate and estimate 2,000 samples of length 500 of a DGP that combines the features of the previous two power experiments.

Once again, the unconditional signal to noise ratio is preserved.

Power when all factors are misspecified



Empirical application

- In theory, the expenditure (GDP) and income (GDI) measures of aggregate (real) production should be equal.
- In practice, they differ because they rely on different sources.
- Their difference, known as the "statistical discrepancy", was regarded by some macroeconomists as a curiosity in the National Income and Product Accounts.
- However, the Great Recession has substantially renewed interest in the possibility of obtaining more reliable GDP growth figures by combining those two measures.
- In the early days, some national statistical offices computed a simple equally weighted average.
- More sophisticated methods would give higher weights to the more precise GDP measures.
- Nowadays, the Australian Bureau of Statistics reports a single official GDP figure, and the US Department of Commerce Bureau of Economic Analysis has seriously considered this possibility.

Empirical application

- Dynamic considerations matter, though, as pointed out by Smith, Weale and Satchell (1998, *REStud*).
- There at least two important reasons:
  (a) The associated measurement errors should be stationary but they may well be serially correlated.
  (b) These two GDP measures should be cointegrated with the true GDP, with cointegrating vector (1,-1).
- A single factor model with unit loadings on an $I(1)$ common factor and covariance stationary specific factors provides a natural way of capturing these features:

$$\begin{bmatrix} y_{1,t} \\ y_{2,t} \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} x_t + \begin{bmatrix} u_{1,t} \\ u_{2,t} \end{bmatrix},$$

$$(1-\alpha_x L)(\Delta x_t - \pi) = f_t,$$

$$(1-\alpha_{u_1} L)(u_{1,t}-\delta_1) = v_{1,t}, \qquad (1-\alpha_{u_2} L)(u_{2,t}-\delta_2) = v_{2,t}.$$

Empirical application

- We allow for systematic biases in the measurement errors through δ_1 and δ_2.
- The difference between those biases determines the mean of the statistical discrepancy while their levels fix the initial conditions.
- The usual assumption that the covariance matrix of the common factor innovations, f_t, and the idiosyncratic factor innovations, v_{1t} and v_{2t}, is diagonal turns out to be non-parametrically just identifying in this case (subject to "admissibility").
- Testing for neglected serial correlation in common and idiosyncratic factors is particularly relevant in this context because the contemporaneously filtered GDP series and the succesive revisions as future data becomes available will depend on the underlying ARMA parameters.
- We look at US data from 1947Q1 to 2012Q3 (263 obs).

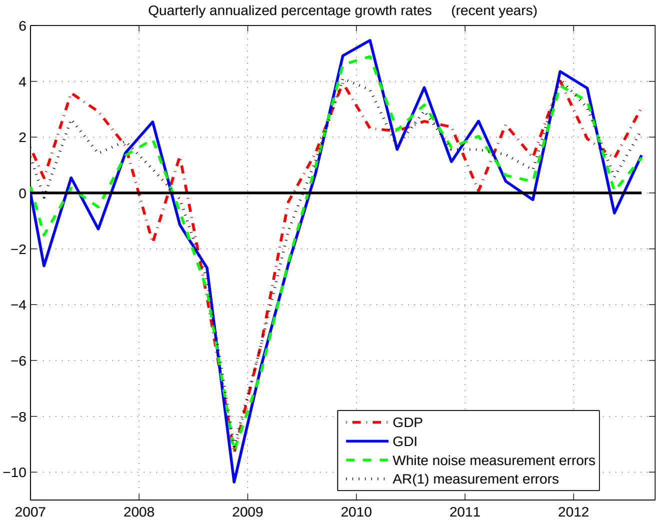
Empirical application

- The difference between (log) GDP and (log) GDI seems covariance stationary, although rather persistent. ▸ plot
- Therefore, in order to work with an invertible model whose spectral density matrix has full rank at all frequencies we combine the statistical discrepancy, $y_{2t} - y_{1t}$, and the equally weighted average of the quarterly rates of growth of GDP and GDI, $(\Delta y_{1t} + \Delta y_{2t})/2$.
- In principle, there are infinitely many other asymptotically equivalent stationarity transformations of the two output measures, but (-1,1) and (1-L)(.5,.5) seems rather natural.
- In the time domain, we can avoid this indeterminacy by computing the log-likelihood function in levels using a diffuse prior for the non-stationary component of the initial observations.

Empirical application

- We initially estimate a model with $\text{AR}(1)$ dynamics in the common factor, but white noise measurement errors.
- The estimated model parameters suggest that GDI provides a much better measure of output than GDP.
- However, LM tests against $\text{AR}(1)$ dynamics in the measurement errors massively reject their null.
- The same is true of $\text{AR}(1)$ vs $\text{AR}(2)$ dynamics in the common factor, but the test statistic is much lower.
- For that reason, we estimate the model with $\text{AR}(1)$ dynamics in both common factor and measurement errors.
- This time we no longer reject when we look at each state variable separately or all of them jointly, although there is weak evidence in favour of higher order dynamics in the idiosyncratic terms.
- Further, the parameter estimates suggest that GDP is far less noisy.
- Reassuringly, this last result does not change when we repeat the exercise allowing for $\text{AR}(2)$ dynamics in the measurement errors.

Smoothed GDP growth series



Extensions

1. Multiple common factors:
   - A notational mess, but otherwise straightforward after dealing with the usual identification issues before estimating the model under the null.
2. Overidentifying restrictions on the dynamic factor loadings
   - Straightforward too, since the corresponding scores can be related to the normal equations in a distributed lag regression of $\mathbf{y}_t$ on $x^K_{t|T}$.
3. Robustness to non-normality
   - We can always resort to the usual time-domain sandwich formulas or their unusual frequency-domain analogues.
   - But given that the serial correlation parameters ψ effectively enter through $\boldsymbol{\mu}_t$ only, where $\boldsymbol{\mu}_t$ denotes the conditional mean of $\mathbf{y}_t$ given its past alone obtained from the Kalman filter prediction equations, the information matrix equality should continue to hold for their scores.
4. Other models
   - Although we have exploited some specificities of dynamic factor models, our procedures can be easily extended to many other linear state space models.

Statistical discrepancy

▸ back