Estimation of Temporal and Spatial Power Law Trends

Peter M. Robinson

London School of Economics

Workshop talk 1, The 6th International Iranian Workshop on Stochastic Processes, May 18 2010

Overview

We consider trends in time series and spatial data.

We model these by power law functions with unknown exponents.

We consider asymptotic properties of parameter estimates.

A more general contribution is to proving consistency in mixed-rate problems.

Plan of talk

1. Motivation

2. Background to polynomial regression model for time series

3. Power law / generalized polynomial model for time series

(4. Background to consistency proof for implicitly-defined extremum estimates)

(5. Generic consistency proof for mixed-rate problems)

6 Spatial/spatio-temporal model and its estimation

7. Asymptotic properties of the estimates

8. Monte Carlo study of finite-sample performance

.

1. Motivation

Polynomial-in-time regression is one of the longest-established tools of time series analysis:

$$y_u = \beta_1 + \beta_2 u + \beta_3 u^2 + ...\beta_p u^{p-1} + x_u, \ u = 1, 2, ..., N,$$

where x_u is an unobservable error, with zero mean and constant, finite variance (but not necessarily serially uncorrelated).

The least squares estimate (LSE) of $\beta = (\beta_1, ..., \beta_P)'$ is

$$\widehat{\beta} = M^{-1} m.$$

where

$$M = \sum_{u=1}^{N} \begin{bmatrix} 1 \\ u \\ \cdot \\ \cdot \\ \cdot \\ u^{p-1} \end{bmatrix} \begin{bmatrix} 1 \\ u \\ \cdot \\ \cdot \\ \cdot \\ u^{p-1} \end{bmatrix}', \quad m = \sum_{u=1}^{N} \begin{bmatrix} 1 \\ u \\ \cdot \\ \cdot \\ \cdot \\ u^{p-1} \end{bmatrix} y_u.$$

If $x_u \sim NID(0, \sigma^2)$ then

$$\widehat{\beta} \sim N(0, \sigma^2 M^{-1}).$$

So we can carry out exact statistical inference.

Moreover, we can rewrite the model in terms of orthogonal polynomials:

$$y_u = \gamma_1\phi_1^N(u) + \gamma_2\phi_2^N(u) + \gamma_3\phi_3^N(u) + ...\gamma_p\phi_p^N(u) + x_u,\ u = 1, 2, ..., N,$$

where

$$\phi_1^N(u) = 1,\ \phi_2^N(u) = u-(N+1)/2, \phi_3^N(u) = \mathrm{u}^2-(N+1)u+(N+1)(N+2)/6, ...,$$

Then the LSE of the γ_i are independent, as well as normally disributed.

Correspondingly, the F-ratio statistic for testing

$$H_0 : \gamma_i = 0,\ i = 1, ..., p,$$

can be decomposed into p independent statistics whih can be used to test the p individal hypotheses

$$H_{0i} : \gamma_i = 0$$

individually,and thereby test for a more parsimonious model.

The $x_u \sim NID(0, \sigma^2)$ assumption is extremely strong, however.

Returning to the original form

$$y_u = \beta_1 + \beta_2 u + \beta_3 u^2 + ...\beta_p u^{p-1} + x_u,\ \ u = 1, 2, ..., N,$$

suppose now that the x_u are not necesarily normally distributed or independent.

Suppose x_u is covariance stationary with *spectral density*

$$F(\lambda) = \frac{1}{2\pi} \sum_{j=-\infty}^{\infty} E(x_u x_{u+j}) e^{ij\lambda},\ \ -\pi < \lambda < \pi.$$

Suppose that $F(\lambda)$ is at least continuous and positive at $\lambda = 0$ (as with autoregressive moving average processes).

Aside (looking ahead to my future taks):

$$0 < F(0) < \infty$$

is one definition of *short memory* or *short range dependence*.If

$$F(0) = \infty, \text{ e.g. } F(\lambda) \sim \lambda^{-2d}, \ 0 < d < 1/2, \text{ as } \lambda \to 0+,$$

we have *short memory* or *short range dependence*.

If

$$F(0) = 0, \text{ e.g. } F(\lambda) \sim \lambda^{-2d}, \ -1/2 < d < 0, \text{ as } \lambda \to 0+,$$

we have *antipersistence* or *negative dependence*.

Under

$$0 < F(0) < \infty$$

and mild additional assumptions,

$$D(\widehat{\beta} - \beta) \rightarrow_d N(0, 2\pi F(0) Q^{-1}), \text{ as } N \rightarrow \infty, .$$

where

$$D = diag(N^{1/2}, N^{3/2}, ..., N^{(p+1)/2})$$

and Q is a known, positive definite matrix.

We can estimate $F(0)$ and thence carry out approximate statistical inference on β.

Moreover despite the dependence in x_u, β is (asymptotically Gauss-Markov) efficient.

3. Power law / generalized polynomial model for time series

Polynomial models have been extended to lattice spatial or spatio-temporal data.

Though not all relevant time series theory has been explicitly extended to the lattice case, this seems substantially possible.

Polynomials are nevertheless restrictive.

The Weierstrass theorem justifies uniform approximation of any continuous function over a compact interval, but seems less practically relevant the longer the data set.

Also, polynomials do not allow for a *decaying* trend.

Nonparametric smoothing may be unreliable in series of moderate length, when instead richer parametric models than polynomials might be considered.

One class that advantageously nests polynomials, and has received little theoretical attention, are "generalized polynomial" or "power law" models.

Consider the more general moddl

$$y_u = \sum_{j=1}^{p} \beta_j u^{\theta_j} + x_u,$$

where the θ_j and β_j are real-valued and all can be unknown, $\theta_j > -1/2$ for all j, and the zero-mean unobservable process x_u is covariance stationary with short memory.

Our original polynomial model is a special case ($\theta_j = j - 1$ for all j), indeed this is a hypothesis that might be tested.

When $0 > \theta_j > -1/2$ thereis a *decaying* trend component u^{θ_j}.

Note that the model is now nonlinear in the parameter vector $\theta = (\ \theta_1, ..., .\theta_p)'$

θ and β can be estimated by the nonlinear least squares estimates (NLSE:)

$$\left(\widehat{\theta}, \widehat{\beta}\right) = \arg\min Q(h, b)$$

where

$$Q(b, h) = \sum_{u=1}^{N} (y_u - \sum_{j=1}^{p} b_j u^{h_j})^2.$$

$\left(\widehat{\theta}, \widehat{\beta}\right)$ are not explicitly defned but require numerical optimization.

Because the b_j are involved only quadratically we can eliminate them and apply numerical optimization to a function of h (with $\widehat{\beta}$ then available by an explicit side calculation).

Thus asymptotic theory, with sample size N increasing, is needed to justify rules of statistical inference even when x_u is NID.

This appears to be unavailable, indeed it can present some difficult features.

Because the estimates are not explicitly defined, asymptotic distribution theory makes use (in application of the mean value theorem) of an initial consistency proof, as is common.

Most such proofs require regressors to be non-trending, whence under suitable additional conditions all parameter estimates are $N^{\frac{1}{2}}$-consistent.

Mixed rates of convergence frequently arise (in the NLSE of our trend model all rates differ, and for implicitly-defined extremum estimates like $\left(\widehat{\theta}, \widehat{\beta}\right)$they are typically associated with difficulty in the initial consistency proof, due to the objective function not converging uniformly to a function that is uniquely optimized over the whole parameter space.

Consistency proofs here have tended to be derived in a somewhat *ad hoc* fashion, geared to the case at hand.

The approach we develop is likely to apply to a quite general class of estimates (not just the NLSE) of a variety of models.

4. Background to consistency proof for implicitly-defined extremum estimates

Wald (1949), Wolfowitz (1949) proved respectively stong and weak consistency for ML estimates (see also Aitchison and Silvey (1958), Huber (1964), Hoadley (1971) etc).

Jennrich (1969), Malinvaud (1970) used similar ideas for NLSE.

These ideas also apply to a large class of implicitly-defined extremum estimates.

Consider objective function $Q_N(h)$, $\mathbb{R}^p \Rightarrow \mathbb{R}$, based on sample size N.

Let $\Theta \subset \mathbb{R}^p$, let

$$\hat{\theta} = \hat{\theta}_N = \arg\min_{h \in \Theta} Q_N(h)$$

exist.

A typical consistency theorem is as follows.

Theorem A *Suppose there exists*

$$\theta \in \Theta,$$

such that we can write

$$Q_N(h) - Q_N(\theta) = U_N(h) + V_N(h),$$

where for any $\varepsilon > 0$ *there exists* $\eta > 0$ *such that*

$$\underset{\|h-\theta\|\geq\varepsilon}{inf} U_N(h) \geq \eta, \ \ \forall \text{ large enough } N,$$

and

$$\sup_{h\in\Theta} |V_N(h)| \to_p 0 \text{ as } N \to \infty.$$

Then

$$\hat{\theta} \to_p \theta, \text{ as } N \to \infty.$$

Correspondingly, much of the literature assumes that $Q_N(h)$ converges (uniformly) to a function that is uniquely minimized over Θ (often assumed compact), as in the nonlinear regression

$$y_u = f(z_u; \theta) + x_u$$

with approximately stationary regressors z_u and errors x_u, and

$$Q_N(h) = \frac{1}{N} \sum_{u=1}^{N} \{y_u - f(z_u; h)\}^2 .$$

Actually "stability" is not required.

Let

$$\mathcal{N} = \{h : h \in \Theta, \|h - \theta\| < \varepsilon\} \ , \quad \bar{\mathcal{N}} = \Theta - \mathcal{N}.$$

Then

$$\begin{aligned} P\left\|\hat{\theta} - \theta\right\| &\geq \varepsilon) \leq P\left(\inf_{\bar{\mathcal{N}}} \{Q_N(h) - Q_N(\theta)\} \leq 0\right) \\ &\leq P\left(\inf_{\bar{\mathcal{N}}} U_N(h) - \sup_{\Theta} |V_N(h)| \leq 0\right) \\ &\leq P\left(\sup_{\Theta} |V_N(h)| / \inf_{\bar{\mathcal{N}}} U_N(h) \geq 1\right), \end{aligned}$$

which $\rightarrow 0$ if

$$\sup_{\Theta} |V_N(h)| / \inf_{\bar{\mathcal{N}}} U_N(h) = o_p(1),$$

so it is <u>relative</u> rates of numerator and denominator that matter (and $U_N(h)$ might be stochastic).

Moreover, there is further "slack" in the method.

E.g if

$$y_u = f(z_u;\theta) + x_u, \quad Q_N(h) = \frac{1}{N}\sum_{u=1}^{N} \{y_u - f(z_u;h)\}^2 \,,$$

then

$$\begin{aligned} U_N(h) &= \frac{1}{N}\sum_{u=1}^{N} \{f(z_u;\theta) - f(z_u;h)\}^2 \,, \\ V_N(h) &= \frac{2}{N}\sum_{u=1}^{N} \{f(z_u;\theta) - f(z_u;h)\} \, x_u. \end{aligned}$$

Then with $\inf_{\bar{\mathcal{N}}} U_N(h) > \eta$ we might also have $\sup_{\Theta} |V_N(h)| = O_p(N^{-1/2})$.

Wu (1981) explicitly focussed on nonstationary possibilities, giving milder sufficient conditions for

$$\underline{\lim}_{N\to\infty} \inf_{\bar{\mathcal{N}}} \{Q_N(h) - Q_N(\theta)\} > 0 \quad (i.p. \text{ or } a.s.)$$

in case of NLSE with iid, finite variance, errors.

These tend to be difficult to check but it seems that his and other proof methods described above are of limited value when θ is a vector.

The proof methods seem able to handle only limited variation of rates over the elements of θ.

5. Generic consistency proof for mixed-rate problems

We give a generic consistency proof that seems to apply in many cases.

It also delivers rates that are not optimal, but nearly so (and our relative rates reflect relative optimal ones).

We consider a neighbourhood $\mathcal{N}$ that is ellipsoidal and shrinks as n increases, with rate varying across dimensions.

This partly exploits the "slack" in the previous approach, because we can afford to make $\inf_{\bar{\mathcal{N}}} U_N(h)$ smaller.

But we also use the fact that $V_N(h)$ is small when h is close to θ, e.g.

$$V_N(h) = \frac{2}{N} \sum_{u=1}^{N} \{f(z_u;\theta) - f(z_u;h)\}\, x_u.$$

We split $\bar{\mathcal{N}}$ into finitely many "donuts", over each of which take sup of $V_N(h)$ and inf of $U_N(h)$.

Let $\Theta_i \subset \mathbb{R}$, $i = 1, ..., p$, and define $\Theta = \Pi_{i=1}^{p} \Theta_i$, so $\Theta \subset \mathbb{R}^p$.

For positive scalars C_{iw}, $i = 1, ..., p$, $w = 1, 2, ...$, such that $C_{iw} \leq C_{i,w+1}$, $i = 1, ..., p$, define $C_w = (C_{1w}, ..., C_{pw})$, and

$$\begin{aligned} \mathcal{N}_i(C_{iw}) &= \{h_i : |h_i - \theta_i| < C_{iw}\}\,,\ \bar{\mathcal{N}}_i(C_{iw}) = \Theta_i \backslash \mathcal{N}_i(C_{iw}), \\ \mathcal{N}(C_w) &= \prod_{i=1}^{p} \mathcal{N}_i(C_{iw}),\ \bar{\mathcal{N}}(C) = \prod_{i=1}^{p} \bar{\mathcal{N}}_i(C_{iw}), \\ \mathcal{S}_w &= \bar{\mathcal{N}}(C_w) \cap \mathcal{N}(C_{w+1}). \end{aligned}$$

Theorem B *Let there exist a finite integer* W; *also* C_{iw}, $w = 1, ..., W + 1$, *depending on* N, *such that*

$$\Theta \subset \mathcal{N}(C_{W+1})$$

for N *sufficiently large, and let*

$$\sup_{\mathcal{S}_w} |V_N(h)| / \inf_{\mathcal{S}_w} U_N(h) \to_p 0, \; as \; N \to \infty, \; w = 1, ..., W.$$

Then

$$\hat{\theta} = \theta + \mathcal{O}_p(C_1), \; as \; N \to \infty,$$

where $O_p(C_1)$ *is a vector with* i*-th element* $O_p(C_{i1})$.

(Also a.s. convergence version.)

To check the conditions in specific cases it is convenient to refer to positive norming sequences s_w, $w = 1, ..., W$, depending on N, such that $s_1 < ... < s_W$ and $s_1 \to \infty$ as $N \to \infty$.

Then we need to show that for some finite W we can choose C_w ,$w = 1, ..., W+1$, and s_w, $w = 1, ..., W$, such that

$$\begin{aligned} p\underline{\lim}_{N\to\infty} \inf_{\mathcal{S}_w} U_N(h)/s_w &> 0, \ w = 1, ..., W, \\ \sup_{\mathcal{S}_w} |V_N(h)| / s_w &\to_p 0, \ as\ N \to \infty, \ w = 1, ..., W, \\ \Theta &\subset \mathcal{N}(C_{W+1}). \end{aligned}$$

6 Spatial/spatio-temporal model and its estimation

Let the integer $d \geq 1$ represent the dimension on which data are observed, where $d = 1$ for time series and $d \geq 2$ for spatial or spatio-temporal data.

Let u now be the $d-$dimensional multi-index $u = (u_1, u_2, ..., u_d)'$.

Denoting $\mathbb{Z} = \{j : j = 0, \pm 1, ...\}$, consider

$$y_u = \sum_{i=1}^{d} \sum_{j=1}^{p_i} \beta_{ij} u_i^{\theta_{ij}} + x_u = f(u;\theta)'\beta + x_u,\ u \in \mathbb{Z}^d,$$

where x_u is described subsequently and

$$\begin{aligned} \beta &= \left(\beta_1', ..., \beta_d'\right)', \beta_i = \left(\beta_{i1}, ..., \beta_{ip_i}\right)', \\ \theta &= \left(\theta_1', ..., \theta_d'\right)', \theta_i = \left(\theta_{i1}, ..., \theta_{ip_i}\right)', \\ f(u;\theta) &= \left(f_1(u_1;\theta_1)', ..., f_d(u_d;\theta_d)'\right)', \\ f_i(u_i;\theta_i) &= \left(u_i^{\theta_{i1}}, ..., u_i^{\theta_{ip_i}}\right)', \end{aligned}$$

for $i = 1, ..., d$.

Defining $p = p_1 + ... + p_d$, the $p \times 1$ vectors β and θ are supposed unknown.

One of the d dimensions could be time, in which case we have a spatio-temporal model.

The spatial or spatio-temporal model potentially involves even more θ parameters than the the series one, exacerbating the "mixed-rate" issue.

But it is an issue even in the time series case $d = 1$.

7. Asymptotic properties of the estimates

Our consistency proof confines the NLSE of θ to a compact subset Θ of

$$\left(-\frac{1}{2}, \infty\right)^p$$

which also contains θ.

We introduce two assumptions which imply identifiability of θ and β.

Assumption 1 $\theta \in \Theta$.

Assumption 2 $\theta_{ij} = 0$ for at most one (i, j); $\beta_{ij} \neq 0$ for all (i, j).

We allow an intercept, but do not specifically include one.

Given

$$N = \prod_{i=1}^{d} n_i$$

observations on y_u, $u \in \mathbb{N} = \mathbb{N}_1 \times ... \times \mathbb{N}_d$, $\mathbb{N}_i = (1, ..., n_i)$, define the NLSE of β, θ by

$$\left(\widehat{\beta}, \widehat{\theta}\right) = \arg \min_{b \in \mathbb{R}^p, h \in \Theta} Q(b, h),$$

where

$$Q(b, h) = \sum_{u \in \mathbb{N}} \left\{ y_u - b' f(u; h) \right\}^2 .$$

Asymptotic normality requires further assumptions.

The first entails short range dependence of the x_u.

(We could cover long range dependence and negative dependence also, leading to different rates.)

Assumption 3 *x_u, $u \in \mathbb{Z}^d$, is covariance stationary with zero mean, and its autocovariance function, $\gamma_u = cov\,(x_t, x_{t+u})$, for the multi-index $t = (t_1, ..., t_d)'$, satisfies $\sum_{u \in \mathbb{Z}^d} |\gamma_u| < \infty$.*

The next assumption, of increase with algebraic rate of observations in all dimensions, is capable of generalization, and employed for simplicity.

Assumption 4 *$n_i \sim B_i N^{b_i}$, $i = 1, ..., d$,* as *$N \to \infty$, where $B_i > 0$, $b_i > 0$, $i = 1, ..., d$, $\prod_{i=1}^{d} B_i = \sum_{i=1}^{d} b_i = 1$.*

Define $\zeta_{ij} = b_i \theta_{ij}$, and with no loss of generality, identify dimension $i = 1$ such that

$$\zeta_{11} = \min_{1 \le i \le d} \{\zeta_{i1}\},$$

where, if two or more i satisfy this, an arbitrary choice is made.

Theorem 1 *If Assumptions 1-4 hold, for* $j = 1, ..., p_i$, $i = 1, ..., d$, *as* $N \to \infty$,

$$\hat{\theta}_{ij} - \theta_{ij} = O_p\left(N^{\chi - \zeta_{ij} - \frac{1}{2}}\right),$$

for any $\chi > 0$.

As is common with initial consistency proofs a sharp rate (corresponding to $\chi = 0$) is not quite delivered (smoothness conditions, in particular, are not exploited).

Theorem 1 is used in the proof of our central limit theorem (CLT), for which we also need consistency, with a rate, for $\hat{\beta}$.

Theorem 2 *If Assumptions 1-4 hold, for* $j = 1, ..., p_i$, $i = 1, ..., d$,

$$\hat{\beta}_{ij} = \beta_{ij} + O_p\left((\log N) N^{\chi - \zeta_{ij} - \frac{1}{2}}\right), \text{ as } N \to \infty.$$

The *relative* rates for the $\hat{\theta}_{ij}$ and $\hat{\beta}_{ij}$ in Theorems 1 and 2 are matched by relative rates that feature in our CLT.

For this we introduce first

Assumption 5 *$x_u = \sum_{v \in \mathbb{Z}^d} \xi_v \varepsilon_{u-v}$, $\sum_{v \in \mathbb{Z}^d} |\xi_v| < \infty$, $u \in \mathbb{Z}^d$, where v is the multi-index $v = (v_1, ..., v_d)'$, $\left\{\varepsilon_u, u \in \mathbb{Z}^d\right\}$ are independent random variables with zero mean and unit variance, $\left\{\varepsilon_u^2, u \in \mathbb{Z}^d\right\}$ are uniformly integrable, and $\sum_{v \in \mathbb{Z}^d} \xi_v \neq 0$.*

Assumption 5 implies Assumption 3, and both imply existence and boundedness of the spectral density

$$F(\lambda) = (2\pi)^{-1} \left| \sum_{v \in \mathbb{Z}^d} \xi_v e^{iv'\lambda} \right|^2$$

of x_u, where λ is the multi-index $\lambda = (\lambda_1, ..., \lambda_d)'$, while Assumption 5 implies also $F(0) > 0$.

Introduce $p \times p$ matrices

$$D = N^{\frac{1}{2}} diag\left\{n_1^{\theta_{11}}, ..., n_1^{\theta_{1p_1}}, ..., n_d^{\theta_{d1}}, ..., n_d^{\theta_{dp_d}}\right\}, L(s) = diag\left\{L_1(s_1), ..., L_d(s_d)\right\}$$

where $L_i(s_i) = (\log s_i) I_{p_i}$, and $2p \times 2p$ matrices

$$D_+ = I_2 \otimes D,\ \ L_+ = diag\left\{I_p, L(n)\right\}.$$

Define $\alpha = (\theta', \beta')'$, $\widehat{\alpha} = (\widehat{\theta}', \widehat{\beta}')'$.

Let B be a certain $p \times 2p$ matrix depending only on β, and Υ a certain nonsingular $p \times p$ matrix depending only on θ.

Denote by $\mathfrak{N}_r(a, A)$ an r-dimensional normal vector with mean vector a and (possibly singular) covariance matrix A.

Theorem 3 *If Assumptions 1, 2 and 5 hold, as* $N \to \infty$

$$D_{+} L_{+}^{-1} (\widehat{\alpha} - \alpha) \to_{d} \mathfrak{N}_{2p} \left(0, 2\pi F(0) B' \Upsilon^{-1} B \right).$$

Comments:

1. With known θ, long-established techniques give

$$D\left(\hat{\beta}-\beta\right)\to_d \mathfrak{N}_p\left(0, 2\pi F(0)\Phi^{-1}\right),$$

for a known, nonsingular matrix Φ, so our ignorance of θ incurs not only efficiency loss in estimating β, but slightly slower convergence.

2. Theorem 3 also implies a singularity in the limit distribution, whose covariance matrix has rank p only.

This is due to bias in $\hat{\beta}$, which on expansion is seen to have a term linear in $\hat{\theta}-\theta$ that dominates the contribution from $\Sigma_{u\in\mathbb{N}} f(u;\theta)x_u$.

But Theorem 3 does provide separate inference on β, though given Assumption 1 we cannot test zero restrictions.

3. If independence of the x_u is not assumed, the limiting covariance matrix in Theorem 3 can be consistently estimated (under additional conditions) by replacing $F(0)$ by a parametric or smoothed nonparametric estimate based on regression residuals.

4. The form of the limiting covariance matrix in Theorem 3, with dependence simply reflected in the factor $2\pi F(0)$, suggests that a generalized NLSE, which corrects parametrically or nonparametrically for correlation in x_u, affords no efficiency improvement (as in the.original polynomial model).

8. Monte Carlo study of finite-sample performance

A Monte Carlo study provides some information on finite sample performance.

Issues of concern, given unknown θ, are bias and variability of the NLSE, and accuracy of large sample inference rules suggested by Theorem 3.

We took $d = 2$, $p_1 = p_2 = 1$, and the x_u iid $\mathfrak{N}_1(0,1)$ normal variables, picking 2 $(\theta_1, \theta_2) = (\theta_{11}, \theta_{21})$ combinations - $(1,1)$, $(0.5, 2)$ - but throughout took $\Theta_{i1} = [-0.45, 4]$, $\beta_i = \beta_{i1} = 1$, $i = 1, 2$.

We varied N absolutely and also the relative n_1, n_2, taking $n_1, n_2 = (8, 12)$, $(10, 10), (11, 20), (15, 15)$.

Tables 1 and 2 report, for the respective parameter combinations, bias (BIAS), mean squared error (MSE), and empirical size at 5% (SIZE5) and 1% (SIZE1) for $\hat{\theta}_i$, $\hat{\beta}_i$, and also $\tilde{\beta}_i$, the LSE of β_i that correctly assumes θ, for $i = 1, 2$, across 1000 replications.

The sizes were proportions of significant estimates, using normal critical values scaled by estimated standard deviations, which in case of the $\hat{\theta}_i$, $\hat{\beta}_i$ were computed on the basis of Theorem 3 with current parameter estimates replacing true values of θ, β, and $2\pi F(0)$ replaced by the sum of squared residuals divided by N (so the spatial independence of the x_u was treated as known, as it was also in the, conventional, scaling used for the $\tilde{\beta}_i$).

The tables reveal a definite inferiority of the NLSE relative to the LSE, but unsurprisingly, as the LSE is exactly unbiased, more efficient, and yields exact critical regions.

Though the NLSE-based tests on β are nearly always over-sized, this phenomenon diminishes with increased N, and overall the discrepancy between the performances of the two classes of β estimate does not seem very serious.

There is also a predominate over-sizing of the tests on θ, but again this falls as N increases, and in Table 2, in particular, it is often modest.

There is a tendency for the NLSE to over-estimate, but for β, biases only exceed 2% of the parameter value when $n_i = 8$, $n_i = 12$, and for θ they never reach 1%, while overall they mostly fall with increasing N, as does the MSE.

Table 1: $\theta_1 = 1$, $\theta_2 = 1$, $\beta_1 = 1, \beta_2 = 1, \sigma^2 = 1$

n_1	n_2		$\widehat{\theta}_1$	$\widehat{\theta}_2$	$\widehat{\beta}_1$	$\widetilde{\beta}_1$	$\widehat{\beta}_2$	$\widetilde{\beta}_2$
8	12	BIAS	0.008	0.007	0.024	0.000	0.017	0.000
		MSE	0.016	0.007	0.080	0.001	0.051	0.000
		SIZE5	0.100	0.125	0.151	0.048	0.166	0.055
		SIZE1	0.044	0.048	0.075	0.010	0.084	0.010
10	10	BIAS	0.005	0.009	0.016	-0.001	0.009	0.002
		MSE	0.010	0.009	0.060	0.006	0.063	0.007
		SIZE5	0.132	0.132	0.180	0.053	0.186	0.051
		SIZE1	0.055	0.050	0.084	0.015	0.090	0.011
11	20	BIAS	-0.002	0.002	0.016	0.000	-0.007	0.000
		MSE	0.003	0.001	0.022	0.000	0.010	0.000
		SIZE5	0.086	0.104	0.115	0.039	0.120	0.051
		SIZE1	0.030	0.039	0.051	0.005	0.049	0.012
15	15	BIAS	0.003	0.002	0.006	0.000	-0.001	0.000
		MSE	0.002	0.002	0.013	0.000	0.013	0.000
		SIZE5	0.074	0.075	0.108	0.043	0.103	0.039
		SIZE1	0.024	0.022	0.033	0.010	0.037	0.010

Table 2: $\theta_1 = 2$, $\theta_2 = 1/2$, $\beta_1 = 1, \beta_2 = 1, \sigma^2 = 1$

n_1	n_2		$\widehat{\theta}_1$	$\widehat{\theta}_2$	$\widehat{\beta}_1$	$\widetilde{\beta}_1$	$\widehat{\beta}_2$	$\widetilde{\beta}_2$
8	12	BIAS	0.008	0.001	0.024	0.003	-0.002	-0.000
		MSE	0.014	0.001	0.071	0.005	0.001	0.000
		SIZE5	0.063	0.060	0.087	0.077	0.053	0.090
		SIZE1	0.028	0.012	0.038	0.029	0.014	0.034
10	10	BIAS	0.008	0.000	0.020	0.004	0.000	-0.000
		MSE	0.013	0.003	0.074	0.004	0.001	0.000
		SIZE5	0.069	0.057	0.101	0.058	0.065	0.039
		SIZE1	0.033	0.013	0.047	0.015	0.017	0.009
11	20	BIAS	0.005	-0.000	-0.001	-0.002	0.000	0.000
		MSE	0.005	0.000	0.028	0.002	0.000	0.000
		SIZE5	0.052	0.054	0.069	0.030	0.059	0.041
		SIZE1	0.017	0.012	0.017	0.012	0.011	0.006
15	15	BIAS	0.002	0.001	0.004	0.001	0.004	0.000
		MSE	0.004	0.001	0.025	0.001	0.000	0.000
		SIZE5	0.058	0.044	0.070	0.081	0.043	0.055
		SIZE1	0.018	0.011	0.019	0.019	0.010	0.020