SEMIPARAMETRIC INFERENCE IN MULTIVARIATE FRACTIONALLY COINTEGRATED SYSTEMS

J. Hualde and P.M. Robinson

Universidad de Navarra and LSE

International Conference on "Breaks and Persistence in Econometrics",

11 December 2006

Semiparametric modelling is popular in cointegration analysis of $I(1)$ time series with $I(0)$ cointegrating errors.

Optimal inference on unknown cointegrating relations loses validity if VAR order is under-specified, or if the process lies outside the VAR class.

One can do as well, in first-order asymptotics, allowing the $I(0)$ inputs to have nonparametric autocorrelation.

Another source of misspecification is the $I(1)/I(0)$ framework itself.

In a fractional setting integration orders are unknown, to non-trivially generalize the $I(1)/I(0)$ assumption.

Optimal estimates have been developed in a fully parametric, and bivariate setting.

We develop partly-optimal inference on cointegrating relations in a semiparametric multivariate fractional setting, with integration orders that can be unknown, and stationary and/or nonstationary.

For a scalar or vector sequence v_t, $t = 0, \pm 1, ...$, denote

$$v_t^{\#} = v_t 1(t > 0),$$

where $1(\cdot)$ is the indicator function.

Define $\Delta = 1 - L$, where L is the lag operator, and, for real α, $\alpha \neq -1, -2, ...$,

$$\Delta^{-\alpha} = \sum_{j=0}^{\infty} a_j(\alpha) L^j, \quad a_j(\alpha) = \frac{\Gamma(j+\alpha)}{\Gamma(\alpha)\Gamma(j+1)}.$$

A scalar process $\zeta_t \sim I\,(d)$, if for any $l \times 1$ ($l < \infty$) covariance stationary process $\xi_t = (\xi_{it})$, whose spectral density matrix is continuous and nonsingular

$$\zeta_t = \sum_{k=1}^{l} \Delta^{-d_k} \xi_{kt}^{\#},$$

for $d = \max_{1 \leq k \leq l} d_k$.

A vector process is $I\,(d)$ if all its components are $I\,(d)$.

For a $p \times 1$ vector $d = (d_1, ..., d_p)'$, denote $\Delta\,(d) = diag\left\{\Delta^{d_1}, ..., \Delta^{d_p}\right\}$.

Let u_t, $t = 0, \pm 1, ...$, be a $p \times 1$ stationary unobservable process with zero mean and nonparametric spectral density matrix $f(\lambda)$,

$$E(u_0 u_j') = \int_{-\pi}^{\pi} e^{ij\lambda} f(\lambda) d\lambda,$$

that is continuous and nonsingular.

For a $p \times p$ nonsingular matrix Υ and a $p \times 1$ vector $\delta = (\delta_1, ..., \delta_p)'$, define the $p \times 1$ vector observable process z_t, $t = 0, \pm 1, ...$ by

$$\Upsilon z_t = \Delta^{-1}(\delta)\, u_t^{\#},$$

where

$$0 \leq \delta_1, \delta_2, ..., \delta_r < \delta_{r+1} = \delta_{r+2} = ... = \delta_p,$$

$$\delta_p - \delta_i \neq 1/2,\ i \in \{1, ..., r\},$$

$$\Upsilon = \begin{pmatrix} \Upsilon_1 & \Upsilon_2 \\ 0_{p-r,r} & I_{p-r} \end{pmatrix},$$

where Υ_1 is $r \times r$ upper-triangular with ones in the main diagonal and Υ_2 is $r \times (p-r)$, with all rows of $\Upsilon_1^{-1}\Upsilon_2$ non-null but Υ_1 and Υ_2 otherwise unknown and unrestricted.

$z_t \sim I(\delta_p)$.

z_t has cointegrating rank r.

Memories of the r cointegrating errors can vary, $u_{it} \sim I(\delta_i)$, $i = 1, ..., r$.

Our results are new even if $r = 1$, for any $p \geq 2$.

Given there no prior restrictions on f, those on Υ and δ ensure identifiability, and imply that (given consistent estimates of Υ and δ) consistent estimation of f is possible.

With

$$\delta_1 = \delta_2 = ... = \delta_r = 0, \;\; \delta_{r+1} = \delta_{r+2} = ... = \delta_p = 1,$$

we have the usual $I(1)/I(0)$ cointegrated system.

We could can transform our setup with upper-triangular Υ_1 to that with $\Upsilon_1 = I_r$, but this reduces the "achievement" of a cointegration analysis, since the relations would generally each have integration order $\max_{i \in \{1,...,r\}} \delta_i$.

While this transformation seems innocuous in the traditional framework where $\delta_i = 0, i = 1, ..., r$, in our fractional setting it could reduce cointegrating gaps.

We include:

stationary cointegration, $\delta_p \in (0, 1/2)$

weak cointegration, $\delta_p - \delta_i < 1/2$

strong cointegration, $\delta_p - \delta_i > 1/2$

We extend nontrivially the bivariate model to a richer multivariate framework,

allow for nonparametric f,

simultaneously cover relations of weak and strong cointegration.

Asymptotics for point estimates for Υ differ significantly across these cases, the same rules of inference prevail throughout, the same Wald test statistic (for a linear hypothesis) having a null limit χ^2 distribution.

We treat the nonparametric autocorrelation in the frequency domain.

We use a ratio of weighted periodogram averages either across all frequencies or within a shrinking neighbourhood of zero frequency.

The weighting is inverse with respect to smoothed estimates of f.

Because of the concentration of spectral mass around zero frequency, where f changes little, computationally simpler estimates, with the same asymptotic properties, use just an estimate of $f(0)$.

Form unknown elements of Υ into the $q \times 1$ vector ν, where $q = r(p - (r+1)/2)$.

Equivalently write

$$vec\Upsilon = C\nu - c,$$

for suitable C and c.

Define the $p^2 \times p$ matrix-valued process z_t^d, whose ith $p \times p$ sub-matrix is diagonal with (j,j)th element $\Delta^{d_j} z_{it}$.

Define the DFT

$$w_\xi\left(\lambda\right) = \frac{1}{(2\pi n)^{\frac{1}{2}}} \sum_{t=1}^{n} \xi_t e^{it\lambda},$$

for real λ.

Denote by $\lambda_j = 2\pi j/n$, $j = 0, ..., [n/2]$, the Fourier frequencies.

Given z_t, $t = 1, ..., n$, a nonsingular estimate $\widehat{f}(\lambda)$ of $f(\lambda)$, for integer m such that

$$m \to \infty \text{ as } n \to \infty, \quad 1 \le m \le n/2,$$

and $s_j = 1$, $j = 0, n/2$, $s_j = 2$, otherwise, define

$$\begin{aligned}
\widehat{a}_m(d) &= C' \sum_{j=0}^{m} s_j \operatorname{Re}\left\{ w_{z^d}(-\lambda_j) \widehat{f}^{-1}\left(\lambda_j\right) w'_{z^d}(\lambda_j) \right\} c, \\
\widehat{b}_m(d) &= C' \sum_{j=0}^{m} s_j \operatorname{Re}\left\{ w_{z^d}(-\lambda_j) \widehat{f}^{-1}\left(\lambda_j\right) w'_{z^d}(\lambda_j) \right\} C, \\
\widehat{a}^{\circ}_m(d) &= C' \sum_{j=0}^{m} s_j \operatorname{Re}\left\{ w_{z^d}(-\lambda_j) \widehat{f}^{-1}(0)\, w'_{z^d}(\lambda_j) \right\} c, \\
\widehat{b}^{\circ}_m(d) &= C' \sum_{j=0}^{m} s_j \operatorname{Re}\left\{ w_{z^d}(-\lambda_j) \widehat{f}^{-1}(0)\, w'_{z^d}(\lambda_j) \right\} C.
\end{aligned}$$

Defining

$$\widehat{\nu}_m\,(d) = \widehat{b}_m^{-1}\,(d)\,\widehat{a}_m\,(d)\,, \quad \widehat{\nu}_m^{\circ}\,(d) = \widehat{b}_m^{\circ}\,(d)^{-1}\,\widehat{a}_m^{\circ}\,(d)\,,$$

consider the estimates of ν :

$$\begin{aligned} \text{W (“weighted”)} \quad &: \quad \widehat{\nu}_m(\delta),\ \widehat{\nu}_m(\widehat{\delta}); \\ \text{Z (“zero-frequency”)} \quad &: \quad \widehat{\nu}_m^{\circ}(\delta),\ \widehat{\nu}_m^{\circ}(\widehat{\delta}), \end{aligned}$$

where $\widehat{\delta}$ estimates δ.

The estimates are of GLS type, and effectively “whiten” the data, as their desirable asymptotic properties will confirm.

$\widehat{\nu}_m(\delta)$, $\widehat{\nu}^{\text{o}}_m(\delta)$, treat δ as known (as is usual in the $I\,(1)\,/I\,(0)$ case), but in our context they are generally regarded as infeasible, and included in part for completeness and to demonstrate that estimation of δ makes no asymptotic difference, as well as to imply, with $\widehat{\nu}_m(\widehat{\delta})$, $\widehat{\nu}^{\text{o}}_m(\widehat{\delta})$, cases where δ is partly known (e.g. $\delta_p = 1$ is known but δ_1,...,δ_r are unknown).

The computational simplicity of the Z over the W estimates is due not only to having to estimate f at only frequency zero, but to

$$\widehat{a}^{\circ}_{[n/2]}(d) = \frac{1}{2\pi} C' \sum_{t=1}^{n} z_t^d \widehat{f}^{-1}(0) \left(z_t^d\right)' c, \; \widehat{b}^{\circ}_{[n/2]}(d) = \frac{1}{2\pi} C' \sum_{t=1}^{n} z_t^d \widehat{f}^{-1}(0) \left(z_t^d\right)' C.$$

In a parametric setting with strong cointegration, Z estimates only do as well as W ones when the cointegrating gap > 1;

when it is 1, a "second-order bias" appears, when it is less than 1 (but greater than 1/2) convergence is slower due to sub-optimal weighting, and in each case the null χ^2 limit distribution of Wald test statistics is lost.

Limiting the increase of m repairs this defect

Consider

$$H_0 : R\nu = v,$$

where v is a given $J \times 1$ vector and R is a given $J \times q$ matrix with full row rank and constraints don't link different rows of Υ (to avoid commutativity problems due to different convergence rates).

We permit linear restrictions among coefficients in the same cointegrating relation, including the possibility that one or more of the ν_k is known *a priori*, and also exclusion restrictions.

The W and Z Wald statistics on which tests are based are

$$\begin{aligned}
W_m(d) &= \left(R\widehat{\nu}_m(d) - v\right)' \left(R b_m^{-1}(d) R'\right)^{-1} \left(R\widehat{\nu}_m(d) - v\right), \\
W_m^{\circ}(d) &= \left(R\widehat{\nu}_m^{\circ}(d) - v\right)' \left(R b_m^{\circ}(d)^{-1} R'\right)^{-1} \left(R\widehat{\nu}_m^{\circ}(d) - v\right),
\end{aligned}$$

with $d = \delta$ or $\widehat{\delta}$.

Assume for the W estimates that

$$u_t = A\,(L)\,\varepsilon_t, \quad A\,(z) = I_p + \sum_{j=1}^{\infty} A_j z^j,$$

where the A_j are $p \times p$ such that

$$\det\{A\,(z)\} \neq 0, \;\; |z| = 1 \text{ (W) or} \det\{A\,(1)\} \neq 0 \text{ (Z)}$$

$A(e^{i\lambda})$ is differentiable in $\lambda \in [-\pi, \pi]$ with derivative in $Lip\,(\eta)$, $\eta > 1/2$, the ε_t are iid with mean 0, pd covariance matrix Ω, and $E\,\|\varepsilon_t\|^{\mu} < \infty$, $\mu \geq 4$, and if $\delta_p - \delta_i > 1/2$ for some $i \in \{1, ..., r\}$, $\mu > 2/(2\min_{i:\delta_p - \delta_i > 1/2}(\delta_p - \delta_i) - 1)$.

This is easily satisfied if u_t is a stationary and invertible ARMA.

The rates in our conditions have implications for the choice of m.

The usual estimates of integration orders and spectral densities have convergence rates no better than $n^{2/5}$, so that not all conditions can hold when m grows as fast as n, and $\delta_p - \delta_i \leq 3/5$.

Since bias increases with m, bias-reduction may be needed.

Various devices are available.

Our ν estimates have mixed normal asymptotics when all cointegrating relations are strong, normal asymptotics when all are weak, and when both kinds of relation are present there is a corresponding combination of limiting distributions.

The final, global, outcome is that Wald statistics the same, standard, limit behaviour:

Under H_0,

$$W_m(\delta), W_m\left(\widehat{\delta}\right) \rightarrow_d \chi^2_J,$$

$$W_m^{\circ}(\delta), W_m^{\circ}\left(\widehat{\delta}\right) \rightarrow_d \chi^2_J.$$

Thus, the limit theory of Wald tests, familiar in many classical situations in econometrics and associated with optimal procedures in the $I(1)/I(0)$ cointegration literature,holds here simultaneously for weak (including stationary) and strong cointegration, and in the possible presence of unknown integration orders of observables and/or cointegrating errors.

Extensions:

Over-identifying restrictions

Variable integration orders of observables.

Parametric counterpart: $f(\lambda) = f(\lambda;\tau)$, $\lambda \in (-\pi, \pi]$, where $f(\lambda;\tau)$ is a known function of λ and an unknown finite-dimensional vector τ (e.g. u_t is a VAR of prescribed degree).

Given an estimate $\widehat{\tau}$ of τ, the $f(\lambda_j;\widehat{\tau})$ can replace the $\widehat{f}(\lambda_j)$ with $m = [n/2]$.

Under strong cointegration the resulting estimate of ν does does no better than ours.

Under weak cointegration, it is, like ours, asymptotically normal but with a faster, $\sqrt{n}$, rate and is not asymptotically independent of $\widehat{\tau}$, or of the estimate of δ, which also needs to be $\sqrt{n}$-consistent.

We can obtain optimal estimates of ν, δ and τ by jointly optimizing a Gaussian likelihood.