Preliminary Test Estimation for Regression Models with Long-Memory Disturbance

Masanobu Taniguchi

School of Fundamental Science and Engineering, Waseda University, Japan

Hiroaki Ogata

School of International Liberal Studies, Waseda University, Japan

Hiroshi Shiraishi

School of Fundamental Science and Engineering, Waseda University, Japan

Short title

Preliminary test for long memory

For a class of time series regression models with long-memory disturbance, we are interested in estimation of a subset of the regression coefficient vector and spectral parameter of the residual process when the complementary subset is suspected to be close to $\mathbf{0}$. In this situation we evaluate the mean square errors of the restricted and unrestricted MLE and a preliminary test estimator when the complementary parameters are contiguous to zero vector. The results are expressed in terms of the regression spectra and the residual spectra. Since we assume long-memory dependence for the disturbance, the asymptotics are much different from the case of i.i.d. disturbance. Numerical studies elucidate some interesting features of regression and long-memory structures.

1   Introduction.

Suppose that one is interested in estimation of a subset $\boldsymbol{\theta}_1 \equiv (\theta_1, \ldots, \theta_{p_1})'$ of the concerned parameters $\theta_1, \ldots, \theta_{p_1+p_2}$ when the complementary subset $\boldsymbol{\theta}_2 = (\theta_{p_1+1}, \ldots, \theta_{p_1+p_2})'$ is suspected to be

1

redundant. In such a situation we are led to use the likelihood ratio test LR which tests $H : \boldsymbol{\theta}_2 = \boldsymbol{0}$ v.s. $A : \boldsymbol{\theta}_2 \neq \boldsymbol{0}$. Then it is natural to introduce a preliminary test (PT) estimator $\hat{\boldsymbol{\theta}}_1^{PT}$ based on LR. Let $\hat{\boldsymbol{\theta}}_1$ be the maximum likelihood estimator (MLE) of $\boldsymbol{\theta}_1$ under H, and $\tilde{\boldsymbol{\theta}}_1$ the MLE of $\boldsymbol{\theta}_1$ under A. Then $\hat{\boldsymbol{\theta}}_1^{PT} = \hat{\boldsymbol{\theta}}_1$ if LR does not reject H, and $\hat{\boldsymbol{\theta}}_1^{PT} = \tilde{\boldsymbol{\theta}}_1$ if LR rejects H. For independent observations, a lot of discussions have been addressed for the mean square error (MSE) comparison of $\hat{\boldsymbol{\theta}}_1$, $\tilde{\boldsymbol{\theta}}_1$ and $\hat{\boldsymbol{\theta}}_1^{PT}$ under $A_n : \boldsymbol{\theta}_2 = O(1/\sqrt{n})$, when n is the sample size. For diverse review of this topic, the reader may refer to Saleh (2006).

For dependent observations, Saleh (1992) established fundamental contributions for this problem. For a Gaussian AR($p_1 + p_2$) model with coefficients $\theta_1, \ldots, \theta_{p_1+p_2}$, he explicitly evaluated the MSE's of $\hat{\boldsymbol{\theta}}_1$, $\tilde{\boldsymbol{\theta}}_1$ and $\hat{\boldsymbol{\theta}}_1^{PT}$ under A_n, and made their comparisons. Maeyama et al (2008) also discussed such problems for linear processes with mean 0.

Recently much of attention has been paid to time series regression models with long-memory disturbances (e.g., Taniguchi and Kakizawa (2000)). There are a lot of long-memory phenomena. It is well known that many important economic variables are aggregates of a very large number of microvariables. In view of this, Granger (1980) gave an example showing that aggregation of short-memory process leads a long-memory time series. This paper deals with a class of time series regression models with regression coefficient vector $\boldsymbol{\beta}$ and fractional ARIMA residual process depending on the unknown coefficient vector $\boldsymbol{\theta}$. We are interested in estimation of a subset vector $\boldsymbol{\rho}_1$ of $(\boldsymbol{\theta}, \boldsymbol{\beta})$ when the complementary subset vector $\boldsymbol{\rho}_2$ is suspected to be close to $\boldsymbol{0}$. In this setting, similarly as the above, we introduce the three estimatiors $\hat{\boldsymbol{\rho}}_1$, $\tilde{\boldsymbol{\rho}}_1$ and $\hat{\boldsymbol{\rho}}_1^{PT}$ of $\boldsymbol{\rho}_1$. Under $A_n : \boldsymbol{\theta}_2 = O(1/\sqrt{n})$, by use of LAN results, we evaluate their MSE's explicitly in terms of the regression spectrum, the residual spectra and the innovation density. The regressor may include the polynomial and harmonic trends, and Gaussianity of the residual process is not assumed, hence, our model is very general. In our long-memory model, the asymptotic distribution and consistency order of the regression coefficient estimators are much different from these for short-memory or i.i.d. model (see, Lemma 1 and Example 1).

We evaluate the MSE's of three estimators by numerical studies in various settings. In the polynomial regression model and harmonic regression model, we can see that $\hat{\boldsymbol{\rho}}_1^{PT}$ is relatively robuster with respect to perturbation of parameters than the other two estimators.

The paper is organized as follows. Section 2 introduces our time series regression model, and describes the LAN theorem. Section 3 addresses the problem of preliminary test estimator, and introduces $\hat{\boldsymbol{\rho}}_1$, $\tilde{\boldsymbol{\rho}}_1$ and $\hat{\boldsymbol{\rho}}_1^{PT}$. By use of LAN theorem, Section 4 derives the MSE's of the three estimators. Section 5 provides some numerical studies and elucidates the features of MSE's of three estimators.

2   Settings.

Suppose that $\boldsymbol{Y}^n = (Y_1, \ldots, Y_n)'$ is generated by

$$Y_t = \boldsymbol{X}_t'\boldsymbol{\beta} + e_t, \quad t \in \boldsymbol{Z}, \tag{2.1}$$

where $\{X_t = (X_{t1}, \ldots, X_{tb})'\}$ is a sequence of b-dimensional nonrandom regressors, $\beta = (\beta_1, \ldots, \beta_b)'$ is the unknown regression coefficient and $\{e_t\}$ is a stationary long-memory process generated by the fractional ARIMA (FARIMA(p, d, q)) model :

$$\sum_{k=0}^{p} \phi_k B^k (1 - B)^d e_t = \sum_{k=0}^{q} \eta_k B^k \epsilon_t, \quad t \in \mathbf{Z}, \quad \phi_0 = \eta_0 = 1. \tag{2.2}$$

Here B stands for the lag operator and

$$\boldsymbol{\theta} = (\theta_1, \ldots, \theta_{p+q+1})' = (d, \phi_1, \ldots, \phi_p, \eta_1, \ldots, \eta_q)'$$

is the unknown parameter of the residual process, and $\{\epsilon_t\}$ is an i.i.d. innovation process with nonvanishing density $g(\cdot)$.

Initially we make the following assumptions.

(A1) The characteristic polynomials $\phi(z) \equiv \sum_{k=0}^{p} \phi_k z^k$ and $\eta(z) \equiv \sum_{k=0}^{q} \eta_k z^k$ have no roots within the unit disc $\mathcal{D} \equiv \{z \in \mathbf{C} : |z| \leq 1\}$.

(A2) $0 < d < 1/2$, and $\Theta = \{\boldsymbol{\theta} : \boldsymbol{\theta} \text{ satisfies } (A2) \text{ and } (A3)\}$.

(A3) $g = g(\cdot)$ satisfies that $\int z g(z)\, dz = 0$, $\sigma^2 \equiv \int z^2 g(z)\, dz < \infty$, and g is absolutely continuous with a.e. derivative g' satisfying

$$0 < \mathcal{F}(g) \equiv \int [g'(z)/g(z)]^2 g(z)\, dz < \infty,$$

and

$$\int [g'(z)/g(z)]^4 g(z)\, dz < \infty.$$

It is seen that (A3) implies $E[g'(\epsilon_t)/g(\epsilon_t)] = 0$, and, from (A1), $\psi(z) \equiv \phi(z)\eta(z)^{-1}(1 - z)^d$ can be expanded as the absolutely convergent sum :

$$\psi(z) = \sum_{k=0}^{\infty} \psi_k z^k, \quad z \in \mathcal{D}. \tag{2.3}$$

Hence, $\{e_t\}$ has the AR(∞) representation

$$\sum_{k=0}^{\infty} \psi_k e_{t-k} = \epsilon_t, \quad t \in \mathbf{Z}. \tag{2.4}$$

The spectral density of $\{e_t\}$ is expressed as

$$f_{\boldsymbol{\theta}}(\lambda) = \frac{\sigma^2 \left| \sum_{k=0}^{q} \eta_k e^{ik\lambda} \right|^2}{2\pi \left| 1 - e^{i\lambda} \right|^{2d} \left| \sum_{k=0}^{p} \phi_k e^{ik\lambda} \right|^2}. \tag{2.5}$$

3

On the regressor $\{X_t\}$, we impose a sort of Grenander's conditions [e.g., Hannan (1970)]. Letting

$$a_{kj}^n(h) = \begin{cases} \sum_{t=1}^{n-h} X_{t+h,k} X_{tj}, & h = 0, 1, \ldots, \\ \sum_{t=1-h}^{n} X_{t+h,k} X_{tj}, & h = -1, -2, \ldots, \end{cases}$$

the following conditions are assumed to hold:

(G1) $a_{kk}^n(0) \to \infty$ as $n \to \infty$, $k = 1, \ldots, b$.

(G2) $\lim_{n\to\infty} X_{n+1,k}^2 / a_{kk}^n(0) = 0$, $k = 1, \ldots, b$.

(G3) $\lim_{n\to\infty} a_{kj}^n(h) / \{a_{kk}^n(0) a_{jj}^n(0)\}^{1/2} = r_{kj}(h)$ exists for every $k, j = 1, \ldots, b$ and $h \in \mathbf{Z}$. Denote by $\mathbf{\Gamma}(h)$ the $b \times b$ matrix $\{r_{kj}(h)\}$.

(G4) $\mathbf{\Gamma}(0)$ is nonsingular.

Then there exists a Hermitian matrix function $\mathbf{M}(\lambda) = \{M_{jk}(\lambda); j, k = 1, \ldots, b\}$ with positive semidefinite increments such that

$$\mathbf{\Gamma}(h) = \int_{-\pi}^{\pi} e^{ih\lambda} d\mathbf{M}(\lambda) \tag{2.6}$$

We further make the following assumptions.

(G5) $a_{ll}^n(0) = O(n^{1+\alpha})$ for some $\alpha \geq 0$, and

$$\max_{1 \leq t \leq n} \frac{X_{tl}^2}{a_{ll}^n(0)} = O(n^{-\delta}), \quad l = 1, \ldots, b,$$

for some $\delta > 1 - 2d$.

(G6) (i) $X_{tk} = t^{k-1}$, $k = 1, \ldots, b_1, 0 \leq b_1 \leq b$, which implies $M_{kk}(0^+) - M_{kk}(0) = 1$, $k = 1, \ldots, b_1$,

  (ii) $0 < M_{kk}(0^+) - M_{kk}(0) < 1$, $k = b_1 + 1, \ldots, b_2$,

  (iii) $M_{kk}(0^+) - M_{kk}(0) = 0$, $k = b_2 + 1, \ldots, b$.

Let $\mathbf{X}_{\cdot k} \equiv (X_{1k}, \ldots, X_{nk})'$, $k = 1, \ldots, b$, and $\mathbf{D}_n \equiv diag\{n^{-d}\|\mathbf{X}_{\cdot 1}\|, \ldots, n^{-d}\|\mathbf{X}_{\cdot b_1}\|, \|\mathbf{X}_{\cdot b_1+1}\|, \ldots, \{\mathbf{X}_{\cdot b}\|\}$, where $\| \cdot \|$ denotes the Euclidean norm. We introduce the local sequences

$$\boldsymbol{\theta}^{(n)} = \boldsymbol{\theta} + n^{-1/2}\mathbf{h}, \quad \boldsymbol{\beta}^{(n)} = \boldsymbol{\beta} + \mathbf{D}_n^{-1}\mathbf{k}, \tag{2.7}$$

where $\mathbf{h} \in \mathbf{R}^{p+q+1}$, $\mathbf{k} \in \mathbf{R}^b$, and $\mathbf{u} = (\mathbf{h}', \mathbf{k}')'$ belongs to some open subset $\mathcal{H}$ of $\mathbf{R}^{p+q+b+1}$. The sequence of statistical experiments is

$$\mathcal{E}_n = \{\mathbf{R}^{\mathbf{Z}}, \mathcal{B}^{\mathbf{Z}}, \{P_{\boldsymbol{\theta},\boldsymbol{\beta}}^{(n)} : (\boldsymbol{\theta},\boldsymbol{\beta}) \in \Theta \times \mathbf{R}^b\}\}, \quad n \in \mathbf{N},$$

4

where $\mathcal{B}^{\boldsymbol{Z}}$ denotes the Borel σ-field on $\boldsymbol{R}^{\boldsymbol{Z}}$ and $P_{\boldsymbol{\theta},\boldsymbol{\beta}}^{(n)}$ the joint distribution of $\{\epsilon_s, s \le 0; Y_1, \ldots, Y_n\}$ characterized by the parameter value $(\boldsymbol{\theta}, \boldsymbol{\beta})$ and the innovation density g. Denote by $H_g^{(n)}(\boldsymbol{\theta}, \boldsymbol{\beta})$ the sequence of simple hypotheses $\{\{P_{\boldsymbol{\theta},\boldsymbol{\beta}}^{(n)}\}, n \in \boldsymbol{N}\}$. We write the log-likelihood ratio for $H_g^{(n)}(\boldsymbol{\theta}^{(n)}, \boldsymbol{\beta}^{(n)})$ with respect to $H_g^{(n)}(\boldsymbol{\theta}, \boldsymbol{\beta})$ as

$$\Lambda_g^{(n)}(\boldsymbol{\theta}, \boldsymbol{\beta}) \equiv \log \frac{dP_{\boldsymbol{\theta}^{(n)}, \boldsymbol{\beta}^{(n)}}^{(n)}}{dP_{\boldsymbol{\theta},\boldsymbol{\beta}}^{(n)}}. \tag{2.8}$$

To describe fundamental results we need the notation :

$$\boldsymbol{W}(\boldsymbol{\theta}) = \frac{1}{2\pi} \begin{pmatrix} \boldsymbol{W}_1(\boldsymbol{\theta}) & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{W}_2(\boldsymbol{\theta}) \end{pmatrix}, \quad (b \times b - matrix) \tag{2.9}$$

where $\boldsymbol{W}_1(\boldsymbol{\theta})$ is the $b_1 \times b_1$-matrix with (k, j)th entry,

$$\frac{\Gamma(k - d)\Gamma(j - d)\{(2k - 1)(2j - 1)\}^{-1/2}}{(\sigma^2/2\pi)|\eta(1)/\phi(1)|^2 \Gamma(k - 2d)\Gamma(j - 2d)(k + j - 1 - 2d)}, \tag{2.10}$$

and $\boldsymbol{W}_2(\boldsymbol{\theta})$ is the $(b - b_1) \times (b - b_1)$-matrix with (k, j)th entry,

$$\int_{-\pi}^{\pi} f_{\boldsymbol{\theta}}(\lambda)^{-1} dM_{k+b_1, j+b_1}(\lambda).$$

The Fisher information matrix of the residual process is given by

$$Q(\boldsymbol{\theta}) = \frac{1}{4\pi} \int_{-\pi}^{\pi} \frac{\partial}{\partial\boldsymbol{\theta}} \log f_{\boldsymbol{\theta}}(\lambda) \cdot \frac{\partial}{\partial\boldsymbol{\theta}'} \log f_{\boldsymbol{\theta}}(\lambda) \, d\lambda. \tag{2.11}$$

The following lemma is due to Hallin, Taniguchi, Serroukh and Choy (1999).

LEMMA 1. *(LAN result)* *Suppose that (A1) − (A3) and (G1) − (G6) hold. Then the sequence of experiments $\mathcal{E}_n$, $n \in \boldsymbol{N}$, is locally asymptotically normal (LAN). That is,*

(i) *For all $\boldsymbol{\theta}$ and $\boldsymbol{\beta}$, the log-likelihood ratio (2.8) admits, under $H_g^{(n)}(\boldsymbol{\theta}, \boldsymbol{\beta})$, as $n \to \infty$, the asymptotic expansion*

$$\Lambda_g^{(n)}(\boldsymbol{\theta}, \boldsymbol{\beta}) = (\boldsymbol{h}', \boldsymbol{k}')\Delta_g^{(n)}(\boldsymbol{\theta}, \boldsymbol{\beta}) - \frac{1}{2}\left[\sigma^2 \mathcal{F}(g)\boldsymbol{h}'Q(\boldsymbol{\theta})\boldsymbol{h} + \mathcal{F}(g)\boldsymbol{k}'\boldsymbol{W}(\boldsymbol{\theta})\boldsymbol{k}\right] + o_p(1), \tag{2.12}$$

where

$$\Delta_g^{(n)}(\boldsymbol{\theta}, \boldsymbol{\beta}) \equiv n^{-1/2} \begin{bmatrix} \sum_{t=1}^{n} \frac{g'(z_t)}{g(z_t)} \sum_{\nu=1}^{t-1} \frac{\partial}{\partial\boldsymbol{\theta}}(\psi_\nu)e_{t-\nu} \\ -\boldsymbol{D}_n^{-1} \sum_{t=1}^{n} \frac{g'(z_t)}{g(z_t)} \sum_{\nu=1}^{t-1} \psi_\nu \boldsymbol{X}_{t-\nu} \end{bmatrix} \tag{2.13}$$

and

$$z_t = z_t(\boldsymbol{\theta}, \boldsymbol{\beta}) \equiv \sum_{k=0}^{t-1} \psi_k(Y_{t-k} - \boldsymbol{X}_{t-k}'\boldsymbol{\beta}).$$

5

(ii) *Under* $H_g^{(n)}(\theta, \beta)$,

$$\Delta_g^{(n)}(\theta, \beta) \xrightarrow{d} N(0, \Psi_g), \tag{2.14}$$

where

$$\Psi_g = \Psi_g(\theta, \beta) \equiv \begin{pmatrix} \sigma^2 \mathcal{F}(g) Q(\theta) & 0 \\ 0 & \mathcal{F}(g) W(\theta) \end{pmatrix}. \tag{2.15}$$

3 The problems.

In this section we discuss estimation of the time series regression model (2.1) with long-memory disturbances when the complementary subset of (θ, β) is suspected to be redundant. Concretely, for the partitions $\theta = (\theta_1', \theta_2')'$ and $\beta = (\beta_1', \beta_2')'$, we are now interested in the estimation of θ_1 and β_1 when θ_2 and β_2 are suspected to be close to 0 (zero vector). In such a situation it is natural to consider an estimatior based on a test for the hypothesis

$$H: \ \theta_2 = 0 \text{ and } \beta_2 = 0 \quad \text{v.s.} \quad A: \ \theta_2 \neq 0 \text{ or } \beta_2 \neq 0 \tag{3.1}$$

For the stretch $(Y_1, \ldots, Y_n, X_1, \ldots, X_n)$, we introduce the following quasi log-likelihood:

$$l_n(\theta, \beta) = \sum_{t=1}^{n} \log g \left\{ \sum_{v=0}^{t-1} \psi_v (Y_{t-v} - X_{t-v}' \beta) \right\}, \tag{3.2}$$

where $\psi_v = \psi_v(\theta)$ is given in (2.3) and (2.4), and $\theta = (\theta_1, \theta_2)$ and $\beta = (\beta_1, \beta_2)$. Define the unrestricted quasi MLE (UQMLE) of θ and β by

$$(\tilde{\theta}_1, \tilde{\theta}_2, \tilde{\beta}_1, \tilde{\beta}_2) = \arg \max_{(\theta_1, \theta_2, \beta_1, \beta_2)} l_n(\theta, \beta), \tag{3.3}$$

and the restricted quasi MLE (RQMLE) of θ_1 and β_1 by

$$(\hat{\theta}_1, \hat{\beta}_1) = \arg \max_{(\theta_1, \beta_1)} l_n((\theta_1, 0), (\beta_1, 0)). \tag{3.4}$$

In what follows we permutate the components of θ and β, and write

$$\rho = \begin{pmatrix} \rho_1 \\ \cdots \\ \rho_2 \end{pmatrix} \begin{matrix} \}a_1 + b_1 \\ \\ \}a_2 + b_2 \end{matrix} = \begin{pmatrix} \theta_1 \\ \beta_1 \\ \cdots \\ \theta_2 \\ \beta_2 \end{pmatrix} \begin{matrix} \}a_1 \\ \}b_1 \\ \\ \}a_2 \\ \}b_2 \end{matrix}$$

and

$$\tilde{l}_n(\rho) = l_n(\theta, \beta),$$

6

where $a_1 + a_2 = p + q + 1$, $b_1 + b_2 = b$. We also rewrite the fundamental quantities $\Lambda_g^{(n)}(\theta, \beta)$, $\Delta_g^{(n)}(\theta, \beta)$ and $\Psi_g = \Psi_g(\theta, \beta)$ in Lemma 1 as $\tilde{\Lambda}_g^{(n)}(\rho)$, $\tilde{\Delta}_g^{(n)}(\rho)$ and $\tilde{\Psi}_g = \Psi_g(\rho)$ corresponding to the permutated vector ρ, respectively.

Now let us return the problem of testing (3.1):

$$H: \ \rho_2 = 0 \quad \text{v.s.} \quad A: \ \rho_2 \neq 0. \tag{3.5}$$

For this we use the following quasi log-likelihood ratio test

$$\mathcal{L}_n = 2\left[\tilde{l}_n(\tilde{\rho}_1, \tilde{\rho}_2) - \tilde{l}_n(\hat{\rho}_1, 0)\right], \tag{3.6}$$

where

$$\tilde{\rho}_1 = \begin{pmatrix} \tilde{\theta}_1 \\ \tilde{\beta}_1 \end{pmatrix}, \ \tilde{\rho}_2 = \begin{pmatrix} \tilde{\theta}_2 \\ \tilde{\beta}_2 \end{pmatrix} \ and \ \hat{\rho}_1 = \begin{pmatrix} \hat{\theta}_1 \\ \hat{\beta}_1 \end{pmatrix}. \tag{3.7}$$

Partition $\tilde{\Delta}_g^{(n)}(\rho)$ and $\tilde{\Psi}_g$ corresponding to ρ_1 and ρ_2, and write them as

$$\tilde{\Delta}_g^{(n)}(\rho) = \begin{pmatrix} Z_1 \\ \cdots \\ Z_2 \end{pmatrix} \begin{matrix} {\scriptstyle \}a_1 + b_1} \\ \\ {\scriptstyle \}a_2 + b_2} \end{matrix}, \tag{3.8}$$

$$\tilde{\Psi}_g = \begin{bmatrix} \overbrace{F_{11}}^{a_1+b_1} & \vdots & \overbrace{F_{12}}^{a_2+b_2} \\ \cdots & & \cdots \\ F_{21} & \vdots & F_{22} \end{bmatrix} \begin{matrix} {\scriptstyle \}a_1 + b_1} \\ \\ {\scriptstyle \}a_2 + b_2} \end{matrix}. \tag{3.9}$$

Similarly as in Taniguchi and Amano (2008) we can see that, under H,

$$\mathcal{L}_n = (Z_2 - F_{21}F_{11}^{-1}Z_1)' F_{22.1}^{-1} (Z_2 - F_{21}F_{11}^{-1}Z_1) + o_p(1), \tag{3.10}$$

where $F_{22.1} = F_{22} - F_{21}F_{11}^{-1}F_{12}$. Since

$$Z_2 - F_{21}F_{11}^{-1}Z_1 \xrightarrow{d} N(0, F_{22.1}), \quad \text{under } H, \tag{3.11}$$

we have

$$\mathcal{L}_n \xrightarrow{d} \chi^2_{a_2+b_2}, \quad \text{under } H. \tag{3.12}$$

Then we can discuss the estimation problem of ρ_1 when ρ_2 is suspected to be close to 0. In this case it is natural to introduce the preliminary test quasi maximum likelihood estimator (PTQMLE)

$$\hat{\rho}_1^{PT} \equiv \tilde{\rho}_1 - (\tilde{\rho}_1 - \hat{\rho}_1)I\{\mathcal{L}_n \leq \chi^2_{a_2+b_2}(\alpha)\}, \tag{3.13}$$

where $I\{\cdot\}$ is the indicatior function, and $\chi^2_{a_2+b_2}(\alpha)$ is the α-level critical value of $\chi^2_{a_2+b_2}$. The implication of $\hat{\rho}_1^{PT}$ is if the test $\mathcal{L}_n$ accept $H: \rho_2 = 0$, then $\hat{\rho}_1^{PT} = \hat{\rho}_1$ (restricted estimatior), and if not so, then $\hat{\rho}_1^{PT} = \tilde{\rho}_1$ (unrestricted estimatior).

4 Asymptotic Theory for Estimators.

This section discusses the asymptotics of $\hat{\rho}_1$, $\tilde{\rho}_1$ and $\hat{\rho}_1^{PT}$ when ρ_2 is contiguous to $\mathbf{0}$. Let

$$H_n \equiv \overbrace{\begin{pmatrix} \begin{matrix} \sqrt{n} & & 0 \\ & \ddots & \\ 0 & & \sqrt{n} \end{matrix} & \mathbf{0} \\ \mathbf{0} & D_n \end{pmatrix}}^{p+q+1 \quad\quad b} \begin{matrix} \}p+q+1 \\ \\ \}b \end{matrix} \quad , \tag{4.1}$$

which is a standardizing matrix for estimators of $(\boldsymbol{\theta}, \boldsymbol{\beta})$. Permutating the diagonal elements of H_n so that the resulting matrix becomes the standardizing matrix for the parameter (ρ_1, ρ_2). Write it as

$$\tilde{H}_n \equiv \overbrace{\begin{pmatrix} \tilde{F}_n & 0 \\ 0 & \tilde{G}_n \end{pmatrix}}^{a_1+b_1 \quad a_2+b_2} \begin{matrix} \}a_1+b_1 \\ \}a_2+b_2 \end{matrix} \tag{4.2}$$

where $\tilde{F}_n$ and $\tilde{G}_n$ correspond to the parameters ρ_1 and ρ_2, respectively.

Consider the problem of testing

$$H : \boldsymbol{\rho}_2 = \mathbf{0}, \quad \text{v.s.} \quad A_n : \boldsymbol{\rho}_2 = \tilde{G}_n^{-1} \cdot \boldsymbol{\xi}, \tag{4.3}$$

where $\boldsymbol{\xi} \in R^{a_2+b_2}$. Recalling the discussion of Section 3, it is seen that the log-likelihood ratio $\tilde{\Lambda}_g^{(n)}(\boldsymbol{\rho})$ between H and A_n has the stochastic expansion

$$\tilde{\Lambda}_g^{(n)}(\boldsymbol{\rho}) = \boldsymbol{\xi}' Z_2 - \frac{1}{2}\boldsymbol{\xi}' F_{22}\boldsymbol{\xi} + o_p(1) \quad \text{under } H. \tag{4.4}$$

Since the asymptotic covariance between $Z_2 - F_{21}F_{11}^{-1}Z_1$ and $\tilde{\Lambda}_g^{(n)}(\boldsymbol{\rho})$ is given by $F_{22.1}\boldsymbol{\xi}$, LeCam's third lemma implies that, under A_n,

$$\mathcal{L}_n \xrightarrow{d} \chi_{a_2+b_2}^2(\boldsymbol{\xi}' F_{22.1}\boldsymbol{\xi}), \tag{4.5}$$

where $\chi_{a_2+b_2}^2(\boldsymbol{\xi}' F_{22.1}\boldsymbol{\xi})$ is the noncentral χ^2-distribution with $(a_2 + b_2)$ degrees of freedom and noncentrality $\boldsymbol{\xi}' F_{22.1}\boldsymbol{\xi}$. Next we derive the contiguous asymptotic distribution of $\tilde{F}_n(\hat{\rho}_1 - \rho_1)$. Since, under H,

$$\tilde{F}_n(\hat{\rho}_1 - \rho_1) = F_{11}^{-1}Z_1 + o_p(1), \tag{4.6}$$

and the asymptotic covariance between $F_{11}^{-1}Z_1$ and $\tilde{\Lambda}_g^{(n)}(\boldsymbol{\rho})$ is $F_{11}^{-1}F_{12}\boldsymbol{\xi}$, hence, LeCam's third lemma yields

$$\tilde{F}_n(\hat{\rho}_1 - \rho_1) \xrightarrow{d} N(F_{11}^{-1}F_{12}\boldsymbol{\xi}, F_{11}^{-1}), \quad \text{under } A_n. \tag{4.7}$$

8

Under H, $F_{11}^{-1}Z_1$ and $Z_2 - F_{21}F_{11}^{-1}Z_1$ are asymptotically independent, which implies that $\tilde{F}_n(\hat{\rho}_1 - \rho_1)$ and $\mathcal{L}_n$ are so. Hence it follows that, under A_n,

$$\lim_{n\to\infty} P[\tilde{F}_n(\hat{\rho}_1 - \rho_1) \overset{(v)}{\le} \boldsymbol{x}, \ \mathcal{L}_n < \chi^2_{a_2+b_2}(\alpha)]$$
$$= \Phi_{F_{11}^{-1}}(\boldsymbol{x} - F_{11}^{-1}F_{12}\boldsymbol{\xi}) \, Chi_{(a_2+b_2)}[\chi^2_{a_2+b_2}(\alpha); \boldsymbol{\xi}'F_{22.1}\boldsymbol{\xi}], \qquad (4.8)$$

where $\boldsymbol{x} \in R^{a_1+b_1}$, $\overset{(v)}{\le}$ implies componentwise inequality $\le$, $\Phi_{F_{11}^{-1}}(\cdot)$ is the (a_1+b_1)-dimensional normal distribution function with mean $\boldsymbol{0}$ and covariance matrix F_{11}^{-1} and $Chi_{(a_2+b_2)}[\ \cdot\ ; \boldsymbol{\xi}'F_{22.1}\boldsymbol{\xi}]$ is the distribution function of $\chi^2_{a_2+b_2}(\boldsymbol{\xi}'F_{22.1}\boldsymbol{\xi})$. Next we derive the contiguous asymptotic distribution of $\tilde{F}_n(\tilde{\rho}_1 - \rho_1)$. Note that

$$\tilde{\Psi}_g^{-1} = \left[\begin{array}{cc} F_{11}^{-1} + F_{11}^{-1}F_{12}F_{22.1}^{-1}F_{21}F_{11}^{-1} & -F_{11}^{-1}F_{12}F_{22.1}^{-1} \\ -F_{22.1}^{-1}F_{21}F_{11}^{-1} & F_{22.1}^{-1} \end{array} \right]. \qquad (4.9)$$

Similarly as in Taniguchi and Amano (2008), we observe that, under H,

$$\left[\begin{array}{c} \tilde{F}_n(\tilde{\rho}_1 - \rho_1) \\ \tilde{G}_n(\tilde{\rho}_2 - \rho_2) \end{array} \right] = \tilde{\Psi}_g^{-1} \left[\begin{array}{c} Z_1 \\ Z_2 \end{array} \right] + o_p(1), \qquad (4.10)$$

leading to, together with (4.9),

$$\begin{aligned} \tilde{F}_n(\tilde{\rho}_1 - \rho_1) &= F_{11}^{-1}Z_1 - F_{11}^{-1}F_{12}F_{22.1}^{-1}\{Z_2 - F_{21}F_{11}^{-1}Z_1\} + o_p(1) \\ &= \mathcal{X}_0 - F_{11}^{-1}F_{12}\mathcal{Z}_0 + o_p(1), \quad (say). \end{aligned} \qquad (4.11)$$

LeCam's third lemma leads to

$$\left(\begin{array}{c} \mathcal{X}_0 \\ \mathcal{Z}_0 \end{array} \right) \sim N\left[\left(\begin{array}{c} F_{11}^{-1}F_{12}\boldsymbol{\xi} \\ \boldsymbol{\xi} \end{array} \right), \left(\begin{array}{cc} F_{11}^{-1} & 0 \\ 0 & F_{22.1}^{-1} \end{array} \right) \right]$$

under A_n. Then the asymptotic mean of $\mathcal{X}_0 - F_{11}^{-1}F_{12}\mathcal{Z}_0$ is equal to $\boldsymbol{0}$ under A_n. Hence we may rewrite $\mathcal{X}_0 - F_{11}^{-1}F_{12}\mathcal{Z}_0$ as $\mathcal{X} - F_{11}^{-1}F_{12}\mathcal{Z}$, where

$$\left(\begin{array}{c} \mathcal{X} \\ \mathcal{Z} \end{array} \right) \sim N\left[\left(\begin{array}{c} 0 \\ 0 \end{array} \right), \left(\begin{array}{cc} F_{11}^{-1} & 0 \\ 0 & F_{22.1}^{-1} \end{array} \right) \right]$$

under A_n. Noting the correspondence $\mathcal{X} \leftrightarrow \boldsymbol{x}$ and $\mathcal{Z} \leftrightarrow \boldsymbol{z}$, we have, under A_n,

$$\lim_{n\to\infty} P[\tilde{F}_n(\tilde{\rho}_1 - \rho_1) \overset{(v)}{\le} \boldsymbol{x}, \ \mathcal{L}_n \ge \chi^2_{a_2+b_2}(\alpha)]$$
$$= \int_{\{z:(z+\xi)'F_{22.1}(z+\xi)\ge\chi^2_{a_2+b_2}(\alpha)\}} \Phi_{F_{11}^{-1}}(\boldsymbol{x} + F_{11}^{-1}F_{12}\boldsymbol{z}) \, d\Phi_{F_{22.1}^{-1}}(\boldsymbol{z}). \qquad (4.12)$$

From (4.8) and (4.12), the contiguous asymptotic distribution of $\tilde{F}_n(\hat{\rho}_1^{PT} - \rho_1)$ is (4.8) + (4.12).

We evaluate the mean square error (MSE) of standardized estimator $(*)$ by $M(*) \equiv E[tr\{(*)(*)'\}]$. From the above arguments, Saleh (1992) and Maeyama et al (2008), we have,

9

Theorem 1. *Under (A1)-(A3), (G1)-(G6), it holds that*

$$(i)\ M(\hat{\boldsymbol{\rho}}_1) = tr\boldsymbol{F}_{11}^{-1} + \boldsymbol{\xi}'\boldsymbol{F}_{21}\boldsymbol{F}_{11}^{-2}\boldsymbol{F}_{12}\boldsymbol{\xi}, \tag{4.13}$$

$$(ii)\ M(\tilde{\boldsymbol{\rho}}_1) = tr\boldsymbol{F}_{11}^{-1} + tr\{\boldsymbol{F}_{11}^{-1}\boldsymbol{F}_{12}\boldsymbol{F}_{22.1}^{-1}\boldsymbol{F}_{21}\boldsymbol{F}_{11}^{-1}\}, \tag{4.14}$$

$$(iii)\ M(\hat{\boldsymbol{\rho}}_1^{PT}) = tr\boldsymbol{F}_{11}^{-1} + tr\{\boldsymbol{F}_{11}^{-1}\boldsymbol{F}_{12}\boldsymbol{F}_{22.1}^{-1}\boldsymbol{F}_{21}\boldsymbol{F}_{11}^{-1}\}\{1 - Chi_{(a_2+b_2+2)}[\chi^2_{a_2+b_2}(\alpha);\boldsymbol{\xi}'\boldsymbol{F}_{22.1}\boldsymbol{\xi}]\}$$

$$+(\boldsymbol{\xi}'\boldsymbol{F}_{21}\boldsymbol{F}_{11}^{-2}\boldsymbol{F}_{12}\boldsymbol{\xi})\{2Chi_{(a_2+b_2+2)}[\chi^2_{a_2+b_2}(\alpha);\boldsymbol{\xi}'\boldsymbol{F}_{22.1}\boldsymbol{\xi}]$$

$$-Chi_{(a_2+b_2+4)}[\chi^2_{a_2+b_2}(\alpha);\boldsymbol{\xi}'\boldsymbol{F}_{22.1}\boldsymbol{\xi}]\}. \tag{4.15}$$

5   Numerical Studies.

In this section we investigate the MSE properties of $\hat{\boldsymbol{\rho}}_1$, $\tilde{\boldsymbol{\rho}}_1$ and $\hat{\boldsymbol{\rho}}_1^{PT}$ numerically.

Example 1 (Polynomial regression model).   Recall our model (2.1):

$$Y_t = (X_{t1},\dots,X_{tb})\boldsymbol{\beta} + e_t, \tag{5.1}$$

where $\{e_t\} \sim FARIMA(p,d,q)$ with spectral density $f_{\boldsymbol{\theta}}(\lambda)$ given in (2.5). Suppose that $\boldsymbol{\beta} = (\boldsymbol{\beta}'_{b_1}, \boldsymbol{\beta}'_{b_2})'$ is unknown, the structure of $f_{\boldsymbol{\theta}}(\lambda)$ is known, and $X_{tk} = t^{k-1}$, $k = 1,\dots,b_1 + b_2$. We are now interested in estimation of $\boldsymbol{\rho}_1 = \boldsymbol{\beta}_{b_1}$ when $\boldsymbol{\rho}_2 = \boldsymbol{\beta}_{b_2}$ is suspected to be close to $\boldsymbol{0}$. In this case, $\tilde{\boldsymbol{F}}_n = diag\left(\cdots, \frac{n^{k-1/2}}{\sqrt{2k-1}}, \cdots\right)$, $k = 1,\dots,b_1$, (e.g. Taniguchi and Kakizawa (2000, p348)). From (2.10) it is seen that the matrix $\tilde{\boldsymbol{\Psi}}_g$ becomes

$$j\left[\cdots\ \underset{k}{\overset{\vdots}{\frac{\Gamma(k-d)\Gamma(j-d)\{(2k-1)(2j-1)^{-1/2}\}}{\sigma^2|\eta(1)/\phi(1)|^2\Gamma(k-2d)\Gamma(j-2d)(k+j-1-2d)}}}\ \cdots \atop \vdots \right] \times \mathcal{F}(g) = \left[\begin{array}{cc} \boldsymbol{F}_{11} & \boldsymbol{F}_{12} \\ \boldsymbol{F}_{21} & \boldsymbol{F}_{22} \end{array}\right]\begin{array}{c}\}b_1 \\ \}b_2\end{array}.$$

When $\boldsymbol{\beta}_2 = \tilde{\boldsymbol{G}}_n^{-1}\boldsymbol{\xi}$, Theorem 1 provided MSE's for the three estimators.

In the following, we see the magnitudes of MSE's of the three estimators in various cases. Suppose that $p = q = 1$, $\phi_1 = 0.4$ and $\eta_1 = 0.2$ in the model (2.2) and that $a_1 = a_2 = 0$, $b_1 = b_2 = 1$, $\alpha = 0.05$, $\sigma^2 = 1$ and g is the Gaussian density function. Figures 1 and 2 show the magnitudes of MSE's of the three estimators as functions of $0 < \xi < 10$ when $d = 0.1$ and $d = 0.4$, respectively. When ξ is samll, the RQMLE is the best and UQMLE is the worst, but when ξ becomes larger, the RQMLE becomes worst and the UQMLE and the PTQMLE has smaller MSE's than that of RQMLE. From this figure, we can see that the PTQMLE has moderate MSE than the other two estimators.

Example 2 (Harmonic regression model).   In the model (5.1), suppose that $X_{tk} = \cos v_k t$, where $0 < v_1 < \cdots < v_{b_1+b_2} < \pi$, and that the long memory parameter d is equal to 0. Here we are

interested in estimation of $\boldsymbol{\rho}_1 \equiv (\boldsymbol{\theta}'_{a_1}, \boldsymbol{\beta}'_b)'$ when $\boldsymbol{\rho}_2 \equiv \boldsymbol{\theta}_{a_2}$ is suspected to be close to $\mathbf{0}$. In this case,

$$\tilde{\boldsymbol{F}}_n = diag(\overbrace{\sqrt{n}, \ldots, \sqrt{n}}^{a_1}, \overbrace{\sqrt{n}/\sqrt{2}, \ldots, \sqrt{n}/\sqrt{2}}^{b}).$$

From Lemma 1, it is seen that

$$
\begin{aligned}
&\tilde{\boldsymbol{\Psi}}_g \\
&= \left[\begin{array}{cc|c}
\frac{\sigma^2 \mathcal{F}(g)}{4\pi} \int_{-\pi}^{\pi} \frac{\partial \log f_{\boldsymbol{\theta}}(\lambda)}{\partial \theta_{a_1}} \frac{\partial \log f_{\boldsymbol{\theta}}(\lambda)}{\partial \theta'_{a_1}} d\lambda & \mathbf{0} & \frac{\sigma^2 \mathcal{F}(g)}{4\pi} \int_{-\pi}^{\pi} \frac{\partial \log f_{\boldsymbol{\theta}}(\lambda)}{\partial \theta_{a_1}} \frac{\partial \log f_{\boldsymbol{\theta}}(\lambda)}{\partial \theta'_{a_2}} d\lambda \\
\mathbf{0} & \frac{\mathcal{F}(g)}{2\pi} \int_{-\pi}^{\pi} f_{\boldsymbol{\theta}}(\lambda)^{-1} d\boldsymbol{M}(\lambda) & \mathbf{0} \\
\hline
\frac{\sigma^2 \mathcal{F}(g)}{4\pi} \int_{-\pi}^{\pi} \frac{\partial \log f_{\boldsymbol{\theta}}(\lambda)}{\partial \theta_{a_2}} \frac{\partial \log f_{\boldsymbol{\theta}}(\lambda)}{\partial \theta'_{a_1}} d\lambda & \mathbf{0} & \frac{\sigma^2 \mathcal{F}(g)}{4\pi} \int_{-\pi}^{\pi} \frac{\partial \log f_{\boldsymbol{\theta}}(\lambda)}{\partial \theta_{a_2}} \frac{\partial \log f_{\boldsymbol{\theta}}(\lambda)}{\partial \theta'_{a_2}} d\lambda
\end{array} \right] \\
&= \left[\begin{array}{c|c}
\boldsymbol{F}_{11} & \boldsymbol{F}_{12} \\
\hline
\boldsymbol{F}_{21} & \boldsymbol{F}_{22}
\end{array} \right] \begin{array}{l} \}a_1 + b \\ \}a_2 \end{array}
\end{aligned} \tag{5.2}
$$

When $\boldsymbol{\theta}_{a_2} = \tilde{\boldsymbol{G}}_n^{-1} \boldsymbol{\xi}$, Theorem 1 provided MSE's for the three estimators.

Figure 3 shows the magnitudes of MSE's of the three estimators as functions $0.5 < \theta_2 < 1$. The other parameters are fixed as $\xi = 1$, $\theta_1 = 0$ and $\nu_1 = \pi/3$. When θ_2 is near from 0, the RQMLE is the best and UQMLE is the worst, but when θ_2 is near from 1, the RQMLE becomes worst and the UQMLE and the PTQMLE has smaller MSE's. From this figure, we can conclude that the PTQMLE is moderate and hardly affected by the parameter θ_2 while the RQMLE and UQMLE is sensitive. Therefore, we can conclude that PTQMLE is robust with respect to perturbation of parameter θ_2.

REFERENCES

[1] Granger, C.W.J. (1980). Long memory relationships and the aggregation of dynamic models. *J. Econometrics.* No **14**, 227-238.

[2] Hallin, M., Taniguchi, M., Serroukh, A. and Choy, K. (1999). Local asymptotic normality for regression models with long-memory disturbance. *Ann. Statist.* No **27-6**, 2054-2080.

[3] Hannan, E.J. (1970). *Multiple Time Series.* Wiley, New York.

[4] Maeyama, Y., Tamaki, K. and Taniguchi, M. (2008). Premilinary test estimation for spectra and its applications to financial hedging problem. *Waseda University Time Series Discussion Paper (WUTS 41).*

[5] Saleh, A. K. Md. E. (1992). On shrinkage estimation of the parameters of an autoregressive Gaussian process. *Theory. Prob. Appl.* **37** 250-260.

[6] Saleh, A. K. Md. E. (2006). *Theory of Preliminary Test and Stein-Type Estimation with Applications.* Wiley-Interscience.

[7] Taniguchi, M. (1985). An asymptotic expansion for the distribution of the likelihood ratio criterion for a Gaussian autoregressive moving average process under a local alternative. *Econometric Theory.* **1** 73-84.

[8] Taniguchi, M. and Amano, T (2007). Systematic approach for portmanteau tests in view of Whittle likelihood ratio. *Waseda University Time Series Discussion Paper (WUTS 37).*

[9] Taniguchi, M. and Kakizawa,Y. (2000). *Asymptotic Theory of Statistical Inference for Time Series.* Springer-Verlag, New York.
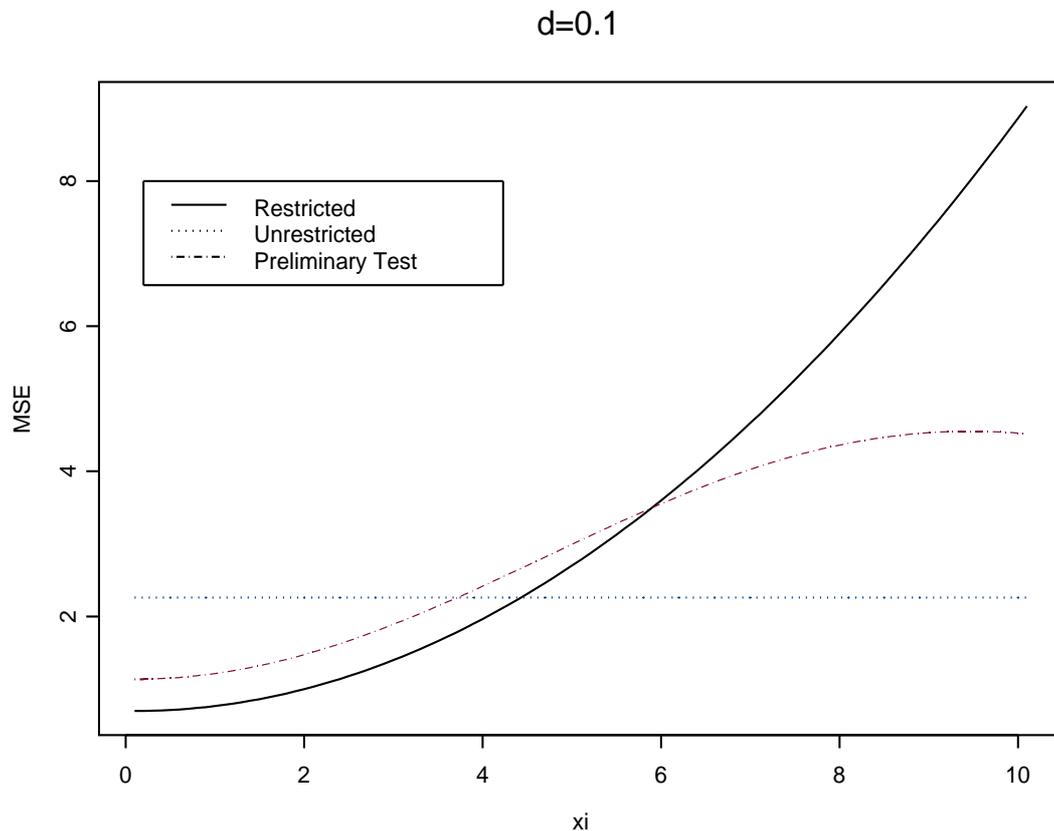
d=0.1

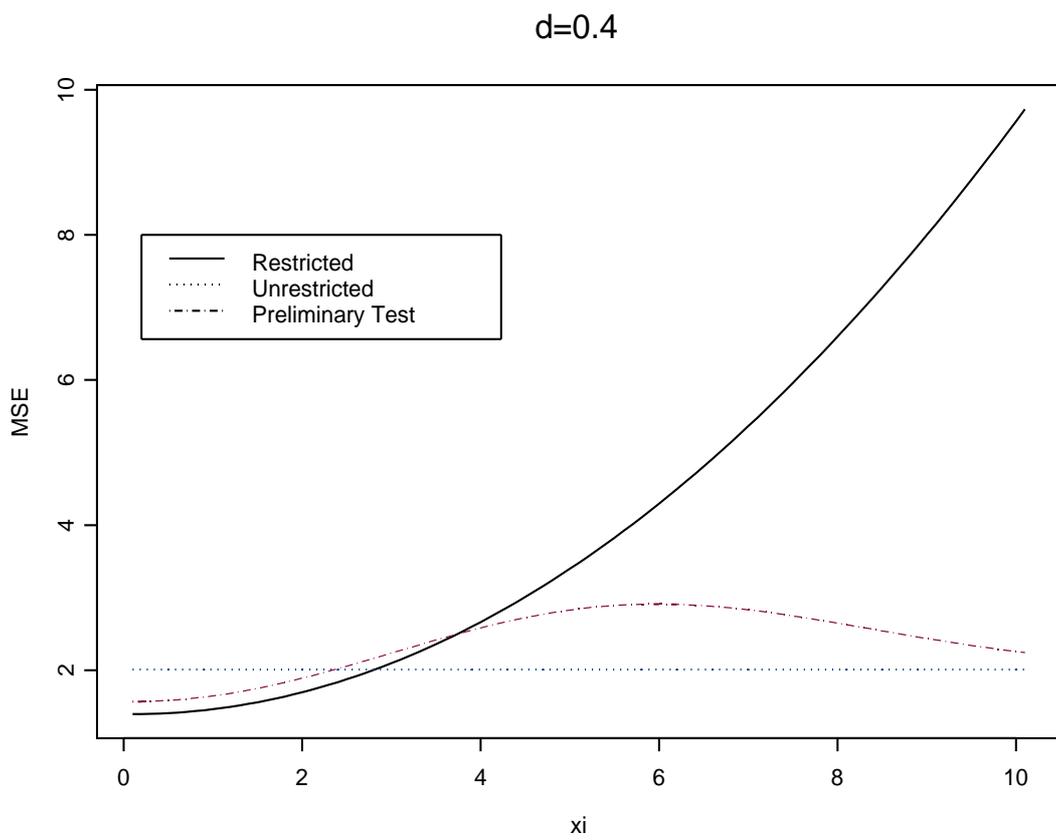Figure 1: MSEs in the case of $d = 0.1$
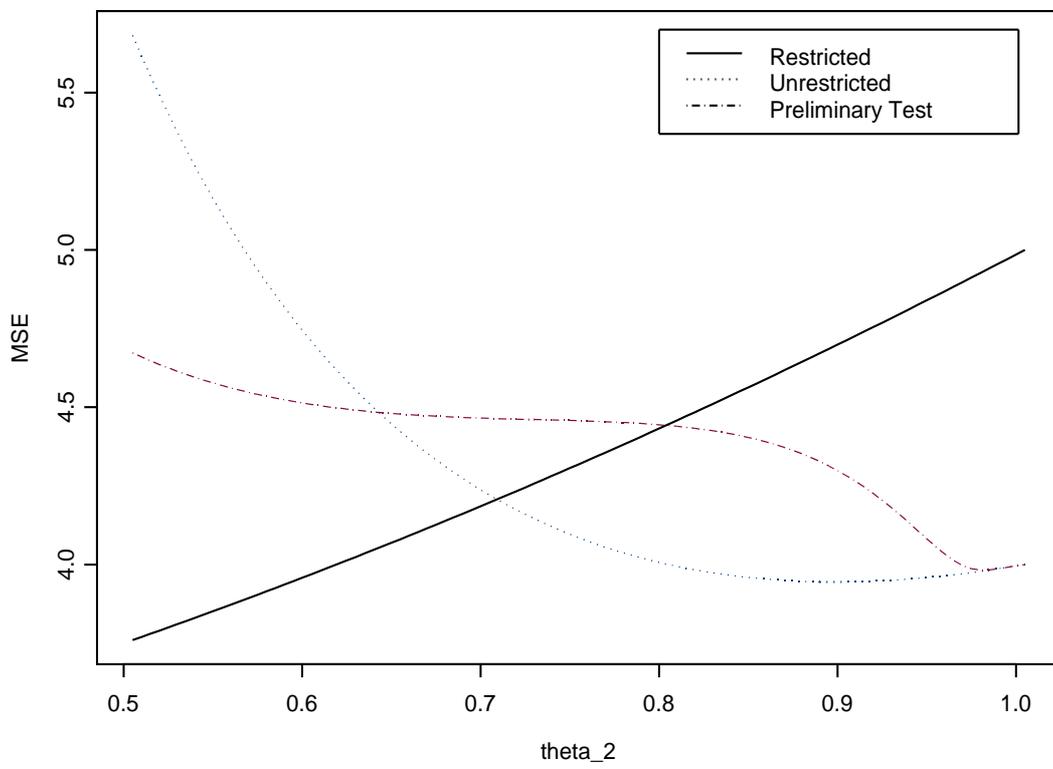
Figure 2: MSEs in the case of $d = 0.4$

theta_1=0, nu=pi/3

Figure 3: MSEs as functions of $0.5 < \theta_2 < 1.0$