Modelling Memory of Economic and Financial Time Series*

by

Peter M Robinson
London School of Economics and Political Science

Contents:

Discussion Paper
No.EM/05/487
March 2005

The Suntory Centre
Suntory and Toyota International Centres for
Economics and Related Disciplines
London School of Economics and Political Science
Houghton Street
London WC2A 2AE
Tel.: 020 7955 6679

---

* This paper is based on a public lecture presented at the University of Western Australia in September 2004. This research was supported by ESRC grant R000239936.

Abstract

Much time series data are recorded on economic and financial variables. Statistical modelling of such data is now very well developed, and has applications in forecasting. We review a variety of statistical models from the viewpoint of 'memory', or strength of dependence across time, which is a helpful discriminator between different phenomena of interest. Both linear and nonlinear models are discussed.

Keywords: Long memory; short memory; stochastic volatility
JEL No.: C22.

Contact details:
London School of Economics, Houghton Street, London WC2A 2AE, United Kingdom. Email: p.m.robinson@lse.ac.uk.

1. INTRODUCTION

Economic and financial time series data are often recorded at (almost) equally-spaced intervals of time, e.g. yearly, quarterly, monthly. Such data can often be viewed as representing observations on a continuous-time process. This might be modelled as a stochastic differential equation, say. But there are generally huge identification problems in trying to fit a continuous time model to discrete data (see e.g. Robinson, 1977). In this paper we consider modelling the discrete observations directly, reviewing a variety of models from the perspective of 'memory'.

Consider observations

$$y_t, \quad t = 1, ..., n,$$

where y_t represents a financial or economic variable (e.g. GNP, asset price) at time t, and the unit interval can represent any constant time interval, e.g. 1 year, 1 second. A general model for y_t is

$$y_t = d_t + s_t \tag{1}$$

where: d_t is a deterministic component, e.g. a polynomial or cyclic function; s_t is a stochastic component, described by random variables. Note that (1) is an additive model. However, it could be obtained by taking logs in an initial multiplicative model.

Typically, both s_t and d_t are specified parametrically or nonparametrically by the econometrician. We will not discuss estimation We focus on the modelling of s_t, and the (somewhat nebulous) issue of memory. We will not discuss d_t further, though there has been controversy as to whether trends are better described stochastically or deterministically.

Let x_t be a generic sequence of random variables, which could represent s_t The notation

$$x_t \sim IID$$

means that the x_t are independent and identically distributed. Further, for $\theta > 0$,

$x_t \sim IID(\theta)$ means that

$$x_t \sim IID \text{ and } E\left|x_t\right|^{\theta} < \infty. \tag{2}$$

For $\theta \geq 1$ we will assume also that

$$E(x_t) = 0,$$

with no loss of generality when $s_t = x_t$ because a non-zero mean could be introduced in d_t. In case (2) holds only for $\theta < 1$, an alternative location of the distribution of x_t would entail a zero median.

2. MODELS WITH SECOND MOMENT MEMORY

Often we assume

$$Ex_t^2 < \infty.$$

Here, a weaker concept than $x_t \sim IID(2)$ is

$$x_t \sim UH,$$

i.e. the x_t are uncorrelated and homoscedastic. This means that

$$var(x_t) \text{ is constant over } t,$$

$$cov(x_t, x_{t+u}) = 0, \quad \text{all } u \neq 0. \tag{3}$$

If $x_t \sim IID(\theta)$, some $\theta > 0$, we can say unambiguously that x_t has zero memory. If $x_t \sim UH$ there is no memory with respect to 2nd moments (cf (3)). However, there could be memory with respect to higher moments, say.

The distinction between "$IID(2)$" and "UH" has become very important in econometrics and finance nowadays. We shall return to this, but we first discuss processes which have memory in 2nd moments.

A process x_t is covariance stationary if

$$\gamma_u = cov(x_t, x_{t+u})$$

depends on u only and is finite for all t.

If $x_t \sim UH$, then $\gamma_u = 0$, all $u \neq 0$. On the other hand, if $\gamma_u \neq 0$ for some $u \neq 0$, x_t has some (2nd moment) memory.

Now define the lag operator L, such that $Lx_t = x_{t-1}$. Our first model example is as follows (see e.g. Box and Jenkins, 1971):

Example 1 *Moving average (MA) process (of order 1)*

$$x_t = (1 + \alpha L)\varepsilon_t, \quad \alpha \neq 0,$$

where $\varepsilon_t \sim UH$. (Often $|\alpha| < 1$ is prescribed for invertibility or identifiability reasons.)

For this process

$$\begin{aligned} \gamma_u &\neq 0, \quad u = 1 \\ &= 0, \quad u > 1. \end{aligned}$$

Our next example (see e.g. Box and Jenkins, 1971) is:

Example 2 *Autoregressive (AR) process (of order 1)*

$$(1 - \alpha L)x_t = \varepsilon_t, \quad 0 < |\alpha| < 1,$$

where $\varepsilon_t \sim UH$.

For this model $\gamma_u \neq 0$, for all u, but γ_u decays exponentially to 0 as $u \to \infty$.

Both Examples 1 and 2 illustrate short memory models. They can be significantly generalized, to allow for additional lags, and combined (to form mixed autoregressive moving average (ARMA) models), but still retain the property of eventual cutout, or exponential decay, of γ_u.

However, we adopt a much less stringent definition of short memory, that covers many other processes. We say x_t has short memory (in 2nd moments) if

$$\sum_{u=-\infty}^{\infty} |\gamma_u| < \infty. \tag{4}$$

It is convenient to consider this restriction alongside properties of the spectral density, which is given by

$$f(\lambda) = \frac{1}{2\pi} \sum_{u=-\infty}^{\infty} \gamma_u \cos(\lambda u), \quad -\pi < \lambda \leq \pi.$$

Clearly, $f(\lambda)$ is well-defined under (4). If $x_t \sim UH$

$$f(\lambda) = \frac{\gamma_0}{2\pi} = \text{constant}.$$

But otherwise $f(\lambda)$ varies. By way of interpretation, if $f(\lambda)$ is large for small λ there are substantial long term movements in the series.

The summability condition (4) on the γ_u is related to smoothness conditions on $f(\lambda)$. In particular if $f(\lambda) \sim Lip(\eta)$, for some $\eta > \frac{1}{2}$ then (4) holds (see Zygmund, 1979); the Lipschitz condition is stronger than continuity of $f(\lambda)$ but weaker than differentiability. Since $f(\lambda)$ is periodic of period 2π, it is implied that $f(\lambda)$ is bounded.

However, there has been considerable interest in processes x_t which do not satisfy (4), and have unbounded $f(\lambda)$. Empirically, smoothed nonparametric estimates of $f(\lambda)$ are sometimes very peaked near $\lambda = 0$, say, possibly suggesting that

$$f(0) = \infty,$$

i.e.

$$\sum_{u=-\infty}^{\infty} \gamma_u = \infty. \tag{5}$$

If (5) holds we say x_t has long memory (in 2nd moments). Note that $f(\lambda)$ could instead diverge at one or more non-zero λ, when there is long memory of a cyclic or seasonal type. However, we will not discuss such phenomena.

An x_t that satisfies (5) is as follows

Example 3 *$I(d)$ model (Adenstedt, 1974)*

$$(1-L)^d x_t = \varepsilon_t, \quad \varepsilon_t \sim UH, \quad |d| < \frac{1}{2}.$$

For such x_t,

$$\begin{aligned} f(\lambda) &= \frac{var(\varepsilon_t)}{2\pi} \left|1 - e^{i\lambda}\right|^{-2d}, \quad -\pi < \lambda \leq \pi \\ &\sim C\lambda^{-2d}, \quad \text{as } \lambda \to 0+. \end{aligned}$$

For $d = 0$, $x_t = \varepsilon_t$, i.e. has short memory, $0 < f(0) < \infty$. For $0 < d < \frac{1}{2}$, x_t has long memory, $f(0) = \infty$. For $-\frac{1}{2} < d < 0$, x_t has negative memory, $f(0) = 0$.

The restriction $d < \frac{1}{2}$ indicates covariance stationarity, the restriction $d > -\frac{1}{2}$ indicates invertibility. The $I(d)$ model can be extended to allow ε_t to be a stationary and invertible AR, MA or ARMA, without affecting this memory classification. Such "fractional" models form a convenient bridge from (short memory) stationary to nonstationary models. There is also interest in fractional nonstationary models (where $d \geq \frac{1}{2}$), as well as fractional noninvertible ones (where $d \leq -\frac{1}{2}$). We will discuss only the former.

For nonstationary models γ_u and $f(\lambda)$ are not strictly defined. However, we can introduce a truncation, modifying Example 3 as follows.

Example 4 *$I(d)$ model, $d \geq \frac{1}{2}$,*

$$\begin{aligned} (1-L)^d x_t &= \varepsilon_t, \quad t \geq 1, \\ x_t &= 0, \quad t \leq 0, \end{aligned}$$

where $\varepsilon_t \sim UH$.

For this model x_t has variance that is finite for all t, but changes as $t \to \infty$. We can say that d measures the memory of x_t; d is sometimes called the memory parameter. As a special case for $d = 1$ we have the familiar unit root process

$$(1 - L)x_t = \varepsilon_t, \ \ t \geq 1.$$

This can also be obtained from the AR

$$(1 - \alpha L)x_t = \varepsilon_t,$$

putting $\alpha = 1$ (to violate the stationarity restriction in Example 2). But the "fractional" class is "smoother" with respect to departures from the unit root, in the sense that asymptotic distributions of, for example, statistics for testing for a unit root directed against fractional alternatives are of standard (χ^2) form, whereas ones directed against autoregressive alternatives are of non-standard form (see Dickey and Fuller, 1979, Robinson, 1994).

We focus on univariate processes x_t, but the vector case is also important. For example, we can cover (fractional) cointegration, between two or more related economic series, e.g. consumption and income (Engle and Granger, 1987). Here, the observable series x_t and y_t both have memory d but for some β

$$y_t - \beta x_t$$

has memory $c < d$.

If x_t is Gaussian and stationary then it suffices to model it in terms of γ_u (or equivalently $f(\lambda)$). But otherwise not all the information is contained in 2nd moments. One way of modelling such non-Gaussian x_t is as follows (see e.g. Hannan, 1970):

Example 5 *Linear process*

$$x_t = \alpha(L)\varepsilon_t,$$

where the ε_t are IID with some <u>*non-normal*</u> *distribution and*

$$\alpha(L) = 1 + \sum_{j=1}^{\infty} \alpha_j L^j.$$

For example, in the MA special case

$$\alpha(L) = 1 + \alpha L.$$

If $\varepsilon_t \sim IID(2)$ then we can include models with either short memory or long memory in 2nd moments. But we can also study other properties. And if $\varepsilon_t \sim IID(\theta)$, $\theta < 2$, this can be a convenient model for <u>heavy-tailed</u> data.

An alternative way of modelling non-Gaussian series is via <u>non-linear</u> models.

Example 6 *Nonlinear AR (e.g. Jones, 1978)*

$$x_t = g(x_{t-1}) + \varepsilon_t,$$

where g is some non-linear function and $\varepsilon_t \sim IID$.

For such models, we can again look at 2nd moment memory, but also at other properties, bringing us to our next topic.

3. MODELS WITH NO SECOND MOMENT MEMORY BUT WITH MEMORY IN NONLINEAR FUNCTIONS

For some financial data, an important class of models starts from the contention that

$$x_t \sim UH$$

may be reasonable, but not

$$x_t \sim IID(2).$$

Example 7 *ARCH model (Engle, 1982)*

$$x_t = \varepsilon_t (1 + \alpha x_{t-1}^2)^{\frac{1}{2}}, \quad 0 < \alpha < 1,$$

where $\varepsilon_t \sim IID(2)$*.*

The ARCH model implies that

$$cov(x_t, x_{t+u}) = 0, \quad \text{all } u \neq 0$$

but

$$var(x_t \,| x_{t-1}) = var(\varepsilon_t)(1 + \alpha x_{t-1}^2).$$

Such a model is said to possess conditional heteroscedasticity. It is implied that the sequence x_t has zero (2nd moment) memory but the sequence x_t^2 has short (2nd moment) memory. Such models have been extended and greatly used in practice. In some versions of the model $var(x_t) = \infty$. In more, $Ex_t^4 = \infty$, agreeing with some empirical evidence. However, ARCH models can be hard to handle theoretically, and they may not explain all features of the data. One such feature is leverage: $cov(x_t^2, x_{t-u}) < 0$, for some $u \geq 1$. Another such feature is long memory in x_t^2.

One model that overcomes both these drawbacks is as follows:.

Example 8 *LARCH (Robinson, 1991):*

$$\begin{aligned} x_t &= \varepsilon_t(\mu + \alpha(L)\varepsilon_t), \\ \varepsilon_t &\sim IID(2), \quad \alpha(L) = \sum_{j=1}^{\infty} \alpha_j L^j. \end{aligned}$$

However, unlike ARCH, the estimation of LARCH has not been adequately discussed, so it is not presently a very viable tool for empirical analysis.

Far more popular alternatives to ARCH are stochastic volatility (SV) models. A particular version that is often studied is as follows.

Example 9 *SV model (Taylor, 1986).*

$$x_t = \varepsilon_t e^{\mu + \alpha \eta_t},$$

where ε_t is IID and η_t is a stationary Gaussian process.

Distributional assumptions are often imposed also on ε_t, and properties are affected depending on whether η_t is independent of ε_s, for all $s > t$, or η_t is independent of ε_s, for all s, t. In any case whereas

$$x_t \sim UH,$$

we have

$$cov\left(|x_t|^\theta, |x_{t+u}|^\theta\right) \neq 0, \quad u \neq 0, \ \theta > 0,$$

(e.g. when $\theta = 2$). Moreover, if we choose η_t to have long memory then $|x_t|^\theta$ can also have long memory. Further, we can generalize to models such as

$$x_t = f_1(\varepsilon_t) f_2(\eta_t)$$

where ε_t and η_t can both be vector processes with long or short memory. We can then study the memory of quantities such as $|x_t|^\theta$ (see Robinson, 2001).

4. FINAL COMMENTS

Versions of the models we have discussed involving finitely many unknown parameters are commonly estimated. But they can also be the basis for nonparametric modelling. In either case, finite-sample properties of estimates and test statistics are generally intractable. However, asymptotic (as sample size $\to \infty$) properties are well developed in some of the models, less so in others. In many cases we have a normal approximation for the estimates, leading to convenient hypothesis testing and interval estimation. An important application of the estimated model is in forecasting.

REFERENCES

[1] Adenstedt, R. (1974). On large-sample estimation for the mean of a stationary random sequence. *Ann Statist.* 2, 1095-1107.

[2] Box, G.E.P. and Jenkins, G.M. (1971). Time Series Analysis. Forecasting and Control. Holden Day, San Francisco.

[3] Dickey, D.A. and Fuller, W.A. (1979). Distribution of the estimators for autoregressive time series with a unit root. *J. Amer. Statist. Assoc.* 74, 427-431.

[4] Engle, R.F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica* 50, 987-1007.

[5] Engle, R.F. and Granger, C.W.J. (1987). Cointegration and error correction: representation, estimation and testing. *Econometrica* 55, 251-276.

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

[7] Jones, D.A. (1978). Nonlinear autoregressive processes. *Proc. R. Soc. Lond. A* 360, 71-95.

[8] Robinson, P.M. (1977). The construction of continuous time models and discrete approximations in econometrics. *J. Econometrics* 6, 173-197.

[9] Robinson, P.M. (1991). Testing for strong serial correlation and dynamic conditional heteroscedasticity in multiple regression. *J. Econometrics* 47, 67-84.

[10] Robinson, P.M. (1994). Efficient tests of nonstationary hypotheses. *J. Amer. Statist. Assoc.* 89, 1420-1437.

[11] Robinson, P.M. (2001). The memory of stochastic volatility models. *J. Econometrics* 101, 195-218.

[12] Taylor, S.J. (1986). *Modelling Financial Time Series.* Chichester, UK.

[13] Zygmund, A. (1979). *Trigonometric Series.* Cambridge University Press, Cambridge.