Forecasting Business and Economic Time Series with Overdifferenced Models*

Nuno Crato**

1. Overview

In economic and business forecasting we deal usually with time series that are not stationary in the levels. One way of analysing and forecasting these series starts with differencing, as a means of obtaining series that are stationary and to which traditional autocorrelation tools can be applied. This process is not exempt of difficulties, as over-differencing is likely to occur introducing problems for forecasting and increasing critically the forecasting error variance (Crato, 1992a).

The new class of fractionally integrated autoregressive moving average models, ARFIMA, introduces more flexibility allowing the differencing parameter to be a noninteger real. We show that this more general classe of models provides an innovative framework for estimating the degree of differencing appropriate for a given time series. We propose a periodogram technique for estimating this degree of differencing and show that this technique is able to handle long-memory disturbances.

Some partial results along these lines were developed in Crato (1992b), but the theoretical results were limited to the ARIMA (0, 1, 0) case, where the differenced noise is white. The theorems presented below are applicable to general ARIMA (*p*, 1, *q*) models, where the differenced processes have short-memory non-white disturbances.

In order to illustrate the procedure, we provide a modelling and forecasting application using financial data.The plan of the paper is as follows. Section 2 introduces the notation for the ARIMA and ARFIMA models. Section 3 introduces the notation and the main concepts for the spectral and periodogram analysis. Section 4 characterizes the behavior of the periodogram of nonstationary ARIMA and of stationary long-memory processes. Section 5 presents an application using exchange rates. Section 6 concludes. Mathematical proofs can be found in a technical report (Crato, 1991).

2. The forecasting models

Let (X_t), $t = 0, \pm 1, \pm 2, \ldots$, represent an observed time series. Let B be the backwards shift operator, $BX_t = X_{t-1}$, and let ∇ represent the differencing operator, $\nabla X_t = (1 - B) = X_t = X_t - X_{t-1}$. Considerer the so-called ARMA (*p, q*) process

$$X_t - \varnothing_1 X_{t-1} - \ldots - \varnothing_p X_{t-p} = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \ldots + \theta_q \varepsilon_{t-q}, \quad (1)$$

written

$$\varnothing (B) X_t = \theta (B) \varepsilon_t,$$

where $\varnothing(.)$ and $\theta(.)$ represent the polynomials $\varnothing(z) = 1 - \varnothing_1 z - \ldots - \varnothing_p z^p$ and $\theta(z) = 1 + \theta_1 z + \ldots + \varnothing_q z^q$. The ARMA process is considered to be driven by *white noise*, i.e., it is assumed that $E[\varepsilon_t] = 0$, $E[\varepsilon_t^2] = \sigma^2 > 0$ and $E[\varepsilon_t \varepsilon_{t+h}] = 0$ for all $h \neq 0$. It is also assumed that $\varnothing(z)$ and $\theta(z)$ have all its roots

*A previous version of this paper was presented at the October 1992 meeting of The Institute of Management Sciences, Myrtle Beach, SC. I am grateful to Howard M. Taylor for various suggestions regarding the problem dealt with in this paper, to Peter Brockwell for providing me with the new version of his program ITSM, and to António A. Costa for providing the exchange rate data.

**Department of Pure and Applied Mathematics, Stevens Institute of Technology, Hoboken, NJ 07030, USA

outside the unit circle [$|z| = 1$]. The process (1) with all the started conditions will be called a *regular ARMA.*

Consider the differenced process $Y_t = \nabla X_t = X_t — X_{t-1}$, where (Y_t) is an ARMA (p, q). The process (X_t) is called an ARIMA (p, 1, q). Higher order differences are defined recursively, $\nabla^{n+1} = \nabla.\nabla^n$. In general, (X_t) is called an ARIMA (*p, d, q*) with *d* any natural number, if $\nabla^d X_t = Y_t$, $Y_t \sim$ ARMA (*p, q*).

ARIMA models can be integrated in a more general set of models where the differencing parameter *d* is not bound to be an integer. More specifically, consider the so-called fractionally differenced ARMA models (FARMA) introduced independently by Granger and Joyeux (1980) and by Hosking (1981). Let *d* be any real number. The fractional difference of the process (X_t) is defined through the binomial expansion

$$\nabla^d = (1 — B)^d = \sum_{k=0}^{\infty} \binom{d}{k} (— B)^k,$$

where $\binom{d}{k}$ is the binomial coefficient $d(d— 1)... (d — k+1)/k!$. Explicitly,

$$\nabla^d X_t = X_t — dX_{t-1} + d \frac{(d — 1)}{2} X_{t-2} — ....$$

The process (X_t) is said to be a *regular FARMA* (*p, d, q*) process with *d* in the open interval]— .5,.5[if $(\nabla^d X_t)$ is a regular ARMA and we write

$$\varnothing (B)\, \nabla^d X_t = \theta (B)\, \varepsilon_t, \quad d \in\ ]\ .5,.5[. \qquad (2)$$

For $d \neq 0$ the autocorrelations of the process (X_t), have an asymptotic hyperbolic decay which is charateristic of *long-memory processes* (see Brockwell and Davis, ch. 13.2).

The particular case $d > 0$ is very important. It can be proved that for $d \in\ ]0,.5[$ the autocovariances $\gamma(h)$ and autocorrelations $\rho(h)$ of (X_t) decay very slowly and are all eventually positive. In this case the process (X_t) is called a *persistent* process. When $d \geq .5$ the model (2) is nonstationary and is called an ARFIMA (*p, d, q*).

3. The spectrum and the periodogram of ARFIMA and ARIMA models

Let (X_t), $t = 0, \pm 1, \pm 2, ...$, represent a stationary time series that has (constant) mean μ_x, autocovariance function $\gamma_x(h) = \mathrm{Cov}\,[X_t, X_{t+h}]$, and autocorrelations $\rho_x(h) = \gamma_x(h)/\gamma_x(0)$. We will assume that there exists an integrable function $f_x(\lambda)$ such that

$$f_x(\lambda) = \sum e^{-ik\gamma} \gamma_x(k) \geq 0, \quad \forall \lambda \in [— \pi, \pi], \qquad (3)$$

and we will call $f_x(\lambda)$ the *spectral density function* of the process (X_t).

Consider a regular ARMA (*p, q*) process (1). Lef $f_\varepsilon(\lambda)$ represent the spectral density function of the noise (ε_t). It is known that $f_\varepsilon(\lambda) = \sigma^2/2\pi$ over $[— \pi, \pi]$, as the noise is white, and that the following relation holds

$$|1 — \varnothing_1 z — ... — \varnothing_p z^p|^2 f_x(\lambda) = |1 + \theta_1 z + \theta_q z^q|^2 f_\varepsilon(\lambda), \qquad (4)$$

with $z = e^{-i\lambda}$, and for all $\lambda \in [— \pi, \pi]$.

Consider the process $Y_t = \nabla X_t = (1 — B) X_t = X_t — X_{t-1}$, where (X_t) is a stationary process. Then,

$$f_y(\lambda) = |1 — z|^2 f_x(\lambda), \quad — \pi \leq \lambda \leq \pi,$$

with $z = e^{-i\lambda}$. The function $|1 + \theta_1 z + ... + \theta_q z^q|^2 / |1 — \varnothing_1 z — ... — \varnothing_p z^p|^2$, that translates the operator $(1 + \theta_1 B + ... + \theta_q B^q) / (1 — \varnothing_1 B — ... — \varnothing_p B^p)$, and the function $|1 — z|^2$, that translates the operator (1 — B) are called *power transfer functions.* But now assume that (X_t) is an ARIMA (*p*, 1, *d*) and still $Y_t = (1 — B) X_t$. Formally, applying the transfer function we could write $f_y(\lambda) = |1 — z|^2 f_x(\lambda)$. This is sometimes done even if it does not make sense since (Y_t) is not stationary and thus $f_y(\lambda)$ is not defined. If we rewrite this relation as

$$f_x(\lambda) = f_y(\lambda) \quad |1 — e^{-i\lambda}|^{-2} \qquad (5)$$

the fact that $f_x(\lambda) \uparrow \infty$ when $\lambda \to 0$ would intuitively follow from the facts $f_y(\lambda) > 0\ \forall \lambda$ and $|1 — e^{-i\lambda}| \to 0$ as $\lambda \to 0$.

Now consider a finite sample realization of the process (Y_t) with $t = 1,2,...n$, and let $I_{y,n}(\omega_j)$ represent its periodogram at a Fourier frequency $\omega_j \neq 0$

$$I_{y,n}(\omega) = \sum_{k=-n+1}^{n-1} e^{-ik\omega}\hat{\gamma}(k), \qquad (6)$$

where $\omega = \omega\, 2\pi j/n \in\,]0,\pi[$ and $\hat{\gamma}(k) = n^{-1}\sum(Y_t - \bar{Y}_n)(Y_{t+k} - \bar{Y}_n)$ is the usual covariance estimator, with $\bar{Y}_n = n^{-1}\sum Y_t$.

The expression (6) is very useful to keep in mind, since it is a sort of finite sample counterpart of (3), showing that in one sense the periodogram is an estimador of the spectral density.

It is "experimentally" known that, if (Y_t) is an ARIMA $(p, 1, q)$, then $I_{y,n}(\omega_j)$ has a sharp peak at lower frequencies, seeming to diverge when n increases and $\omega_j \to 0$. This fact would be implied by the relation (5) if we could assume its validity.

Practitioners, when facing a spectral peak at zero, have the temptation of taking differences of the process in order to get a smoother periodogram that would reveal an achievement of stationarity. This blind practice of differencing, however, has been receiving generalized criticism in the econometric time series literature since it can lead to overdifferencing; introducing a unit root on the moving average polynominal of (ε_t) creates noninvertibility and prevents optimal forecasts. Consequently, one would like to have a more precise idea of the reliability of the relation (5).

To complicate matters further, it has recently been shown that a behavior similar to nonstationarity is reproduced by stationary long-memory models such as FARMA. These models will yield noninvertible processes when differenced.

By the power transfer function technique it is seen that the FARMA process has the spectral density function.

$$f_x(\lambda) = f_{\nabla x}(\lambda)\ \ |1 - e^{-i\lambda}|^{-2d} \qquad (7)$$

where $f_{\nabla x}(\lambda) = |\theta(z)|^2 / |\varnothing(z)|^2 \sigma^2 / 2\pi$ is the spectral density function of a regular ARMA process. For a persistent FARMA (d > 0) this relation (7) implies that $f_x(\lambda) \uparrow \infty$ as $\lambda \to 0$. In fact, since $(\nabla^d X_t)$ is a regular ARMA, $f_{\nabla x}(\lambda) > 0$ for all $\lambda > 0$, and $|1 - e^{-i\lambda}| \to 0$ as $\lambda \to 0$.

Thus, both a nonstationary ARIMA $(p, 1, q)$ and a stationary FARMA (p, d, q) with $d > 0$ generate periodograms that diverge at $\lambda = 0$. Differencing is in order in the first case and not in the second.

4. The test

A spectral estimator of the differencing parameter d has been suggested by Geweke and Porter-Hudak (1983). Their estimator is based on the fact that, for low nonzero Fourier frequencies ω_j, the periodogram of a regular FARMA is dominated by the function $|1 - e^{-i\omega}|^{2d}$, as (7) suggests. If we take logarithms on both sides of (7), and we replace the spectral density f_x by the periodogram $I_{x,n}$ and λ by the Fourier frequencies ω_j, then the following relation will hold for low-order Fourier frequencies $\omega = \omega_j$, $j = 1,2,..., m \ll n$.

$$\ln I_{x,n}(\omega) = \ln f\, f_{\nabla x}(0) - d \ln|1 - e^{-i\omega}|^2.$$

Noting that $|1 - e^{-i\omega}|^2 = 4\sin^2{}^{\omega/2}$ and $\omega = \omega_j = 2\pi j/n$ and introducing the disturbance term e_j, we can write the linear regression equation

$$\ln I(j) = a + d \ln\left[1/4 \sin^{-2}\pi j\right] + e_j, \qquad (8)$$

for $j = 1,2,..., m_u < n$, where m_u is some function of the sample size n. For details see Brockwell and Davis (1991, pp. 529 ff.). If we add a lower truncation m_L at the very first frequencies, then the consistency of the least-squares estimator for d is based on the following theorem that uses recent results of Yajima (1989) and Robinson (1993) and generalizes previous results.

Theorem 1 *Let* (X_t) *be a regular FARMA* (p, d, q) *with* $d \in\,]-.5,.5[$ *and let* $m_L = [n^\beta]$, $m_u = [n^a]$, *where* $\alpha \in]0,1[$. $0 < \beta < \alpha$, and [.] *is the greatest integer function. Then, the least-squares estimator of d from the linear equation (8) is asymptotically unbiased and normal.*

In practice we can take $\beta = .1$ $\alpha = .5$ so that the regression can be performed over the frequencies ω_j, $j = [n^{-1}], \ldots, [\sqrt{n}]$.

Consider now the ARIMA (*p,* 1*, q*). We know that the spectrum of such a nonstationary model is not defined and that the variance of the process diverges to infinity. Nevertheless, in practice and since we are using a finite sample, we can overcome these problems. First, instead of working with the spectral density of the process, we can work with the finite sample periodogram and study its behavior in medium to large samples. Second, in order to be able to obtain theoretical results, we can condition the expectations on the first observation, say X_0, and then $\mathrm{Var}[X_t \mid X_0]$ exists for all finite *t*. Third, we express the periodogram of $(X_t)^n_{t=1}$, in terms of the periodogram of $(\nabla X_t)^n_{t=1}$, so that the well known behavior of $I_{\nabla x,n}$ can help in explaining the behavior of $I_{x,n}$.

The basic result is contained in the next theorem.

Theorem 2 *Let* (X_t) *be an ARIMA* (*p*, 1, *q*). *Consider the realization sample* $(X_t)^n$. *Then the periodograms of* $(X_t)^n$ *say* $I_{x,n}(\omega)$, and of $(\nabla X_t)^n$ *say* $I_{\nabla x,n}(\omega)$, are related through the identity

$$|1 - e^{-i\omega}|^2 \, I_{x,n}(\omega) = I_{\nabla x,n}(\omega) + n^{-1}(X_0 - X_n)^2 - R_n(\omega), \quad (9)$$

where the remainder term R_n is bounded and, for low-order Fourier frequencies ω_j,

$$E\,|R_n(\omega_j)| \leq M_n \rightarrow 2\pi f_{\nabla x}(0). \quad (10)$$

Then, if (X_t) is an ARIMA (*p*, 1, *q*), since $|1 - e^{-i\omega}|^2 \rightarrow 0$ as $\omega \rightarrow 0$ we can think that, for large n, $I_{x,n}(\omega)$ will tend to have a singularity of order 2 at zero. In contrast, if (X_t) is an ARMA or a FARMA, $I_{x,n}(\omega)$ would tend to have a zero ($d < 0$), assume a finite nonzero value (d=0), or have a singularity ($d > 0$) of order less than one when ω approaches zero.

Consequently, a spectral approach can help distinguish the two processes and choose the appropriate degree of differencing for a forecasting model. Detailed simulations reported in Crato (1991) show that a test based on this estimation of *d* has quite good properties. Moreover, this is a much more direct method than the traditional analysis of the decay of the autocorrelation function (see Brockwell and Davis 1991, p. 276). Furthermore, the slow decay of the autocorrelations, traditionally suggesting differencing, can just result from long-memory characteristics of a stationary process, as shown in Crato (1992a), and such process should not be differenced.

5. Exchange rates forecasting

As an illustration, we will now study the real exchange rates of the Spanish peseta, the British pound, the U. S. dollar, and the Japanese yen, against the Portuguese escudo, during the flexible regime period. The observations are monthly, begin in June 1973 and end in June 1992. The exchange rates are *end-of-the-period.* Consumer price indexes are used to compute the real rates. All raw data were found in the *International Financial Statistics* and at the *Banco de Portugal* publications.

The question whether the real exchange rates are a stationary random process or a nonstationary one has been extensively discussed, given its implication for the Purchasing Power Parity hypothesis. Recently, Witt (1992) provided some tests that conclude for the stationary of the dollar based rates. Our tests arrive at the opposite conclusion (see a much more detailed discussion of the problem in Crato and Costa 1993) leading to the stationarity in the first differences.

Table 1

Spectral Estimates of the Differenced Parameter *d* by Geweke and Porter-Hudak Method.

Countries	$m = \alpha^{.5}$	$m = \alpha^{.55}$	$m = \alpha^{.5}$
SPAIN	$\hat{d} = 1.11$	$\hat{d} = 1.06$	$\hat{d} = .94$
	$s = .15$	$s = .15$	$s = .14$
U. K.	$\hat{d} = 1.06$	$\hat{d} = 1.95$	$\hat{d} = .83$
	$s = .26$	$s = .23$	$s = .17$
USA	$\hat{d} = 1.12$	$\hat{d} = 1.15$	$\hat{d} = 1.20$
	$s = .16$	$s = .21$	$s = .17$
JAPAN	$\hat{d} = 1.52$	$\hat{d} = 1.26$	$\hat{d} = 1.26$
	$s = .22$	$s = .20$	$s = .17$
Estimation period: June 1973-June 1992.			

In Table 1 we present the d estimates obtained with the regression (8) resulting from the extension of the Geweke and Porter-Hudak method (theorem 2) and using different values of α. The standard deviations s are as reported from the regression.

The estimates of the differencing order point consistently to $d = 1$ and t-tests, performed as the most sensible approximation in lack of finite-sample statistical results, *reject the hypothesis of stationarity* (H_0: $d < .5$).

The estimates also conclude that taking $d = 2$ (overdifferencing) would be inappropriate for forecasting this series. Accordingly, long-term forecasts of the exchange rates should eventually level off near the present values, while using an overdifferenced model the forecasts would extrapolate the recent changes in levels and, eventually, diverge. This is perfectly consistent with the random walk theory of exchange rates´ behavior. This theory states that real exchange rates´ changes are essentially *unpredictable* i.e., that the best prediction for the periods ahead is the level of the present period.

As a forecasting illustration we selected two sets of models using the AICC criterion (see Brockwell and Davis 1991, p. 304). With $d = 1$, AICC allways selected the random walk ARIMA (0,1,0), for all currencies. With $d = 2$, AICC allways selected an ARIMA (0,1,1), for all currencies.

The estimations of the parameters was done with the maximum-likelihood routine of ITSM of Brockwell and Davis (1991). The results of the forecasts, in each case, are shown in Table 2. Details are available from the author.

Table 2
Mean Square Forecasting Error
with Competing ARIMA Models

Country	ARIMA (0, 1, 0)	ARIMA (0, 2, 1)
SPAIN	.002172	.004235
U. K.	.001088	.016478
U.S.A	.009864	.004706
JAPAN	.001603	.014648

Estimation period: June 1973-June 1990; forecasting period: July 1990-June 1992.

The results are instructive. In three out of four cases, the overdifferenced models present larger mean square forecasting errors. In the case where the overdifferenced model performs better (U.S.) we are finding a period of continuous apreciation of the U. S. dollar, where the trend estimated by the ARIMA (0,2,1) persists for a long period.
In one case, the Japanese yen, we see clearly how the overdifferenced model can diverge on the long run and give predictions much worse than those from the random walk model.

6. Final Remarks

Traditional autocorrelation methods can be misleading for identifying the appropriate degree of differencing of an ARIMA forecasting model. We propose a decision making rule based on a periodogram analysis.
We have shown that nonstationary ARIMA models have periodograms with a singularity at the zero Fourier frequency. Moreover, we have characterized this singularity. For an ARIMA with a simple unit root, i.e., an ARIMA (d, 1 q), the order of the spectral singularity is 2, while it is zero for an ARMA (p, q), and less than 1 for a FARMA process. An extension of the spectral regression method of Geweke and Porter-Hudak (1983) was presented, providing an estimator of the appropriate degree of differencing. This estimator can be directly applied both to stationary and to nonstationary processes.

An application with real exchange rates was developed and the tests suggested models stationary in their first differences. As a forecasting illustration, we compared such models wiith overdifferenced ARIMA (p, 2, q), showing how overdifferencing can increase the forecasting errors.

REFERENCES

BROCKWELL, Peter J. and Davis, Richard A., (1991). *Time Series: Theory and Methods*, second edition (New York: Springer-Verlag, 1991).

CRATO, Nuno, (1991). «Periodogram analysis of nonstationary random variables», Department of Mathematics, University of Delaware.

CRATO, Nuno, (1992a). «Long-memory time series misspecified as nonstationary ARIMA», *1992 Proceedings of the American Statistical Association*, Business and Economic Statistics Section, pp. 82-87.

CRATO, Nuno (1992b). «Spectral analysis of nonstationary economic variables: a new result and an application», *Economia*, vol. XVI, No. 1, pp. 1-19.

CRATO, Nuno and Costa, António A. (1993) «Estacionaridade e reversão nas taxas de câmbio reais: O caso português», submitted for publication.

GEWEKE, John and Porter-Hudak, Susan, (1983). «The estimation and application of long memory time series models», *Journal of Time Series Analysis*, vol. 4, No. 4, pp. 221-238.

GRANGER, C. W. J. and Joyeux, Roselyne, (1980). «An introduction to long-memory time series models and fractional differencing», *Journal of Time Series Analysis*, vol. 1, No. 1, pp. 15.29.

HOSKING, J. R. M., (1981). «Fractional differencing», *Biometrika*, vol. 68, No. 1, pp. 165-176.

ROBINSON, Peter M. (1983). «Log-periodogram regression of time series with long range dependence,» London School of Economics.

WITT, Joseph A., Jr., (1992). «The long-run behavior of the real exchange rate: a reconsideration», *Journal of Money, Credit and Banking*, vol. 24, No. 1, pp. 72-82.

YAJIMA, Yoshihiri, (1989). «A central limit theorem of Fourier transforms of strongly dependent stationary processes», *Journal of Time Series Analysis*, vol. 10, No. 4, pp. 374-383.