CEMA WORKING PAPER SERIES

On Model Approximation for Long-Memory Processes:
A Cautionary Result

Clive W. J. Granger
Department of Economics
University of California, San Diego
La Jolla, CA 92093-0508

Working Paper 45

May 2000

Working Paper 45, 1–4

On Model Approximation for Long-Memory Processes: A Cautionary Result

An Essay in Honor of Professor G.S. Maddala

Clive W. J. Granger

Department of Economics
University of California, San Diego
La Jolla, CA 92093-0508
E-mail: cgranger@weber.ucsd.edu

It is pointed out that if the generating mechanism is a fraction integrated process $I(d)$, where d can be less than 1/2, but a simple ARMA model is fitted, a consistent estimation procedure is likely to produce a unit root. Thus the properties of the fitted model will be quite unlike those of the generating mechanism.

1. INTRODUCTION

It is possible to build an econometric model and then to reach deductions from it that are not strictly plausible. This can occur, for example, when the true data generating process does not lie within the class of models being considered, which will occur frequently. A consequence is that the properties of the fitted model will not necessarily be equal to the properties of the data in several important aspects. Professor G.S. Maddala was well known for his careful modeling, not just becoming too deeply embedded in the mathematical aspects of the model, and then providing links to appropriate data. This small note provides an illustration of how the wrong model can be achieved with quite incorrect properties. I hope that he would approved of its aims, at least.

2. DEFINING A LONG-MEMORY PROCESS

If X_t is a stationary process with unit variance and spectrum $f(w)$, given by

$$f(w) = 1 + c\sum_{j=0}^{\infty} \rho_j e^{ijw} \tag{2.1}$$

where $\rho_j = corr(X_t, X_{t-j})$ and $c = 1/\pi$, then it has become common practice to say that X_t has "long-memory" if $f(w)$ tends to infinity as w goes to zero. As, after a simple affine transformation,

$$f(0) = \lim_{m\to\infty} \sum_{j=0}^{m} \rho_j \tag{2.2}$$

the long-memory property implies that the sequence ρ_j produces a divergent series. Consider the case where $\rho_k > 0$ and suppose that $\rho_{k+1} \neq \rho_k$ so that the autocorrelations decline with the lag but remain positive. It follows that a necessary condition for divergence is

$$\lim_{k} \frac{\rho_{k+1}}{\rho_k} = 1. \tag{2.3}$$

As any long-memory process will have a spectrum that is unbounded at zero frequency its autocorrelations will obey (2.3). An example is the pure fractional unit root process $I(d)$ generated by $(1-B)^d X_t = \varepsilon_t$, ε_t iid, B the backward operator for which $\rho_k \sim ck^{2d-1}$ for large, positive k as in Beran (1994) and elsewhere. This autocorrelation sequence clearly obeys (2.3). The same will be true for any $ARIMA(p, d, q)$ process with $|d| < 1/2$.

3. MODEL APPROXIMATION

Now suppose that an $ARMA(1, q)$ model is used to approximate a long-memory process, so that a model of the form

$$X_t = \phi X_{t-1} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_q \varepsilon_{t-q} \tag{3.1}$$

is fitted to data from the stationary, zero-mean, long-memory process X_t. A form of the methods of moments or Yule-Walker equations can be used to estimate the coefficients ϕ, θ_j, by multiplying the equation by X_{t-k} and taking expectations, using $k = 1, \ldots, q+1$, sequentially giving q equations to solve most of which will be both non-linear and complicated. However the last equation will be $E[X_t, X_{t-q-1}] = \phi E[X_{i-1}, X_{t-q-1}]$, i.e. $\rho_{q+1} =$

$\phi\rho_q$, so that

$$\hat{\phi} = \frac{\hat{\rho}_{q+1}}{\hat{\rho}_q}.$$

Provided consistent estimates of autocorrelations are used, and q is large, then asymptotically $\hat{\phi} \cong 1$, according to (2.3). There is no reason to believe that there will be a compensating unit root in the moving average term; in fact there cannot be if q is finite, although there may be a large root. The estimation method is simple but consistent. I expect that if more sophisticated methods were analyzed they would reach the same conclusion.

4. IMPLICATIONS

If only a limited class of models is presented to the data for consideration, the best one can hope for is the best approximation, in some sense, to the true data generation process. If the DGP has a dominant property, such as long-memory, and the class of models being considered does not include this case, a poor approximation may result. The theorem says that if the data has the long-memory property, then a fitted $ARMA(1, q)$ model will produce, asymptotically, a perfect unit root, or a near unit root for a large sample. Thus an $I(d), 0 < d < 1$ process will be "found" to be $I(1)$, by the data analysis. Although only one type of estimation procedure is considered in the proof, it is consistent, and I suspect that other consistent estimation methods can be shown to produce something similar; it just happens to be simple using the simple method of moments. It should be noted that both $I(1)$ and $I(d), 0 < d < 1$ processes do have the same dominant property of the infinite peak at zero frequency. Other properties are quite different, such as the shape of the estimate correlogram and the expansion of the variance as the sample size increases. It is well known, for example, that the variance is proportional to t^2 for the $I(1)$ process, to $\log t$ for the $I(1/2)$? process and converges to a constant for $0 < d < 1/2$ Thus, the fitted, unit root model, will have quite different properties than the DGP. When it is claimed that a "unit root has been found" it does not follow, necessarily, that the series has all of the properties associated with a unit root.

The theorem is concerned just with $ARMA(1, q)$ processes, but a similar but more complicated argument holds for $ARMA(p, q)$, with p fixed and q increasing. As an approximation, the argument will also hold for $ARMA(2, 1)$ written as

$$(1 - \alpha B)(1 - \beta B)X_t = (1 - \gamma B)\varepsilon_t$$

and thus as

$$(1-\alpha B)X_t = (1-\gamma B)(1-\alpha B)^{-1}\varepsilon_t$$

which is $ARMA(1,q)$ of a constrained form.

As a simple numerical example, Bollerslev, Engle, and Nelson (1994) report fitting various $ARMA$ models to $\log \sigma_t$ where σ_t is a time-changing measure of volatility, including $ARMA(1,q), q = 1,2,3,4$ and 5 and $ARMA(2,1)$. They only report the first autoregressive parameter for the $ARMA(2,1)$ model using three sets of long daily stock index data, finding values of 0.964, 0.951, and 0.970 for samples for periods 1885–1914, 1925–1952, and 1953–1990, respectively. These values are by no means a test of the theorem state above but are supportive of it.

The $I(d)$ process is long-memory, but there are plenty of other models that have similar properties, including those with breaks (discussed by Granger and Hyung, (2000)) and possibly some non-linear processes.

The result reported in Section 2 is for a rather simple, specific situation but could be symptomatic of many cases where the data contains a long-memory type property but an inappropriate class of models is fitted, leading to a unit root being "found" even though a unit root process has other properties, such as an exploding variance, not seen in the data.

REFERENCES

Beran, J., 1994, Statistics for long-memory statistics. Chapman and Hall, London.

Bollerslev, T., R. F. Engle, and D. B. Nelson, 1994, ARCH models. In: *Handbook of Econometrics*, R. F. Engle and D. McFadden eds., Volume IV, chapter 9, p2967-3039. Elsevier Scientific Publishers.

Granger, C. W. J. and N. Hyung, 2000, Occasional stochastic breaks and long-memory. Working paper, UCSD Dept. of Economics