Consequences of Fractionally Integrated Regressors

Werner Ploberger
Department of Economics, University of Rochester,
Rochester, NY 14627-0156, USA.

E-mail: wplo@troi.cc.rochester.edu

August 1999

Preliminary version

Abstract

This paper analyzes linear models. It investigates the difference between the sum of squares of the residuals and the sum of squares of the prediction errors when the parameter is estimated consecutively: In case the regressors are "fractionally integrated" (in a very broad sense) it is shown that the asymptotic behavior of this difference is determined by the order of integration of the regressors.

1 Introduction

Most of the papers of the conference focus on some aspect of the analysis of fractionally integrated processes themselves, i.e. one wants to make inference on the parameter of integration or test hypotheses concerning the degree of integration.

This paper focuses on the consequences of regressors being "fractionally integrated": Suppose one has given a standard linear model

$$y_t = x_t'\beta + u_t$$

with all the usual assumptions fulfilled. Then it is well-known that $\widehat{y_t} = x_t'\beta$ is the best predictor for y_t given x_t: The problem, however, is that in general we do not know β and therefore have to estimate it: Hence - in almost all realistic situations - one has to replace $\widehat{y_t}$ by some other function $\widehat{\widehat{y_t}}$ depending on the available information only:

$$\widehat{\widehat{y_t}} = \widehat{\widehat{y_t}}(x_t, x_{t-1}, ..., x_1, y_{t-1}, ..y_1)$$

We will analyze the use of

$$\widehat{\widehat{y_t}} = x_t'\widehat{\beta_{t-1}} \tag{1}$$

where $\widehat{\beta_{t-1}}$ is the OLS-estimator for β given $x_{t-1}, ..., x_1, y_{t-1}, ..y_1$.

If the usual conditions for consistency of the OLS-estimator are satisfied it is easily seen that

$$\widehat{y_t} - \widehat{\widehat{y_t}} \to 0$$

So asymptotically the difference between theoretically best and realistic predictor should converge to zero. This is, however, not true anymore for the *compound* prediction error (This result goes back to some work of Dawid: For a rather complete list of references one should consult e.g. Gerencer&Rissanen(1992)): Indeed one can show in the stationary case (i.e. if

$$\frac{1}{n}\sum_{t\leq n} x_t x_t' \to R,$$

where R is nonsingular) that when using the predictor (1)

$$\frac{\sum_{t\leq n}\left(y_t - \widehat{\widehat{y_t}}\right)^2 - \sum_{t\leq n}\left(y_t - \widehat{y_t}\right)^2}{p\log n} \to \frac{1}{2} \tag{2}$$

where p is the dimension of the vector β. The qualitative nature of this result sounds plausible: The more parameters we have to estimate, the bigger our "loss" in predictive accuracy will be. Moreover, results like the one above are often used to justify criteria for order estimation like e.g. BIC or FPE.

We are now going to generalize this result. We have, however, to bear in mind that (2) is wrong in the case of nonstationary regressors: This is an easy consequence of the results derived in Ploberger&Phillips(1998): Their findings show that the nature of the nonstationarity influences the limiting relationship (2).

So let us now assume that our processes x_t and u_t satisfy the following assumptions:

1. The x_t fulfill the following requirement: There exist diagonal matrices $D_n = diag(n^{\alpha_1}, ..., n^{\alpha_p})$ so that

$$D_n^{-1} \sum_{t \leq n} x_t x_t' D \rightarrow_D R \tag{3}$$

where $\rightarrow_D$should denote convergence in distribution and R is a random matrix which is a.s. positive definite

2. The u_t are i.i.d $G(0, \sigma^2)$ and independent of the x_t.

3. The first differences $x_t - x_{t-1}$ are stationary, ergodic, have nonsingular variance-covariance matrix and there exists a $\kappa > 0$ so that

$$E \left\| x_t - x_{t-1} \right\|^{2+\kappa} < \infty$$

Remark 1.1 *Assumption 1 is fulfilled if the processes are stationary - in this case simply take all the α to be equal $\frac{1}{2}$. Moreover, we also cover the case where the components are independent fractionally integrated process: for a recent survey cf Robinson(1994): in these cases one has $\alpha > \frac{1}{2}$ depending on the degree of integration. Our assumption may not be fulfilled if the process is "fractionally cointegrated", but then we may be able to "rotate" our x_t so that we end up with a process following our assumption.*

Remark 1.2 *Assumption 2 is relatively restrictive, especially the Gaussianity of the error terms seems very restrictive: On the other hand, this special form of the error terms makes it possible to use the result from Phillips-Ploberger, which simplifies the proof enormously.*

Remark 1.3 *The last of our conditions is of only technical nature: I think it would be possible to live without it - but only at the expense of a more difficult proof: Moreover, I think the restriction to processes where the first differences are stationary is not that stringent, Moreover, the proof below shows is easily generalized to integrated processes of higher order.*

2 The Main Theorem

Let us now formulate and prove the following theorem:

Theorem 2.1 *Assume all of the above assumptions hold true. Then*

$$P - \lim_{n\to\infty} \frac{\sum_{t\le n}\left(y_t - \widehat{\widehat{y}}_t\right)^2 - \sum_{t\le n}\left(y_t - \widehat{y}_t\right)^2}{\sigma^2 \log n \cdot \sum \alpha_i} = 1 \tag{4}$$

So let us prove (4):
We will do this in three steps:
Let us first define

$$\Delta_n = \sum_{t\le n}\left(y_t - \widehat{\widehat{y}}_t\right)^2 - \sum_{t\le n}\left(y_t - \widehat{y}_t\right)^2$$

First we establish that for all $\varepsilon > 0$

$$P\left[\frac{\Delta_n}{\sigma^2 \log n \cdot \sum \alpha_i} < 1 - \varepsilon\right] \to 0 \tag{5}$$

Then we construct random variables A_n so that

$$\Delta_n + A_n \ge 0 \tag{6}$$

and

$$A_n / \log n \to 0 \tag{7}$$

in probability and

$$E(\Delta_n + A_n / x_t, x_{t-1}, ..., x_1)/(\log n \cdot \sum \alpha_i) \to 1 \tag{8}$$

in probability;
This would prove our theorem: (5) and (7) imply

$$P\left[\frac{\Delta_n + A_n}{\sigma^2 \log n \cdot \sum \alpha_i} < 1 - \varepsilon\right] \to 0 \tag{9}$$

Moreover, if for some $\eta > 0$

$$\delta = \limsup P\left[\sigma^2 > 1 + \eta\right] > 0$$

then we could easily see that - because of (6) and (9)

$$\limsup P\left[E(\Delta_n + A_n/x_t, x_{t-1}, ..., x_1)/(\sigma^2 \log n \cdot \sum \alpha_i) > 1 + \frac{\delta\eta}{2}\right] > 0$$

which would contradict (8).

So let us first establish (5): Let us denote for $\beta \in \mathbf{R}^p$ by P_β the probability measure on the space of $x_n, x_{n-1}, ..., x_1, y_{n-1}, ..y_1$ when $y_t = x_t'\beta + u_t$. Keeping in mind that the u_t are Gaussian, it is an easy exercise of using the results of sections 4 and 5 of Ploberger and Phillips(1998) to show that for all $\varepsilon, \alpha > 0$ and each compact set K the Lebesgue measure of

$$\left\{\beta \in K : P_\beta\left(\left[\frac{\Delta_n}{\sigma^2 \log\det \sum_{t\le n} x_t x_t'} \ge \frac{1-\varepsilon}{2}\right]\right) \ge \alpha\right\}$$

converges to zero. It is now an elementary exercise to show that the distribution of P_β is invariant of β. Therefore we may easily conclude that $P\left(\left[\frac{\Delta_n}{\sigma^2 \log\det \sum_{t\le n} x_t x_t'} \ge \frac{1-\varepsilon}{2}\right]\right) \to 0.$

Moreover, directly form (3) we see that

$$\frac{\log\det \sum_{t\le n} x_t x_t'}{\log n} \to 2\sum \alpha_i \tag{10}$$

(as R was assumed to be nonsingular!).

Therefore it remains to establish (6),(7),(8): For that purpose, the following Lemma is helpful:

Lemma 2.2 *Let* $R_t = \sum_{s\le t} x_s x_s'$. *Then*

$$\frac{\sum_{t\le n} x_t' R_{t-1}^{-1} x_t}{\log\det R_t} \to 1$$

Let us prove the Lemma: We have

$$\begin{aligned} \log\det R_t - \log\det R_{t-1} &= \\ \log\det \sqrt{R_{t-1}^{-1}}\left(R_{t-1}^{-1} + x_t x_t'\right)\sqrt{R_{t-1}^{-1}} &= \log\det(I + \sqrt{R_{t-1}^{-1}} x_t x_t' \sqrt{R_{t-1}^{-1}}) \end{aligned}$$

Elementary analysis shows that $\log\det(I + A) - tr(A) = o(A)$ uniformly in A for $A \to 0$: Hence for every $\eta > 0$ there exists a $\delta = \delta(\eta) > 0$ so that for all A with $\|A\| < \delta$

$$(1-\eta)\, trA \le \log\det(I + A) \le (1+\eta) trA$$

So suppose we could show that

$$tr(\sqrt{R_{t-1}^{-1}}\,(x_t x_t')\,\sqrt{R_{t-1}^{-1}}) \to 0 \tag{11}$$

(as $\sqrt{R_{t-1}^{-1}}\left(R_{t-1}^{-1} + x_t x_t'\right)\sqrt{R_{t-1}^{-1}}$ is nonnegative definite this proves that the norm of the matrix converges to zero, too): then for every $\eta > 0$ there exists a $K = K(\eta)$ so that for $t > K$ $\left\|\sqrt{R_{t-1}^{-1}}\left(R_{t-1}^{-1} + x_t x_t'\right)\sqrt{R_{t-1}^{-1}}\right\| < \delta\left(\eta\right)$ and therefore

$$\begin{aligned}
&(1-\eta)\left(\log\det R_t - \log\det R_K\right) \\
&\leq \quad \sum_{t\leq n} x_t' R_{t-1}^{-1} x_t \\
&\leq \quad (1+\eta)\left(\log\det R_t - \log\det R_K\right) + \log\det R_K
\end{aligned}$$

This would prove our lemma, since we know from (10) that $\log\det R_t \to \infty$.

Therefore it remains to establish (11): first observe that

$$tr(\sqrt{R_{t-1}^{-1}}\,(x_t x_t')\,\sqrt{R_{t-1}^{-1}} = x_t' R_{t-1}^{-1} x_t \tag{12}$$

It is easily seen that

$$(x_t - x_{t-1})(x_t - x_{t-1})' \leq 2\left(x_t x_t' + x_{t-1} x_{t-1}'\right)$$

so that

$$\frac{1}{4}\frac{1}{n}\sum_{t\leq n}(x_t - x_{t-1})(x_t - x_{t-1})' \leq \frac{1}{n} R_n$$

According to our assumptions $x_t - x_{t-1}$ is ergodic and variance-covariance matrix of its components is nonsingular: Therefore the left-handside of the above inequality converges to some nonsingular matrix and we may conclude that there exists a nonsingular matrix S so that for all but a finite number of $n \in \mathbf{N}$

$$R_n \geq nS \tag{13}$$

Now let us suppose (12) is false: Then there exists an $\varepsilon > 0$ and an infinite sequence $J \subset \mathbf{N}$ so that

$$x_n' R_{n-1}^{-1} x_n \geq \varepsilon \tag{14}$$

for $n \in I$. With the help of (13) we may conclude that $x_n' x_n \geq \varepsilon n$ for $n \in J$. Let us now define M to be a natural number so that $M \geq 2/\varepsilon$. We have postulated that

$$E\left\{((x_t - x_{t-1})(x_t - x_{t-1})')\right\}^{1+\alpha} < \infty$$

Then it is a standard exercise (Borel-Cantelli-Lemma and Chebyshev's inequality) that there exists a $\gamma < \frac{1}{2}$ so that $(x_n - x_{n-1})/n^\gamma \to 0$. Consequently

$$\sup_{m \leq M} \|x_n - x_{n-m}\| / n^\gamma \to 0$$

Let us now analyze

$$R_{n-1} = R_{n-m-1} + \sum_{n-1 \geq i \geq n-M} x_i x_i' \tag{15}$$

It is easily seen that for all p-vectors a, b and all real $\lambda \in (0,1)$ we have $(a+b)(a+b)' \geq aa'(1-\lambda) - bb'(\frac{1}{\lambda}+1)$. Let us now for $n \in I$ choose $a = x_n$, $b = x_i - x_n$ and $\lambda = n^{-\kappa}$, where κ is positive and so that $\kappa + 2\gamma < 1$: then we may conclude that

$$\begin{aligned}
\sum_{n-1 \geq i \geq n-M} x_i x_i' &\geq M x_n x_n' (1 - n^{-\kappa}) - n^\kappa \sum_{n-1 \geq i \geq n-M} (x_n - x_{n-i})(x_n - x_{n-i})' \\
&\geq M x_n x_n' (1 - n^{-\kappa}) - n^{\kappa+2\gamma} I \text{ for all but finitely many } n \in J \\
&\geq \frac{2}{3} M x_n x_n' - n^{\kappa+2\gamma} I \text{ for all but finitely many } n \in J
\end{aligned}$$

Moreover, from (13) we can easily see that $R_{n-M-1} \geq nS/2$ for all but finitely many $n \in J$; As $\kappa + 2\gamma < 1$ $n^{\kappa+2\gamma} I < nS/4$ for all but finitely many $n \in J$ and - consequently we have (for all but finitely many $n \in J$)

$$R_{n-1} \geq \frac{2M}{3} x_n x_n' + \frac{n}{4} S$$

and therefore $x_n' R_{n-1}^{-1} x_n \leq \frac{3}{2}\frac{1}{M} \leq \frac{3}{4}\varepsilon$, which directly contradicts (14)

Having established the lemma, it is now relatively easy to prove (6),(7),(8): observe that

$$\begin{aligned}
\Delta_n &= \sum_{t \leq n} \left(y_t - x_t' \widehat{\beta_{t-1}}\right)^2 - \sum_{t \leq n} (y_t - x_t' \beta)^2 = \\
& 2 \sum_{t \leq n} u_t x_t' \widehat{\beta_{t-1}} + \sum_{t \leq n} \left(x_t' \widehat{\beta_{t-1}}\right)^2
\end{aligned}$$

Then we can define $A_n = 2\sum_{t\leq n} u_t x_t' \widehat{\beta_{t-1}}$: the above equation immediately shows that condition (6) is fulfilled: For proving (7) it can easily be seen that $E(A_n^2/x_1, ..., x_n) = 4\sum_{t\leq n} x_t' R_{t-1}^{-1} x = O(\log\det R_n) = O(\log n)$, (cf. the above lemma and (10)) therefore .

$$E(\left(\frac{A_n}{\log n}\right)^2 / x_1, ..., x_n) \to 0$$

which shows (7). It remains to show (8): But $\Delta_n - A_n = \sum_{t\leq n}\left(x_t'\widehat{\beta_{t-1}}\right)^2$ and it is now an elementary task to show that

$$E(\Delta_n - A_n/x_1, .., x_n) = \sum_{t\leq n} x_t' R_{t-1}^{-1} x_t$$

Then (8) is an immediate consequence of the above Lemma.

References

1. Gerencer, L. and J. Rissanen (1992): Asymptotics of predictive stochastic complexity, in: New Directions in Time Series 2, Brillinger-Caines-Geweke-Parzen-Rosenblatt-Taqqu (eds.), Springer Verlag, New York, p. 93-112.

2. Ploberger, W. and Phillips, P.C.B. (1998): An extension of Rissanen's bound on the best empirical DGP, mimeo, Yale University.

3. Robinson, P. (1994): Time series with strong dependence, in C. Sims (ed.), Advances in Econometrics: 6th World Congress, Cambridge, Cambridge University Press.