

Parametric Inference on Strong Dependence

Peter M. Robinson

London School of Economics

Plenary talk, The 6th International Iranian Workshop on
Stochastic Processes, May 19 2010

(Based on joint work with **Javier Hualde**)

Background

Short memory parametric time series models (having bounded spectral density or summable autocovariances), such as stationary and invertible ARMA, can be estimated by several, asymptotically equivalent, "Gaussian" or "Whittle" methods.

Asymptotic theory for the estimates (consistency, asymptotic normality with \sqrt{n} rate, and maybe efficiency) was long-ago established.

"Long memory time series" often refers to stationary and invertible ones, with spectral density diverging (as a power law) at zero frequency, or autocorrelations being non-summable.

This divergence makes establishing asymptotic normality harder.

But due essentially to the way the exploding periodogram and spectral density compensate for one another, "Gaussian" estimates have the same desirable asymptotic properties as in the short memory case.

Nonstationary processes such as unit root ones *a fortiori* have long memory.

Fractional models can cover stationary and nonstationary processes.

They can also cover non-invertible processes.

This entails a "memory parameter" that can take on any real value.

We would like a method and theory of estimation that is agnostic to the location of the memory parameter, i.e. does not require knowledge of whether the process is stationary or nonstationary, invertible or non-invertible.

Plan of talk

1. Fractional model
2. Consistency
3. Asymptotic normality
4. Empirical example
5. Finite-sample performance
6. Multivariate extension
7. Further comments and extensions

Fractional model

Whittle estimates of parameters in univariate stationary long memory or fractional processes have been shown to be \sqrt{n} -consistent and asymptotically normal, for sample size n .

In many time series, e.g. macroeconomic ones, the possibility of non-stationarity must be taken seriously.

Unit root models occupy a similarly specialized position relative to fractionally non-stationary processes as short memory ones do relative to fractional stationary ones.

$$x_t = \Delta^{-\delta_0} \{u_t \mathbf{1}(t > 0)\}, t = 0, \pm 1, \dots,$$

$$u_t = \theta(L; \varphi_0) \varepsilon_t, t = 0, \pm 1, \dots$$

L is the lag operator; $\Delta = 1 - L$ is the difference operator;

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

$\mathbf{1}(\cdot)$ is the indicator function; δ_0 is an unknown real number and φ_0 is an unknown $p \times 1$ real vector;

$$\theta(s; \varphi) = \sum_{j=0}^{\infty} \theta_j(\varphi) s^j,$$

where for all φ , $\theta_0(\varphi) = 1$, $\theta(s; \varphi): \mathbb{R}^1 \times \mathbb{R}^p$ is continuous in s and $|\theta(s; \varphi)| \neq 0$, $|s| = 1$; ε_t is a zero-mean unobservable white noise sequence.

θ describes parametric short memory autocorrelation.

E.g. $\theta(s; \varphi)$ is a rational function of s , whose denominator and numerator are polynomials in s of degrees p_1 and p_2 respectively, so u_t is $\text{ARMA}(p_1, p_2)$, and x_t is $\text{FARIMA}(p_1, \delta_0, p_2)$.

Due to the truncation x_t is actually non-stationary for all δ_0 .

But for $\delta_0 < 1/2$, $\Delta^{-\delta_0} u_t$ is stationary and x_t is “asymptotically stationary”.

For $\delta_0 \geq 1/2$, x_t is non-stationary in a more substantial sense, in particular $\text{Var}(x_t)$ diverges as $t \rightarrow \infty$, so the truncation is needed to avoid explosion.

For $\delta_0 = 1$ x_t has a unit root.

We wish to estimate $\tau_0 = (\delta_0, \varphi'_0)'$ from $x_t, t = 1, \dots, n$.

For any $\tau = (\delta, \varphi)'$, define

$$\varepsilon_t(\tau) = \Delta^\delta \theta^{-1}(L; \varphi) x_t, \quad t \geq 1,$$

noting that $x_t = 0, t \leq 0$.

Define

$$\hat{\tau} = \arg \min_{\tau \in \mathcal{T}} R_n(\tau),$$

where

$$R_n(\tau) = \frac{1}{n} \sum_{t=1}^n \varepsilon_t^2(\tau)$$

and $\mathcal{T} = [\nabla_1, \nabla_2] \times \Psi$, where $\nabla_1 < \nabla_2$, and Ψ is a compact subset of \mathbb{R}^p .

$\hat{\tau}$ is sometimes termed a “conditional sum of squares” estimate (“truncated sum of squares” might be more suitable).

It has the anticipated advantage of having the same limit distribution as the MLE of τ_0 under Gaussianity with \sqrt{n} rate, and thereby being asymptotically efficient (though we do not assume Gaussianity).

It has been used in estimation of non-fractional ARMA models (when δ_0 is a given integer), and in stationary and non-stationary FARIMA models.

There is a large gap in inferential theory when the possibility of non-stationarity ($\delta_0 \geq 1/2$) and/or non-invertibility ($\delta_0 \leq -1/2$) is to be allowed.

$R_n(\tau)$ doesn't converge uniformly on \mathcal{T} that entails a δ -interval, containing δ_0 , of length greater than $1/2$, so usual consistency proof doesn't work.

Consistency is used in CLT.

For invertible processes, Velasco and R established consistency, and thence asymptotic normality, of an alternative estimate of τ_0 , under an alternative definition of fractional nonstationarity and using tapering and “skipping” of Fourier frequencies, achieving a CLT with \sqrt{n} rate but with an inflated variance.

Consistency

A1. (i)

$$|\theta(s; \varphi)| \neq |\theta(s; \varphi_0)|,$$

for all $\varphi \neq \varphi_0$, $\varphi \in \Psi$, on a set $S \subset \{s : |s| = 1\}$ of positive measure;

(ii) for all φ , $\theta(e^{i\lambda}; \varphi)$ is differentiable in λ with derivative in $Lip(\varsigma)$, $\varsigma > 1/2$;

(iii) for all λ , $\theta(e^{i\lambda}; \varphi)$ is continuous in φ ;

(iv) for all $\varphi \in \Psi$,

$$|\theta(s; \varphi)| \neq 0, |s| = 1.$$

Condition (i) provides identification while (ii) and (iv) ensure that u_t is an $I(0)$ process (e.g. a stationary and invertible ARMA process).

A2. The ε_t are stationary and ergodic with finite fourth moment, and

$$E(\varepsilon_t | \mathcal{F}_{t-1}) = 0, \quad E(\varepsilon_t^2 | \mathcal{F}_{t-1}) = \sigma_0^2, \text{ a.s.},$$

where \mathcal{F}_t is the σ -field of events generated by ε_s , $s \leq t$, and conditional (on \mathcal{F}_{t-1}) third and fourth moments of ε_t equal the corresponding unconditional moments.

A3. For all $\delta \in [\nabla_1, \nabla_2]$, $\varphi \in \Psi$ such that $\delta_0 - \delta > 1/2$, there exists a random variable $D(\tau) > 0$ a.s. such that $n^{-2(\delta_0 - \delta)} \sum_{t=1}^n \varepsilon_t^2(\tau)$ converges weakly to $D(\tau)$ in the space $D[0, 1]$.

A2 avoids requiring independence or identity of distribution of ε_t , but rules out conditional heteroskedasticity.

A3 can be checked under a higher-moment type condition.

Theorem 1 *Let A1-A3 hold. Then as $n \rightarrow \infty$,*

$$\hat{\tau} \rightarrow_p \tau_0.$$

The proof reflects the fact that $R_n(\tau)$ converges in probability to a well-behaved function when $\delta > \delta_0 - \frac{1}{2}$, and diverges when $\delta < \delta_0 - \frac{1}{2}$, while the need to establish uniform convergence and behaviour in a neighbourhood of $\delta = \delta_0 - \frac{1}{2}$ requires additional special treatment.

We split the admissible δ -interval $[\nabla_1, \nabla_2]$ into four disjoint intervals

$$\begin{aligned} \mathcal{I}_1 &= \left[\nabla_1, \delta_0 - \frac{1}{2} - \eta \right], \mathcal{I}_2 = \left(\delta_0 - \frac{1}{2} - \eta, \delta_0 - \frac{1}{2} \right), \\ \mathcal{I}_3 &= \left[\delta_0 - \frac{1}{2}, \delta_0 - \nabla \right], \mathcal{I}_4 = (\delta_0 - \nabla, \nabla_2], \end{aligned}$$

for arbitrarily small positive η and for $\nabla \subset (1/3, \frac{1}{2})$.

Asymptotic normality

A4. (i)

$$\tau_0 \in \text{int } \mathcal{T};$$

(ii) for all λ , $\theta(e^{i\lambda}; \varphi)$ is twice continuously differentiable in φ on a closed neighbourhood $N_\epsilon(\varphi_0)$ of radius $0 < \epsilon < 1/2$ about φ_0 ;

(iii) the matrix

$$\mathbf{A} = \begin{pmatrix} \pi^2/6 & -\sum_{j=1}^{\infty} \mathbf{b}'_j(\varphi_0)/j \\ -\sum_{j=1}^{\infty} \mathbf{b}_j(\varphi_0)/j & \sum_{j=1}^{\infty} \mathbf{b}_j(\varphi_0) \mathbf{b}'_j(\varphi_0) \end{pmatrix}$$

is non-singular, where

$$\mathbf{b}_j(\varphi_0) = \sum_{k=0}^{j-1} \theta_k(\varphi_0) \frac{\partial \phi_{j-k}(\varphi_0)}{\partial \varphi},$$
$$\theta^{-1}(s; \varphi) = \sum_{j=0}^{\infty} \phi_j(\varphi) s^j.$$

Again, this holds for FARIMAs.

Theorem 2 *Let A1-A4 hold. Then as $n \rightarrow \infty$,*

$$\sqrt{n}(\hat{\boldsymbol{\tau}} - \boldsymbol{\tau}_0) \rightarrow_d N(\mathbf{0}, \mathbf{A}^{-1}).$$

Multivariate extension

When observations on several related time series are available joint modelling can achieve efficiency gains.

Consider a vector $\mathbf{x}_t = (x_{1t}, \dots, x_{rt})'$ given by

$$\mathbf{x}_t = \Lambda_0^{-1} \{\mathbf{u}_t \mathbf{1}(t > 0)\}, \quad t = 0, \pm 1, \dots,$$

where $\mathbf{u}_t = (u_{1t}, \dots, u_{rt})'$,

$$\mathbf{u}_t = \Theta(L; \varphi_0) \boldsymbol{\varepsilon}_t, \quad t = 0, \pm 1, \dots,$$

in which $\boldsymbol{\varepsilon}_t = (\varepsilon_{1t}, \dots, \varepsilon_{rt})'$, φ_0 is (as in the univariate case) a $p \times 1$ vector of short-memory parameters,

$$\Theta(s; \varphi) = \sum_{j=0}^{\infty} \Theta_j(\varphi) s^j, \quad \Theta_0(\varphi) = \mathbf{I}_r \text{ for all } \varphi,$$

and

$$\Lambda_0 = \text{diag} \left(\Delta^{\delta_{01}}, \dots, \Delta^{\delta_{0r}} \right),$$

where the memory parameters δ_{0i} are unknown real numbers.

In general, all δ_{0i} can be distinct but for the sake of parsimony we allow for the possibility that they are known to lie in a set of dimension $q < r$.

For example, perhaps as a consequence of pre-testing, we might believe some or all the δ_{0i} are equal, and imposing this restriction in the estimation could further improve efficiency.

Introduce known functions $\delta_i = \delta_i(\boldsymbol{\delta})$, $i = 1, \dots, r$, of $q \times 1$ vector $\boldsymbol{\delta}$, such that for some $\boldsymbol{\delta}_0$ we have $\delta_{0i} = \delta_i(\boldsymbol{\delta}_0)$, $i = 1, \dots, r$.

Denote $\boldsymbol{\tau} = (\boldsymbol{\delta}', \boldsymbol{\varphi}')$ and define

$$\boldsymbol{\varepsilon}_t(\boldsymbol{\tau}) = \boldsymbol{\Theta}^{-1}(L; \boldsymbol{\varphi}) \boldsymbol{\Lambda}(\boldsymbol{\delta}) \mathbf{x}_t, \quad t \geq 1,$$

where $\boldsymbol{\Lambda}(\boldsymbol{\delta}) = \text{diag}(\Delta^{\delta_1}, \dots, \Delta^{\delta_r})$.

Gaussian likelihood considerations suggest the objective function

$$R_n^*(\boldsymbol{\tau}) = \det \{ \boldsymbol{\Sigma}_n(\boldsymbol{\tau}) \} ,$$

where

$$\boldsymbol{\Sigma}_n(\boldsymbol{\tau}) = \frac{1}{n} \sum_{t=1}^n \boldsymbol{\varepsilon}_t(\boldsymbol{\tau}) \boldsymbol{\varepsilon}_t'(\boldsymbol{\tau}),$$

assuming that no prior restrictions link $\boldsymbol{\tau}_0$ with the covariance matrix of $\boldsymbol{\varepsilon}_t$.

Unfortunately our consistency proof for the univariate case does not straightforwardly extend to an estimate minimizing $R_n^*(\boldsymbol{\tau})$, at least if $q > 1$.

Also $R_n^*(\boldsymbol{\tau})$ is liable to pose a severe computational challenge since p is liable to be larger in the multivariate case and q may exceed 1; it may be difficult to locate an approximate minimum as a preliminary to iteration.

We avoid both these problems by taking a single Newton step from an initial \sqrt{n} -consistent estimate $\tilde{\tau}$. Defining

$$\mathbf{H}_n(\tau) = \frac{1}{n} \sum_{t=1}^n \left(\frac{\partial \varepsilon_t(\tau)}{\partial \tau'} \right)' \Sigma_n^{-1}(\tau) \frac{\partial \varepsilon_t(\tau)}{\partial \tau'},$$

$$\mathbf{h}_n(\tau) = \frac{1}{n} \sum_{t=1}^n \left(\frac{\partial \varepsilon_t(\tau)}{\partial \tau'} \right)' \Sigma_n^{-1}(\tau) \varepsilon_t(\tau),$$

we consider the estimate

$$\hat{\tau} = \tilde{\tau} - \mathbf{H}_n^{-1}(\tilde{\tau}) \mathbf{h}_n(\tilde{\tau}).$$

A5. (i) For all φ , $\Theta(e^{i\lambda}; \varphi)$ is differentiable in λ with derivative in $Lip(\varsigma)$, $\varsigma > 1/2$;

(ii) for all φ ,

$$\det \{ \Theta(s; \varphi) \} \neq 0, \quad |s| = 1;$$

(iii) the ε_t are stationary and ergodic with finite fourth moment,

$$E(\varepsilon_t | \mathcal{F}_{t-1}) = 0, \quad E(\varepsilon_t \varepsilon_t' | \mathcal{F}_{t-1}) = \Sigma_0$$

almost surely, where Σ_0 is positive definite, \mathcal{F}_t is the σ -field of events generated by ε_s , $s \leq t$, and conditional (on \mathcal{F}_{t-1}) third and fourth moments and cross-moments of elements of ε_t equal the corresponding unconditional moments;

(iv) for all λ , $\Theta(e^{i\lambda}; \varphi)$ is twice continuously differentiable in φ on a closed neighbourhood $N_\epsilon(\varphi_0)$ of radius $0 < \epsilon < 1/2$ about φ_0 ;

(v) the matrix B having (i, j) th element

$$\sum_{k=1}^{\infty} tr \left\{ \left(\mathbf{d}_k^{(i)}(\varphi_0) \right)' \Sigma_0^{-1} \mathbf{d}_k^{(j)}(\varphi_0) \Sigma_0 \right\}$$

is non-singular, where $\mathbf{d}_k^{(i)}(\varphi_0)$ is

$$-\frac{\partial \delta_i(\boldsymbol{\delta}_0)}{\partial \delta_i} \sum_{l=1}^k \frac{1}{l} \sum_{m=0}^{k-l} \Phi_m^{(i)}(\varphi_0) \Theta_{k-l-m}(\varphi_0),$$

for $1 \leq i \leq r$, and

$$\sum_{l=1}^k \frac{\partial \Phi_l(\varphi_0)}{\partial \varphi_i} \Theta_{k-l}(\varphi_0), \quad r+1 \leq i \leq r+p,$$

for $r+1 \leq i \leq r+p$, the $\Phi_j(\varphi)$ being coefficients in the expansion

$$\Theta^{-1}(s; \varphi) = \Phi(s, \varphi) = \sum_{j=0}^{\infty} \Phi_j(\varphi) s^j,$$

where $\Phi_m^{(i)}(\varphi_0)$ is an $r \times r$ matrix whose i -th column is the i -th column of $\Phi_i(\varphi_0)$ and whose other elements are all zero;

(vi) $\delta_i(\boldsymbol{\delta})$ is twice continuously differentiable in $\boldsymbol{\delta}$,
for $i = 1, \dots, r$;

(vii) $\tilde{\boldsymbol{\tau}}$ is a \sqrt{n} -consistent estimate of $\boldsymbol{\tau}_0$.

Theorem 3 Let A5 hold. Then as $n \rightarrow \infty$

$$\sqrt{n}(\hat{\boldsymbol{\tau}} - \boldsymbol{\tau}_0) \rightarrow_d N(\mathbf{0}, \mathbf{B}^{-1}).$$

Finite-sample performance

We used data generated from a FARIMA(1, δ_0 , 0), for 8 stationary, nonstationary, invertible and non-invertible δ_0 , and 3 AR coefficients φ_0 , and $n = 64, 128, 256$.

We compared our estimate $\hat{\tau} = (\hat{\delta}, \hat{\varphi})'$ one, $\hat{\tau}_W = (\hat{\delta}_W, \hat{\varphi}_W)'$, that "cheats" by carrying out the correct degree of integer-differencing needed to shift the process to the stationary/invertible region, estimates the memory and AR parameters from the filtered sequence, then adds to or subtracts from the memory estimate the appropriate integer.

With $\varphi_0 = \pm 0.5$, $\hat{\delta}$ was the more biased, but with $\varphi_0 = 0$, $\hat{\delta}_W$ was the more biased.

In the great majority of cases, $\hat{\delta}$ was the more precise.

In the great majority of cases, $\hat{\varphi}$ was both less biased and more precise than $\hat{\varphi}_W$.

Empirical examples

US Quarterly income and consumption 1947Q1-1981Q2
($n = 138$)

Evidence of unit root previously found.

For each series, we first estimated δ_0 semiparametrically, then filtered accordingly, then applied Box-Jenkins-type procedures to identify ARMA orders.

Then for the resulting FARIMAs we estimated our $\hat{\tau}$ from the original data.

We then computed t -ratios, and applying our asymptotic theory, strongly rejected $\delta_0 = 1$ for both series.

Further comments and extensions

Our univariate and multivariate structures cover a wide range of parametric models for stationary and nonstationary time series, with memory parameters allowed to lie in a set that can be arbitrarily large.

Unit root series are a special case, but unlike in the bulk of the large literature on these models we do not have to assume knowledge that memory parameters are 1.

As the nondiagonal structure of \mathbf{A} and \mathbf{B} suggests, there is efficiency loss in estimating φ_0 if memory parameters are unknown, but on the other hand if these are misspecified φ_0 will in general be inconsistently estimated.

Our limit distribution theory can be used to test hypotheses on the memory and other parameters, after straightforwardly forming consistent estimates of \mathbf{A} or \mathbf{B} .

Our multivariate model does not cover fractionally cointegrated systems because Σ_0 is required to be positive definite.

On the other hand our theory for univariate estimation should cover estimation of individual memory parameters of observations.

Moreover, again on an individual basis, it should be possible to derive analogous properties of estimates of memory parameters of cointegrating errors based on residuals that use simple estimates of cointegrating vectors, such as least squares.

In a more standard regression setting, for example with deterministic regressors such as polynomial functions of time, it should be possible to extend our theory for univariate and multivariate models to residual-based estimates of memory parameters of errors.

Nonstationary fractional series can be defined in many ways.

Our definition is a leading one in the literature, and has been termed “Type II”.

Another popular one is “Type I”: it seems likely that the asymptotic theory can also be established in a “Type I” setting.