

**CONDITIONAL MEANS OF TIME SERIES PROCESSES AND
TIME SERIES PROCESSES FOR CONDITIONAL MEANS***

Gabriele Fiorentini and Enrique Sentana**

WP-AD 97-17

* We have benefitted from discussions with Manuel Arellano, Manuel Balmaseda, Ian Domowitz, Lucrezia Reichlin and George Tauchen. We would also like to thank John Campbell for kindly allowing us to use his data, and seminar audiences at CEMFI and UPV (Bilbao) for very useful comments and suggestions.

** G. Fiorentini: University of Alicante; E. Sentana: CEMFI.

**Editor: Instituto Valenciano de
Investigaciones Económicas, S.A.**
Primera Edición Junio 1997.
ISBN: 84-482-1518-4
Depósito Legal: V-2222-1997
Impreso por Key, S.A.
C/. Cardenal Benlloch, 69, 46021-Valencia.
Impreso en España.

**CONDITIONAL MEANS OF TIME SERIES PROCESSES
AND TIME SERIES PROCESSES FOR
CONDITIONAL MEANS**

Gabriele Fiorentini & Enrique Sentana

A B S T R A C T

We study the processes for the conditional mean and variance given a specification of the process for the observed time series. We derive general results for the conditional mean of univariate and vector linear processes, and then apply it to various models of interest. We also consider the joint process for a subvector and its expected value conditional on the whole information set. In this respect, we derive necessary and sufficient conditions for one of the variables in a bivariate VAR(1) to have a white noise univariate representation while its conditional mean follows an AR(1) with a high autocorrelation coefficient. We also compare the persistence of shocks to the conditional mean relative to the observed variable using measures of total and iterim persistence of shocks for stationary processes based on the impulse response function. We apply our results to post-war US monthly real stock market returns and dividend yields. Our findings seem to confirm that stock returns are very close to white noise, while expected returns are well represented by an AR(1) process with a first-order autocorrelation of .9755. We also find that small unexpected variations in expected returns have a large negative immediate impact on observed returns, which is thereafter compensated by a slowly diminishing positive effect on expected returns.

KEYWORDS: Time Series Processes; Conditional Moments; Expected Returns; Persistence.

Introduction

The first and second conditional moments of economic and financial time series (given past behaviour) are often identified with important economic concepts. For instance, consider the stochastic process for stock market excess returns, r_t , whose first two conditional moments given the information set I_{t-1} are:

$$\mu_t = E(r_t \mid I_{t-1})$$

$$\sigma_t^2 = V(r_t \mid I_{t-1})$$

In this context, μ_t is usually associated with the risk premium of the stock market as a whole, σ_t^2 with its volatility, and μ_t/σ_t^2 with the market price of risk.

In this paper we study the time series properties of the processes for the unobserved conditional mean and variance, μ_t, σ_t^2 , given a specification of the process for the observed time series, r_t . Apart from providing useful insights into the statistical features of time series models, the properties of a process and its conditional mean often have relevant economic implications. For example, the fact that stock market returns have negligible autocorrelations was traditionally regarded as evidence in favour of the present value model with constant expected returns. More recently, though, Shiller (1984), Summers (1986), Poterba and Summers (1988) and Fama and French (1988) showed that near white noise behaviour for observed returns is compatible with a smoothly time-varying expected return whose first-order autocorrelation is high (see also Campbell (1991)). Obviously, from the point of view of explaining movements in asset prices, there is a substantial difference between constant and time-varying expected returns.

Although the information set I_{t-1} generally includes variables other than past values of r_t , for simplicity we start with univariate analysis. In this respect, we derive a general result for the conditional mean of univariate linear processes satisfying standard regularity conditions. Then, we apply this result to various models of interest used in the analysis of economic and financial time series, such as stationary ARMA, ARIMA, multiplicative seasonal ARIMA, and ARFIMA models. In order to apply our general result to the conditional

variance, we use the fact that σ_t^2 is the conditional mean of the squared innovation, and that conditional heteroskedasticity models often have a straightforward interpretation as linear processes for the squared innovations. We present examples for GARCH, GARCH-M and QARCH models.

We also look at the persistence of shocks in the conditional mean process as compared to the persistence of shocks in the process for the observed variable. However, most persistence measures put forward in the literature imply that shocks to stationary variables have zero persistence, despite the fact that the response of a variable to a shock varies substantially across different covariance stationary processes. For that reason, we use a measure of persistence of shocks for stationary processes based on the impulse response function, which captures the importance of the deviations of a series from its unperturbed path following a single shock.

The univariate framework, though, is often too restrictive for the analysis of such issues, as there is only one shock that drives the processes for the observed variable and its conditional mean. In other words, the joint process for r_t and its conditional mean is reduced-rank with a singular covariance matrix for the innovations. This has been long realized, and two main alternative approaches have been proposed. The first one specifies directly a stochastic process for the conditional moment with “its own” innovation. In this way, the stochastic volatility literature often assumes that the (log) conditional variance follows a univariate AR(p) process. Similarly, Campbell (1990) assumes that the expected stock return follows a univariate AR(p) process, and then derives the implied process for observed returns. Here, we follow the opposite route, which is more in line with the tradition in Rational Expectations econometrics. That is, we start from an observed multivariate process for the variable of interest and other variables that Granger-cause it, and then derive the implicit process for its expected value conditional on past information. In this multivariate framework, we also compare the persistence of shocks to a series and its conditional mean.

As an empirical illustration we look at post-war US monthly real stock market returns. Since several studies have found some predictability in returns using lagged dividend yields, we estimate a bivariate model for these two variables. Then, we obtain the implied joint process for actual and ex-

pected returns, as well as their univariate representations.

The rest of the paper is organized as follows. In Section 2 we present the results related to univariate analysis. A measure of persistence for univariate stationary processes is introduced in Section 3. The next two sections are multivariate extensions of the previous ones. In Section 6 we derive conditions under which white-noise behaviour for a variable is compatible with a persistent stochastic process for its conditional mean. The results of the empirical application are discussed in Section 7. Finally, our conclusions are presented in Section 8.

2 The Conditional Mean of a Univariate Process

In this section we derive the time series processes for the conditional mean of some commonly used univariate time series processes. We begin by stating the general result for linear processes, and then we analyze several cases of interest such as ARMA and ARIMA processes, univariate GARCH and GARCH in mean processes.

Let L denote the lag (or backshift) operator, $Lx_t = x_{t-1}$, and let $a(L)$ and $b(L)$ denote (non-normalized) polynomials in L . A linear stochastic process of order k and h can be written as¹

$$[1 - a(L)]x_t = [1 - b(L)]\epsilon_t \quad (1)$$

or

$$(1 - a_1L - \dots - a_kL^k)x_t = (1 - b_1L - \dots - b_hL^h)\epsilon_t$$

where ϵ_t is the innovation in the process at time t , and the roots of $a(L) = 1$ and $b(L) = 1$ lie on or outside the unit circle. This includes integrated and invertible processes (whether strictly or not) but rules out explosive as well as non-invertible processes. However, if some of the roots of $b(L) = 1$ lie inside the unit circle the process admits the invertible representation $[1 - a(L)]x_t = [1 - b'(L)]\epsilon'_t$, where $\epsilon'_t = [1 - b'(L)]^{-1}[1 - b(L)]\epsilon_t$.

¹The notation usually adopted for linear process is $a(L)x_t = b(L)\epsilon_t$. Our choice is based on convenience, as it makes the algebra and exposition much simpler.

Let $\mu_t = E_{t-1}(x_t)$ denote the conditional mean² of x_t , i.e. its minimum mean square error one-period ahead forecast.

Proposition 1: *A linear process of order k and h for x_t , $[1 - a(L)]x_t = [1 - b(L)]\epsilon_t$, implies that μ_t follows another linear process given by $[1 - a(L)]\mu_t = [a(L) - b(L)]\epsilon_t$.*

If $a(L) = 1$ and $a(L) - b(L) = 0$ do not share roots in common, the conditional mean follows a linear process of order k and $m - 1$ ($m = \max(k, h)$). However, k and $m - 1$ should be interpreted as maximum orders because cancellation of common factors often occurs, as will be illustrated in some examples below. Nevertheless, the common factors will never involve a reduction in the order of integration since x_t and its conditional mean are always cointegrated. Notice also that no assumption has been made regarding k and h . Therefore, the result holds for processes of infinite order. Finally, note that the innovation in the process for μ_t is proportional to the innovation in the lagged value of x_t . In the rest of this section we shall apply the above result to several models of practical interest.

2.1 ARMA-type Processes

Autoregressive Integrated Moving Average (ARIMA) models are the best known linear processes. An ARIMA(p, d, q) process can be represented as

$$[1 - \phi(L)](1 - L)^d x_t = [1 - \theta(L)]\epsilon_t$$

or

$$[1 - \Phi(L)]x_t = [1 - \theta(L)]\epsilon_t$$

where $1 - \Phi(L) = [1 - \phi(L)](1 - L)^d$ and the roots of $\phi(L) = 1$ lie outside the unit circle. From Proposition 1, it is easy to see that the conditional mean of an ARIMA model follows a process that also has the autocorrelation function (ACF) of an ARIMA process. Specifically,

Result 1: *An ARIMA(p, d, q) process for x_t implies that $(1 - L)^d \mu_t$ displays the ACF of an ARMA($p, m - 1$) process, with $m = \max(p + d, q)$, the i^{th} AR coefficient given by ϕ_i , and the i^{th} MA coefficient given by $(\Phi_i - \theta_i)/(\Phi_1 - \theta_1)$ if $\Phi_1 \neq \theta_1$.*

²In this paper the terms conditional mean and linear projection are treated as equivalent unless otherwise specified.

As a simple example, take the ARMA(1,1) model

$$x_t = \phi x_{t-1} + \epsilon_t - \theta \epsilon_{t-1}$$

In this case, the process for the conditional mean is the following AR(1)

$$\mu_{t+1} = \phi \mu_t + (\phi - \theta) \epsilon_t$$

If we let $\phi - \theta$ go to zero, we can make the ARMA(1,1) process as close as desired to white noise, and yet keep the first autocorrelation of the conditional mean equal to ϕ . However, the variance of the mean goes to zero with $\phi - \theta$, so that it actually converges to a constant in the limit.

As a second example, consider the AR(2) process

$$[1 - \phi(L)]x_t = (1 - \phi_1 L - \phi_2 L^2)x_t = \epsilon_t$$

In this model, the conditional mean follows an ARMA(2,1) process, unless $\phi_1 = \sqrt{\phi_2^3/(1 + \phi_2)}$, in which case it reduces to an AR(1).

With some minor modifications, Result 1 is readily applicable to Multiplicative seasonal ARIMA models. The purely seasonal ARIMA_S(P,D,Q) model takes the form

$$[1 - \Phi_s(L)]x_t = [1 - \phi_s(L)](1 - L^s)^D x_t = [1 - \theta_s(L)]\epsilon_t$$

where $\phi_s(L)$ and $\theta_s(L)$ are polynomials in L^s , and typically the value of s is 4 or 12 for quarterly or monthly data. The equation $\Phi_s(L) = 1$ has $D \times s$ roots on the unit circle. Multiplicative models combine features of purely seasonal and ordinary ARIMA models. The general ARIMA(p,d,q) \times_s (P,D,Q,) model takes the form

$$[1 - \Phi(L)][1 - \Phi_s(L)]x_t = [(1 - \theta(L))[1 - \theta_s(L)]\epsilon_t$$

Result 1 can be modified accordingly:

Result 2: A multiplicative ARIMA(p,d,q) \times_s (P,D,Q) process for x_t implies that the stationary transformation of the conditional mean, $(1-L)^d(1-L^s)^D \mu_t$, displays the ACF of an ARMA(p+P,m-1) process, with $m = \max(p+d+Ps+Ds, q+Qs)$.

In this case, the process for the conditional mean has an expression which is of the ARIMA type but, in general, it will not display a multiplicative moving average part, unless the model is purely seasonal. As an example, consider the quarterly airline model

$$(1 - L)(1 - L^4)x_t = (1 - \theta_1 L)(1 - \theta_4 L^4)\epsilon_t$$

This yields as conditional mean

$$(1 - L)(1 - L^4)\mu_{t+1} = (1 - \theta_1)\epsilon_t + (1 - \theta_4)\epsilon_{t-3} + (\theta_1\theta_4 - 1)\epsilon_{t-4}$$

Another class of linear processes which has been increasing popular recently are Autoregressive Fractionally Integrated Moving Average (ARFIMA) models. They were introduced to represent stochastic process which do not display the typical exponential decay in the correlogram associated with ARMA models. Following Granger and Joyeux (1980) and Hosking (1981), the simple ARFIMA(0, γ , 0) takes the form

$$(1 - L)^\gamma x_t = \epsilon_t$$

where γ is a real number, and

$$(1 - L)^\gamma = \sum_{k=0}^{\infty} \binom{\gamma}{k} (-L)^k = 1 - \gamma L - \frac{1}{2}\gamma(\gamma-1)L^2 - \frac{1}{6}\gamma(\gamma-1)(\gamma-2)L^3 - \dots$$

On the basis of Proposition 1, it is straightforward to show that the conditional mean also follows a fractionally integrated process of order γ , but with an infinite order moving average part. Specifically,

$$(1 - L)^\gamma \mu_t = \phi(L)\epsilon_t = [\gamma L + \frac{1}{2}\gamma(\gamma-1)L^2 + \frac{1}{6}\gamma(\gamma-1)(\gamma-2)L^3 + \dots]\epsilon_t$$

Again, the observed process and its conditional mean are fractionally cointegrated, so that no reduction in the order occurs.

2.2 ARCH-type Processes

Proposition 1 can readily be applied to autoregressive conditional heteroskedastic processes. Let ϵ_t denote the innovation in a stochastic process.

Then ϵ_t is said to follow a (semi-strong) GARCH(p,q) process if $E_{t-1}(\epsilon_t) = 0$ and $E_{t-1}(\epsilon_t^2) = \sigma_t^2$ with

$$[1 - \beta(L)]\sigma_t^2 = \alpha_0 + \alpha(L)\epsilon_t^2$$

A GARCH(p,q) process can be represented as an ARMA(m,p) for the squared error process (with $m=\max[p,q]$), that is

$$[1 - \alpha(L) - \beta(L)]\epsilon_t^2 = \alpha_0 + [1 - \beta(L)]v_t$$

where $v_t = \epsilon_t^2 - \sigma_t^2$. Given that σ_t^2 is the conditional mean of ϵ_t^2 , a straightforward application of Result 1 leads to an ARMA(m,m-1) representation for σ_t^2 . But since $[\alpha(L) + \beta(L)] - \beta(L) = \alpha(L)$, Result 1 simplifies as follows:³

Result 3: *A fourth-moment stationary GARCH(p,q) process for ϵ_t implies that σ_t^2 displays the ACF of an ARMA(m,q-1) process, with $m=\max(p,q)$, the i^{th} AR coefficient given by $\alpha_i + \beta_i$, and the i^{th} MA coefficient given by α_i/α_1 if $\alpha_1 \neq 0$. If the GARCH process is not fourth-moment stationary, σ_t^2 will behave as an ARMA process with infinite variance innovations.*

In most empirical applications, the simple GARCH(1,1) specification is adopted

$$\sigma_t^2 = \alpha_0 + \alpha_1\epsilon_{t-1}^2 + \beta_1\sigma_{t-1}^2$$

and Result 3 gives

$$\sigma_t^2 = \alpha_0 + (\alpha_1 + \beta_1)\sigma_{t-1}^2 + \alpha_1 v_{t-1}$$

that is, an AR(1) process for the conditional variance with autoregressive parameter equal to $\alpha_1 + \beta_1$ and variance of innovations proportional to α_1^2 .

Also, the following result will prove useful:

Result 4: *A GARCH(p,q) process for ϵ_t implies that the ACF of σ_t^2 can only take non-negative values.*

Our results can be extended to GARCH in mean (GARCH-M) processes. A variable x_t is said to follow a GARCH-M model of orders p and q, if

$$x_t = \delta\sigma_t^2 + \epsilon_t \tag{2}$$

³Result 3 can also be found in Fiorentini and Maravall (1996)

where σ_t^2 is the conditional variance of ϵ_t which, in turn, follows a GARCH(p,q). Hong (1989) investigates the ACF of x_t for the GARCH-M(1,1) case. Results 3 and 4 allow us to generalize his findings. First, notice that the conditional mean of x_t is proportional to its conditional variance. Hence, according to Result 3, μ_t follows an ARMA(m,q-1) process since the constant δ only affects the variance of the innovation in the ARMA process for the conditional mean, but not its autocorrelation structure. Second, provided that $E(\epsilon_t^3) = 0$, equation (2) states that x_t is the sum of two components uncorrelated at all leads and lags: a white-noise component (ϵ_t) and a constant times the conditional variance. Therefore, since $m \geq q$, we have that:

Result 5: *A GARCH-M(p,q) process for x_t with $E(\epsilon_t^3) = 0$ implies that x_t displays the ACF of an ARMA(m,m) process, with $m = \max(p,q)$, and the i^{th} AR coefficient given by $\alpha_i + \beta_i$.*

To derive the MA coefficients, one computes the variance and the first m autocovariances of the moving average part of the model, and then solves a standard system of nonlinear equations. The GARCH-M model, however, is an example in which the difference between conditional means and linear projections is important.

Also, Results 4 and 5 imply that:

Corollary 1: *If x_t follows a GARCH-M(p,q) process, its ACF can only take non-negative values. In particular, for $k \neq 0$, $Cov(x_t, x_{t-k}) = \delta^2 Cov(\sigma_t^2, \sigma_{t-k}^2)$.*

This last result may explain the poor empirical performance of GARCH-M models. It tells us that regardless of the parameter values, and in particular regardless of the sign of δ , the autocorrelations of x_t implied by the model are all positive.

Proposition 1 can also be applied to other ARCH models in the literature, such as the Generalized Quadratic ARCH (GQARCH) model of Sentana (1995). The conditional variance of a GQARCH(1,1) model is given by:

$$\sigma_t^2 = \alpha_0 + \psi_1 \epsilon_{t-1} + \alpha_1 \epsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2$$

It is straightforward to see that the process for the conditional variance is an AR(1) given by

$$(1 - (\alpha_1 + \beta_1)L)\sigma_{t+1}^2 = \alpha_0 + (\alpha_1 v_t + \psi_1 \epsilon_t)$$

while for the squared innovations we have

$$(1 - (\alpha_1 + \beta_1)L)\epsilon_t^2 = \alpha_0 + (1 - \beta_1L)v_t + \psi_1L\epsilon_t = (1 - \beta_1^*L)a_t$$

i.e. an ARMA(1,1). Exact expressions for the parameters of these processes can be obtained by solving a nonlinear system of two covariance equations. Notice that when $\psi_1 = 0$ the result is exactly equal to the GARCH case.

3 Persistence of Shocks in Covariance Stationary Time Series

As we mentioned in the introduction, the comparison between the persistence of shocks in a variable and its conditional mean often bears relevant economic implications. However, the persistence of economic shocks is usually measured by looking at the long-run effect of an innovation on the level of a series (e.g. Campbell and Mankiw, 1987). As a consequence, shocks to stationary processes are usually assigned zero persistence. At the same time, however, stationary processes are often referred to as showing “high” or “low” persistence to shocks. For instance, a stationary AR(1) process is labelled highly persistent when the value of the autoregressive parameter is close to 1, since such a process will take a long time to revert to its mean following a shock. But, how persistent is an ARMA(1,1) whose autoregressive and moving average coefficients are both close to 1? In what follows, we introduce a measure of persistence of shocks that can be applied to any covariance stationary process.

Let $x_t = \Psi(L)\epsilon_t$ denote the Wold representation of the unperturbed process, where $\Psi(L)$ is square-summable. Let's now define the perturbed process $x_t^* = \Psi(L)\epsilon_t^*$, where $\epsilon_s^* = \epsilon_s$ ($\forall s \neq t$), and $\epsilon_t^* = \epsilon_t + 1 \times \sigma_\epsilon$. We want a measure of how much x_t^* deviates from x_t . Obviously, since the process is stationary, the net effect on x_{t+k}^* of a shock to ϵ_t is zero in the limit. However, the route taken by x_{t+k}^* to go back to its original path x_{t+k} may differ substantially across different models. For instance, if x_t is white noise (i.e. $x_t = \epsilon_t$), the original unperturbed level of the series is restored after one period. In contrast, if x_t follows an AR(1) process ($x_t = \phi x_{t-1} + \epsilon_t$) with $\phi = .95$, x_{t+k}^* will stay significantly “far away” from x_{t+k} for a long period of time. In other words,

the deviation from the original path in response to a shock will be substantial. On the contrary, when $\phi = .1$ the shock will effectively exhaust its impact very briefly and the deviation of x_{t+k}^* from x_{t+k} will be inappreciable. In the case of an ARMA(1,1) process ($x_t = \phi x_{t-1} + \epsilon_t - \theta \epsilon_{t-1}$) with $\phi = .95$ and $\theta = .9$, the shock provokes little variation on x_{t+k}^* but the series will take a long time to go back to its original level.

Since $x_{t+k}^* - x_{t+k} = \psi_k \sigma_\epsilon$, any “reasonable” measure of the persistence of shocks must be based on the impulse response function (IRF). The mean or median lags are potential candidates. However, they are only valid for non-negative impulse response functions, when the IRF can be interpreted as a probability distribution for time. For instance, the mean and median lag give sensible answers for the model $x_t = .45x_{t-1} + \epsilon_t$, but not for $x_t = -.45x_{t-1} + \epsilon_t$ or $x_t = .45x_{t-1} + \epsilon_t - .8\epsilon_{t-1}$ or $x_t = -.45x_{t-1} + \epsilon_t + .8\epsilon_{t-1}$, even though their impulse response functions are all identical in magnitude (see Figure 1).

For that reason, we propose the use of

$$P_\infty(x_t | \epsilon_t) = \sum_{j=0}^{\infty} \psi_j^2$$

as a measure of the persistence of shocks. In principle, $\sum_{j=0}^{\infty} |\psi_j|$ could play a similar role except that not all covariance stationary processes have Wold representations with absolute-summable coefficients. Besides the algebra of our measure is simpler, and its interpretation straightforward since

$$P_\infty(x_t | \epsilon_t) = \frac{V(x_t)}{V(\epsilon_t)}$$

i.e. the ratio of the variance of the process to the variance of the shocks.

Diebold and Rudebusch (1989) have forcefully argued that sometimes it is more interesting to look at the effect of a shock on a variable k periods after its occurrence. For this purpose, we suggest to use

$$P_k(x_t | \epsilon_t) = \sum_{j=0}^k \psi_j^2$$

as a measure of the interim persistence of shocks. Again, the interpretation of the interim measure is also immediate, since

$$P_k(x_t | \epsilon_t) = \frac{V(x_{t+k} - \hat{x}_{t+k|t-1})}{V(\epsilon_t)}$$

i.e. the ratio of the variance of the $(k+1)$ -period-ahead forecast error to the variance of the shock. Obviously, for covariance stationary processes, $\hat{x}_{t+k|t-1}$ converges to $E(x_{t+k})$, and $P_k(x_t | \epsilon_t)$ to $P_\infty(x_t | \epsilon_t)$. But unlike $P_\infty(x_t | \epsilon_t)$, the k -period measure $P_k(x_t | \epsilon_t)$ can be used and interpreted for non-stationary processes as well.

Let's consider some examples to appreciate how such measures work in practice. In ARMA(1,1) models, $x_t = \phi x_{t-1} + \epsilon_t - \theta \epsilon_{t-1}$, we obtain

$$P_k(x_t | \epsilon_t) = 1 + (\phi - \theta)^2 \frac{1 - \phi^{2k}}{(1 - \phi^2)} \quad P_\infty(x_t | \epsilon_t) = 1 + (\phi - \theta)^2 \frac{1}{(1 - \phi^2)}$$

In particular, for AR(1) models ($\theta = 0$), $P_\infty(x_t | \epsilon_t) = \frac{1}{(1 - \phi^2)}$ is a monotonic transformation of the absolute value of ϕ . Therefore, our measure would say that the process is more persistent when $\phi = .95$ than when $\phi = .1$, which is in agreement with widely held views. Notice that our measure of persistence for white noise (i.e. $\phi = \theta$) is 1, and this represents its lower bound. There is no upper bound, of course, since it will be infinite for a non-stationary IMA(1,1) process. However, if the moving average parameter is close to one, say $\theta = .98$, the persistence of a shock after 400 periods (a century of quarterly data !) is only $P_{400}(x_t | \epsilon_t) = 1.16$, well below the persistence of a stationary AR(1) with autoregressive parameter equal to .5.

Similarly, for a purely fractionally integrated process $(1 - L)^\gamma x_t = \epsilon_t$

$$P_k(x_t | \epsilon_t) = \sum_{i=0}^k \left(\frac{\Gamma(i+\gamma)}{\Gamma(i+1)\Gamma(\gamma)} \right)^2 \quad P_\infty(x_t | \epsilon_t) = \frac{\Gamma(1-2\gamma)}{[\Gamma(1-\gamma)]^2}$$

where $\Gamma()$ is the gamma function. Note that although a fractionally integrated process has a much longer memory than an AR(1) process, the persistence of shocks is not necessarily larger. For instance, if γ is 0.1, $P_\infty(x_t | \epsilon_t) = 1.02$, which is smaller than the persistence of an AR(1) with $\phi = 0.5$ ($P_\infty(x_t | \epsilon_t) = 1.333$).

We are now in a position to compare the persistence of shocks in the conditional mean vis a vis the persistence of shocks in the observed variable. Let

$$x_t = \epsilon_t + \psi_1 \epsilon_{t_1} + \psi_2 \epsilon_{t_2} + \psi_3 \epsilon_{t_3} + \dots$$

be the moving average representation of the process for x_t . Since its conditional mean can be expressed as

$$\mu_{t+1} = \psi_1 \epsilon_t + \psi_2 \epsilon_{t_1} + \psi_3 \epsilon_{t_2} + \dots$$

it follows that

$$P_{\infty}(\mu_{t+1} \mid \epsilon_t) = P_{\infty}(x_t \mid \epsilon_t) - 1$$

That is, the persistence of the only shock that drives the joint process on the observed variable is 1 plus the persistence of the same shock on the conditional mean. Therefore, the lower bound on the persistence of shocks to the mean process is zero, corresponding to a model with constant mean.

The ARMA(1,1) model provides some intuition for the above result. As we saw in section 2.1, if $\phi - \theta$ is very small, it is possible to find examples in which the process for the observed series is very close to white noise, while the process for the conditional mean is an AR(1) with a very high autoregressive parameter. However, the effect of a shock on the conditional mean is also very small, and the deviation of the conditional mean from its original path is negligible. In the limit, the observed series is white noise only if the conditional mean is constant. This fact is behind the traditional misconception that white noise behaviour for stock returns requires constant expected returns. As we shall see in section 6.2, this is no longer necessarily so in a multivariate framework.

As our last example, consider a GARCH(1,1) process $\sigma_t^2 = \alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2$, which can also be written as $\sigma_t^2 = \alpha_0 + (\alpha_1 + \beta_1) \sigma_{t-1}^2 + \alpha_1 v_t$. Given what we have just seen, the persistence of the conditional variance is

$$P_{\infty}(\sigma_{t+1}^2 \mid v_t) = P_{\infty}(\epsilon_t^2 \mid v_t) - 1 = \frac{\alpha_1^2}{1 - (\alpha_1 + \beta_1)^2}$$

Thus, in a GARCH(1,1) process, the persistence of shocks to the conditional variance depends not only on the value of $\alpha_1 + \beta_1$, but also on the value of α_1 (see also Engle and Mustafa, 1992). In particular, as it happens with the conditional mean in ARMA(1,1) models, the conditional variance process will display little persistence to shocks when α_1 is small.

4 The Conditional Mean of a Vector Process

The results of the previous sections can be readily generalized to multivariate processes. Let \mathbf{x}_t denote a vector process of order n . In this section we study the marginal processes for \mathbf{x}_t and its conditional mean μ_t . Furthermore, in section 5 we extend our measure of persistence to multivariate processes.

Finally, in section 6 we shall also consider the joint process for a subvector \mathbf{x}_{it} and its expected value conditional on the whole information set, μ_{it} ⁴.

Let's consider a multivariate linear vector process of order k and h

$$[I - A(L)]\mathbf{x}_t = [I - B(L)]\epsilon_t$$

where ϵ_t is a $n \times 1$ white noise process, with 0 mean and covariance matrix Σ , I is the identity matrix of order n , $A(L)$ is a $n \times n$ matrix whose typical element is a polynomial in L of order k , and $B(L)$ is analogously defined. More explicitly,

$$\mathbf{x}_t = A_1\mathbf{x}_{t-1} + \dots + A_k\mathbf{x}_{t-k} + \epsilon_t + B_1\epsilon_{t-1} + \dots + B_h\epsilon_{t-h}$$

where A_i and B_i denote $n \times n$ matrices of coefficients.

Define $\mu_t = E_{t-1}(\mathbf{x}_t)$ as the $n \times 1$ conditional mean vector. The following result generalizes proposition 1:

Proposition 2: *A vector linear process of order k and h for \mathbf{x}_t , $[1 - A(L)]\mathbf{x}_t = [1 - B(L)]\epsilon_t$, implies that μ_t follows another vector linear process given by $[1 - A(L)]\mu_t = [A(L) - B(L)]\epsilon_t$; the elements of $A(L) - B(L)$ are in general polynomials of degree $m - 1$ with $m = \max(k, h)$.*

As examples, we shall consider standard VARMA and vector GARCH processes.

4.1 VARMA Processes

A VARMA(p, q) process is expressed as

$$[I - \Phi(L)]\mathbf{x}_t = [I - \Theta(L)]\epsilon_t$$

or,

$$\mathbf{x}_t = \Phi_1\mathbf{x}_{t-1} + \dots + \Phi_p\mathbf{x}_{t-p} + \epsilon_t + \Theta_1\epsilon_{t-1} + \dots + \Theta_q\epsilon_{t-q} ;,$$

Direct application of Proposition 2 leads to the following result:

⁴Our results also apply to the analysis of the marginal processes for all possible partitions of \mathbf{x}_t into subvectors $\mathbf{x}_{1t}, \dots, \mathbf{x}_{jt}$ of dimension n_1, \dots, n_j , and their means conditional on the marginalized information sets. The reason is that if the $n \times 1$ vector \mathbf{x}_t follows a linear process, the marginal process for any $n_1 \times 1$ subvector \mathbf{x}_{1t} will also be a linear process. Thus, if n_1 is bigger than 1 the results of this section apply, while if $n_1 = 1$ we can use the results of Section 2 on univariate process.

Result 6: A VARMA(p, q) process for \mathbf{x}_t implies that μ_t displays the ACF of a VARMA($p, m-1$) process, with $m=\max(p, q)$, the i^{th} AR matrix of coefficients given by Φ_i , and the i^{th} MA matrix of coefficients given by $(\Phi_1 - \Theta_1)^{-1} \times (\Phi_i - \Theta_i)$, provided $|\Phi_1 - \Theta_1| \neq 0$.

As an example consider a bivariate VARMA(1,1). According to the above result, the process for the conditional mean is the following VAR(1):

$$\begin{pmatrix} \mu_{1,t+1} \\ \mu_{2,t+1} \end{pmatrix} = \begin{pmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{pmatrix} \begin{pmatrix} \mu_{1,t} \\ \mu_{2,t} \end{pmatrix} + \begin{pmatrix} \phi_{11} - \theta_{11} & \phi_{12} - \theta_{12} \\ \phi_{21} - \theta_{21} & \phi_{22} - \theta_{22} \end{pmatrix} \begin{pmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{pmatrix}$$

Notice that since the matrix of autoregressive coefficients is shared by the observed time series and their conditional means, they will have the same (co-)integration properties.

4.2 Vector GARCH Processes

Let ϵ_t denote a $k \times 1$ vector of innovations. Then ϵ_t is said to follow a multivariate (semi-strong) GARCH(p, q) process if $E_{t-1}(\epsilon_t) = 0$ and $E_{t-1}(\epsilon_t \epsilon_t') = \Sigma_t$, with

$$[1 - \beta(L)] \text{vech} \Sigma_t = \alpha_0 + \alpha(L) \text{vech}(\epsilon_t \epsilon_t')$$

A multivariate GARCH(p, q) process can be represented as the following VARMA(m, p) (with $m=\max[p, q]$) on $\text{vech}(\epsilon_t \epsilon_t')$,

$$[1 - \alpha(L) - \beta(L)] \text{vech}(\epsilon_t \epsilon_t') = \alpha_0 + [1 - \beta(L)] \mathbf{v}_t$$

where $\mathbf{v}_t = \text{vech}(\epsilon_t \epsilon_t' - \Sigma_t)$.

Then, proposition 2 simplifies to

Result 7: A multivariate GARCH(p, q) process for ϵ_t implies that $\text{vech}(\Sigma_t)$ displays the ACF of an VARMA($m, q-1$) process, with $m=\max(p, q)$, the i^{th} AR matrix of coefficients given by $\alpha_i + \beta_i$, and the i^{th} MA matrix of coefficients given by $\alpha_1^{-1} \alpha_i$ provided $|\alpha_1| \neq 0$.

5 Persistence of Shocks in Multivariate models

The same notion of persistence of shocks employed in univariate analysis can be extended to multivariate models. That is, the persistence of a given

shock on a variable can be measured by the variation of the series with respect to the original unperturbed process provoked by that shock.

For simplicity, let's consider a covariance stationary bivariate model⁵

$$[I - A(L)]\mathbf{x}_t = [I - B(L)]\epsilon_t$$

where A and B are 2×2 matrices of polynomials in L , and ϵ_t is 2×1 white noise process with zero mean and covariance matrix Σ . Let

$$\mathbf{x}_t = \Psi(L)\epsilon_t$$

denote its Wold representation, and define a matrix Σ^* such that $\Sigma^*\Sigma^{*\prime} = \Sigma$. Then, the infinite moving average representation of \mathbf{x}_t in terms of the standardized orthogonal innovations $\epsilon_t^* = \Sigma^{*-1}\epsilon_t$ is

$$\mathbf{x}_t = \Psi^*(L)\epsilon_t^*$$

where $\Psi_i^* = \Psi_i\Sigma^*$ and the covariance matrix of ϵ_t^* is the identity matrix.

We can then define the persistence of a shock to $\epsilon_{i,t}^*$ on the j^{th} variable as

$$P_\infty(x_{j,t} \mid \epsilon_{i,t}^*) = \sum_{k=0}^{\infty} \psi_{ji,k}^{*2}$$

where $j = 1, 2$ and $i = 1, 2$.

However, it is well known that the decomposition of the covariance matrix of the one-period ahead prediction errors in the Wold representation is not unique and, thus, that the orthogonal shocks are not identified. More specifically, let Ω be an orthonormal basis of \Re^2 (see appendix B). Then, any orthonormal transformation of Σ^* will provide another infinite MA representation with orthogonal shocks. In particular,

$$\mathbf{x}_t = \Psi^{**}(L)\epsilon_t^{**}$$

where $\Psi_i^{**} = \Psi_i\Sigma^*\Omega$ ⁶. The persistence of the "new" i^{th} shock on $x_{j,t}$ will be

$$P_\infty(x_{j,t} \mid \epsilon_{i,t}^{**}) = \sum_{k=0}^{\infty} \psi_{ji,k}^{**2}$$

⁵The generalization to processes of more variables is straightforward

⁶In fact, there are many more MA representations of a covariance stationary process in terms of "non-fundamental" orthogonal shocks

which in general is different from $P_\infty(x_{j,t} \mid \epsilon_{i,t}^*)$. Therefore, any attempt to define a single measure of persistence for a given variable irrespectively of the shock is largely futile. For that reason, our measure of persistence is conditional on a given specification of the shocks.

Ideally, such a specification should be largely guided by economic theory considerations. Nevertheless, when there is no a priori identification of the orthogonal shocks ϵ_t^* , it is nowadays customary to look at the different Cholesky decompositions of Σ . In particular, when the model is bivariate, researchers analyze two "leading" cases based on the Cholesky decomposition of Σ , and the Cholesky decomposition of the covariance matrix of the re-ordered system, in the hope of finding the results robust to the specification of the shocks. This is what Koop, Pesaran and Potter (1996) call generalized variance decompositions. However, an example in appendix C shows that the different Cholesky decompositions are not necessarily representative.

6 Time Series Processes for a Variable and its Conditional Mean in a Multivariate Model

In Section 2, we saw that in a univariate model, the joint process for an observed variable and its conditional mean is singular, since there is only one shock driving both series. In this section we shall obtain the joint process for a subvector of \mathbf{x}_t and its conditional mean in a multivariate setting.

For this purpose, let $(\mathbf{x}_t, \mu_{t+1})$ be a $2n \times 1$ vector⁷ stochastic process. We want to obtain from it the marginal process for the $2n_1 \times 1$ subvector $(\mathbf{x}_{1,t}, \mu_{1,t+1})$. Importantly, the mean $\mu_{1,t+1}$ is conditional on the full information set I_t .

Consider a VARMA(p,q) process

$$\mathbf{x}_t = A_1 \mathbf{x}_{t-1} + \dots + A_p \mathbf{x}_{t-p} + \epsilon_t + B_1 \epsilon_{t-1} + \dots + B_q \epsilon_{t-q}$$

On the basis of Result 6, we can write

$$\begin{pmatrix} \mathbf{x}_t \\ \mu_{t+1} \end{pmatrix} = \begin{pmatrix} 0 & I \\ 0 & A_1 \end{pmatrix} \begin{pmatrix} \mathbf{x}_{t-1} \\ \mu_t \end{pmatrix} + \sum_{i=2}^p \begin{pmatrix} 0 & 0 \\ 0 & A_i \end{pmatrix} \begin{pmatrix} \mathbf{x}_{t-i} \\ \mu_{t-i+1} \end{pmatrix} +$$

⁷Notice that the temporal phase-shift between the two vectors is only apparent as both \mathbf{x}_t and μ_{t+1} belong to the information set I_t .

$$\begin{pmatrix} I \\ C_1 \end{pmatrix} \epsilon_t + \sum_{j=2}^m \begin{pmatrix} 0 \\ C_j \end{pmatrix} \epsilon_{t-j+1} \quad (3)$$

where I is an $n \times n$ identity matrix, $m = \max(p, q)$ and $C_i = A_i + B_i$ (with $A_i = 0$ if $i > p$ and $B_i = 0$ if $i > q$). This gives us the joint process for $(\mathbf{x}_t, \mu_{t+1})$ directly.

Let's partition \mathbf{x}_t and μ_{t+1} into two subvectors $(\mathbf{x}_{1,t}, \mathbf{x}_{2,t})$ and $(\mu_{1,t+1}, \mu_{2,t+1})$ respectively. Marginalizing with respect to $\mathbf{x}_{1,t}$ and $\mu_{1,t+1}$ results in the joint process for this set of variables and their expected values. However, given that the conditional means $\mu_{1,t}$ are based on the information contained on past values of both $\mathbf{x}_{1,t}$ and $\mathbf{x}_{2,t}$, the innovations to the joint process $(\mathbf{x}_{1,t}, \mu_{1,t+1})$ are not linearly dependent in general, since they are a full rank linear transformation of the innovations in all the observed variables.

As a simple example, let's consider a bivariate VAR(1) model for some variable, r_t say, and some predictor variable, δ_t say, which helps explain $\mu_{r,t}$ ⁸.

$$\begin{pmatrix} r_t \\ \delta_t \end{pmatrix} = A \begin{pmatrix} r_{t-1} \\ \delta_{t-1} \end{pmatrix} + \begin{pmatrix} u_t \\ v_t \end{pmatrix}$$

In this case, representation (3) is simply

$$\begin{pmatrix} r_t \\ \mu_{r,t+1} \\ \delta_t \\ \mu_{\delta,t+1} \end{pmatrix} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & a_{11} & 0 & a_{12} \\ 0 & 0 & 0 & 1 \\ 0 & a_{21} & 0 & a_{22} \end{pmatrix} \begin{pmatrix} r_{t-1} \\ \mu_{r,t} \\ \delta_{t-1} \\ \mu_{\delta,t} \end{pmatrix} + \begin{pmatrix} 1 & 0 \\ a_{11} & a_{12} \\ 0 & 1 \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} u_t \\ v_t \end{pmatrix}$$

which, for this particular model, coincides with the (re-arranged) Akaike (1974) state space representation.

Marginalizing with respect to r_t and its conditional mean yields

$$\begin{pmatrix} r_t \\ \mu_{r,t+1} \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 0 & tr(A) \end{pmatrix} \begin{pmatrix} r_{t-1} \\ \mu_{r,t} \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ 0 & -|A| \end{pmatrix} \begin{pmatrix} r_{t-2} \\ \mu_{r,t-1} \end{pmatrix} + \begin{pmatrix} u_t \\ w_t \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ -|A| & 0 \end{pmatrix} \begin{pmatrix} u_{t-1} \\ w_{t-1} \end{pmatrix}$$

⁸The change of notation is made for consistency with the empirical section.

where $tr(A)$ and $|A|$ denote the trace and the determinant of the matrix A , and $w_t = a_{11}u_t + a_{12}v_t$. Thus, we obtain a (reduced rank) VARMA(2,1) model with a full rank covariance matrix for the innovations u_t and w_t , whose covariance is

$$E(u_t w_t) = a_{11}\sigma_u^2 + a_{12}\sigma_{uv}$$

and their correlation

$$\rho_{uw} = \frac{a_{11}\sigma_u^2 + a_{12}\sigma_{uv}}{\sigma_u \sqrt{a_{11}^2\sigma_u^2 + a_{12}^2\sigma_v^2 + 2a_{11}a_{12}\sigma_{uv}}}$$

Therefore, its Wold decomposition will be given by

$$\begin{pmatrix} u_t \\ w_t \end{pmatrix} + \begin{pmatrix} 0 & 1 \\ -|A| & tr(A) \end{pmatrix} \begin{pmatrix} u_{t-1} \\ w_{t-1} \end{pmatrix} + \sum_{j=2}^{\infty} \begin{pmatrix} -|A|g_{j-2} & g_{j-1} \\ -|A|g_{j-1} & g_j \end{pmatrix} \begin{pmatrix} u_{t-j} \\ w_{t-j} \end{pmatrix}$$

where $g_j = tr(A)g_{j-1} - |A|g_{j-2}$ with $g_0 = 1, g_1 = tr(A)$.

Note that as expected, the effect of u_{t-j} and w_{t-j} on r_t for $j > 0$ is exactly the same as their effect on $\mu_{r,t}$. As a consequence, whatever the orthogonalization of the shocks, the persistence of a given shock on the observed process is at least as large as its persistence on the conditional mean. Unlike in the univariate case, though, it is possible for both effects to be equal in size if a shock does not have any contemporaneous impact on r_t .

In this general case the marginal processes for r_t and its conditional mean are

$$(1 - tr(A)L + |A|L^2)\mu_{r,t+1} = a_{11}u_t - |A|u_{t-1} + a_{12}v_t = (1 - \pi L)\eta_t$$

and

$$(1 - tr(A)L + |A|L^2)r_t = u_t - a_{22}u_{t-1} + a_{12}v_{t-1} = (1 - \theta L)\xi_t$$

where the values of π , θ , σ_η and σ_ξ can be easily obtained by solving a simple quadratic equation.

Our next exercise is to investigate in a multivariate setup the response to shocks of r_t and its conditional mean. In particular, we analyze shocks that affect the observed variables r_t and δ_t directly, as they may be simpler to interpret in practice. In this respect, it is convenient to re-write the model in terms of u_t and v_t as

$$\begin{aligned}
\begin{pmatrix} r_t \\ \mu_{r,t+1} \end{pmatrix} &= \begin{pmatrix} 0 & 1 \\ 0 & \text{tr}(A) \end{pmatrix} \begin{pmatrix} r_{t-1} \\ \mu_{r,t} \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ 0 & -|A| \end{pmatrix} \begin{pmatrix} r_{t-2} \\ \mu_{r,t-1} \end{pmatrix} + \\
&\quad \begin{pmatrix} 1 & 0 \\ a_{11} & a_{12} \end{pmatrix} \begin{pmatrix} u_t \\ v_t \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ -|A| & 0 \end{pmatrix} \begin{pmatrix} u_{t-1} \\ v_{t-1} \end{pmatrix} = \\
&\quad \begin{pmatrix} 1 & 0 \\ a_{11} & a_{12} \end{pmatrix} \begin{pmatrix} u_t \\ v_t \end{pmatrix} + \begin{pmatrix} a_{11} & a_{12} \\ \text{tr}(A)a_{11} - |A| & \text{tr}(A)a_{12} \end{pmatrix} \begin{pmatrix} u_{t-1} \\ v_{t-1} \end{pmatrix} + \\
&\quad \sum_{j=2}^{\infty} \begin{pmatrix} a_{11}g_{j-1} - |A|g_{j-2} & a_{12}g_{j-1} \\ a_{11}g_j - |A|g_{j-1} & a_{12}g_j \end{pmatrix} \begin{pmatrix} u_{t-j} \\ v_{t-j} \end{pmatrix}
\end{aligned}$$

To keep the algebra as simple as possible, we only consider in detail those special cases that lead to an AR(1) process for $\mu_{r,t}$. The rest of the section is devoted to a detailed analysis of such cases.

6.1 Case A: $a_{12} = 0$

When $a_{12} = 0$ the joint process for r_t and its conditional mean is a reduced-rank VAR(1) with a singular covariance matrix for the innovations. That is

$$\begin{pmatrix} r_t \\ \mu_{r,t+1} \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 0 & a_{11} \end{pmatrix} \begin{pmatrix} r_{t-1} \\ \mu_{r,t} \end{pmatrix} + \begin{pmatrix} 1 \\ a_{11} \end{pmatrix} u_t$$

Therefore, the marginal processes are

$$(1 - a_{11}L)\mu_{r,t+1} = a_{11}u_t$$

and

$$(1 - a_{11}L)r_t = u_t$$

The reason is obvious. When $a_{12} = 0$, δ_t does not Granger-cause r_t , so that we are in effect back to the univariate case. As we saw in Section 2, it is impossible to achieve a white noise representation for a series with time-varying conditional mean in the context of linear models (see Granger, 1983).

6.2 Case B: r_t white-noise

Given that the marginal process for r_t is ARMA(2,1), r_t cannot be exactly white-noise unless one of the roots of $(1 - \text{tr}(A)L + |A|L^2) = 0$ is zero. But this requires $|A| = 0$, so that the VAR(1) for the observed variables r_t and δ_t has to be of reduced rank.

In this case we can distinguish several different possibilities, namely

B1) $a_{11} = 0$ and $a_{12} = 0$

B2) $a_{11} = 0$ and $a_{21} = 0$

B3) $a_{22} = 0$ and $a_{12} = 0$

B4) $a_{22} = 0$ and $a_{21} = 0$

B5) $a_{11}a_{22} = a_{12}a_{21}$

First notice that B1 is nested into case A. In particular, in case B1 r_t is white noise and the conditional mean is constant. Similarly, case B3 is in effect the same as case A.

Consider case B4. The joint process is now

$$\begin{pmatrix} r_t \\ \mu_{r,t+1} \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 0 & a_{11} \end{pmatrix} \begin{pmatrix} r_{t-1} \\ \mu_{r,t} \end{pmatrix} + \begin{pmatrix} 1 & 0 \\ a_{11} & a_{12} \end{pmatrix} \begin{pmatrix} u_t \\ v_t \end{pmatrix}$$

while the marginal processes are

$$(1 - a_{11}L)\mu_{r,t+1} = a_{12}v_t + a_{11}u_t$$

and

$$(1 - a_{11}L)r_t = u_t + a_{12}Lv_t$$

However, when $a_{22} = a_{21} = 0$ the variable δ_t in the original VAR(1) is white noise. This makes this case empirically uninteresting and we will not analyze it further.

6.2.1 Case B2

When $a_{11} = 0$ and $a_{21} = 0$ the joint process is

$$\begin{pmatrix} r_t \\ \mu_{r,t+1} \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 0 & a_{22} \end{pmatrix} \begin{pmatrix} r_{t-1} \\ \mu_{r,t} \end{pmatrix} + \begin{pmatrix} 1 & 0 \\ 0 & a_{12} \end{pmatrix} \begin{pmatrix} u_t \\ v_t \end{pmatrix}$$

while the marginal processes are

$$(1 - a_{22}L)\mu_{r,t+1} = a_{12}v_t$$

and

$$(1 - a_{22}L)r_t = (1 - a_{22}L)u_t + a_{12}Lv_t = (1 - \theta L)\eta_t$$

Therefore, the process for r_t is an ARMA(1,1), which reduces to white noise if cancellation occurs⁹. This case was first analyzed by Campbell (1991). Note that here the conditional mean is exactly proportional to the observed process δ_t , so that $w_t = a_{12}v_t$. This simplifies the analysis considerably.

The covariance equation system gives

$$\frac{-\theta}{1 + \theta^2} = \frac{-a_{22}\sigma_u^2 + a_{12}\sigma_{uv}}{(1 + a_{22}^2)\sigma_u^2 + a_{12}^2\sigma_v^2 - 2a_{22}a_{12}\sigma_{uv}} \quad (4)$$

Let $\gamma = \sigma_u/\sigma_v$ and $\rho = \sigma_{uv}/(\sigma_u\sigma_v)$, and, without loss of generality, take $a_{12} = 1$ as scaling normalization.

Equation (4) can then be written as

$$\frac{-\theta}{1 + \theta^2} = \frac{-a_{22}\gamma + \rho}{(1 + a_{22}^2)\gamma + 1/\gamma - 2a_{22}\rho} \quad (5)$$

In principle a_{22} can take any value below 1, but for r_t to be white noise, it must be the case that $\theta = a_{22}$. Such a restriction has implications for all the other parameters, and in particular for the covariance matrix of the innovations. Imposing this restriction in (5) we get the following relationship between γ and ρ .

$$\gamma = \frac{-a_{22}}{1 - a_{22}^2} \frac{1}{\rho} \quad (6)$$

It is more interesting, though, to look at the implications of $\theta = a_{22}$ for the relationship between the R^2 of the first equation, i.e. the proportion of variance of r_t explained by its conditional mean ($\text{var}(\mu_{r,t})/\text{var}(r_t)$), and the correlation between the innovations. Since R^2 is related to γ through

$$R^2 = \frac{1}{1 + (1 - a_{22}^2)\gamma^2} \quad (7)$$

⁹Note that if r_t is white noise, so is any temporal aggregate such as $r_t + r_{t-1}$.

we can obtain the mapping between R^2 and ρ compatible with white noise behaviour for r_t by combining equations (7) and (6). Such a mapping is presented in the first plot of Figure 2 for values a_{22} ranging from .98 to -.98. As shown by Campbell (1991), exact white noise behaviour for r_t can be obtained with a_{22} positive as long as shocks to r_t and shocks to its "expected value" δ_t are negatively correlated. Notice that the closer a_{22} is to 1, the larger the correlation must be in absolute value.

To gain some intuition on this result, it is convenient to look at the impulse response functions of the variables with respect to the different shocks.

Let's write

$$Cov \begin{pmatrix} u_t \\ v_t \end{pmatrix} = \sigma_v^2 \Sigma = \sigma_v^2 \begin{pmatrix} \gamma^2 & \rho\gamma \\ \rho\gamma & 1 \end{pmatrix}$$

We only consider two kinds of shocks: those that affect r_t directly through u_t , and those that affect δ_t directly through v_t . To study the response to a shock in u_t we use the Cholesky decomposition of Σ in the original system, i.e.

$$\Sigma^* = \begin{pmatrix} \gamma & 0 \\ \rho & \sqrt{1-\rho^2} \end{pmatrix} = \begin{pmatrix} k/\rho & 0 \\ \rho & \sqrt{1-\rho^2} \end{pmatrix}$$

where $k = -a_{22}/(1 - a_{22}^2)$.

The corresponding impulse response functions are

$$\text{IRF}_0(r_t) = 1; \text{IRF}_j(r_t) = (\rho/\gamma)a_{22}^{j-1}, \text{ for } j > 0.$$

$$\text{IRF}_j(\mu_{r,t+1}) = (\rho/\gamma)a_{22}^j, j = 0, \dots, \infty.$$

These impulse response function are displayed in the lower plots of Figure 2, for $a_{22} = 0.7$ and $\rho = -0.9$, which correspond to $\gamma = \sigma_u/\sigma_v = 1.52$ and $R^2 = 0.46$. Note that since r_t is white noise, the initial positive effect of a shock to u_t is slowly compensated by the negative impact on δ_t .

We can also compute the persistence of a shock to u_t on r_t and its conditional mean.

$$P_\infty(r_t | u_t) = 1 + 1/\gamma^2 \left(\frac{\rho^2}{1 - a_{22}^2} \right) \quad P_\infty(\mu_{r,t+1} | u_t) = 1/\gamma^2 \left(\frac{\rho^2}{1 - a_{22}^2} \right)$$

Perhaps more interesting in the study of the effects of a shock to the conditional mean, v_t . To do so, we use the Cholesky decomposition of Σ in the

re-ordered system

$$\Sigma^{**} = \begin{pmatrix} \gamma\sqrt{1-\rho^2} & \rho\gamma \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} k(\sqrt{1-\rho^2}/\rho) & k \\ 0 & 1 \end{pmatrix}$$

Now we get

$$\text{IRF}_0(r_t) = k; \text{IRF}_j(r_t) = a_{22}^{j-1}, \text{ for } j > 0.$$

$$\text{IRF}_j(\mu_{r,t+1}) = a_{22}^j, j = 0, \dots, \infty.$$

Note that for a_{22} close to 1, k is be very large and negative. Therefore, a positive shock to v_t has a very negative immediate impact on r_t , which is then slowly reversed by the positive and slowly decaying effect on its conditional mean. Such a pattern is a direct consequence of the restrictions that guarantee a white noise marginal process for r_t .

Again, it is easy to compute the persistence of a shock to v_t on r_t and its conditional mean.

$$P_\infty(r_t | v_t) = k^2 + \frac{1}{1 - a_{22}^2} : \quad P_\infty(\mu_{r,t+1} | v_t) = \frac{1}{1 - a_{22}^2} :$$

6.2.2 Case B5

Case B5 nests all the previous ones. Apart from $|A| = 0$, we require $|tr(A)| < 1$ for the stability of the VAR. The joint process is now

$$\begin{pmatrix} r_t \\ \mu_{r,t+1} \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 0 & a_{11} + a_{22} \end{pmatrix} \begin{pmatrix} r_{t-1} \\ \mu_{r,t} \end{pmatrix} + \begin{pmatrix} 1 & 0 \\ a_{11} & a_{12} \end{pmatrix} \begin{pmatrix} u_t \\ v_t \end{pmatrix}$$

and the marginal processes

$$(1 - [a_{11} + a_{22}]L)\mu_{r,t+1} = a_{11}u_t + a_{12}v_t = \xi_t$$

and

$$(1 - [a_{11} + a_{22}]L)r_t = (1 - a_{22}L)u_t + a_{12}Lv_t = (1 - \theta L)\eta_t$$

The covariance equation system becomes

$$\frac{-\theta}{1 + \theta^2} = \frac{-a_{22}\sigma_u^2 + a_{12}\sigma_{uv}}{(1 + a_{22}^2)\sigma_u^2 + a_{12}^2\sigma_v^2 - 2a_{22}a_{12}\sigma_{uv}} \quad (8)$$

which if we set the scale parameter a_{12} to 1, and impose $a_{11} + a_{22} = \theta$ to achieve white noise behaviour for r_t yields

$$\frac{-(a_{11} + a_{22})}{1 + (a_{11} + a_{22})^2} = \frac{-a_{22}\gamma + \rho}{(1 + a_{22}^2)\gamma + 1/\gamma - 2a_{22}\rho}$$

where γ and ρ are defined as above. The relationship between R^2 and γ is now given by

$$R^2 = \frac{1 + \gamma^2 a_{11}^2}{1 + \gamma^2(1 - a_{22}^2 - 2a_{11}a_{22})}$$

For particular values of a_{11} and a_{22} we can get the mapping between R^2 and ρ consistent with white noise behaviour for r_t . Figure 3 shows such a mapping for four values of θ . For instance, in the first plot $\theta = 0.9$, and each curve corresponds to a value of a_{11} and $a_{22} = \theta - a_{11}$. Compared to case B2, now we can get univariate white noise behaviour for r_t and AR(1) behaviour for $\mu_{r,t}$ with zero or even positive correlation between the innovations to r_t and δ_t .

We derive the impulse response function for $\rho = 0.0$, which is compatible with white noise behaviour for r_t when $\theta > 0$ if $a_{11} < 0$. Here the covariance matrix of the innovations is diagonal, so we only get one Cholesky decomposition, namely

$$\Sigma^* = \begin{pmatrix} \gamma & 0 \\ 0 & 1 \end{pmatrix}$$

The impulse response functions with respect to u_t are

$$\text{IRF}_0(r_t) = 1; \text{IRF}_j(r_t) = a_{11}tr^{j-1}(A), \text{ for } j > 0.$$

$$\text{IRF}_j(\mu_{r,t+1}) = a_{11}tr^j(A), j = 0, \dots, \infty.$$

Again, the initial positive impact of u_t on r_t is slowly compensated by its negative impact on μ_t . Notice, however, that since the shock originates in r_t , $P_\infty(r_t | u_t) = 1 + P_\infty(\mu_{r,t} | u_t)$ as in the univariate case. In this particular example

$$P_\infty(r_t | u_t) = 1 + \frac{a_{11}^2}{1 - tr^2(A)} \quad P_\infty(\mu_{r,t+1} | u_t) = \frac{a_{11}^2}{1 - tr^2(A)} \quad .$$

Similarly, the impulse response functions with respect to v_t are

$$\text{IRF}_0(r_t) = 0; \text{IRF}_j(r_t) = tr(A)^{j-1}, \text{ for } j > 0.$$

$$\text{IRF}_j(\mu_{r,t+1}) = tr(A)^j, j = 0, \dots, \infty.$$

Therefore, here we have a situation in which the persistence of the shock will be the same on r_t and its conditional mean because $\rho = 0$. More specifically

$$P_{\infty}(r_t | v_t) = \frac{1}{1 - tr^2(A)} \quad P_{\infty}(\mu_{r,t+1} | v_t) = \frac{1}{1 - tr^2(A)}$$

7 Empirical Application to US Stock Market Returns

As we mentioned in the introduction, the fact that stock market returns have almost negligible autocorrelations was traditionally regarded as evidence in favour of the present value model with constant expected returns. More recently, though, several authors showed that near white noise behaviour for observed returns is compatible with a smoothly time-varying expected return whose first-order autocorrelation is high (see Campbell (1991) and the references therein). Obviously, from the point of view of explaining movements in asset prices, there is a substantial difference between constant and time-varying expected returns.

In order to throw some light on this issue, we apply the results of the previous section to post-war US monthly stock market returns. Since several studies have found some predictability in returns using lagged dividend yields, we estimate the following bivariate VAR(1)

$$\begin{pmatrix} r_t \\ \delta_t \end{pmatrix} = \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} + \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} r_{t-1} \\ \delta_{t-1} \end{pmatrix} + \begin{pmatrix} u_t \\ v_t \end{pmatrix}$$

where r_t is the (continuously compounded) real stock market return, and δ_t is the corresponding dividend-yield (see chapter 7 of Campbell, Lo and MacKinlay (1996) for details). The sample covers 516 monthly observations from January 1952 to December 1994.

Parameter estimates and heteroskedasticity-robust standard errors are presented in the first column of Table 1. As expected, the predictability of r_t is very small ($R^2 = .0226$). In contrast, dividend yields are highly predictable, especially on the basis of its own lagged values ($R^2 = 0.9961$).

These estimates imply that $tr(A)$ is 1.0695 and $|A| = 0.0952$, so that the roots of the characteristic equation for the second order autoregressive polynomial $(1 - tr(A)L + |A| L^2)$ are (.9714,.0980). We also have that the moving

average parameter for observed returns is .9916, while the standard deviation of ξ is .042. As a result, the implied theoretical first order autocorrelation equals 0.0837, which is very close to the sample value of the 0.0859.

As we saw in Section 6, it is impossible for the univariate representation of r_t to be exactly white-noise in a VAR(1) unless the companion matrix is reduced rank. For that reason, we also estimate by maximum likelihood a restricted VAR(1) model in which $a_{21} = a_{11}a_{22}/a_{12}$, or equivalently, $|A| = 0$. The results are presented in the second column of Table 1. Notice that the reduced rank restriction can only be rejected at the 5.92% level, despite the large number of observations.

Using the results in Section 6, it is then straightforward to obtain the joint process for actual and expected stock returns implied by the restricted parameter estimates, as well as their univariate representations. First of all, note that the correlation between innovations to returns and dividend yields (ρ_{uv}) is .0713. In contrast, the correlation between the bivariate innovations to observed and expected returns, ρ_{uw} , is -.9466. Therefore, it is not surprising that the implicit univariate representation of r_t obtained on the basis of the restricted parameter estimates is essentially white noise, with a negligible theoretical first autocorrelation (-.011). On the other hand, we find that the implicit univariate representation of expected returns is given by an AR(1) with coefficient .9755. However, the standard deviation of the univariate innovations to expected returns is 0.0010, which is 42 times smaller than the corresponding standard deviation for observed returns. Notice though, that the standard deviation of expected returns is only nine times smaller than the standard deviation of actual returns, because their autocorrelation coefficients are widely different.

The univariate representations, though, only give a partial picture, which is clearly insufficient for gauging the effect on r_t and its conditional mean of shocks to the bivariate process. In particular, we are interested in analyzing those shocks that affect r_t directly through u_t , and those that affect it indirectly through the innovation in $\mu_{r,t}$, w_t .

The impulse response functions are presented in Figure 4. Note that as in Section 6.2.1, the initial positive effect on returns of a shock to u_t is later reversed by the very slowly decaying negative effect on expected re-

turns. Similarly, a shock to expected returns has a large negative immediate impact on returns, and then it is compensated by the slowly diminishing positive effect on expected returns. However, the effects of shocks on expected returns are very small compared to the effect on actual returns. This is confirmed by our persistence measure. For the estimated parameter values, $P_\infty(r_t | u_t) = 1.0105$, while $P_\infty(\mu_{r,t} | u_t) = 0.0105$. Similarly, $P_\infty(r_t | w_t) = 1600.00$, while $P_\infty(\mu_{r,t} | w_t) = 20.64$. These results are in line with the argument in Campbell (1991) that a small unexpected variation in expected returns can have dramatic consequences on observed returns when the covariance between the innovations to actual and expected returns is large in absolute value but negative. Campbell (1991) provides an economic intuition for such a high negative correlation.

8 Conclusions

In this paper we study the time series properties of the processes for the (unobserved) conditional mean and variance, $\mu_t \sigma_t^2$, given a specification of the process for the observed time series. We first derive a general result for the conditional mean of univariate linear processes, and then apply it to various models of interest used in the analysis of economic and financial time series, such as stationary ARMA, ARIMA, multiplicative seasonal ARIMA and ARFIMA models, GARCH, GARCH-M and QARCH models.

We also look at the persistence of shocks to the conditional mean process, and compare it to the persistence of shocks to the observed variable. To do so, we use a measure of persistence of shocks for stationary processes which captures the importance of the deviations of a series from its unperturbed path following a single shock. Our measure is based on the impulse response function, and can be interpreted as the ratio of the variance of the series to the variance of the shock. We also propose a way of gauging the interim persistence of shocks that can be applied to non-stationary series as well.

The univariate framework, however, is often too restrictive, as there is only one shock that drives the processes for the observed variable and its conditional mean. For that reason we generalize our results to a multivariate setting. We start from an observed multivariate process for \mathbf{x}_t , and then derive the implicit

process for the conditional means. In this multivariate framework, we also look at the persistence of shocks. Finally, we consider the joint process for a subvector \mathbf{x}_{it} and its expected value conditional on the whole information set. In this respect, we derive necessary and sufficient conditions for one of the variables in a bivariate VAR(1) to have a white noise univariate representation while its conditional mean follows an AR(1) with a high autocorrelation coefficient.

We apply our results to US monthly real stock market returns and dividend yields over the period 1952-1994 to throw some light on the issue of whether white noise behaviour for returns is compatible with smooth, highly correlated time-varying expected returns. Our findings seem to confirm that stock returns are very close to white noise, while expected returns are well represented by an AR(1) process with a first-order autocorrelation of .9755. Furthermore, the standard deviation of the univariate innovations in the expected return series is over 42 times smaller than the corresponding standard deviation for the observed variables. Our results also indicate that innovations to observed and expected returns are negatively correlated, with a correlation coefficient of -.9466. As a result, a shock to expected returns has a large negative immediate impact on returns, which is thereafter compensated by a slowly diminishing positive effect on expected returns. However, the effects of shocks on expected returns are very small compared to their effect on actual returns. In this respect, our results confirm that a small unanticipated variation in expected returns can have dramatic consequences on observed returns.

Table 1: VAR(1) Estimation Results
US real stock returns and dividend yields
1952:1-1994:12

Par (White std. errors)	Unrestricted ($ A \neq 0$)	Restricted ($ A = 0$)
c_1	-.0188 (.0086)	-.0169 (.0086)
a_{11}	.0708 (.0438)	-.0231 (.0086)
a_{12}	.6455 (.2281)	.6074 (.2285)
c_2	2.12e-4 (1.09e-4)	2.14e-4 (1.09e-4)
a_{21}	-.0379 (5.58e-4)	-.0379
a_{22}	.9986 (.0029)	.9985 (.0029)
σ_u	.0419	.0421
σ_v	5.34e-4	5.34e-4
ρ_{uv}	.0717	0.0719

Wald test
 $H_0: |A| = 0$ $\chi^2_1 = 3.558$
p-value .0592

Appendices

A Proofs of Results

A.1 Proposition 1

Let's write $\mu_t = a(L)x_t - b(L)\epsilon_t$. Multiplying both sides by $1 - a(L)$ yields $[1 - a(L)]\mu_t = a(L)[1 - a(L)]x_t - b(L)[1 - a(L)]\epsilon_t = (a(L)[1 - b(L)] - b(L)[1 - a(L)])\epsilon_t$, which, after simplification, becomes $[1 - a(L)]\mu_t = [a(L) - b(L)]\epsilon_t$ \square

A.2 Proposition 2

In this case $\mu_t = A(L)\mathbf{x}_t - B(L)\epsilon_t$. Premultiplying both sides by $[I - A(L)]$ yields $[I - A(L)]\mu_t = [I - A(L)]A(L)\mathbf{x}_t - [I - A(L)]B(L)\epsilon_t$. Then, since $[I - A(L)]A(L) = A(L)[I - A(L)]$, it follows that $[I - A(L)]\mu_t = A(L)[I - B(L)]\epsilon_t - [I - A(L)]B(L)\epsilon_t = [A(L) - B(L)]\epsilon_t$ \square

A.3 Result 4

First notice that any covariance stationary GARCH(p,q) can be written in an ARCH(∞) form as

$$\sigma_t^2 = \alpha_0^* + \alpha^*(L)\epsilon_t^2$$

where

$$\alpha_0^* = \frac{\alpha_0}{[1 - \beta(1)]} \quad \text{and} \quad \alpha^*(L) = \alpha(L)[1 - \beta(L)]^{-1}$$

with

$$\alpha_i^* \geq 0 \quad \forall i, \quad \sum_{i=1}^{\infty} \alpha_i^* < 1$$

This implies that we can write a stationary AR(∞) for ϵ_t^2 ,

$$[1 - \alpha^*(L)]\epsilon_t^2 = \alpha_0^* + v_t$$

Since the infinite moving average representation of ϵ_t^2 is

$$\epsilon_t^2 = \alpha_0^{**} + \psi(L)v_t$$

with $\psi_0 = 1$ and $\psi_j = \sum_{i=1}^j \alpha_i^* \psi_{j-i}$, it is easy to verify that $\psi_j \geq 0$ for $j = 0, 1, \dots, \infty$, so that all the autocovariances of ϵ_t^2 will be non-negative.

Then, using the fact that σ_t^2 is a linear combination of the ϵ_t^2 with positive coefficients, we have that,

$$\text{Cov}(\sigma_t^2 \sigma_{t-k}^2) = \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} \alpha_i^* \alpha_j^* \text{Cov}(\epsilon_t^2 \epsilon_{t-[k+(i-j)]}^2) : .$$

which is non-negative for every k . □

B Rotation Matrices: Some Examples

Consider the 2×2 orthonormal matrix

$$\Omega(\omega) = \begin{pmatrix} \cos \omega & \sin \omega \\ -\sin \omega & \cos \omega \end{pmatrix}$$

First, it can be proved by straightforward multiplication that

$$\Omega'(\omega) = \begin{pmatrix} \cos \omega & -\sin \omega \\ \sin \omega & \cos \omega \end{pmatrix} = \Omega^{-1}(\omega)$$

as required. Second, note that

$$\Omega(-\omega) = \begin{pmatrix} \cos(-\omega) & -\sin(-\omega) \\ \sin(-\omega) & \cos(-\omega) \end{pmatrix} = \begin{pmatrix} \cos \omega & -\sin \omega \\ \sin \omega & \cos \omega \end{pmatrix} = \Omega'(\omega)$$

so that $\Omega(-\omega)$ produces the inverse effect that $\Omega(\omega)$. Also, note that

$$\Omega(\omega + \pi) = \begin{pmatrix} \cos(\omega + \pi) & \sin(\omega + \pi) \\ -\sin(\omega + \pi) & \cos(\omega + \pi) \end{pmatrix} = \begin{pmatrix} -\cos \omega & -\sin \omega \\ \sin \omega & -\cos \omega \end{pmatrix} = -\Omega(\omega)$$

so that, if we do not care about the sign of the rotation, we can concentrate on the interval $0 \leq \omega < \pi$. Analogously, it can be easily verified that $\Omega(\omega + 2\pi) = \Omega(\omega)$ and $\Omega(\omega - \pi) = -\Omega(-\omega) = -\Omega'(\omega)$. Therefore, $\Omega(\omega)$ and $\Omega(\omega - \pi)$ cancel each other (up to sign).

Finally note that all these properties come from

$$\Omega(\omega_1)\Omega(\omega_2) = \begin{pmatrix} \cos \omega_1 & \sin \omega_1 \\ -\sin \omega_1 & \cos \omega_1 \end{pmatrix} \times \begin{pmatrix} \cos \omega_2 & \sin \omega_2 \\ -\sin \omega_2 & \cos \omega_2 \end{pmatrix} =$$

$$\begin{pmatrix} \cos(\omega_1 + \omega_2) & \sin(\omega_1 + \omega_2) \\ -\sin(\omega_1 + \omega_2) & \cos(\omega_1 + \omega_2) \end{pmatrix} = \Omega(\omega_1 + \omega_2)$$

together with

$$\Omega(0) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad \text{and} \quad \Omega(\pi) = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}$$

Some examples of $\Omega(\omega)$ are

$$\Omega(\pi/4) = \begin{pmatrix} \sqrt{1/2} & \sqrt{1/2} \\ -\sqrt{1/2} & \sqrt{1/2} \end{pmatrix} = \Omega(3\pi/4)$$

and

$$\Omega(\pi/2) = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$$

C Cholesky Decompositions and Orthonormal Rotations

Let's now see what effect a rotation has on the orthogonalizations of a covariance matrix Σ , where

$$\Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{pmatrix}$$

Let's start from the first Cholesky decomposition of Σ given by

$$\Sigma = \begin{pmatrix} \sqrt{\sigma_{11}} & 0 \\ \sigma_{12}/\sqrt{\sigma_{11}} & \sqrt{\sigma_{22} - \frac{\sigma_{12}^2}{\sigma_{11}}} \end{pmatrix} \begin{pmatrix} \sqrt{\sigma_{11}} & \sigma_{12}/\sqrt{\sigma_{11}} \\ 0 & \sqrt{\sigma_{22} - \frac{\sigma_{12}^2}{\sigma_{11}}} \end{pmatrix} = \Sigma_L \Sigma'_L$$

In this case we are assigning relatively more variance to the first shock than to the second one. In contrast, in the other Cholesky decomposition, $\Sigma = \Sigma_U \Sigma'_U$ the opposite happens.

Let's find $\Omega(\omega)$ such that $\Sigma = \Sigma_L \Omega(\omega) \Omega'(\omega) \Sigma'_L = \Sigma_U \Sigma'_U$. Since we require $\Sigma_L \Omega(\omega)$ to be upper triangular

$$\sigma_{12}/\sqrt{\sigma_{11}} \cos \omega - \sqrt{\sigma_{22} - \frac{\sigma_{12}^2}{\sigma_{11}}} \sin \omega = 0,$$

which gives

$$\omega = \arctan \left[\frac{\sigma_{12}/\sqrt{\sigma_{11}}}{\sqrt{\sigma_{22} - \frac{\sigma_{12}^2}{\sigma_{11}}}} \right] = \arctan \left[\frac{\sigma_{12}}{\sqrt{|\Sigma|}} \right] = \arcsin \rho$$

where ρ is the correlation coefficient. In this context, rotations in the range $(0, \arcsin \rho)$ transfer variance from the first shock to the second. However, the Cholesky decompositions are by no means the limits of the variance that can be assigned to each shock, unless $\rho = \pm 1$. In general, the variance attributed to the second shock can be made even greater for some $\omega \in (\rho, \pi)$.

As an extreme example, consider the case in which $\rho = 0$ so that the two Cholesky decompositions coincide. By letting $\omega = \pi/2$, we can assign all the variance of the first (second) variable to the second (first) shock. This may lead to significant differences in interpretation of apparently straightforward processes such as

$$\begin{pmatrix} x_{1,t} \\ x_{2,t} \end{pmatrix} = \begin{pmatrix} a_{11} & 0 \\ 0 & a_{22} \end{pmatrix} \begin{pmatrix} x_{1,t-1} \\ x_{2,t-1} \end{pmatrix} + \begin{pmatrix} u_{1,t} \\ u_{2,t} \end{pmatrix}$$

with $\sigma_{12} = 0$.

Figure 1: **Impulse Response Functions**

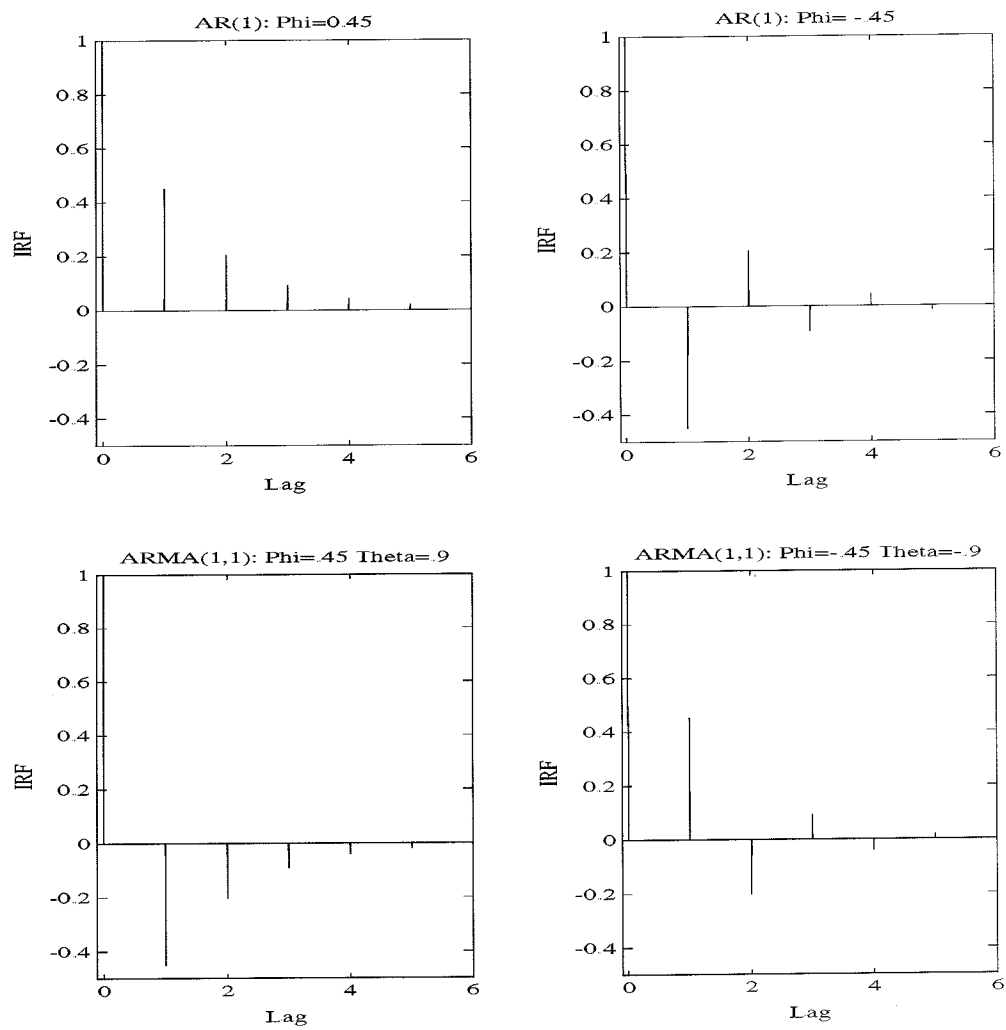


Figure 2: **Case B2**

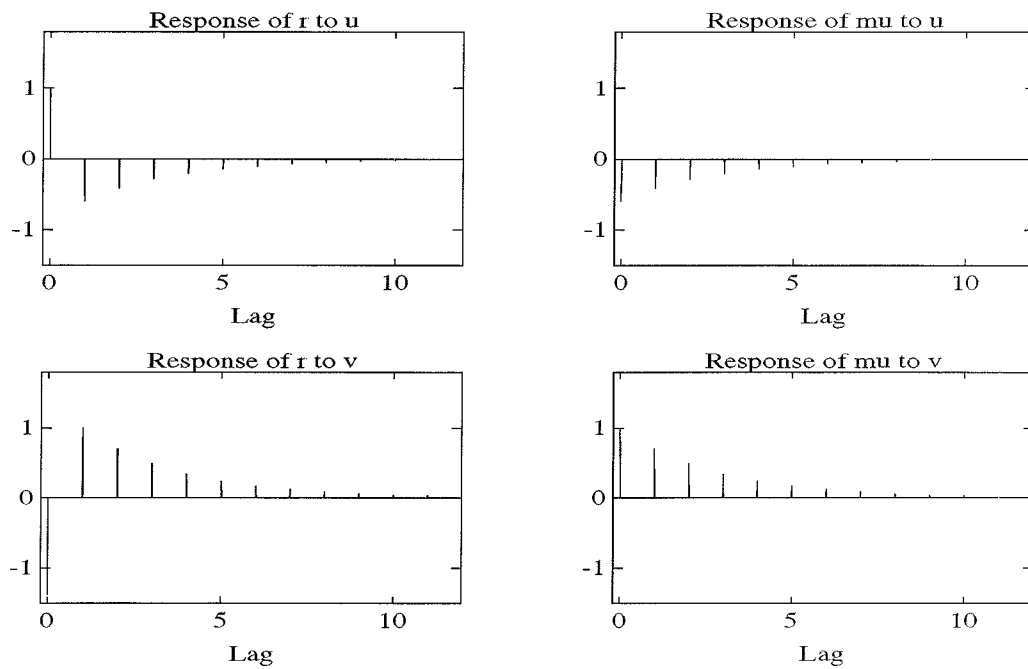
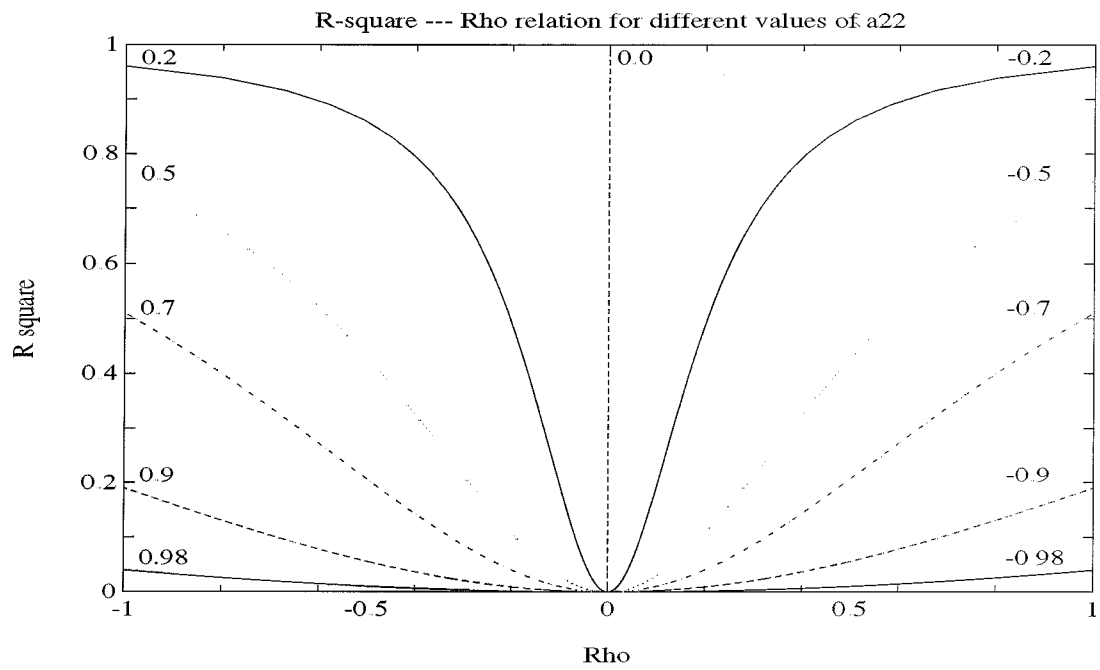


Figure 3: **Case B5**

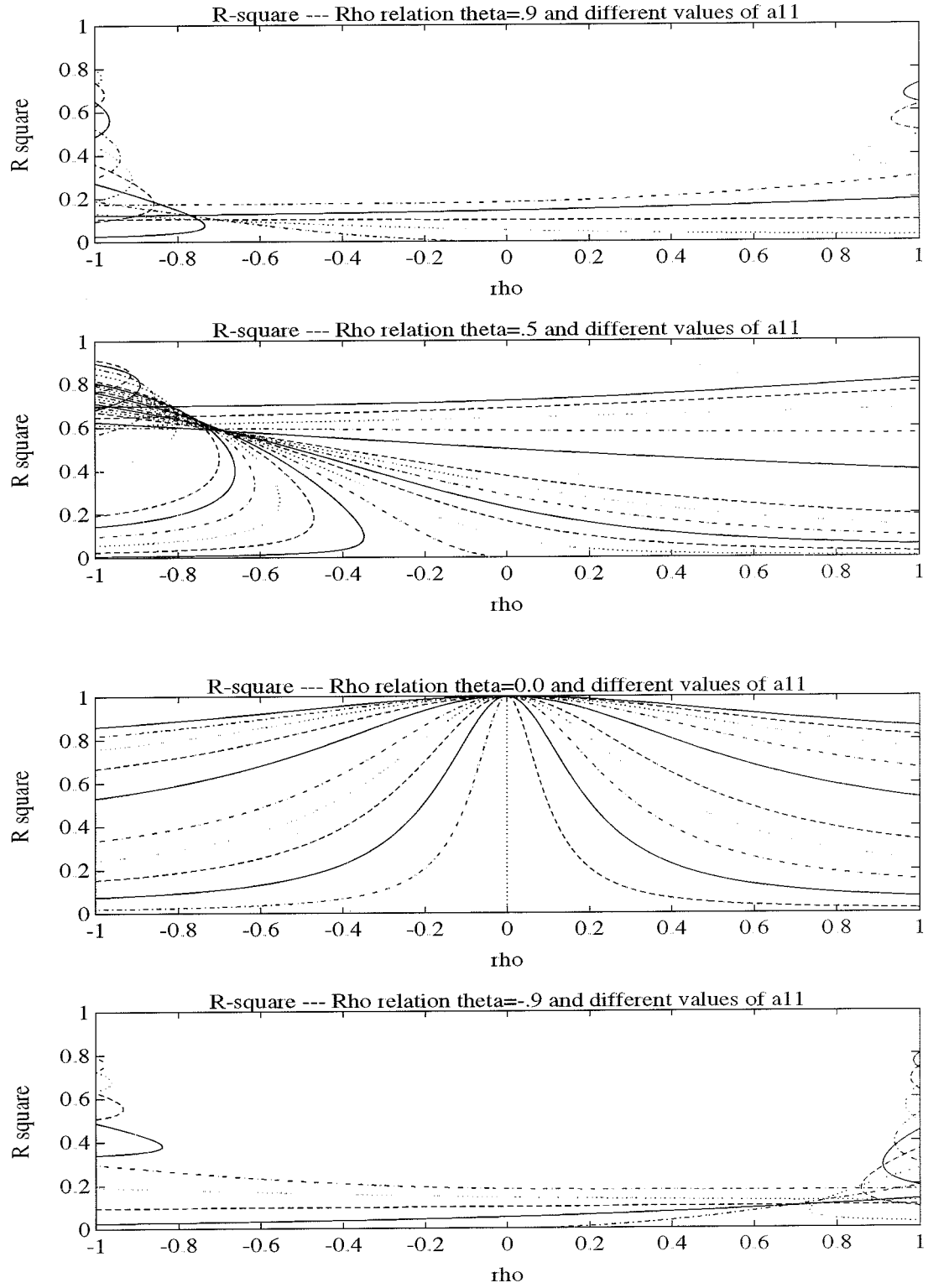
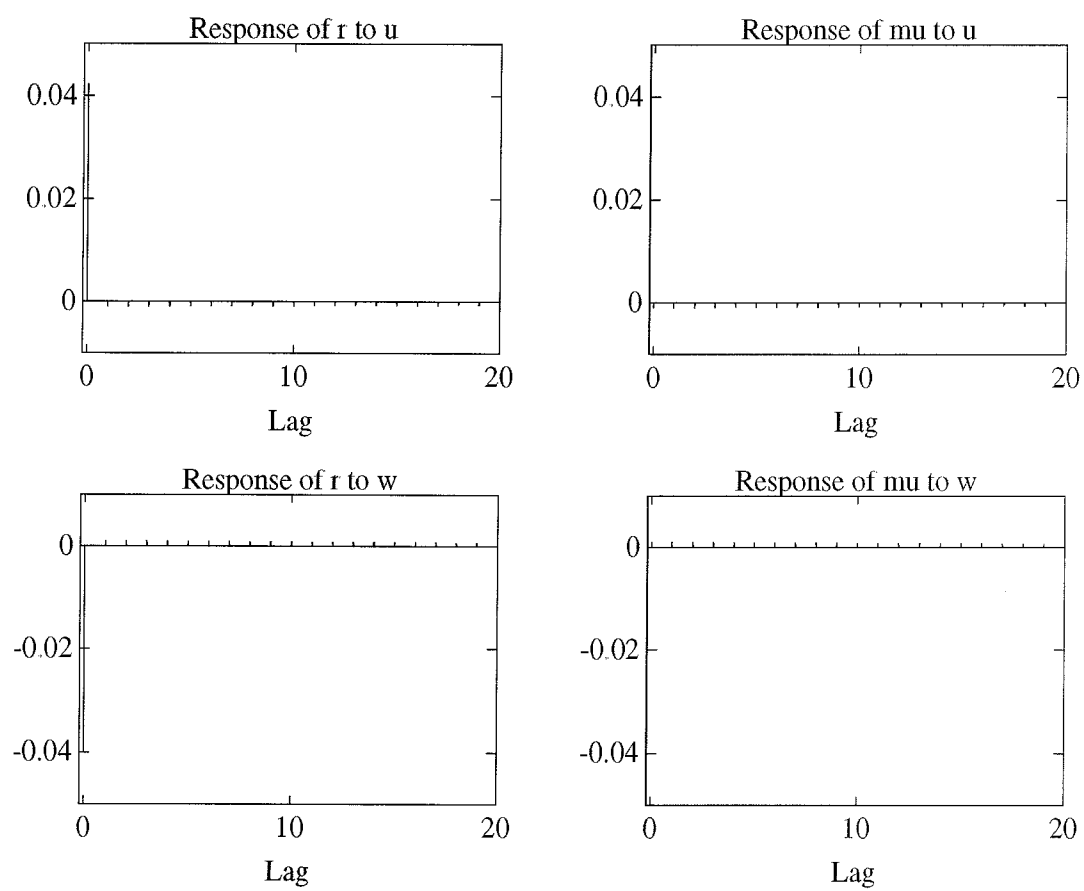


Figure 4: **IRF** for returns and expected returns



References

- AKAIKE, H. (1974): "Markovian Representation of Stochastic Processes and its Application to the Analysis of Autoregressive Moving Average Processes". *Annals of the Institute of Statistical Mathematics*, 26, 363-387.
- CAMPBELL, J.Y. (1990): "Measuring the Persistence of Expected Returns". *American Economic Review Papers and Proceedings*, May 1990.
- CAMPBELL, J.Y. (1991): "A Variance Decomposition for Stock Returns". *Economic Journal*, 101, 157-179.
- CAMPBELL, J.Y., A.W. LO AND A.C. MACKINLAY (1996): *The Econometrics of Financial Markets*, Princeton University Press, forthcoming.
- CAMPBELL, J.Y. AND N.G. MANKIW (1987): "Are Output Fluctuations Transitory ?". *Quarterly Journal of Economics*, 102, 875-880.
- DIEBOLD, F.X. AND G.D. RUDEBUSCH (1989): "Long Memory and Persistence in Aggregate Output". *Journal of Monetary Economics*, 24, 189-209.
- ENGLE, R.F. AND C. MUSTAFA (1992): "Implied ARCH Models from Option Prices". *Journal of Econometrics*, 52, 289-311.
- FAMA, E.F. AND K.R. FRENCH (1988): "Permanent and Temporary Components of Stock Prices". *Journal of Political Economy*, 96, 246-273.
- FIorentini, G. AND A. MARAVALL (1996): "Unobserved Components in ARCH Models: An Application to Seasonal Adjustment". *Journal of Forecasting*, 15, 175-201.
- GRANGER, C.W.J. (1983): "Forecasting White Noise". in A. ZELLNER, ed. *Applied Time Series Analysis of Economic Data*. US Department of Commerce, Bureau of the Census, Washington.
- GRANGER, C.W.J. AND R. JOYEUX (1980): "An Introduction to Long Memory Time Series and Fractional Differencing". *Journal of Time Series Analysis*, 1, 15-39.
- HONG, E.P. (1991): "The autocorrelation structure for the GARCH-M process". *Economics Letters*, 37, 129-132.
- HOSKING, J.R.M. (1981): "Fractional Differencing". *Biometrika*, 68, 165-176.
- KOOP G., M.H. PESARAN AND S.M. POTTER (1996): "Impulse Response Analysis in Nonlinear Multivariate Models". *Journal of Econometrics*, 74, 119-147.

POTERBA, J. AND SUMMERS L.H. (1988): "Mean Reversion in Stock Returns: Evidence and Implications". *Journal of Financial Economics*, 22, 27-60.

SENTANA, E. (1995): "Quadratic ARCH Models". *Review of Economic Studies*, 62, 639-661.

SHILLER, R.J. (1984): "Stock Prices and Social Dynamics". *Brookings Papers on Economic Activity*, 2, 457-498.

SUMMERS, L.H. (1986): "Does the Stock Market Rationality Reflect Fundamental Values ?". *Journal of Finance*, 41, 591-601.

PUBLISHED ISSUES*

- WP-AD 95-01 "Cost Monotonic Mechanisms"
M. Ginés, F. Marhuenda. March 1995.
- WP-AD 95-02 "Implementation of the Walrasian Correspondence by Market Games"
L. Corchón, S. Wilkie. March 1995.
- WP-AD 95-03 "Terms-of-Trade and the Current Account: A Two-Country/Two-Sector Growth Model"
M.D. Guilló. March 1995.
- WP-AD 95-04 "Exchange-Proofness or Divorce-Proofness? Stability in One-Sided Matching Markets"
J. Alcalde. March 1995.
- WP-AD 95-05 "Implementation of Stable Solutions to Marriage Problems"
J. Alcalde. March 1995.
- WP-AD 95-06 "Capabilities and Utilities"
C. Herrero. March 1995.
- WP-AD 95-07 "Rational Choice on Nonfinite Sets by Means of Expansion-Contraction Axioms"
M.C. Sánchez. March 1995.
- WP-AD 95-08 "Veto in Fixed Agenda Social Choice Correspondences"
M.C. Sánchez, J.E. Peris. March 1995.
- WP-AD 95-09 "Temporary Equilibrium Dynamics with Bayesian Learning"
S. Chatterji. March 1995.
- WP-AD 95-10 "Existence of Maximal Elements in a Binary Relation Relaxing the Convexity Condition"
J.V. Llinares. May 1995.
- WP-AD 95-11 "Three Kinds of Utility Functions from the Measure Concept"
J.E. Peris, B. Subiza. May 1995.
- WP-AD 95-12 "Classical Equilibrium with Increasing Returns"
A. Villar. May 1995.
- WP-AD 95-13 "Bargaining with Claims in Economic Environments"
C. Herrero. May 1995.
- WP-AD 95-14 "The Theory of Implementation when the Planner is a Player"
S. Baliga, L. Corchón, T. Sjöström. May 1995.
- WP-AD 95-15 "Popular Support for Progressive Taxation"
F. Marhuenda, I. Ortuño. May 1995.
- WP-AD 95-16 "Expanded Version of Regret Theory: Experimental Test"
R. Sirvent, J. Tomás. July 1995.
- WP-AD 95-17 "Unified Treatment of the Problem of Existence of Maximal Elements in Binary Relations. A Characterization"
J.V. Llinares. July 1995.

* Please contact IVIE's Publications Department to obtain a list of publications previous to 1995.

- WP-AD 95-18 "A Note on Stability of Best Reply and Gradient Systems with Applications to Imperfectly Competitive Models"
L.C. Corchón, A. Mas-Colell. July 1995.
- WP-AD 95-19 "Redistribution and Individual Characteristics"
I. Iturbe-Ormaetxe. September 1995.
- WP-AD 95-20 "A Mechanism for Meta-Bargaining Problems"
M^a. C. Marco, J. E. Peris, B. Subiza. September 1995.
- WP-AD 95-21 "Signalling Games and Incentive Dominance"
G. Olcina, A. Urbano. September 1995.
- WP-AD 95-22 "Multiple Adverse Selection"
J.M. López-Cuñat. December 1995.
- WP-AD 95-23 "Ranking Social Decisions without Individual Preferences on the Basis of Opportunities"
C. Herrero, I. Iturbe, J. Nieto. December 1995.
- WP-AD 95-24 "The Extended Claim-Egalitarian Solution across Cardinalities"
M^a C. Marco. December 1995.
- WP-AD 95-25 "A Decent Proposal"
L. Corchón, C. Herrero. December 1995.
- WP-AD 96-01 "A Spatial Model of Political Competition and Proportional Representation"
I. Ortuño. February 1996.
- WP-AD 96-02 "Temporary Equilibrium with Learning: The Stability of Random Walk Beliefs"
S. Chatterji. February 1996.
- WP-AD 96-03 "Marketing Cooperation for Differentiated Products"
M. Peitz. February 1996.
- WP-AD 96-04 "Individual Rights and Collective Responsibility: The Rights-Egalitarian Solution"
C.Herrero, M. Maschler, A. Villar. April 1996.
- WP-AD 96-05 "The Evolution of Walrasian Behavior"
F. Vega-Redondo. April 1996.
- WP-AD 96-06 "Evolving Aspirations and Cooperation"
F. Vega-Redondo. April 1996.
- WP-AD 96-07 "A Model of Multiproduct Price Competition"
Y. Tauman, A. Urbano, J. Watanabe. July 1996.
- WP-AD 96-08 "Numerical Representation for Lower Quasi-Continuous Preferences"
J. E. Peris, B. Subiza. July 1996.
- WP-AD 96-09 "Rationality of Bargaining Solutions"
M. C. Sánchez. July 1996.
- WP-AD 96-10 "The Uniform Rule in Economies with Single Peaked Preferences, Endowments and Population-Monotonicity"
B. Moreno. July 1996.
- WP-AD 96-11 "Modelling Conditional Heteroskedasticity: Application to Stock Return Index "IBEX-35""
A. León, J. Mora. July 1996.

- WP-AD 96-12 "Efficiency, Monotonicity and Rationality in Public Goods Economies"
M. Ginés, F. Marhuenda. July 1996.
- WP-AD 96-13 "Simple Mechanism to Implement the Core of College Admissions Problems"
J. Alcalde, A. Romero-Medina. September 1996.
- WP-AD 96-14 "Agenda Independence in Allocation Problems with Single-Peaked Preferences"
C. Herrero, A. Villar. September 1996.
- WP-AD 96-15 "Mergers for Market Power in a Cournot Setting and Merger Guidelines"
R. Faulí. September 1996.
- WP-AD 96-16 "Consistent Beliefs, Learning and Different Equilibria in Oligopolistic Markets"
G. Fernández de Córdoba. October 1996.
- WP-AD 96-17 "Pigouvian Taxes: A Strategic Approach"
J. Alcalde, L. Corchón, B. Moreno. October 1996.
- WP-AD 96-18 "Differentiated Bertrand Duopoly with Variable Demand"
M. Peitz. October 1996.
- WP-AD 96-19 "Stability of Tâtonnement Processes of Short Period Equilibria with Rational Expectations"
T. Hens. December 1996.
- WP-AD 96-20 "Convergence of Aspirations and (Partial) Cooperation in the Prisoner's Dilemma"
F. Vega, F. Palomino. December 1996.
- WP-AD 96-21 "Testing Non-Nested Semiparametrics Models: An Application to Engel Curves Specification"
M. Delgado, J. Mora. December 1996.
- WP-AD 96-22 "Managerial Incentives for Takeovers"
R. Faulí, M. Motto. December 1996.
- WP-AD 96-23 "Migration and the Evolution of Conventions"
F. Vega-Redondo, V. Bhaskar. December 1996.
- WP-AD 96-24 "Political Compromise and Endogenous Formation of Coalitions"
I. Ortuño-Ortín, A. Gerber. December 1996.
- WP-AD 97-01 "Land Reform and Individual Property Rights"
A. Díaz. February 1997.
- WP-AD 97-02 "Models à la Lancaster and à la Hotelling when they are the Same"
M. Peitz. February 1997.
- WP-AD 97-03 "On Merger Profitability in a Cournot Setting"
R. Faulí. February 1997.
- WP-AD 97-04 "Core Selections in Economies with Increasing Returns and Public Goods"
M. Ginés. February 1997.
- WP-AD 97-05 "Condorcet Choice Correspondences for Weak Tournaments"
B. Subiza, J. E. Peris. February 1997.
- WP-AD 97-06 "Equal-Loss Solution for Monotonic Coalitional Games"
M^a. C. Marco, B. Subiza. February 1997.

- WP-AD 97-07 "Income Taxation, Uncertainty and Stability"
F. Marhuenda, I. Ortuño-Ortín. February 1997.
- WP-AD 97-08 "A Non-Cooperative Approach to Meta-Bargaining Theory"
E. Naeve-Steinweg. February 1997.
- WP-AD 97-09 "Bargaining, Reputation and Strikes"
V. Calabuig, G. Olcina. February 1997.
- WP-AD 97-10 "Hiring Procedures to Implement Stable Allocations"
J. Alcalde, D. Pérez-Castrillo, A. Romero-Medina. April 1997.
- WP-AD 97-11 "Global Stability in Spite of "Local Instability" with Learning in General Equilibrium Models"
S. Chatterji, S. Chattopadhyay. April 1997.
- WP-AD 97-12 "A Three Factor Agricultural Production Function: The Case of Canada"
C. Echevarría. April 1997.
- WP-AD 97-13 "A Tobit Model with Garch Errors"
G. Calzolari, G. Fiorentini. April 1997.
- WP-AD 97-14 "Nonhomothetic Preferences, Growth, Trade and Land"
C. Echevarría. April 1997.
- WP-AD 97-15 "The Differentiation Triangle"
M. Peitz, M. Canoy. April 1997.
- WP-AD 97-16 "Forward Induction in a Wage Repeated Negotiation"
V. Calabuig, G. Olcina. June 1997.
- WP-AD 97-17 "Conditional Means of Time Series Processes and Time Series Processes for Conditional Means"
G. Fiorentini, E. Sentana. June 1997.
- WP-AD 97-18 "Adverse Selection Under Complete Ignorance"
J.M. López-Cuñat. June 1997.