Conditional Means of Time Series Processes and Time Series Processes for Conditional Means*

Gabriele Fiorentini (University of Alicante)

Enrique Sentana (CEMFI)

April 1996
Current Version: 20 January 1998

Abstract

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.

*We have benefitted from discussions with M. Arellano, M. Balmaseda, I. Domowitz, F. Diebold, L. Reichlin, G. Tauchen and A. Timmermann. We would also like to thank J. Campbell for kindly allowing us to use his data, as well as an anonymous referee and seminar audiences at CEMFI and UPV (Bilbao) for very useful comments and suggestions. Financial support from IVIE is gratefully acknowledged. Correspondence should be addressed to E. Sentana, CEMFI, Casado del Alisal 5, 28014 Madrid, Spain (e-mail: sentana@cemfi.es).

1 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} we denote by $\mu_t = E(\mathrm{r}_t \mid I_{t-1})$ and $\sigma_t^2 = V(\mathrm{r}_t \mid I_{t-1})$. 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.

A univariate framework, though, is 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 particular, we obtain a general result for the conditional mean of multivariate linear processes satisfying standard regularity conditions.[1] Then, we apply this result to various models of interest used in the analysis of economic and financial time series, such as univariate (seasonal) ARIMA and ARFIMA models. We also apply our result to conditional second moments by using the fact that conditional heteroskedasticity models often have a straightforward interpretation as linear processes for the squared innovations. We present examples for multivariate GARCH, and univariate GARCH-M 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.

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 expected returns, as well as their univariate representations.

The rest of the paper is organized as follows. In Section 2 we present our basic result. A measure of persistence for stationary processes is introduced in Section 3. In Section 4 we derive conditions under which white-noise behaviour for a variable is compatible with a serially correlated stochastic process for its conditional mean. The results of the empirical application are discussed in Section 5. Finally, our conclusions are presented in Section 6.

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

2 The Conditional Mean of a Vector Process

A linear stochastic process of orders k and h for the $n \times 1$ vector $\mathbf{x}_t$ can be written as

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

where $\boldsymbol{\epsilon}_t$ is a $n \times 1$ white noise process of one-period ahead prediction errors, 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 of order k in the lag (or backshift) operator L, and $B(L)$ is analogously defined, with the roots of $|I - A(L)| = 0$ and $|I - B(L)| = 0$ on or outside the unite circle. This includes (co-)integrated and invertible processes (whether strictly or not) but rules out explosive as well as non-invertible processes. Define $\boldsymbol{\mu}_t = E_{t-1}(\mathbf{x}_t)$ as the $n \times 1$ conditional mean vector, so that $\mathbf{x}_t = \boldsymbol{\mu}_t + \boldsymbol{\epsilon}_t$. Then

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

Notice that k and $m - 1$ should be interpreted as maximum orders because cancellation often occurs thorough common roots. Nevertheless, the common factors will never change the order of (co-)integration. Note also that the innovation in the process for $\boldsymbol{\mu}_{t+1}$ is proportional to the innovation in $\mathbf{x}_t$. In the rest of this section we shall apply the above result to several models of practical interest.

2.1 Univariate ARIMA-type Processes

An Autoregressive Integrated Moving Average ARIMA(p,d,q) process can be represented as

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

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,

Corollary 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 $\Phi_1 \neq \theta_1$.*

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 $\mu_{t+1} = \phi\mu_t + (\phi-\theta)\epsilon_t$, i.e. an AR(1). If we let $\phi - \theta$ go to zero, (having chosen the initial conditions from the stationary distribution), we can make the ARMA(1,1) process as close as desired to white noise. Correspondingly, the variance of the mean goes to zero with $\phi-\theta$, so that it actually converges to a constant in the limit.

Proposition 1 is also readily applicable to Multiplicative seasonal ARIMA models. In this case, the process for the conditional mean has an expression which, in general, 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. 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,

$$(1-L)^\gamma\mu_{t+1} = [\gamma + \frac{1}{2}\gamma(1-\gamma)L + \frac{1}{6}\gamma(1-\gamma)(2-\gamma)L^2 + \ldots]\epsilon_t$$

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

2.2 Vector GARCH Processes

A vector of innovations, $\boldsymbol{\epsilon}_t$, in a multivariate stochastic process is said to follow a multivariate (semi-strong) GARCH(p,q) process if $E_{t-1}(\boldsymbol{\epsilon}_t) = 0$ and $E_{t-1}(\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t') = \Sigma_t$, with

$$[I - \beta(L)]vech\Sigma_t = \alpha_0 + \alpha(L)vech(\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t')$$

It is well known that a multivariate GARCH(p,q) process with bounded fourth moments can be represented as the following VARMA(m,p) on $vech(\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t')$ (with $m = max[p, q]$),

$$[I - \alpha(L) - \beta(L)]vech(\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t') = \alpha_0 + [I - \beta(L)]\mathbf{v}_t$$

where $\mathbf{v}_t = vech(\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t' - \Sigma_t)$. In this case, Proposition 1 simplifies to[2]

Corollary 2 *A multivariate* GARCH *(p,q) process for* $\boldsymbol{\epsilon}_t$ *implies that* $vech(\Sigma_t)$ *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* $\mid \alpha_1 \mid \neq 0$*.*

In most empirical applications, the simple GARCH(1,1) specification is adopted, so that Corollary 2 gives

$$vech(\Sigma_{t+1}) = \alpha_0 + (\alpha_1 + \beta_1)vech(\Sigma_t) + \alpha_1\mathbf{v}_t$$

that is, a VAR(1) for the conditional variance, with companion matrix equal to $\alpha_1 + \beta_1$ and variance of innovations $\alpha_1 V(\mathbf{v}_t)\alpha_1'$.

2.3 Univariate GARCH-M Processes

A variable x_t is said to follow a GARCH in mean (GARCH-M) process of orders p and q, if

$$x_t = \delta\sigma_t^2 + \epsilon_t$$

where σ_t^2 is the conditional variance of ϵ_t, which in turn, follows a GARCH(p,q) process. The GARCH-M model is an example in which the difference between conditional means and linear projections is important. Notice that the conditional

[2] An analogous result can be found in Fiorentini and Maravall (1996) for the univariate case.

mean of x_t is proportional to its conditional variance. Hence, according to Corollary 2, 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. Also, provided that $E(\epsilon_t^3) = 0$, 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≥q, we have that:[3]

Lemma 1 *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 $\alpha_i + \beta_i$.*

We can derive, in fact, further results about the autocorrelation structure of x_t. The following lemma will prove useful:

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

Then, Lemmas 1 and 2 imply that:

Lemma 3 *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.

3 Persistence of Shocks in Covariance Stationary Time Series

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

[3] Hong (1989) investigates the ACF of x_t in the GARCH-M(1,1) case.

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 covariance stationary processes.

For clarity of exposition, we shall begin with a univariate time series. Let $x_t = \Psi(L)\epsilon_t$ denote the Wold representation of the unperturbed process. Let's now define the perturbed process $x^\star_t = \Psi(L)\epsilon^\star_t$, where $\epsilon^\star_s = \epsilon_s$ ($\forall s \neq t$), and $\epsilon^\star_t = \epsilon_t + 1 \times \sigma_\epsilon$. We want a measure of how much $x^\star_t$ deviates from x_t. Obviously, since the process is stationary, the net effect on $x^\star_{t+k}$ of a shock to ϵ_t is zero in the limit. However, the route taken by $x^\star_{t+k}$ to go back to its original path x_{t+k} may differ substantially across different models. For instance, if x_t follows an AR(1) process ($x_t = \phi x_{t-1} + \epsilon_t$) with $\phi = .95$, $x^\star_{t+k}$ 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 $x^\star_{t+k}$ 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 $x^\star_{t+k}$ but the series will take a long time to go back to its original level.

Although persistence refers to time, since $x^\star_{t+k} - x_{t+k} = \psi_k \sigma_\epsilon$, 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 $x_t = .45x_{t-1} + \epsilon_t - .9\epsilon_{t-1}$ or $x_t = -.45x_{t-1} + \epsilon_t + .9\epsilon_{t-1}$, even though their impulse response functions are all identical in magnitude. For that reason, we propose the use of

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

as a measure of the persistence of shocks. In principle, $\sum_{j=0}^{\infty} \mid \psi_j \mid$ 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

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

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 as a measure of the interim persistence of shocks

$$P_k(x_t \mid \epsilon_t) = \sum_{j=0}^{k} \psi_j^2 = \frac{V(x_{t+k} - \hat{x}_{t+k|t-1})}{V(\epsilon_t)}$$

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 $P_k(x_t \mid \epsilon_t)$ to $P_\infty(x_t \mid \epsilon_t)$. But unlike $P_\infty(x_t \mid \epsilon_t)$, the k-period measure $P_k(x_t \mid \epsilon_t)$ 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

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

In particular, for AR(1) models, $P_\infty(x_t \mid \epsilon_t)$ is a monotonic transformation of the absolute value of ϕ. In this respect, therefore, it is similar to the mean lag, $\phi/(1-\phi)$, which, however, is only well defined when $\phi \geq 0$. 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 $P_{400}(x_t \mid \epsilon_t) = 1.16$, well below the persistence of a stationary AR(1) with autoregressive parameter equal to .5.

We are now in a position to compare the persistence of shocks in the conditional mean μ_{t+1} vis a vis the persistence of shocks in the observed variable, x_t. Since

μ_{t+1} can be expressed as $\mu_{t+1} = \psi_1 \epsilon_t + \psi_2 \epsilon_{t-1} + \ldots$, 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, this is no longer necessarily so in a multivariate framework.

As our last example, consider a GARCH(1,1) process, which corresponds to an ARMA(1,1) process for ϵ_t^2. 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.

The same notion of persistence of shocks 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.

Let $\mathbf{x}_t = \Psi(L)\boldsymbol{\epsilon}_t$ denote the Wold representation of the vector process $\mathbf{x}_t$, and define a matrix $\Sigma^\star$ such that $\Sigma^\star \Sigma^{\star\prime} = \Sigma$. Then, the infinite moving average representation of $\mathbf{x}_t$ in terms of the standardized orthogonal innovations $\boldsymbol{\epsilon}_t^\star = \Sigma^{\star-1}\boldsymbol{\epsilon}_t$ is $\mathbf{x}_t = \Psi^\star(L)\boldsymbol{\epsilon}_t^\star$, where $\Psi_i^\star = \Psi_i \Sigma^\star$ and the covariance matrix of $\boldsymbol{\epsilon}_t^\star$ is the identity matrix.

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

$$P_{\infty}(x_{j,t} \mid \epsilon^{\star}_{i,t}) = \sum_{k=0}^{\infty} \psi^{\star 2}_{ji,k}$$

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 - be an orthonormal basis of $\mathbb{R}^n$. Then, any orthonormal transformation of $\Sigma^{\star}$ will provide another infinite MA representation with orthogonal shocks. In particular, $\mathbf{x}_t = \Psi^{\star\star}(L)\boldsymbol{\epsilon}^{\star\star}_t$ where $\Psi^{\star\star}_i = \Psi_i \Sigma^{\star}$- .[4] The persistence of the "new" i^{th} shock on $x_{j,t}$ will be

$$P_{\infty}(x_{j,t} \mid \epsilon^{\star\star}_{i,t}) = \sum_{k=0}^{\infty} \psi^{\star\star 2}_{ji,k}$$

which in general is different from $P_{\infty}(x_{j,t} \mid \epsilon^{\star}_{i,t})$. 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.

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

In this section, we shall obtain the joint process for an element of $\mathbf{x}_t$ and its conditional mean in a multivariate setting. Importantly, the mean is conditional on the full information set I_t, which means that the innovations to the joint process 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 $\mu_{r,t+1}$.[5] In this case,

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

[5] Notice that the temporal phase-shift between the two variables is only apparent as both r_t and $\mu_{r,t+1}$ belong to the information set I_t.

the joint process 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

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

$$\begin{pmatrix} u_t \\ w_t \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ -|A| & 0 \end{pmatrix} \begin{pmatrix} u_{t-1} \\ w_{t-1} \end{pmatrix}$$

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 correlation is

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

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. To keep the algebra as simple as possible, we only consider in detail those special cases that lead to an AR(1) process for $\mu_{r,t+1}$.

4.1 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. 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. It is then impossible to achieve a white noise representation for a series with time-varying conditional mean in the context of linear models (see Granger, 1983).

4.2 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 $(1 - tr(A)L + |A|\,L^2) = 0$ 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.[6]

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 $a_{21} = 0$

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

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. In case B4, the variable δ_t in the original VAR(1) is white noise. This makes this case empirically uninteresting. Case B2 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.

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

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

and

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

The quadratic equation for θ becomes

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

which if we set the scale parameter a_{12} to 1 without loss of generality, and impose $a_{11} + a_{22} = \theta$ to achieve white noise behaviour for r_t yields

$$\frac{-(a_{11} + a_{22})}{1 + (a_{11} + a_{22})^2} = \frac{-a_{22}\gamma_{uv} + \rho_{uv}}{(1 + a_{22}^2)\gamma_{uv} + 1/\gamma_{uv} - 2a_{22}\rho_{uv}}$$

where $\gamma_{uv} = \sigma_u/\sigma_v$ and $\rho_{uv} = \sigma_{uv}/(\sigma_u\sigma_v)$. It is more interesting, though, to look at the implications of $\theta = a_{11} + a_{22}$ 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 $(var(\mu_{r,t})/var(r_t))$, and the correlation between the innovations. The relationship between R^2 and γ_{uv} is given by

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

For particular values of a_{11} and a_{22} we can get the mapping between R^2 and ρ_{uv} consistent with white noise behaviour for r_t. Figure 1 shows such a mapping for several parameter configurations. For instance, in the first plot $a_{11} = 0$ while a_{22} ranges from .98 to -.98. As shown by Campbell (1991), in this case 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. As the other plots show, though, in general 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.

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 consider two kinds of shocks: those that affect r_t directly through u_t, and those that affect $\mu_{r,t+1}$ directly through w_t. To study the response to a shock in u_t we use the Cholesky decomposition of $V(u_t, w_t)$, i.e.

$$\Sigma^\star = \sigma_w \begin{pmatrix} \gamma_{uw} & 0 \\ \rho_{uw} & \sqrt{1-\rho_{uw}^2} \end{pmatrix} = \sigma_w \begin{pmatrix} k/\rho_{uw} & 0 \\ \rho_{uw} & \sqrt{1-\rho_{uw}^2} \end{pmatrix}$$

where $\gamma_{uw} = \sigma_u/\sigma_w$ and $k = -tr(A)/(1-tr^2(A))$.

The corresponding impulse response functions are

$\mathrm{IRF}_0(r_t) = 1$; $\mathrm{IRF}_j(r_t) = (\rho_{uw}/\gamma_{uw})tr^{j-1}(A)$, for $j > 0$.

$\mathrm{IRF}_j(\mu_{r,t+1}) = (\rho_{uw}/\gamma_{uw})tr^j(A)$, $j = 0, \ldots, \infty$.

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 $\mu_{r,t+1}$.

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

$$P_\infty(r_t \mid u_t) = 1 + \frac{1}{\gamma_{uw}^2}\left(\frac{\rho_{uw}^2}{1-tr^2(A)}\right) \qquad P_\infty(\mu_{r,t+1} \mid u_t) = \frac{1}{\gamma_{uw}^2}\left(\frac{\rho_{uw}^2}{1-tr^2(A)}\right)$$

Perhaps more interesting in the study of the effects of a shock to the conditional mean, w_t. To do so, we use the Cholesky decomposition of $V(w_t, u_t)$

$$\Sigma^{\star\star} = \sigma_w \begin{pmatrix} \gamma_{uw}\sqrt{1-\rho_{uw}^2} & \rho_{uw}\gamma_{uw} \\ 0 & 1 \end{pmatrix} = \sigma_w \begin{pmatrix} k(\sqrt{1-\rho_{uw}^2}/\rho_{uw}) & k \\ 0 & 1 \end{pmatrix}$$

Now we get

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

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

Note that for $tr(A)$ close to 1, k will be very large and negative. Therefore, a positive shock to w_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 w_t on r_t and its conditional mean

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

5 Empirical Application to US Stock 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 a bivariate VAR(1) for r_t and δ_t, 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 (1997) for data definitions and sources). 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 associated with 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 0.0859.

As we saw in Section 4, it is impossible for the univariate representation of r_t to be exactly white-noise in a $\text{V}\textsc{ar}(1)$ unless the companion matrix has reduced rank. For that reason, we also estimate by maximum likelihood a restricted $\text{V}\textsc{ar}(1)$ 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 4, 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 implied 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 $\text{A}\textsc{r}(1)$ 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 $\mu_{r,t+1}$, w_t.

The impulse response functions are presented in Figure 2. Note that as in Section 4.2, the initial positive effect on returns of a shock to u_t is later reversed by the very slowly decaying negative effect on expected returns. 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, $P_\infty(r_t \mid u_t) = 1.0105$, while $P_\infty(\mu_{r,t+1} \mid u_t) = 0.0105$. Similarly, $P_\infty(r_t \mid w_t) = 1600.00$, while $P_\infty(\mu_{r,t+1} \mid w_t) = 20.64$. 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.

6 Conclusions

In this paper we study the time series properties of the processes for the (unobserved) conditional mean and variance, given a specification of the process for the observed time series. We first derive a general result for the conditional mean of a multivariate linear processes, and then apply it to various models of interest used in the analysis of economic and financial time series, such as (seasonal) ARIMA and ARFIMA models, multivariate GARCH and univariate GARCH-M models. Proposition 1 can also be applied to other ARCH models in the literature, such as the Generalized Quadratic ARCH (GQARCH) model of Sentana (1995) (see Fiorentini and Sentana (1996) for details).

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.

We finally consider the joint process for a single variable 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.

From a risk management perspective, it would be useful to extend our results to cover the dynamics of conditional quantiles. Given that monotonic transformations of the observed series produce monotonic transformations of the order statistics, some progress can be made if we restrict ourselves to the Box-Cox quantile regression framework (see Buchinsky, 1995). If the conditional quantiles of a Box-Cox transformation of the observed series are linear, their autocorrelation structure can be derived from the autocorrelation structure of the transformed series, which in turn can be obtained from Granger and Newbold's (1976) results on instantaneous data transformations. The derivation of a more general result constitutes an interesting avenue for future research.

Appendix

A Proofs of Results

A.1 Proposition 1

Let's write $\boldsymbol{\mu}_t = A(L)\mathbf{x}_t - B(L)\boldsymbol{\epsilon}_t$. Premultiplying both sides by $[I - A(L)]$ yields $[I - A(L)]\boldsymbol{\mu}_t = [I - A(L)]A(L)\mathbf{x}_t - [I - A(L)]B(L)\boldsymbol{\epsilon}_t$. Then, since $[I - A(L)]A(L) = A(L)[I - A(L)]$, it follows that $[I - A(L)]\boldsymbol{\mu}_t = A(L)[I - B(L)]\boldsymbol{\epsilon}_t - [I - A(L)]B(L)\boldsymbol{\epsilon}_t = [A(L) - B(L)]\boldsymbol{\epsilon}_t$ □

A.2 Lemma 2

First notice that any covariance stationary GARCH(p,q) can be written in an ARCH(∞) form as $\sigma_t^2 = \alpha_0^\star + \alpha^\star(L)\epsilon_t^2$, where $\alpha_0^\star = \alpha_0[1 - \beta(1)]^{-1}$ and $\alpha^\star(L) = \alpha(L)[1 - \beta(L)]^{-1}$with $\alpha_i^\star \geq 0\ \forall i$, $\sum_{i=1}^{\infty} \alpha_i^\star < 1$. This implies that we can write a stationary AR(∞) for ϵ_t^2, $[1 - \alpha^\star(L)]\epsilon_t^2 = \alpha_0^\star + v_t$

Since the infinite moving average representation of ϵ_t^2 is $\epsilon_t^2 = \alpha_0^{\star\star} + \psi(L)v_t$ with $\psi_0 = 1$ and $\psi_j = \sum_{i=1}^{j} \alpha_i^\star \psi_{j-i}$, it is easy to verify that $\psi_j \geq 0$ for $j = 0, 1, \ldots, \infty$, 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 $Cov(\sigma_t^2 \sigma_{t-k}^2) = \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} \alpha_i^\star \alpha_j^\star Cov(\epsilon_t^2 \epsilon_{t-k+i-j}^2)$, which is non-negative for every k. □

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.

BUCHINSKY, M. (1995): "Quantile regression, Box-Cox transformation model, and the U.S. wage structure, 1963-1987", *Journal of Econometrics*, 65, 109-154.

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 (1997): *The Econometrics of Financial Markets*, Princeton University Press.

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.

FIORENTINI, G. AND E. SENTANA (1996): "Conditional means of time series models and time series models for conditional means", CEMFI Working Paper 9617.

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.

GRANGER, C.W.J. AND P. NEWBOLD (1976): "Forecasting transformed series", *Journal of the Royal Statistical Society B*, 38, 189-203.

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.

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.

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	-.0169
	(.0086)	(.0086)
a_{11}	.0708	-.0231
	(.0438)	(.0086)
a_{12}	.6455	.6074
	(.2281)	(.2285)
c_2	2.12e-4	2.14e-4
	(1.09e-4)	(1.09e-4)
a_{21}	-.0379	-.0379
	(5.58e-4)	
a_{22}	.9986	.9985
	(.0029)	(.0029)
σ_u	.0419	.0421
σ_v	5.34e-4	5.34e-4
ρ_{uv}	.0717	.0719

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

Note: Estimated model:

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

$$\begin{pmatrix} u_t \\ v_t \end{pmatrix} \sim \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_u^2 & \sigma_u\sigma_v\rho_{uv} \\ \sigma_u\sigma_v\rho_{uv} & \sigma_v^2 \end{pmatrix} \right]$$

Figure 1: **Case B**

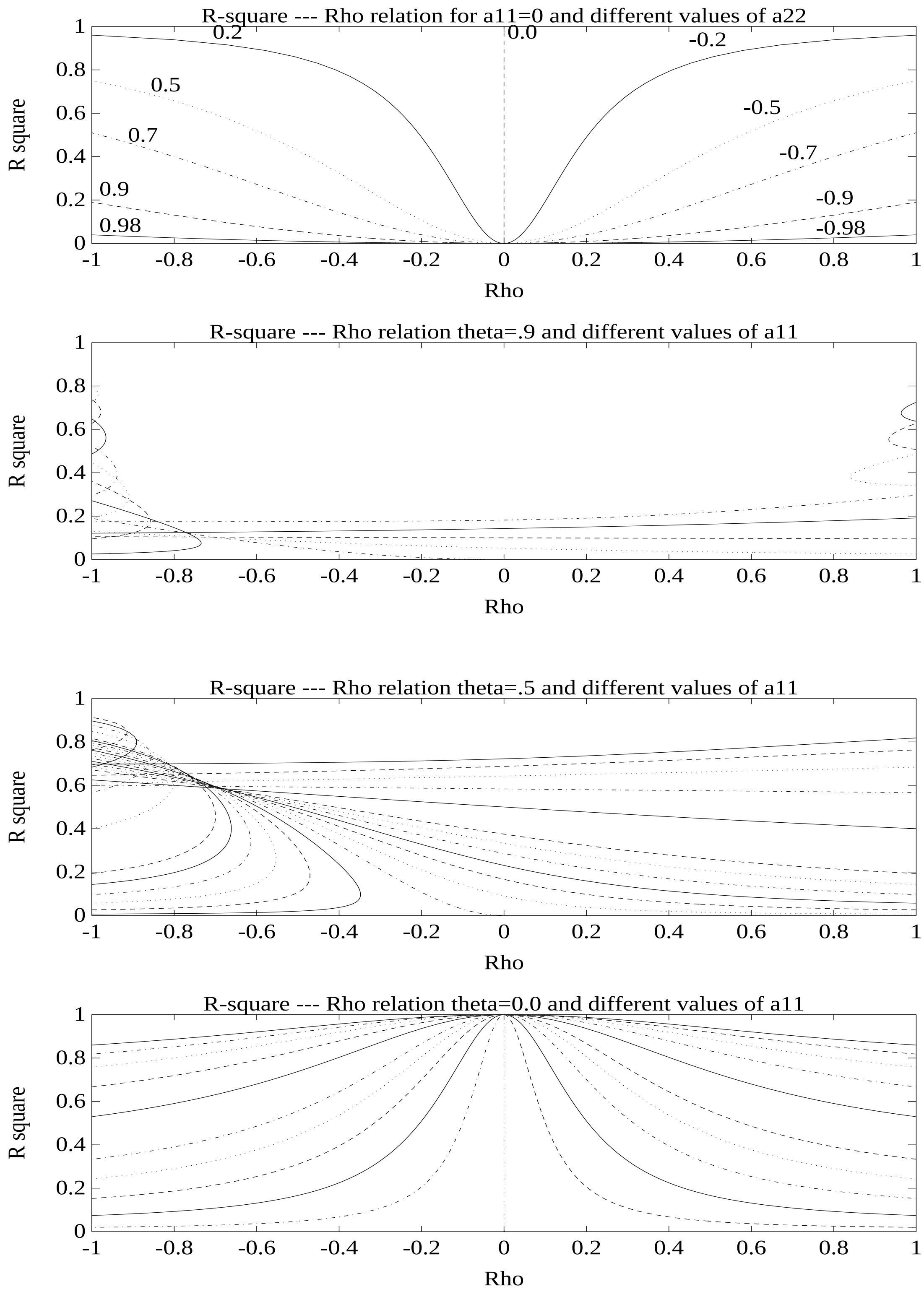

Figure 2: **IRF for returns and expected returns**