How does the Future Change our Past Views of the Present?[1]

Antonis Demos

(Athens University of Economics and Business)

Enrique Sentana

(CEMFI)

Current Version: June 2000

Preliminary

Comments welcome

[1]We have benefited from discussions with Gabriele Fiorentini and Neil Shephard. Of course, the usual caveat applies. We would also like to thank John Campbell for kindly allowing us to use his data. This work was initiated when the authors were based at the LSE Financial Markets Group, whose financial support as part of the ESRC project "The Efficiency and Regulation of Financial Markets" is gratefully acknowledged. The second author also acknowledges research support from CICYT through Project PB 97-0155. (email addresses: demos@aueb.gr and sentana@cemfi.es).

Abstract

We make inferences about expected stock market returns when we, as econometricians, only have the values of observed returns. We explicitly acknowledge that agents use information other than past returns in forming their expectations, but exploit the fact that we have information on what contemporaneous and future returns turned out to be. Given a bivariate process for returns and expected returns, we employ a Bayesian Kalman filter to extract the information on expected returns contained in the whole sample, explicitly taking into account parameter uncertainty. We apply our results to post-war US monthly real stock market returns.

Keywords: Expected Stock Returns, Gibbs Sampling, Bayesian Kalman Filter.

1 Introduction

The fact that stock returns have negligible autocorrelations was traditionally regarded as evidence in favour of the present value model with constant expected returns. Nevertheless, Shiller (1984), Summers (1986), Poterba and Summers (1988), and Fama and French (1988) showed that near white noise behaviour for observed returns, r_t, is compatible with a smooth but mean-reverting time-varying expected return process, $\mu_{rt} = E_{t-1}(r_t)$, 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.

As is well known, a univariate framework is too restrictive for the analysis of such an issue, because there is only one shock that drives the joint process for the observed variable and its conditional mean. In other words, the joint process for returns and its expected value is of reduced rank, with a singular covariance matrix for the innovations. This has long been realised, and two main approaches have been proposed. On the one hand, and by analogy with the stochastic volatility literature, we could directly specify a univariate stochastic process for the conditional moment, μ_{rt}, with "its own" innovation, and then derive the implied process for observed returns, r_t. For instance, following Poterba and Summers (1988), Fama and French (1988), or Campbell (1991), we could assume that expected stock returns follow a univariate AR(1) process, which would imply that r_t must follow an ARMA(1,1) process itself. Alternatively, we could follow the opposite route, and start from an observed multivariate process for the variable of interest, r_t, and another variable that Granger-causes it, δ_t say, and then derive the implied process for the expected return series, μ_{rt}. In this respect, Fiorentini and Sentana (1998) estimate a bivariate VAR(1) process for post-war US real stock market returns and dividend yields, and study the properties of the

implicit expected return process, as well as the impact of the innovations in the conditional mean process on observed returns. Obviously, the second approach reduces to the first one when the Granger-causal variable δ_t is precisely μ_{rt}, and there is no feedback from r_t to μ_{rt+1}.

The main practical limitation of both approaches, however, is that in order to describe the actual temporal evolution of expected returns, we have to assume that the agent's information set, I_{t-1}, and the econometrician's information set, J_{t-1}, effectively coincide. In contrast, the purpose of our analysis is to make inferences about μ_{rt} when we, as econometricians, only observe the values of r_t, so that $I_{t-1} \supseteq J_{t-1} = \{r_{t-1}, ..., r_1\}$. The main point of departure of our approach with respect to the existing literature, though, is that we explicitly take into account the fact that our information set includes not only past returns, $\{r_{t-1}, ..., r_1\}$, but also the present and the future of r_t, i.e. $J_T = \{r_T, r_{T-1}, ..., r_1\}$.

Given the underlying bivariate process for r_t and μ_{rt}, we can use the Kalman filter to extract the information about μ_{rt} contained in J_T in a well defined "optimal" sense. If we knew the parameter values characterising such a process, this would be an easy task. In practice, however, we are faced with the problem of recovering the structural parameters of the joint bivariate process from the reduced form log-likelihood function of the marginalised univariate process for observed returns. Unfortunately, without additional restrictions, the structural parameters are not separately identified in general. Nevertheless, given that we are not really interested in the values of the parameters, but rather on the values of the expected returns, a Bayesian approach provides a rather natural way of recovering μ_{rt}. In addition, one could argue that a Bayesian perspective would also be appropriate even if the parameters were identified in a classical sense, in order to take into account the effects of estimation uncertainty.

Unfortunately, the conditional distribution of μ_{rt} given J_T is generally un-

known. As a result, we must resort to a Gibbs sampling procedure, whereby draws from the joint distribution of expected returns and parameters, conditional on the observed process J_T, are produced by cycling over the following two steps: (i) draw from the conditional distribution of expected returns, given the observed sample and a vector of parameter values, and (ii) draw from the posterior distribution of the parameters, given our priors, the observed sample and expected returns.

We apply our proposed procedure to post-war US monthly real stock returns. In particular, we obtain a series of estimates of the agents' expected returns given all the information available at the end of the sample, together with a measure of the uncertainty surrounding those estimates. We also look at how our views on the returns expected by the agents are affected by contemporaneous and future information.

The paper is divided as follows. We analyse the theoretical issues involved in section 2. The results of the empirical application are discussed in section 3. Finally, our conclusions are presented in section 4. Proofs and auxiliary results are gathered in the appendix.

2 Theoretical Set-Up

2.1 The conditional mean of a vector process

Let us consider a multivariate linear stochastic process of orders k and h

$$[I - \Phi(L)]\, x_t = [I + \Theta(L)]\, u_t$$

or

$$x_t = \Phi_1 x_{t-1} + ... + \Phi_k x_{t-k} + u_t + \Theta_1 u_{t-1} + ... + \Theta_h u_{t-h} \tag{1}$$

where u_t is a $n \times 1$ white noise process of one-period ahead prediction errors, with zero mean and covariance matrix Σ, I is the identity matrix of order n, $\Phi(L)$ is a $n \times n$ matrix whose typical element is a polynomial in the lag operator L of order k, Φ_i are $n \times n$ matrices of coefficients, with $\Theta(L)$ and Θ_i analogously defined. We assume that the roots of $|I - \Phi(L)| = 0$ and $|I + \Theta(L)| = 0$ are on or outside the unit circle, which allows for (co-)integrated and invertible processes (whether strictly or not), but rules out explosive as well as non-invertible processes.

Such a formulation includes many models of interest widely used in the analysis of economic and financial time series, such as univariate and multivariate ARIMA models. It also nests models for conditional second moments, since dynamic conditional heteroskedastic processes often have a straightforward interpretation as linear processes for the squared innovations.[1] Finally, some important non-linear models can also be expressed in this way.[2]

Define $\mu_{t+1} = E_t(x_{t+1})$ as the $n \times 1$ conditional mean vector. The purpose of our analysis is to make inferences about a single element of μ_{t+1}, μ_{1t+1} say, when we, as econometricians, only observe the values of the corresponding variable, x_{1t}. Importantly, though, μ_{1t+1} is conditional on the agents' multivariate information set $X_t = \{x_t, x_{t-1}, x_{t-2}, \ldots\}$.

For the sake of concreteness, suppose that we are interested in learning about $\mu_{1\tau+1}$ given $X_{1t} = \{x_{1t}, x_{1t-1}, x_{1t-2}, \ldots\}$. If we only observe the values of x_{1t} up to, and including time t, with $\tau \geq t$, the inference problem is a prediction one, which under quadratic loss has the trivial solution $E(\mu_{1\tau+1}|X_{1t}) = E(x_{1\tau+1}|X_{1t})$ by the law of iterated expectations. In general, though, if $t > \tau$, so that we observe

[1] For instance, the univariate GQARCH(1,1) model of Sentana (1995) $\sigma_t^2 = E_{t-1}\left(y_t^2\right) = \psi + \beta\sigma_{t-1}^2 + \alpha y_{t-1}^2 + \delta y_{t-1}$ implies a bivariate ARMA(1,1) representation for y_t, y_t^2. Similarly, a multivariate GARCH(1,1) model for ε_t can be written as a VARMA(1,1) process for $vech(\varepsilon_t\varepsilon_t')$ (see e.g. Nijman and Sentana (1996)).

[2] For example, the univariate Bilinear(1,1,1,1) model $z_t = \alpha z_{t-1} + \vartheta z_{t-1}\eta_{t-1} + \eta_t - \beta\eta_{t-1}$, where $\eta_t \sim i.i.d.\left(0, \sigma^2\right)$ can be written as a bivariate ARMA(1,1) model for z_t and $\varepsilon_t = z_t\eta_t - \sigma^2$.

not only the past, but also the present and future of x_{1t}, it becomes a filtering problem, which can be conveniently solved using the Kalman filter.

2.2 Kalman filter smoothers

The results in Fiorentini and Sentana (1998) allow us to write the joint process for x_t and μ_{t+1} as:

$$\begin{pmatrix} x_t \\ \mu_{t+1} \end{pmatrix} = \sum_{i=1}^{k} \begin{pmatrix} 0 & I \\ 0 & \Phi_i \end{pmatrix} \begin{pmatrix} x_{t-i} \\ \mu_{t+1-i} \end{pmatrix} + \begin{pmatrix} I \\ C_1 \end{pmatrix} u_t + \sum_{i=2}^{m} \begin{pmatrix} 0 \\ C_i \end{pmatrix} u_{t-i} \quad (2)$$

where $m = max(k, h)$ and $C_i = \Phi_i + \Theta_i$ (with $\Phi_i = 0$ if $i > k$ and $\Theta_i = 0$ if $i > h$). If we rewrite equation (2) as an extended VAR(1) model in the usual way, i.e.

$$y_t = Ay_{t-1} + Bu_t \quad (3)$$

and take y_t as the state variable, we can then write the model in state-space form, with (3) as the transition equation, and

$$x_{1t} = e_1' y_t + 0 \quad (4)$$

as the measurement equation, where e_1 is the first vector of the orthonormal basis.

Let us call ψ the vector of model parameters, which contains the unrestricted elements of A, B and Σ. The Kalman filter is a recursive algorithm for estimating the state-vector y_t on the basis of the observed variables x_{1t}, and the parameter vector ψ. It usually involves two sets of equations: the prediction equations, which produce $y_{t|t-1} = E(y_t|X_{1t-1}, \psi)$ and the associated mean square error $P_{t|t-1}$, and the updating equations, which produce $y_{t|t} = E(y_t|X_{1t}, \psi)$ and the associated mean square error $P_{t|t}$. In practice, though, we can use the Ricatti equations to obtain estimates of $y_{t|t}$ and $P_{t|t}$ from estimates of $y_{t-1|t-1}$ and $P_{t-1|t-1}$ directly. Specifically,

$$y_{t|t} = \left[A - K_{t-1|t-1} e_1' A\right] y_{t-1|t-1} + K_{t-1|t-1} x_{1t}$$

$$P_{t|t} = \left[I - K_{t-1|t-1}e_1'\right]\left[AP_{t-1|t-1}A^{'} + B\Sigma B^{'}\right]$$

where

$$K_{t-1|t-1} = \left[AP_{t-1|t-1}A^{'} + B\Sigma B^{'}\right]e_1\left[e_1'\left(AP_{t-1|t-1}A^{'} + B\Sigma B^{'}\right)e_1\right]^{-1}$$

is known as the gain.

In fact, if we assume that conditioning on past information, the vector u_t is normally distributed, i.e. $u_t|X_{t-1}, \psi \backsim N(0, \Sigma)$, we will have that $y_t|X_{1t-1}, \psi \backsim N(y_{t|t-1}, P_{t|t-1})$, and $y_t|X_{1t}, \psi \backsim N(y_{t|t}, P_{t|t})$, which means that we can fully characterise the conditional distributions of the states given the observations. Such a characterisation is particularly important in our set-up, because as we shall see below, in general some of the elements of y_t, and in particular the vector of conditional means μ_{t+1}, cannot be fully recovered from the observed series.

In principle, the estimates of $P_{t|t}$ will depend on t. However, at the steady state (see Harvey (1989) for sufficient existence conditions), these estimates are by definition independent of time. In that case, the corresponding Ricatti equations can be written as

$$P = (I - Ke_1')\left[APA^{'} + B\Sigma B^{'}\right] \tag{5}$$

where

$$K = \left[APA^{'} + B\Sigma B^{'}\right]e_1\left[e_1'\left(APA^{'} + B\Sigma B^{'}\right)e_1\right]^{-1} \tag{6}$$

which can then be solved for K and P.

In this well-known set-up, both $E(\mu_{1\tau+1}|X_{1t}, \psi)$ and the associated mean square error can be obtained by means of a smoothing algorithm. In particular, we shall use a fixed point smoother, which, conditional on the parameter values, allows us to see how the estimates of $\mu_{1\tau+1}$ change as we observe more and more observations on x_{1t}.

Let $y_t^* = y_{t-1}^*$ for $t = \tau + 1, ..., T$, with $y_\tau^* = y_\tau$ as initial condition, and

consider the following augmented state-space model:

$$x_t = e_1^{+\prime} y_t^+$$

$$y_t^+ = \begin{pmatrix} y_t \\ y_t^* \end{pmatrix} = \begin{pmatrix} A & 0 \\ 0 & I \end{pmatrix} \begin{pmatrix} y_t \\ y_t^* \end{pmatrix} + \begin{pmatrix} B \\ 0 \end{pmatrix} u_t$$

where $e_1^{+\prime} = (e_1', 0')$.

The Kalman Filter equations for the augmented model can be broken into two parts. The first part are the standard Kalman filter equations given above, while for $t = \tau + 1, ..., T$ the second part is as follows:

$$\begin{aligned} y^*_{t+1|t} &= y^*_{t|t-1} + K^*_{t|t-1}\left[x_{1t} - e_1' y_{t|t-1}\right] \\ P^{**}_{t+1|t} &= P^{**}_{t|t-1} - P^*_{t|t-1} e_1 K^{*'}_{t|t-1} \end{aligned}$$

with

$$\begin{aligned} K^*_{t|t-1} &= P^*_{t|t-1} e_1 \left[e_1' P_{t|t-1} e_1\right]^{-1} \\ P^*_{t+1|t} &= P^*_{t|t-1}\left[A - K_{t|t-1} e_1'\right]' \\ K_{t-1|t-1} &= \left[A P_{t-1|t-1} A' + B\Sigma B'\right] e_1 \left[e_1'\left(A P_{t-1|t-1} A' + B\Sigma B'\right) e_1\right]^{-1} \end{aligned}$$

where the initial state vector is $y^*_{\tau|\tau-1} = y_{\tau|\tau-1}$, and the initial mean square error matrix $P^{**}_{\tau|\tau-1} = P^*_{\tau|\tau-1} = P_{\tau|\tau-1}$.

In this context, a particularly interesting question is whether we can fully recover $\mu_{1\tau+1}$ from the whole past, present and future history of x_{1t}, given a parameter configuration ψ. In the above notation, this is equivalent to asking whether the appropriate element of $P^{**}_{t+1|t}$ goes to 0 as t increases. In principle, we would expect this not to be the case. Nevertheless, the following proposition, which is proved in the appendix, gives necessary and sufficient conditions in this respect:

Proposition 1 $\lim_{t\to\infty} P^{**}_{t+1|t} = 0$ *if and only if the joint spectral density matrix of* x_{1t} *and* μ_{1t+1} *is singular at all frequencies.*

In the next subsection, we present a simple example in which this condition is always satisfied, and another one in which it seldom is.

2.3 Examples

2.3.1 Non-Invertible MA(1)

Consider the following MA(1) process:

$$x_t = \varepsilon_t + \beta\varepsilon_{t-1}$$

where $\varepsilon_t|X_{t-1}, \psi \sim i.i.d.\ (0, \sigma^2)$. If the agents observe the underlying shocks, ε_t, then $\mu_{t+1} = \beta\varepsilon_t$. However, if $|\beta| > 1$, we, as econometricians, cannot recover ε_t from observations on x_t, x_{t-1}, x_{t-2} alone. Consequently, $E(x_{t+1}|X_t, \psi) = \beta^{-1}(1 + \beta^{-1}L)^{-1}x_t \neq \mu_{t+1}$.

In order to apply the techniques described in the previous subsection, it is convenient to re-write the univariate MA(1) model as the following bivariate VAR(1):

$$\begin{pmatrix} x_t \\ \mu_{xt+1} \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} x_{t-1} \\ \mu_{xt} \end{pmatrix} + \begin{pmatrix} 1 \\ \beta \end{pmatrix} \varepsilon_t$$

and

$$x_t = (1, 0) \begin{pmatrix} x_t \\ \mu_{xt+1} \end{pmatrix}$$

Since the eigenvalues of the companion matrix are trivially less than one, the model is stable. Besides, since $P_{1|0} = \begin{pmatrix} \sigma^2 + \{P_{0|0}\}_{22} & \sigma^2\beta \\ \sigma^2\beta & \sigma^2\beta^2 \end{pmatrix}$ is a positive semidefinite matrix for any β, the steady state exists.

To find the steady-state values, we have to solve the Ricatti equations, which are obtained by substituting the appropriate values of A, B and Σ in equations (5) and (6) (see the appendix). Not surprisingly, we get two solutions, either:

$$P = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} \text{ and } K = \begin{pmatrix} 1 \\ \beta \end{pmatrix}$$

or

$$P = \begin{pmatrix} 0 & 0 \\ 0 & \sigma^2\left(\beta^2 - 1\right) \end{pmatrix} \text{ and } K = \begin{pmatrix} 1 \\ \beta^{-1} \end{pmatrix}$$

The first solution corresponds to $|\beta| \leq 1$, and simply says that the ε_t's can be fully recovered from present and past values of x_t. In contrast, the second one applies when $|\beta| > 1$, and implies that the uncertainty associated with the estimation of ε_t's is higher the higher the absolute value of β.[3]

If we start the smoothing recursions from the steady-state matrices, we obtain

$$P^{**}_{t+1|t} = P = \sigma^2 \begin{pmatrix} 0 & 0 \\ 0 & \beta^2 - 1 \end{pmatrix}$$

$$P^{**}_{t+2|t+1} = \sigma^2 \begin{pmatrix} 0 & 0 \\ 0 & \left(\beta^2 - 1\right)\beta^{-2} \end{pmatrix}$$

and by induction

$$P^{**}_{t+k+1|t+k} = \sigma^2 \begin{pmatrix} 0 & 0 \\ 0 & \left(\beta^2 - 1\right)\beta^{-2k} \end{pmatrix}$$

Hence, in this case $\lim_{k\to\infty} P^{**}_{t+k+1|t+k} = 0$, which simply reflects the well known fact that we can fully recover the values of the shocks ε_t, and consequently, the values of μ_{xt+1}, from the future values of the observed series.[4]

[3] An alternative way to obtain the same result is as follows. It is well known that the Wold decomposition of the above model is $x_t = u_t + \beta^{-1}u_{t-1}$, where $u_t = \left(1 + \beta^{-1}L\right)^{-1}(1 + \beta L)\,\varepsilon_t \sim i.i.d.\ \left(0, \sigma_u^2\right)$ with $\sigma_u^2 = \beta^2\sigma^2$. Hence $E\left(x_{t+1}|X_t\right) = E\left(\varepsilon_{t+1} + \beta\varepsilon_t|X_t\right) = \beta^{-1}u_t$. But since $E\left(\varepsilon_{t+1}|X_t\right) = 0$, this implies that $E\left(\varepsilon_t|X_t\right) = \beta^{-2}u_t$. Now $Var\left(x_{t+1}|X_t\right) = \sigma_u^2$ and $Var\left(\varepsilon_{t+1}|X_t\right) = \sigma^2$. Hence $Var\left(\varepsilon_t|X_t\right) = \beta^{-2}\left(\sigma_u^2 - \sigma^2\right) = \beta^{-2}\left(\beta^2\sigma^2 - \sigma^2\right) = \sigma^2\left(1 - \beta^{-2}\right)$. Consequently, $Var\left(\mu_{xt+1}|X_t\right) = Var\left(\beta\varepsilon_t|X_t\right) = \sigma^2\left(\beta^2 - 1\right)$.

[4] Note that the result also follows directly from Proposition 1, since the joint spectral density matrix is singular at all frequencies, given that the covariance matrix of the innovations is singular.

2.3.2 Marginalised Bivariate VAR(1) Models

Consider the following bivariate VAR(1) model for some variable, r_t say, and some other variable, δ_t say, which helps predict r_t:

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

where $(u_t, v_t)|X_{t-1}, \psi \sim i.i.d.\ (0, \Omega)$.

We can write the general state space representation (3) and (4) as

$$r_t = \begin{pmatrix} 1 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} r_t \\ \delta_t \\ \mu_{rt+1} \\ \mu_{\delta t+1} \end{pmatrix}$$

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

which coincides with Akaike's (1974) state space representation. In this case, though, we can reduce the dimension of the state vector without loss of information by marginalising with respect to r_t and μ_{rt+1}. Following Fiorentini and Sentana (1998), this yields

$$\begin{pmatrix} r_t \\ \mu_{rt+1} \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} r_{t-1} \\ \mu_{rt} \end{pmatrix} + \begin{pmatrix} u_t \\ w_t \end{pmatrix} \tag{7}$$

as transition equation, and

$$r_t = \begin{pmatrix} 1 & 0 \end{pmatrix} \begin{pmatrix} r_t \\ \mu_{rt+1} \end{pmatrix} \tag{8}$$

as measurement equation, where $w_t = \alpha_{11}u_t + \alpha_{12}v_t$, a_{22} and a_{21} are the trace and minus determinant, respectively, of the companion matrix of the original VAR, and $Var(u_t, w_t)' = \Sigma$.

In order to find the steady-state values, we have to solve again the Ricatti equations, which are obtained by substituting the appropriate values of A, B and Σ in equations (5) and (6). In this case, the equations have only one admissible solution:

$$P = \begin{pmatrix} 0 & 0 \\ 0 & p_3 \end{pmatrix} \text{ and } K = \begin{pmatrix} 1 \\ \frac{a_{22}p_3 + \sigma_{uw}}{p_3 + \sigma_u^2} \end{pmatrix}$$

where p_3 is the positive root of:[5]

$$p_3^2 + p_3(\sigma_u^2 + 2a_{22}\sigma_{uw} - a_{22}^2\sigma_u^2 - \sigma_w^2) + \sigma_{uw}^2 - \sigma_w^2\sigma_u^2 = 0 \tag{9}$$

If we start the fixed point smoother from the steady state matrices, we get:

$$P^{**}_{t+1|t} = P = \begin{pmatrix} 0 & 0 \\ 0 & p_3 \end{pmatrix}$$

$$P^{**}_{t+2|t+1} = \begin{pmatrix} 0 & 0 \\ 0 & p_3\frac{\sigma_u^2}{p_3+\sigma_u^2} \end{pmatrix},$$

$$P^{**}_{t+3|t+2} = \begin{pmatrix} 0 & 0 \\ 0 & p_3\sigma_u^2\frac{p_3(\sigma_u^2+\sigma_w^2)+\sigma_u^2+\sigma_w^2\sigma_u^2-\sigma_{uw}^2}{(p_3+\sigma_u^2)^3} \end{pmatrix},$$

etc. We prove in the appendix that as $k \to \infty$ the residual variance converges to

$$\frac{\sigma_w^2\sigma_u^2 - \sigma_{uw}^2}{\sqrt{\sigma_w^2(1-\rho^2) + [\rho\sigma_w + (1-a_{22})\sigma_u]^2}\sqrt{\sigma_w^2(1-\rho^2) + [\rho\sigma_w - (1+a_{22})\sigma_u]^2}}$$

which means that in this case $\lim_{k\to\infty} P^{**}_{t+k+1|t+k} = 0$ if and only if the correlation between shocks to returns and shocks to expected returns is either 1 or minus

[5] Notice that equation (9) has two real roots, one positive and one negative, as $\sigma_{uw}^2 - \sigma_w^2\sigma_u^2 \leq 0$, but only the first one is admissible.

1. Notice that this unlikely event corresponds to a situation in which p_3 is 0, so that the process for x_{1t} is effectively univariate, in which case there is no extra gain from the present and future observations on x_{1t}. Note also that the above expressions do not depend on the feedback parameter a_{21}.

2.4 Parameter uncertainty

If we knew the values of the structural parameters characterising A and Σ in the VAR(1) example above, we could directly apply the procedures developed in the previous sections to the data. In practice, though, those parameter values are unknown. More importantly, there may not be enough information in the data alone to estimate them, since the likelihood function of the observed series depends on the set of reduced form parameters characterising the marginal process for r_t. Specifically, given that r_t follows the following univariate ARMA(2,1) process

$$(1 - a_{22}L - a_{21}L^2)r_t = (1 - \theta L)\xi_t$$

where $(1 - \theta L)\xi_t$ is the Wold decomposition of the MA(1) process $u_t - a_{22}u_{t-1} + w_{t-1}$, at best we can only estimate the "reduced form" parameters a_{22}, a_{21}, θ and σ_ξ^2 from the data, assuming that there is no cancellation between the MA and AR polynomials.

The specific relationship between both set of parameters is given by the following equations

$$\frac{\theta}{1+\theta^2} = \frac{(a_{22}\gamma_{uw} - \rho_{uw})\gamma_{uw}}{(1 + a_{22}^2)\gamma_{uw}^2 + 1 - 2a_{22}\rho_{uw}\gamma_{uw}} \tag{10}$$

$$\theta\sigma_\xi^2 = (a_{22}\gamma_{uw} - \rho_{uw})\gamma_{uw}\sigma_w^2 \tag{11}$$

where $\gamma_{uw} = \sigma_u/\sigma_w$ and $\rho_{uw} = \sigma_{uw}/\sigma_u\sigma_w$. Hence, the structural parameters are only partly identified, since in the best situation we have only two equations for three unknowns. In particular, for a given a_{22} and θ, all admissible γ_{uw} and ρ_{uw}

must lie along the intersection of the hyperbola

$$[(1+a_{22}^2)\theta - a_{22}(1+\theta^2)]\gamma_{uw}^2 - (1+\theta^2-2a_{22}\theta)\gamma_{uw}\rho_{uw} + \theta = 0 \qquad (12)$$

and the rectangular region $\gamma_{uw} \geq 0$, $-1 \leq \rho_{uw} \leq 1$.[6]

In fact, we can solve equations (10) and (11) to get γ_{uw} and ρ_{uw} as functions of the "free" parameter σ_w^2. Specifically

$$\gamma_{uw} = \sqrt{\frac{\phi\theta\sigma_\xi^2 + \theta\sigma_w^2}{(a_{22}\phi - \lambda)\sigma_w^2}} \qquad (13)$$

$$\rho_{uw} = \theta\frac{a_{22}\sigma_w^2 + \lambda\sigma_\xi^2}{\sqrt{(\phi\theta\sigma_\xi^2 + \theta\sigma_w^2)(a_{22}\phi - \lambda)\sigma_w^2}} \qquad (14)$$

where

$$\lambda = (1+a_{22}^2)\theta - a_{22}(1+\theta^2)$$

$$\phi = 1+\theta^2-2a_{22}\theta$$

Note again that these expressions do not depend on the feedback parameter a_{21}.

In principle, we could fix σ_w^2 to some admissible positive value (see footnote 6), and then apply the Kalman filter after obtaining γ_{uw} and ρ_{uw} from the above expressions, with a_{22}, a_{21}, θ and σ_ξ^2 replaced by their maximum likelihood estimates. Nevertheless, there are at least three problems with this approach. First, it is not clear a priori which value we should choose for σ_w^2. Second, even if it were, it would be rather difficult to modify the expression for $P_{t+k+1|t+k}^{**}$ in order to account for the uncertainty surrounding the maximum likelihood estimators of the reduced form parameters. This problem is particularly important in practice,

[6] In this respect, it can be proved that the admissible γ_{uw}'s will be between the maximum and the minimum of

$$\left|\frac{1-a_{22}\theta}{(1+a_{22}^2)\theta - a_{22}(1+\theta^2)}\right| \text{ and } \left|\frac{(\theta-\alpha)\theta}{(1+a_{22}^2)\theta - a_{22}(1+\theta^2)}\right|$$

because as we shall see below, some of the reduced form parameters are estimated rather imprecisely. Third, even if we knew how to do it, if we then decided to repeat the exercise for two or more values of σ_w^2, it is not obvious what the right way to combine the results obtained would be.

Therefore, given that our prime interest lies on the expected return series and not on the parameters, a Bayesian perspective appears as the natural solution. For instance, if we are interested in point predictions of $\mu_{1\tau+1}$, we can compute the classical Kalman filter smoother $E(\mu_{1\tau+1}|X_{1T}, \psi)$ for every possible value of the structural parameters, and then average with respect to their posterior probability distributions $\pi(\psi|X_{1T})$. More formally,

$$E(\mu_{1\tau+1}|X_{1T}) = \int E(\mu_{1\tau+1}|X_{1T}, \psi)\ \pi(\psi|X_{1T})\ d\psi$$

with

$$\pi(\psi|X_{1T}) = \frac{\pi(X_{1T}|\psi)\ \pi(\psi)}{\int \pi(X_{1T}|\psi)\ \pi(\psi)\ d\psi}$$

where $\pi(\psi)$ is the prior distribution of the parameters, and $\pi(X_{1T}|\psi)$ the usual likelihood function. In addition, we can also obtain any other characteristic of the conditional distribution of $\mu_{1\tau+1}$ given X_{1T} alone, as explained in the next section.

2.4.1 Gibbs sampling procedure

Let us temporarily assume that we have a sample of size T on both returns and expected returns. In order to specify prior distributions, it is convenient to re-write the state-space form of the bivariate model in the following way:

$$r_t = \mu_{rt} + u_t \tag{15}$$

$$\mu_{rt+1} = a_{21} r_{t-1} + a_{22}\mu_{rt} + \beta u_t + \zeta_t \tag{16}$$

where $\beta = \sigma_{uw}/\sigma_u^2$, $E(\zeta_t) = 0$, and $Var(\zeta_t) = \delta^2 = (\sigma_w^2 - \sigma_{uw}^2/\sigma_u^2)$. Assume that the parameters σ_u^2 and $\varphi = \left(a_{21}, a_{22}, \beta, \delta^2\right)'$ are independently prior distributed,

and have priors of the inverted gamma and normal-inverted gamma conjugate variety, respectively.

Specifically, let us assume that the prior for σ_u^2 is $\sigma_u^{-2} \sim \gamma(\underset{-}{s}^{-2}, \underset{-1}{\nu})$, which means that the inverse of σ_u^2, known as the precision, is distributed as a gamma variate with parameters $\underset{-}{s}^{-2}$ and $\underset{-1}{\nu}$. Notice that this implies that $E(\sigma_u^{-2}) = \underset{-}{s}^{-2}$ and $V(\sigma_u^{-2}) = 2\ \underset{-}{s}^{-4} \ / \ \underset{-1}{\nu}$, so that the variance of the precision is inversely related to $\underset{-1}{\nu}$, the so-called "degrees of freedom" parameter of the gamma distribution. Furthermore, let us assume that the prior of φ and δ^{-2}, where $\varphi = (a_{21}, a_{22}, \beta)'$, is normal-gamma $NG\left(\underset{-}{\varphi}, \underset{-}{Q}, \underset{-}{d}^{-2}, \underset{-2}{\nu}\right)$, independently of σ_u^2, with probability density function

$$f(\varphi, \delta^2) = \phi_4(\varphi|\ \underset{-}{\varphi}, \underset{-}{Q}, \delta^{-2})\gamma\left(\delta^{-2}|\ \underset{-}{d}^{-2}, \underset{-2}{\nu}\right)$$

where $\phi_4(\cdot)$ is the trivariate normal density function, and $\gamma(\cdot)$ is a gamma density with parameters $\underset{-}{d}^{-2}$ and $\underset{-2}{\nu}$.

Then, it is well-known that the posterior distribution of σ_u^{-2} will be given by

$$\sigma_u^{-2}|R_T, M_{rT} \sim \gamma(\bar{s}^{-2}, \bar{\nu}_1)$$

where $R_T' = (r_T, r_{T-1}, ...r_2, r_1)$ and $M_{rT} = (\mu_{rT+1}, \mu_{rT}, .., \mu_{r2}, \mu_{r1})$ are the "observed" data,

$$\bar{\nu}_1 = \underset{-1}{\nu} + T$$

and

$$\bar{\nu}_1 \bar{s}^{-2} = \underset{-1}{\nu}\ \underset{-}{s}^{-2} + \hat{\nu}_T \hat{s}_T^{-2} + T\left(\bar{u}_T\right)^2$$

where

$$\begin{aligned}
\bar{u}_T &= \frac{1}{T}\sum_{t=1}^{T} u_t \\
\hat{s}_T^2 &= \frac{1}{T-1}\sum_{t=1}^{T}\left(u_t - \bar{u}_T\right)^2 \\
\hat{\nu}_T &= T - 1
\end{aligned}$$

(see e.g. Poirier (1995)). On the other hand, the posterior distribution of φ, δ^{-2} is:

$$\varphi, \delta^{-2}|R_T, M_{rT} \sim NG(\bar{\varphi}, \bar{Q}, \bar{d}^{-2}, \bar{\nu}_2)$$

with

$$\begin{aligned}
\bar{\varphi} &= \bar{Q}\left(\underset{-}{Q}^{-1}\underset{-}{\varphi} + X'X\hat{\varphi}_T\right) \\
\hat{\varphi}_T &= (X'X)^{-1}X'y \\
\bar{Q} &= \left(\underset{-}{Q}^{-1} + X'X\right)^{-1} \\
\bar{\nu}_2 &= \underset{-2}{\nu} + T \\
\bar{\nu}_2\bar{d}^{-2} &= \underset{-2}{\nu}\underset{-}{d}^{-2} + \underset{-}{\zeta}'\left(I_T + X\underset{-}{Q}X'\right)\underset{-}{\zeta}
\end{aligned}$$

where $X' = (x_1', x_2', .., x_T')$, $x_t = (r_{t-1}, \mu_{rt}, u_t)$, $\underset{-}{\zeta} = \left(\underset{-1}{\zeta}, \underset{-2}{\zeta}, ..., \underset{-T}{\zeta}\right)'$, and $\underset{-t}{\zeta} = r_t - x_t'\underset{-}{\varphi}$ (see e.g. Poirier (1995)).

Unfortunately, the elements of M_{rT} are unknown. Nevertheless, we can employ a Gibbs sampling procedure to obtain draws from $\pi(M_{rT}, \psi|R_T)$ by cycling over the steps $M_{rT} \sim \pi(M_{rT}|R_T, \psi)$ and $\psi \sim \pi(\psi|R_T, M_{rT})$, where each draw serves to redefine the conditioning set for the next step. In this respect, we use a slightly modified version of the multi-stage simulation smoother developed in de Jong and Shephard (1995) (see the appendix) to draw M_{rT} from $\pi(M_{rT}|R_T, \psi)$.

3 Empirical Application to US stock market returns

3.1 A first look at the log-likelihood function

We apply the procedures explained in the previous section to post-war U.S. real stock market returns. Our sample covers 516 monthly observations on (continuously compounded) returns from January 1952 to December 1994 (see chapter

7 of Campbell, Lo and MacKinlay (1997) for data definitions, transformations and sources).

But before proceeding to our final analysis, it is convenient to have a preliminary look at the log-likelihood function of the reduced form model. As we saw before, when the bivariate representation for r_t and δ_t corresponds to a VAR(1) model, the marginalised process for r_t is an ARMA(2,1). In this respect, the first thing to note is that the (exact) maximum likelihood estimate of the second order autoregressive coefficient a_{21} is very small (.0384), and that the log-likelihood function is rather flat in its neighbourhood. Specifically, the likelihood ratio statistic for $a_{21} = 0$ is only .68. Moreover, if we set a_{21} to 0, and estimate an ARMA(1,1) model instead, the autoregressive and moving average coefficients turn out to be both negative but very close to each other. Consequently, the observed series is close to white noise, a well documented fact for stock market returns in general, and for the dataset under consideration in particular (see e.g. Campbell (1991), or Fiorentini and Sentana (1998)). Importantly, the concentrated log-likelihood as a function of a_{22} and θ is very flat around the optimum, as depicted in Figure 1, which means that there is considerable uncertainty surrounding those point estimates.

3.2 Our current views on expected returns

We initially consider the regression equations (15) and (16), with a_{21} set to 0 for simplicity, but with a constant included, so that $\psi = (\sigma_u^2, c, a_{22}, \beta, \delta^2)'$ is the relevant vector of parameters.

In order to implement our proposed Bayesian approach to smoothing, we first have to specify the hyperparameters that characterise the prior distributions of those structural parameters. In this respect, our aim was to use informative priors that would be in accordance with the "received wisdom". In particular, since there

is a common belief that expected returns follow a smooth process whose first order autocorrelation is high (see Campbell (1991), and the references therein), we choose the prior mean of a_{22} to be .9. We also set the prior mean for c equal to .05, which implies an average real return of 6% on an annual basis. Similarly, given that there is a widespread impression that innovations in returns are highly negatively correlated with innovations in expected returns (see Campbell (1991) for a theoretical justification), we choose the prior mean for β to be -.317. Finally, we specify the prior means of σ_u^{-2} and δ^{-2} to be .04 and 19.24 respectively, which in terms of the original parametrisation correspond to $\rho_{uw} = -.95$ and $\gamma_{uw} = 6.85$. Such parameter values imply a standard deviation for real returns of 5.27%, and an R^2 for equation (15) of around 10%, both on a monthly basis.

As for the variances of the inverted-gamma priors for σ_u^2 and δ^2, we set the "degrees of freedom" $\underset{-1}{\nu}$ and $\underset{-2}{\nu}$ to 270, which is just over half the actual sample size. Finally, we must specify our choice of $\underset{-}{Q}$, i.e. the variance of the prior normal distribution of the parameter vector $\varphi = (c, a_{22}, \beta)'$. One attractive possibility is to set its inverse to $T^* E(x_t x_t' | \psi)$, where $x_t = (1, \mu_{rt}, u_t)'$, T^* is a scalar and ψ is evaluated at its prior mean. This yields

$$\underset{-}{Q}^{-1} = T^* \begin{pmatrix} 1 & c/(1-a_{22}) & 0 \\ c/(1-a_{22}) & c^2/(1-a_{22})^2 + \sigma_w^2/(1-a_{22}^2) & 0 \\ 0 & 0 & \sigma_u^2 \end{pmatrix}$$

Such a choice for $\underset{-}{Q}$ is compatible with the usual interpretation that a conjugate prior can be viewed as the posterior distribution that could have been obtained from a fictitious sample from the same population (see e.g. Poirier (1995)). Thus, T^* could be interpreted as the number of observations, or more precisely, the degrees of freedom in the fictitious sample. For that reason, we also set T^* to 270. In any case, we will perform a sensitivity analysis with respect to the variance hyperparameters.

We initially set the parameters to their prior means in order to start the Gibbs sampler recursions. Each subsequent iteration involves the following steps:

1) Given the current set of parameters, draw values of expected returns by means of the simulation smoother in the appendix.

2) Given the existing simulated expected returns, draw a new set of parameters from their posterior distributions, as explained in section 2.4.1.

As a practical rule, we stop once 1,000 such iterations have been completed. In this respect, it is important to mention that since the Gibbs sampler is just a particular example of a Markov chain Monte Carlo simulator, successive draws are not generally independent. For that reason, we compute inefficiency ratios[7] to assess the extent of the autocorrelation in the drawings.

Figure 2 plots the actual real return series over the whole of the sample period. As is well known, the performance of the U.S. stock market over the sample period has been impressive, with an average return of .546% on a monthly basis. At the same time, it has been very volatile, with a monthly standard deviation of 4.24%, and few large spikes that correspond to well known episodes. For each and every time period, observed returns can be compared with the average across replications of our smoothed expected returns, which are depicted in Figure 3. As anticipated, smoothed expected returns are markedly less volatile than actual returns, but they still show significant variation over time. In this respect, the most distinctive episode is perhaps the period that followed the first oil crisis, when it would seem that at the time agents were holding rather pessimistic views about the future prospects of the US stock market. In the same graph, we also present the first and third quartiles of the simulated smoothed distributions, whose separation provides a measure of the uncertainty about μ_{rt} which still remains once

[7] That is, the ratio of the variance of the sample mean of the drawings to the variance of the drawings divided by the number of replications. Under random sampling, the innefficiency ratio is 1.

the whole sample has been observed. In this respect, note that the interquartile range is generally fairly small, although it increases as we get closer to the end of the sample. The obvious reason is that the amount of future information becomes very scarce at that point.

We can also look at a particular month, and analyse how future information would have changed the views held by a contemporaneous econometrician endowed with the relevant part of the dataset. In particular, assuming that she was employing the same techniques as ourselves, including the same priors, we can fully characterise her views by drawing from the distribution of μ_{rt+1} given J_t (the predictive density), and compare it to the distribution of μ_{rt+1} given J_{t+1} (the updated density) and J_T (the smoothed density). The extra information has two separate effects: on the one hand, it modifies the estimates of μ_{rt+1} for each possible parameter configuration; on the other, it affects the posterior densities of the parameters.

We report the results of such an exercise for two of the most dramatic episodes in the recent history of the US stock market: the exceptional 12.78% rise in the market in January 1975, which consolidated its recovery after the very turbulent period in 73-74 (Figure 4), and the (in)famous 25.73% crash of October 1987 (Figure 5). As can be seen from Table 1, in the first case, the average expected return is in fact revised upwards as the subsequent returns are observed. In contrast, the evolution around October 87 is rather different. A positive mean for the predictive distribution is transformed into a negative one for the updated distribution once the return for October is observed, but then it reverts somewhat towards its unconditional mean level as subsequent observations accumulate. Importantly, the uncertainty that we, as econometricians, have about the views presumably held by stock market participants just before those events occurred, is significantly reduced as the future unravels. In particular, the standard deviation is reduced by

10-15% as we move from the predictive to the updated density, and by a further 50-60% when we move from the updated to the smoothed one. Although the different posterior distributions of μ_{rt+1} are non-normal, similar results are obtained if we replace means and standard deviations by medians and interquartile ranges, respectively.

4 Conclusions

In this paper we make inferences about expected stock market returns when we, as econometricians, only have access to the values of observed returns. The main point of departure with respect to the existing literature is that we explicitly take into account that our information set differs from the agents' information set in two crucial respects. First, when at time $t-1$ agents form their expectations about stock returns over period t, they observe many other facts in addition to the series of past returns. In contrast, we have information on what returns over period t and after turned out to be. Therefore, given an underlying bivariate process for returns, r_t, and expected returns, μ_{rt}, we could use the standard Kalman filter to extract the information on μ_{rt} contained in the whole sample, as long as we also knew the relevant parameter values. Unfortunately, without additional restrictions, the structural parameters cannot be separately identified from the reduced form log-likelihood function of the marginalised univariate process for observed returns. In particular, we show that there is at least one degree of underidentification in the structural parameters. Nevertheless, given that we are not really interested in the values of the parameters, but rather in the values of the expected returns, we employ a Bayesian version of the Kalman filter. A significant advantage of taking a Bayesian perspective in this context is that it explicitly takes into account the effects of estimation uncertainty, which would constitute a very relevant issue even if all the structural parameters could be separately identified.

In order to simulate from the posterior distribution of expected returns given the observed series, we resort to a Gibbs sampling algorithm, whereby draws from the joint distribution of expected returns and parameters, conditional on observed returns, are produced by sampling from the two conditional distributions.

We apply our proposed procedure to post-war US monthly real stock returns. Specifically, we obtain a series of estimates of the agents' expected returns given all the information available at the end of our sample, together with a measure of the uncertainty surrounding those estimates. In this respect, we find that smoothed expected returns are markedly less volatile than actual returns, but they still show significant variation over time. We also find that the uncertainty surrounding expected returns is generally fairly small once the whole sample has been observed, although it increases as we approach the end of the sample.

Finally, we look at how our views on the returns expected by the agents are affected by contemporaneous and future information. We concentrate on two important episodes in the recent history of the US stock market: January 75 and October 87. Our results suggest that the views on expected returns held by a contemporaneous econometrician endowed with observations on returns up to, and including the previous month, would have been significantly altered not only when returns over those months were observed, but also as subsequent observations materialised. Such an effect is noticeable not only on the location of the posterior distributions of expected returns, but also on the corresponding dispersions, which are substantially reduced.

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. (1991) "A Variance decomposition for Stock Returns", *Economic Journal* 101, 157-179.

Campbell, J.Y., Lo, A. W. and MacKinlay, A. C. (1997) *The Econometrics of Financial Markets*, Princeton University Press.

Fama, E.F. and French, K.R. (1988): "Permanent and Temporary Components of Stock Prices", *Journal of Political Economy* 96, 246-273.

Fiorentini, G. and Sentana, E. (1998) "Conditional Means of Time Series Processes and Time Series Processes for Conditional Means", *International Economic Review* 39, 1101-1118.

Harvey, A.C. (1989): *Forecasting, Structural Time Series Models and the Kalman Filter*, Cambridge University Press.

de Jong, P. and N. Shephard (1995) "The Simulation Smoother for Time Series Models", *Biometrika* 82, 339-350.

Koopman, S.J., Shephard, N. and Doornik, J.A. (1999) "Statistical Algorithms for Models in State Space using SsfPack 2.2", *Econometrics Journal* 2, 107-160.

Nijman, T. and Sentana, E. (1996): "Marginalization and contemporaneous aggregation of multivariate GARCH processes", *Journal of Econometrics* 71, 71-87.

Priestley, M.B. (1981): *Spectral Analysis and Time Series* vol. I and II, Academic Press.

Poirier, D.J. (1995) *Intermediate Statistics and Econometrics. A Comparative Approach*, MIT Press.

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.

Appendix

Proof of Proposition 1

Let $\mu_{1\tau+1|\infty} = E(\mu_{1\tau+1}|X_{1\infty}) = \sum_{j=-\infty}^{\infty} \delta_j x_{1t-j}$ be the best (in the mean square error sense) "two-sided" estimate of $\mu_{1\tau+1}$ given a complete realisation of x_{1t}. It is well known (see e.g. Priestley, 1981, section 10.3) that the corresponding "residual variance" is given by

$$\int_{-\pi}^{\pi} \left(h_{\mu\mu}(\omega) - \frac{|h_{\mu x}(\omega)|^2}{h_{xx}(\omega)} \right) d\omega$$

where $H(\omega)$ is the joint spectral density matrix of μ_1 and x_1. Since the integrand is always non-negative, the integral will be 0 if and only if $|h_{\mu x}(\omega)|^2 = h_{\mu\mu}(\omega) \cdot h_{xx}(\omega)$ for all ω, as required. □

MA(1) Example

If we start the smoothing recursions from the steady-state matrices, we obtain

$$P^{**}_{t+1|t} = P = \sigma^2 \begin{pmatrix} 0 & 0 \\ 0 & \beta^2 - 1 \end{pmatrix}$$

$$K^*_{t|t-1} = K = \begin{pmatrix} 1 \\ \beta^{-1} \end{pmatrix}$$

$$K_{t|t-1} = \begin{pmatrix} \beta^{-1} \\ 0 \end{pmatrix}$$

and

$$P^*_{t|t-1} = P_{t|t-1} = \sigma^2 \begin{pmatrix} \beta^2 & \beta \\ \beta & \beta^2 \end{pmatrix}$$

Consequently

$$P^*_{t+1|t} = \sigma^2 \begin{pmatrix} \beta^2 & \beta \\ \beta & \beta^2 \end{pmatrix} \left[\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} - \begin{pmatrix} \beta^{-1} \\ 0 \end{pmatrix} (1,0) \right]' = \sigma^2 \begin{pmatrix} 0 & 0 \\ \beta^2 - 1 & 0 \end{pmatrix}$$

$$K^*_{t+1|t} = P^*_{t+1|t} \begin{pmatrix} 1 \\ 0 \end{pmatrix} \left[(1,0)\sigma^2 \begin{pmatrix} \beta^2 & \beta \\ \beta & \beta^2 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} \right]^{-1} = \begin{pmatrix} 0 \\ \left(\beta^2 - 1\right)\beta^{-2} \end{pmatrix}$$

so that

$$P^{**}_{t+2|t+1} = P^{**}_{t+1|t} - P^*_{t+1|t} \begin{pmatrix} 1 \\ 0 \end{pmatrix} \left(0, \left(\beta^2 - 1\right)\beta^{-2}\right) = \sigma^2 \begin{pmatrix} 0 & 0 \\ 0 & \left(\beta^2 - 1\right)\beta^{-2} \end{pmatrix}$$

Furthermore,

$$P^*_{t+2|t+1} = \sigma^2 \begin{pmatrix} 0 & 0 \\ \left(\beta^2 - 1\right)(-\beta)^{-1} & 0 \end{pmatrix}$$

$$K^*_{t+2|t+1} = P^*_{t+2|t+1} \begin{pmatrix} 1 \\ 0 \end{pmatrix} \left(\sigma^2\beta^2\right)^{-1} = \begin{pmatrix} 0 \\ \left(\beta^2 - 1\right)(-\beta)^{-3} \end{pmatrix}$$

and consequently

$$P^{**}_{t+3|t+2} = \sigma^2 \begin{pmatrix} 0 & 0 \\ 0 & \left(\beta^2 - 1\right)\beta^{-4} \end{pmatrix}$$

In fact, we can prove by induction that

$$P^*_{t+k+1|t+k} = \sigma^2 \begin{pmatrix} 0 & 0 \\ \left(\beta^2 - 1\right)(-\beta)^{-k} & 0 \end{pmatrix}$$

and

$$P^{**}_{t+k+2|t+k+1} = \sigma^2 \begin{pmatrix} 0 & 0 \\ 0 & \left(\beta^2 - 1\right)\beta^{-2(k+1)} \end{pmatrix}$$

The unconditional variance of a VAR(1)

It is well known that the unconditional covariance matrix of a VAR(1), Γ say, satisfies the equation

$$\Gamma = A\Gamma A' + \Sigma$$

In our case, in particular, this yields

$$\begin{aligned}
\begin{pmatrix} \gamma_{rr} & \gamma_{r\mu} \\ \gamma_{r\mu} & \gamma_{\mu\mu} \end{pmatrix} &= \begin{pmatrix} 0 & 1 \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} \gamma_{rr} & \gamma_{r\mu} \\ \gamma_{r\mu} & \gamma_{\mu\mu} \end{pmatrix} \begin{pmatrix} 0 & a_{21} \\ 1 & a_{22} \end{pmatrix} + \begin{pmatrix} \sigma_u^2 & \sigma_{uw} \\ \sigma_{uw} & \sigma_w^2 \end{pmatrix} \\
&= \begin{pmatrix} \gamma_{\mu\mu} & a_{21}\gamma_{r\mu} + a_{22}\gamma_{\mu\mu} \\ a_{21}\gamma_{r\mu} + a_{22}\gamma_{\mu\mu} & a_{21}^2\gamma_{rr} + 2a_{21}a_{22}\gamma_{r\mu} + a_{22}^2\gamma_{\mu\mu} \end{pmatrix} + \begin{pmatrix} \sigma_u^2 & \sigma_{uw} \\ \sigma_{uw} & \sigma_w^2 \end{pmatrix}
\end{aligned}$$

If we solve the system of three linear equations, we finally get:

$$\begin{aligned}
\gamma_{rr} &= \gamma_{\mu\mu} + \sigma_u^2 \\
\gamma_{r\mu} &= \frac{a_{22}\gamma_{\mu\mu} + \sigma_{uw}}{1 - a_{12}} \\
\gamma_{\mu\mu} &= \frac{(1 - a_{21})\left(a_{21}^2\sigma_u^2 + \sigma_w^2\right) + 2a_{21}a_{22}\sigma_{uw}}{(1 + a_{21})\left(1 - a_{21} + a_{22}\right)\left(1 - a_{21} - a_{22}\right)}
\end{aligned}$$

Smoothing and maximum gains in a bivariate VAR(1)

In terms of smoothing, if we start from the steady state matrices, we get:

$$P_{t+1|t}^{**} = P = \begin{pmatrix} 0 & 0 \\ 0 & p_3 \end{pmatrix}$$

$$K_{t|t-1}^{*} = K = \begin{pmatrix} 1 \\ \frac{a_{22}p_3 + \sigma_{uw}}{p_3 + \sigma_u^2} \end{pmatrix}$$

$$K_{t|t-1} = \begin{pmatrix} \frac{a_{22}p_3 + \sigma_{uw}}{p_3 + \sigma_u^2} \\ \frac{a_{21}p_3 + a_{21}\sigma_u^2 + a_{22}^2 p_3 + a_{22}\sigma_{uw}}{p_3 + \sigma_u^2} \end{pmatrix}$$

and

$$P^*_{t|t-1} = P_{t|t-1} = \begin{pmatrix} p_3 + \sigma_u^2 & a_{22}p_3 + \sigma_{uw} \\ a_{22}p_3 + \sigma_{uw} & a_{22}^2 p_3 + \sigma_w^2 \end{pmatrix}$$

Consequently

$$P^*_{t+1|t} = \begin{pmatrix} 0 & 0 \\ p_3 & a_{22}p_3 \end{pmatrix}$$

by virtue of equation (9) and

$$K^*_{t+1|t} = P^*_{t+1|t} \begin{pmatrix} 1 \\ 0 \end{pmatrix} \left[p_3 + \sigma_u^2\right]^{-1} = \begin{pmatrix} 0 \\ \frac{p_3}{p_3+\sigma_u^2} \end{pmatrix}$$

Thus

$$P^{**}_{t+2|t+1} = \begin{pmatrix} 0 & 0 \\ 0 & p_3\frac{\sigma_u^2}{p_3+\sigma_u^2} \end{pmatrix}$$

Now

$$P^*_{t+2|t+1} = P^*_{t+1|t} \left[\begin{pmatrix} 0 & 1 \\ a_{21} & a_{22} \end{pmatrix} - \begin{pmatrix} \frac{a_{22}p_3+\sigma_{uw}}{p_3+\sigma_u^2} \\ \frac{a_{21}p_3+a_{21}\sigma_u^2+a_{22}^2p_3+a_{22}\sigma_{uw}}{p_3+\sigma_u^2} \end{pmatrix} (1,0) \right]'$$

$$= \begin{pmatrix} 0 & 0 \\ p_3 & a_{22}p_3 \end{pmatrix} \begin{pmatrix} -\frac{a_{22}p_3+\sigma_{uw}}{p_3+\sigma_u^2} & a_{21} - \frac{a_{21}p_3+a_{21}\sigma_u^2+a_{22}^2p_3+a_{22}\sigma_{uw}}{p_3+\sigma_u^2} \\ 1 & a_{22} \end{pmatrix}$$

$$= \begin{pmatrix} 0 & 0 \\ a_{22}p_3 - p_3\frac{a_{22}p_3+\sigma_{uw}}{p_3+\sigma_u^2} & p_3a_{21} - p_3\frac{a_{21}p_3+a_{21}\sigma_u^2+a_{22}^2p_3+a_{22}\sigma_{uw}}{p_3+\sigma_u^2} + a_{22}^2p_3 \end{pmatrix}$$

so

$$P^*_{t+2|t+1} = \begin{pmatrix} 0 & 0 \\ p_3\frac{a_{22}\sigma_u^2-\sigma_{uw}}{p_3+\sigma_u^2} & p_3a_{22}\frac{a_{22}\sigma_u^2-\sigma_{uw}}{p_3+\sigma_u^2} \end{pmatrix}$$

Also

$$K^*_{t+2|t+1} = P^*_{t+2|t+1} \begin{pmatrix} 1 \\ 0 \end{pmatrix} \left[p_3 + \sigma_u^2\right]^{-1} = \begin{pmatrix} 0 \\ p_3\frac{a_{22}\sigma_u^2-\sigma_{uw}}{(p_3+\sigma_u^2)^2} \end{pmatrix}$$

Hence

$$P^{**}_{t+3|t+2} = P^{**}_{t+2|t+1} - P^{*}_{t+2|t+1} \begin{pmatrix} 1 \\ 0 \end{pmatrix} \begin{pmatrix} 0 & p_3 \frac{a_{22}\sigma_u^2 - \sigma_{uw}}{(p_3+\sigma_u^2)^2} \end{pmatrix}$$

$$= \begin{pmatrix} 0 & 0 \\ 0 & p_3 \frac{\sigma_u^2}{p_3+\sigma_u^2} \end{pmatrix} - \begin{pmatrix} 0 & 0 \\ p_3 \frac{a_{22}\sigma_u^2 - \sigma_{uw}}{p_3+\sigma_u^2} & p_3 a_{22} \frac{a_{22}\sigma_u^2 - \sigma_{uw}}{p_3+\sigma_u^2} \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} \begin{pmatrix} 0 & p_3 \frac{a_{22}\sigma_u^2 - \sigma_{uw}}{(p_3+\sigma_u^2)^2} \end{pmatrix}$$

$$= \begin{pmatrix} 0 & 0 \\ 0 & \frac{p_3}{p_3+\sigma_u^2} \left(\sigma_u^2 - p_3 \frac{(a_{22}\sigma_u^2 - \sigma_{uw})^2}{(p_3+\sigma_u^2)^2} \right) \end{pmatrix}$$

so

$$P^{**}_{t+3|t+2} = \begin{pmatrix} 0 & 0 \\ 0 & p_3 \sigma_u^2 \frac{p_3(\sigma_u^2+\sigma_w^2)+\sigma_u^2+\sigma_w^2\sigma_u^2-\sigma_{uw}^2}{(p_3+\sigma_u^2)^3} \end{pmatrix}$$

Given that the model is:

$$\begin{pmatrix} r_t \\ \mu_{rt+1} \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} r_{t-1} \\ \mu_{rt} \end{pmatrix} + \begin{pmatrix} u_t \\ w_t \end{pmatrix}$$

where $E \begin{pmatrix} u_t \\ w_t \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$ and $V \begin{pmatrix} u_t \\ w_t \end{pmatrix} = \Sigma$, we can write it in a compact form as:

$$\mathbf{x}_t = \mathbf{A}\mathbf{x}_{t-1} + \boldsymbol{\eta}_t$$

Hence the spectral matrix $H_{\mathbf{x}_t}(\omega)$ of the $\mathbf{x}_t$ process will be given by:

$$H_{\mathbf{x}_t}(\omega) = \frac{1}{2\pi} \boldsymbol{\alpha}^{-1}(e^{-i\omega}) \Sigma \left\{ \boldsymbol{\alpha}^{-1}(e^{-i\omega}) \right\}^*$$

where $\boldsymbol{\alpha}^{-1}(z) = \{I - \mathbf{A}z\}^{-1}$, and an asterisk denotes both conjugation and transposition.

Then, we have that:

$$\boldsymbol{\alpha}^{-1}(z) = \frac{1}{1 - za_{22} - z^2 a_{21}} \begin{pmatrix} 1 - za_{22} & z \\ za_{21} & 1 \end{pmatrix}$$

so that

$$H_{\mathbf{x}_t}(\omega) = \begin{pmatrix} h_{rr}(\omega) & h_{r\mu}(\omega) \\ h_{\mu r}(\omega) & h_{\mu\mu}(\omega) \end{pmatrix}$$

where

$$h_{rr} = c\{\sigma_w^2 + (1+a_{22}^2)\sigma_u^2 - 2a_{22}\sigma_{uw} + \sigma_{uw}(e^{i\omega}+e^{-i\omega}) - a_{22}\sigma_u^2(e^{i\omega}+e^{-i\omega})\}$$

$$h_{r\mu} = c\{\sigma_u^2 a_{21}e^{i\omega} - a_{22}a_{21}\sigma_u^2 + (a_{21}+1)\sigma_{uw} - a_{22}\sigma_{uw}e^{-i\omega} + \sigma_w^2 e^{-i\omega}\}$$

$$h_{\mu\mu} = c\{\sigma_u^2 a_{21}^2 + \sigma_{uw}a_{21}(e^{i\omega}+e^{-i\omega}) + \sigma_w^2\}$$

with

$$c = \frac{1}{2\pi}\frac{1}{\left|1 - e^{-i\omega}a_{22} - e^{-2i\omega}a_{21}\right|^2}$$

The residual variance bound is given by

$$\int_{-\pi}^{\pi}\left(h_{\mu\mu}(\omega) - \frac{|h_{\mu r}(\omega)|^2}{h_{rr}(\omega)}\right)d\omega$$

(see e.g. Priestley, 1981, section 10.3), which after tedious operations can be shown to be

$$\begin{aligned}\int_{-\pi}^{\pi}\Bigg[&\frac{1}{2\pi}\frac{1}{\left|1 - e^{-i\omega}a_{22} - e^{-2i\omega}a_{21}\right|^2}(\sigma_w^2\sigma_u^2 - \sigma_{uw}^2)\\ &\times\{(a_{21}^2 + a_{22}^2 + 1) + (a_{21}-1)a_{22}(e^{i\omega}+e^{-i\omega}) - a_{21}(e^{2i\omega}+e^{-2i\omega})\}\\ &\times\{\sigma_w^2 + (1+a_{22}^2)\sigma_u^2 - 2a_{22}\sigma_{uw} + (\sigma_{uw} - a_{22}\sigma_u^2)(e^{i\omega}+e^{-i\omega})\}^{-1}\Bigg]\,d\omega\end{aligned}$$

But since

$$\left|1 - e^{-i\omega}a_{22} - e^{-2i\omega}a_{21}\right|^2 = 1 + a_{22}^2 + a_{21}^2 + (a_{21}-1)a_{22}(e^{-i\omega}+e^{i\omega}) - a_{21}(e^{-2i\omega}+e^{2i\omega})$$

we have that

$$\int_{-\pi}^{\pi}\left(h_{\mu\mu}(\omega) - \frac{|h_{\mu r}(\omega)|^2}{h_{rr}(\omega)}\right)d\omega = \frac{(\sigma_w^2\sigma_u^2 - \sigma_{uw}^2)}{2\pi}\int_{-\pi}^{\pi}\{\gamma + \delta(e^{i\omega}+e^{-i\omega})\}^{-1}d\omega$$

where

$$\gamma = \sigma_w^2 + (1 + a_{22}^2)\sigma_u^2 - 2a_{22}\sigma_{uw}$$

and

$$\delta = (\sigma_{uw} - a_{22}\sigma_u^2)$$

After solving the integral, we finally obtain

$$\int_{-\pi}^{\pi} \left[h_{\mu\mu}(\omega) - \frac{|h_{\mu r}(\omega)|^2}{h_{rr}(\omega)} \right] d\omega$$

$$= \frac{\sigma_w^2\sigma_u^2 - \sigma_{uw}^2}{\sqrt{\sigma_w^2(1-\rho^2) + [\rho\sigma_w + (1-a_{22})\sigma_u]^2}\sqrt{\sigma_w^2(1-\rho^2) + [\rho\sigma_w - (1+a_{22})\sigma_u]^2}}$$

The Simulation Smoother

Given that the simulation smoother is a backward recursion which requires the Kalman filter output, it is convenient to rewrite the state-space in equations (7) and (8) with:

$$r_t = Z \begin{pmatrix} r_{t-1} \\ \mu_{rt} \end{pmatrix} + G \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix}$$

as measurement equation, and

$$\alpha_{t+1} = \begin{pmatrix} r_t \\ \mu_{rt+1} \end{pmatrix} = \begin{pmatrix} 0 \\ c \end{pmatrix} + T\alpha_t + H \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix}$$

as transition equation, where $\varepsilon_t \backsim N(0, I_2)$, $Z = \begin{pmatrix} 0 & 1 \end{pmatrix}$, $T = \begin{pmatrix} 0 & 1 \\ a_{21} & a_{22} \end{pmatrix}$, $G = \begin{pmatrix} \sigma_u & 0 \end{pmatrix}$ and $H = \begin{pmatrix} \sigma_u & 0 \\ \frac{\sigma_{uw}}{\sigma_u} & \sqrt{\frac{\sigma_w^2\sigma_u^2 - \sigma_{uw}^2}{\sigma_u^2}} \end{pmatrix}$ is the Choleski decomposition of Σ.

As initial condition for the state we can either chose $\alpha_1 \backsim N(c^*, P_1)$ where

$$\begin{aligned} c^* &= \frac{1}{1-a_{21}-a_{22}} \begin{pmatrix} 0 \\ c \end{pmatrix} \\ P_1^{11} &= P_1^{22} + \sigma_u^2 \\ P_1^{12} &= \frac{a_{22} P_1^{22} + \sigma_{uw}}{1-a_{21}} \\ P_1^{22} &= \frac{(1-a_{21})\left(a_{21}^2 \sigma_u^2 + \sigma_w^2\right) + 2a_{21}a_{22}\sigma_{uw}}{(1-a_{21})^2 (1-a_{21}+a_{22})(1-a_{21}-a_{22})} \end{aligned}$$

are the values obtained from the stationary distribution, or a diffuse prior, in which case $\alpha_1 \backsim N(c^*, P_1)$, with $c^* = \begin{pmatrix} 0 \\ c \end{pmatrix}$ and $P_1 = \kappa I_2$, where κ is an arbitrary large number, such as 10^7.

For $t > 1$ we have the following Kalman filter equations:

$$v_t = r_t - a_t^{(2)}, \quad F_t = P_t^{22} + \sigma_u^2,$$

$$K_t = \begin{pmatrix} 1 \\ \frac{a_{22} P_t^{22} + \sigma_{uw}}{F_t} \end{pmatrix}$$

$$a_{t+1} = \begin{pmatrix} r_t \\ c + a_{21} r_{t-1} + a_{22} a_t^{(2)} + K_t^{(2)} v_t \end{pmatrix}$$

$$P_{t+1} = T P_t T^{/} + H H^{/} - K_t F_t K_t^{/} = \begin{pmatrix} 0 & 0 \\ 0 & P_{t+1}^{22} \end{pmatrix}$$

where $P_{t+1}^{22} = a_{22}^2 P_t^{22} + \sigma_w^2 - \left(K_t^{(2)}\right)^2 F_t$. Hence, all elements of the variance matrices P_{t+1} $(t = 1, .., T)$ are zero except the (2,2) one, which simply reflects the fact that the first element of the state vector is observed. The quantities stored are v_t, K_t, and F_t.

Following de Jong and Shephard (1995), the multi-stage simulation smoother starts by setting $N_T = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$, $y_T = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$ and forms

$$e_T = F_T^{-1} v_T - y_T^{(1)} - K_T^{(2)} y_T^{(2)},$$

$$C_T = \Gamma^*(\Omega - \Omega N_T^* \Omega)\Gamma^{*\prime}$$

and

$$\begin{aligned} N_T^* &= \begin{pmatrix} N_T & -N_T K_T \\ -K_T^{/} N_T & D_T \end{pmatrix} \\ y_t^* &= \begin{pmatrix} y_t \\ e_t \end{pmatrix} \end{aligned}$$

where $D_T = F_T^{-1} + K_T' N_T K_T$, $\Gamma^* = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$, so that $(\Gamma^*)' \Gamma^* = \Gamma$, where $\Gamma = diag(1,1,0)$ is the selection matrix, and $\Omega = \begin{pmatrix} H \\ G \end{pmatrix} \begin{pmatrix} H' & G' \end{pmatrix} = \begin{pmatrix} \Sigma & HG' \\ GH' & GG' \end{pmatrix}$.

Now for $t = T, T-1, ..., 1$ the algorithm updates

$$N_{T-1} = \Phi'(N_T^* + \Delta_T' \Delta_T)\Phi \quad and \quad y_{t-1} = \Phi'(y_t^* - \Delta_t' \pi_t) \tag{17}$$

where the random vector π_t is drawn from $N(0, I_2)$ and Δ_T is the solution of the linear system of equations

$$B_T * \Delta_T = \Gamma^* \Omega N_T^* \tag{18}$$

where B_T is the Choleski decomposition of C_T, i.e. $C_T = B_T B_T'$. In this way, we avoid the inversion of C_T (see Koopman et al (1999)). The weight-matrices Δ_t and B_t are kept for the state simulation.

Now notice that C_t is singular for $t = T-1, ..., 1$, although $\Gamma^* \Omega \Gamma^{*\prime} = \Sigma$ is non-singular since $rank(\Gamma^* \Omega \Gamma^{*\prime}) = 2$ (see de Jong and Shephard (1995) and Koopman et al (1999)). Again the singularity of C_t, for all $t's$ except $t = T$, is a direct consequence of the first element of the bivariate state vector being observed. This singularity makes necessary the adjustment of the formulae in Koopman et al (1999). As a result, the Choleski decomposition of $C_t = B_t B_t'$ yields $B_t = \begin{pmatrix} b_t^{11} & 0 \\ b_t^{21} & 0 \end{pmatrix}$, whose Moore-Penrose (MP) inverse is $B_t^+ = \begin{pmatrix} b_t^{+11} & b_t^{+12} \\ 0 & 0 \end{pmatrix}$.

Hence, a convenient solution of (18) is:

$$\Delta_t = B_t^{+}\Gamma^{*}\Omega N_t^{*} = \begin{pmatrix} 0 & \alpha_t^{12} & \alpha_t^{13} \\ 0 & 0 & 0 \end{pmatrix} \quad for \quad t = T-1, ..., 1$$

Now for $t = T, ..., 1$ the updating of N_{t-1} is given by

$$N_{t-1} = \Phi' N_t^{*} \Phi + \Phi' N_t^{*} \Omega (\Gamma^{*})' C_t^{+} \Gamma^{*} \Omega N_t^{*} \Phi$$

where C_t^{+} the MP inverse of C_t, as opposed to equation (17) above. In this way, we avoid storing the matrix Δ_t, but in contrast, we still have to find the MP inverse of the matrix C_t. Nevertheless, if the dimension of C_t is relatively small, the numerical cost is very low.

Following de Jong and Shephard (1995), the simulated state will then be given by:

$$\alpha s_1 = c^{*} + P_1 y_0$$

with updates

$$\alpha s_{t+1} = \begin{pmatrix} 0 \\ c \end{pmatrix} + \begin{pmatrix} 0 & 1 \\ a_{21} & a_{22} \end{pmatrix} \alpha s_t + u_t \qquad (t = 1, 2, .., T)$$

where $u_t = \Sigma y_t + HG' e_t + B_t \pi_t$.

Now $HG' e_t = \begin{pmatrix} \sigma_u^2 e_t \\ \sigma_{uw} e_t \end{pmatrix}$ and $B_t \pi_t = \begin{pmatrix} B_t^{11} \pi_t^{(1)} \\ B_t^{21} \pi_t^{(1)} + B_t^{22} \pi_t^{(2)} \end{pmatrix}$. Finally, we can form

$$u_t = \begin{pmatrix} \sigma_u^2 y_t^{(1)} + \sigma_{uw} y_t^{(2)} + \sigma_u^2 e_t + B_t^{11} \pi_t^{(1)} \\ \sigma_{uw} y_t^{(1)} + \sigma_w^2 y_t^{(2)} + \sigma_{uw} e_t + B_t^{21} \pi_t^{(1)} + B_t^{22} \pi_t^{(2)} \end{pmatrix}$$

In this respect, note that B_t^{22} is zero for $t \neq T$. Therefore, for these $t's$ the whole system is only driven by the first element of the random vector π_t.

Table 1

Descriptive statistics of the distributions for expected returns (%)

January 1975

	Mean	Standard Deviation	First Quartile	Median	Third Quartile	Interquartile Range	Inefficiency Ratio
Predicted	1.3761	2.3893	-.2202	1.4132	3.0711	3.2913	1.6772
Updated	2.8511	2.0350	1.4957	2.9924	4.2560	2.7603	1.4780
Smoothed	3.6087	0.7915	3.0824	3.5732	4.1158	1.0334	1.3522

October 1987

	Mean	Standard Deviation	First Quartile	Median	Third Quartile	Interquartile Range	Inefficiency Ratio
Predicted	0.2453	1.7443	-0.9560	0.1621	1.3046	2.2606	1.1740
Updated	-2.8175	1.5385	-3.8244	-2.7646	-1.7282	2.0962	1.1307
Smoothed	-2.2571	0.7280	-2.7565	-2.2724	-1.7963	0.9602	1.2002

Figure 1: Concentrated Log-Likelihood Function for an ARMA(1,1) Model for Real Returns

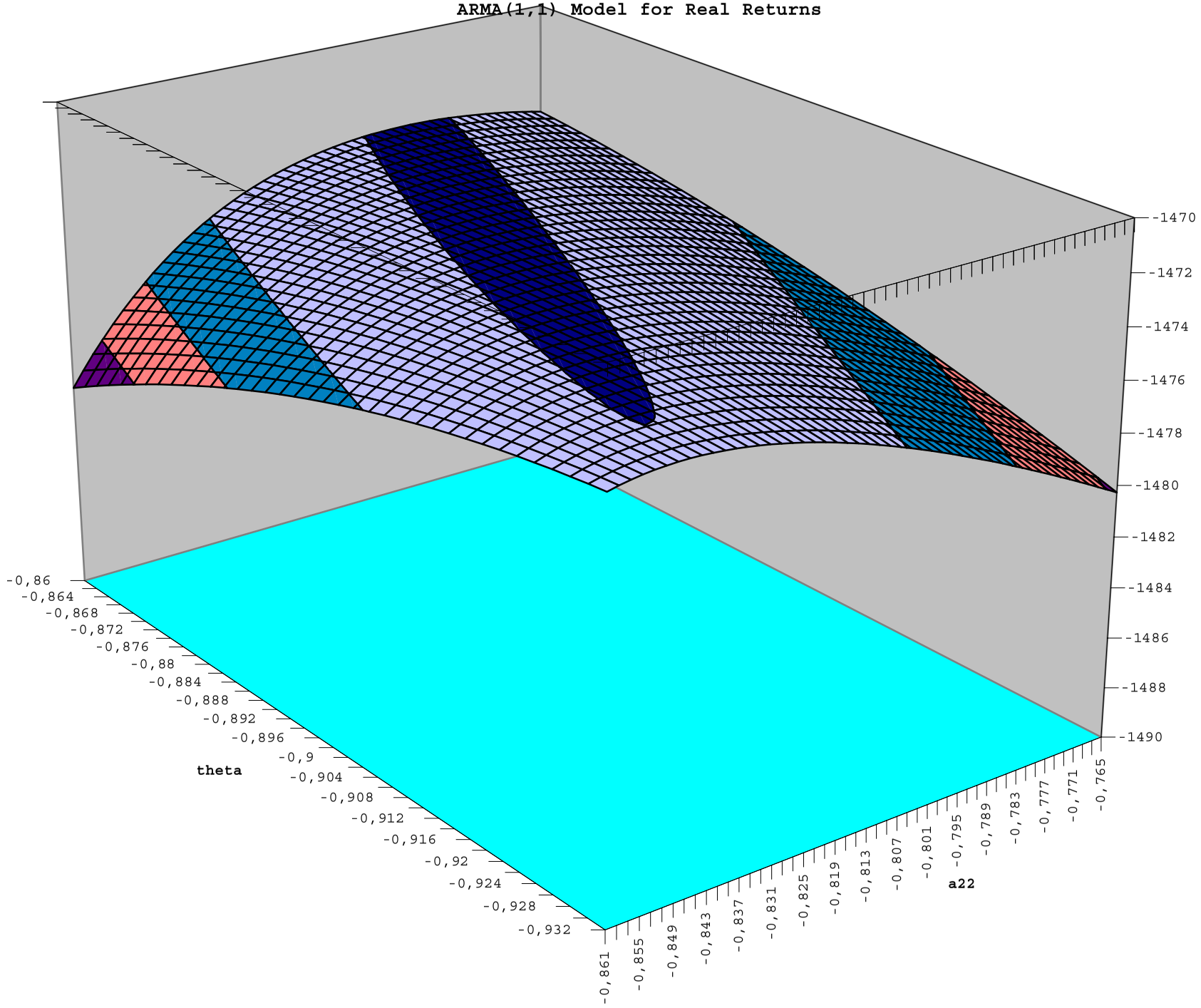

Figure 2: US Monthly Real Returns (%)

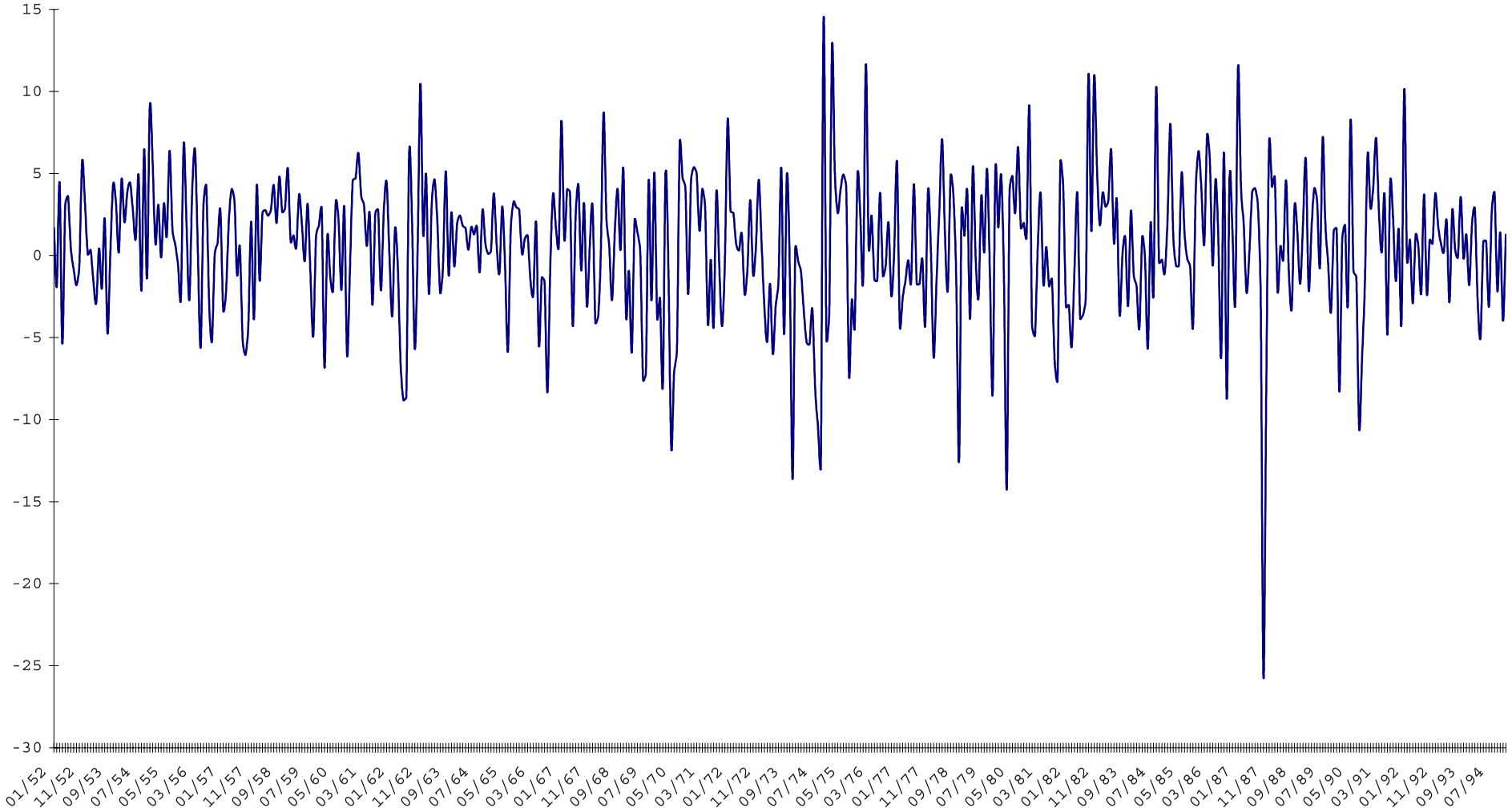

Figure 3: US Monthly Expected Returns (%)

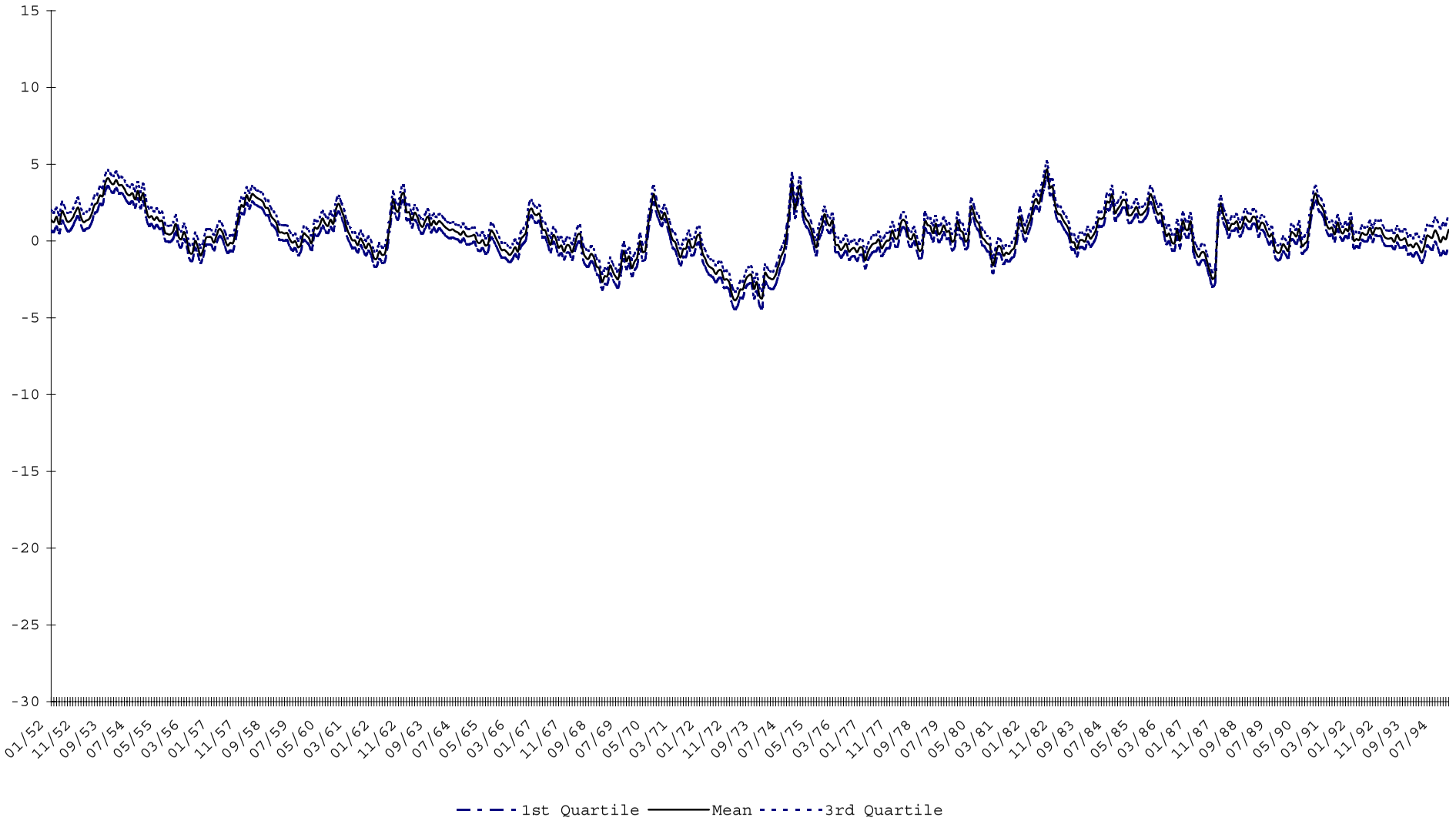

Figure 4: Estimated Densities for Expected Returns (%)
January 1975

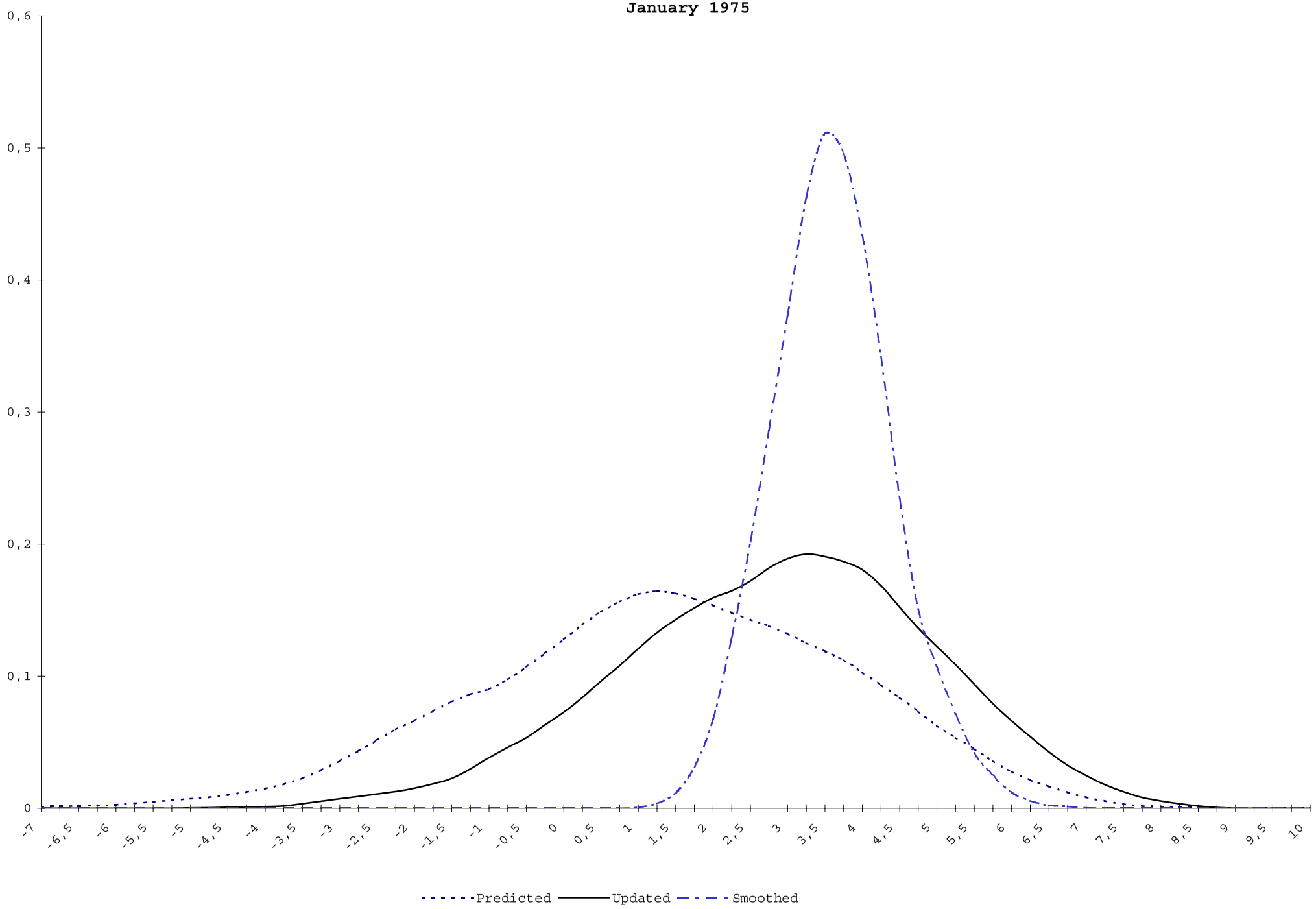

Figure 5: Estimated Densities for Expected Returns (%)
October 1987