BVARTEC - BAYESIAN VECTOR AUTO REGRESSIONS WITH TIME VARYING ERROR-COVARIANCES*

Harald Uhlig
Princeton University

Econometric Research Program
Research Memorandum No. 366

August 1991
Revised July 1992

* I am indebted to C. A. Sims. I am also grateful to Dan Nelson for a helpful conversation and Gregory Chow, Robert Engle, Robin Lumsdaine and Pierre Perron, Neil Shaphard, Arnold Zellner for useful comments on an earlier version of the paper.

Econometric Research Program
Princeton University
203 Fisher Hall
Princeton, NJ 08544-1021, USA

Abstract

This paper introduces Bayesian vector autoregressions with time-varying error covariances (BVARTEC). The method uses an unobserved components version of modelling ARCHness, where the shock to the error-precision matrix is drawn from a multivariate beta distribution. Extending the standard Kalman Filter analysis to the non-linear filtering problem at hand, simple closed-form exact updating formulas for the posterior are derived. These formulas turn out to be similar to the formulas for a multivariate IGARCH(1,1) specification without an intercept. The method is easy to use and particularly suitable for applications in macroeconomics and finance.

1 Introduction

This paper introduces Bayesian vector autoregressions with time-varying error covariances (BVARTEC). The idea in this paper is to extend Bayesian vector autoregressions by allowing for time-variation in the error variance-covariance matrix. Rather than using a multivariate version of the popular ARCH specification, we model the error precision matrix as an unobserved component with idiosynchratic shocks drawn from a multivariate beta distribution. Such a setup allows to interpret a sudden large movement in the data as the result of a draw from a distribution with a randomly increased but unobserved variance, rather than the traditional ARCH-interpretation of an unusually large draw from a distribution with a presently small variance.

The calculations are analytically tractable, exploiting a conjugacy between Wishart distributions and beta distributions, which has recently been established in Uhlig (1992c) for the singular beta distributions needed in this paper. Building on and extending the usual Kalman-Filter analysis to the nonlinear filtering problem at hand, closed-form, exact updating formulas for the posterior are obtained, providing an easy-to-use generalization of Bayesian vector autoregressions as well as an useful analytic benchmark against which other Bayesian versions of ARCHness can be compared. The updating formulas turn out to resemble the formulas for a multivariate IGARCH(1,1) specification without intercept, except that the covariances of the one-step ahead forecast errors themselves rather than the true, but unobservable errors are used to compute the prior mean error precision matrix for the innovations of the next step. The speed of time variation is treated as a hyperparameter. The difficulty in determining that parameter by a quasi-maximum posterior procedure are explained. The method is applied to a four-variable system of US time series, consisting of stock prices, federal funds rates, real GNP and a price index. Mean impulse response functions as well as mean forecasts up to 1996 are constructed with error bands reflecting the possible skewness of the posterior distributions.

Bayesian vector autoregressions have been studied and popularized by Litterman (1979,1980,1981,1984a,1984b,1985), Doan Litterman and Sims (1984),

Doan's RATS Manual (1990), Sims (1986,1987,1989) and Todd (1988), see also the reviews in Lütkepohl (1990) and Canova (1991), and have recently been reexamined in slightly different form in the context of the unit root debate, see Koop (1990), DeJong (1992) and Kleibergen and Van Dijk (1992). ARCH models have been introduced by Engle (1982) and have generated a huge literature, see the review in Bollerslev, Chou and Kroner (1992). Rather than build on multivariate ARCH-models as in Engle (1987b), Baba, Engle, Kraft and Kroner (1987), Baillie and Bollerslev (1990), Bollerslev, Engle and Wooldridge (1988), Diebold and Nerlove (1990), Engel and Rodrigues (1989), Engle, Ito and Lin (1990), Engle, Ng and Rothschild (1990b) and Lin, Engle and Ito (1991), we proceed similarly to the literature on stochastic variance as in Harvey, Ruiz and Shephard (1992), Harvey and Shephard (1992), Watanabe (1992) and in particular Shephard (1990), who uses a setup like the one proposed here in a univariate, non-Bayesian context. In contrast to previous attempts at building ARCHness into Bayesian vectorautoregressions, see Geweke (1988,1989), Sims (1989) and West and Harrison (1989), chapter 10.8, our method results in exact closed-form updating formulas for the posterior.

Section 2 derives some intuition in a simple univariate case. Section 3 introduces the multivariate model and the exact updating formulas. Section 4 suggests and discusses reasonable priors. Section 5 discusses a method for determining the hyperparameter governing the degree of time variation. Section 6 derives interesting variations, including a method for tracking time-varying covariances only suitable for applications in finance, and compares the general method to standard BVAR methodology. Section 8 Section 9 concludes. The appendix A lists some of the distributions and their properties, which are used in the main body of the text. Appendix B contains the proofs for all theorems.

2 A Simple Case

Consider the following simple version of the model studied in this paper:

$$y_t = \beta y_{t-1} + \epsilon_t, \qquad (1)$$

$$\epsilon_t \sim \mathcal{N}\left(0, \sigma_t^2\right), \tag{2}$$

$$\sigma_{t+1}^{-2} = \psi \sigma_t^{-2} u_t,\ u_t \sim \mathcal{B}_1\left(\nu/2, 1/2\right), \tag{3}$$

with the u_t's and ϵ_t's drawn independently, where $t = 1, \ldots, T$ denotes time, $y_t \in \mathbf{R}, t = 0, \ldots, T$ is data and observable, $\psi > 0$, $\nu > 0$ are parameters and $\mathcal{B}_1(p, q)$ denotes the (one-dimensional) beta-distribution on the interval [0,1].

Equation (3) specifies the unobserved precision σ_t^{-2} to be time-varying. The model captures autocorellated heteroskedasticity, a feature often found especially in financial data series. Another popular specification to do so are the ARCH-models. An IGARCH(1,1) model, for example, replaces (3) with

$$\sigma_{t+1}^2 = \iota + (1 - \lambda)\sigma_t^2 + \lambda \epsilon_t^2, \tag{4}$$

where ι and λ are parameters, and thus ties the innovation in the variance to the size of the current shock ϵ_t. While an unusually large ϵ_t in (3) is interpreted as a large draw from a distribution with a randomly increased σ_t, the more traditional IGARCH-interpretation according to (4) just interprets it as an unusually large draw. In both specifications will one expect increased variance later on. In other words, both models are capable of explaining persistent high or low volatility, but the unobserved components specification (3) has a somewhat better chance explaining sudden large movements in the data without resorting to fat tails (see e.g. the discussion of the stock market crash in Nelson (1991)).

Nonetheless, there is a tight connection between these two specifications. Adapting the Bayesian updating formulas (16), (17) and (18) derived below in section 3 to the simple model considered above with $\psi = (\nu + 1)/\nu$ results in

$$n_t = n_{t-1} + y_{t-1}^2 \tag{5}$$

$$\bar{b}_t = \left(\bar{b}_{t-1} n_{t-1} + y_t y_{t-1}\right)/n_t \tag{6}$$

$$s_t^2 = (1 - \lambda) s_{t-1}^2 + \lambda e_t^2 \left(1 - y_{t-1}^2/n_t\right), \tag{7}$$

where $\lambda = 1/(\nu + 1)$ and

$$e_t = y_t - \bar{b}_{t-1} y_{t-1}, \tag{8}$$

delivering the Normal-Gamma[1] posterior $f_{NG}(\beta, \sigma_{t+1}^{-2} \mid \bar{b}_T, n_T, s_T^2, \nu)$ for the Normal-Gamma prior $f_{NG}(\beta, \sigma_1^{-2} \mid \bar{b}_0, n_0, s_0^2, \nu)$, $\bar{b}_0 \in \mathbf{R}$, $n_0, s_0^2 > 0$ (the hyperparameter ν is treated as fixed in the computation of the posterior). The updating equation (7) for the inverse mean precision s_t^2 turns out to resemble equation (4) with the one-step ahead prediction error e_t replacing the true innovation ϵ_t, with a degrees-of-freedom correction and with $\iota = 0$.

For $\psi = (\nu+1)/\nu$, the precision σ_t^{-2} is a martingale $E[\sigma_{t+1}^{-2}] = E[\sigma_t^{-2}]$ on the positive part of the real axis, since $E[u_t] = \nu/(\nu+1)$. As Shephard (1990) remarked, it follows that $\sigma_t^{-2} \to 0\, a.s.$ for similar reasons as in Nelson (1990). Shephard, using essentially the same model, therefore suggests to set $\psi = e^r$, where

$$r = -E[\log u_t]. \tag{9}$$

As a result, $\log(\sigma_t)$ becomes a random walk and the "Nelson-Problem" is removed. In the context of the Bayesian analysis performed here, we will not use this modification, however, for the following robustness reason. Modifying (7) to allow for arbitrary ψ results in

$$s_t^2 = \frac{\nu+1}{\psi\nu}\left((1-\lambda)s_{t-1}^2 + \lambda e_t^2\left(1 - y_{t-1}^2/n_t\right)\right). \tag{10}$$

Consider now the possibility that (3) is misspecified and that σ_t^{-2} (as well as y_t) is actually stationary with $\bar{s}^2 = E[\epsilon_t^2]$ unconditionally. For large t, y_{t-1}/n_t becomes negligble and $\bar{b}_t \to \beta$, so that upon taking expectations on both sides of (10), one gets approximately

$$E[s_t^2] = \frac{\nu+1}{\psi\nu}\left((1-\lambda)E[s_{t-1}^2] + \lambda\bar{s}^2\right). \tag{11}$$

For $\psi = (\nu+1)/\nu$ and $\lambda \neq 0$, s_t^2 will therefore fluctuate around its true, stationary value $\bar{s}^2$, whereas for any other value of ψ, s_t^2 will either diverge or converge to zero. From the perspective of robustness of the Bayesian estimation focussed on in this paper, it is therefore sensible and desirable to choose $\psi = (\nu+1)/\nu$ rather than $\psi = e^r$ for our purposes here despite the "Nelson-Problem". If forward simulation of long time series is the issue,

[1]See Appendix A for the definition of a Normal-Gamma distribution.

however, Shephards (1990) specification with $\psi = e^r$ is superior to the one chosen here.

Figure 1 shows the densities for the multiplicative disturbance $((\nu + 1)/\nu)u_t$ for various choices of ν for the precision σ_t^{-2}, whereas figure 2 plots the corresponding densities for $\nu/(\nu+1)1/u_t$, which are the multiplicative disturbances to the variance σ_t^2. Figure 3 and 4 compare our BVARTEC model to the corresponding IGARCH specification: figure 3 uses $\nu = 9$ and the corresponding $\lambda = .1$, whereas figure 4 uses $\nu = 19$ and $\lambda = 0.05$. The top row in each of these figures shows a realization of length 200 for the standard deviations σ_t arising from equation (3) with $\psi = (\nu+1)/\nu$ as well as a time series of shocks $\epsilon_t \sim \mathcal{N}(0, \sigma_t^2)$. The bottom row uses the IGARCH specification (4) with $\iota = .001$. In each figure, we have used the same standard normal draws for both specification as the basis to calculate the ϵ_t's. Note in particular that the two time series for the ϵ_t in figure 4 look rather similar since σ_t does not fluctuate a lot (in fact, for $\nu = 99$, the two pictures would look practically identical). Inspection of simulations like these may give some intuition on how suitable these processes are for describing e.g. stock return data.

3 The General Model

Consider the VAR(k)-model with time-varying error precision matrices

$$Y_t = B_{(0)}C_t + B_{(1)}Y_{t-1} + B_{(2)}Y_{t-2} + \ldots + B_{(k)}Y_{t-k} + \epsilon_t, \tag{12}$$

$$\epsilon_t \sim \mathcal{N}\left(0, H_t^{-1}\right), \tag{13}$$

$$H_{t+1} = \frac{\nu+1}{\nu}\mathcal{U}(H_t)'Q_t\mathcal{U}(H_t),\ Q_t \sim \mathcal{B}_m\left(\nu/2, 1/2\right), \tag{14}$$

where $t = 1, \ldots, T$ denotes time, Y_t, $t = 1-k, \ldots, T$, size $m \times 1$, is observable data, C_t, size $c \times 1$, denotes deterministic regressors such as a constant and a time trend, the coefficient matrix $B_{(0)}$ is of size $m \times c$, the coefficient matrices $B_{(i)}$, $i = 1, \ldots, k$ are of size $m \times m^2$, $\nu > m - 1$ is a constant and

[2]Time-varying coefficients were not considered, since they do not seem to improve things much in standard BVARs, see Doan, Litterman and Sims (1984). It is not hard to

all $\epsilon_t, t = 1, \ldots, T$, size $m \times 1$, and $Q_t, t = 1, \ldots, T$, size $m \times m$ are independently distributed. $\mathcal{U}(H)$ denotes the upper Cholesky factor of a positive definite matrix H and $\mathcal{B}_m(p, q)$ denotes the multivariate beta-distribution on the $m(m+1)/2$ distinct parameters of positive definite $m \times m$ matrices. The distribution $\mathcal{B}_m(p, q)$ is traditionally only defined for $p > (m-1)/2$ and $q > (m-1)/2$, see Muirhead (1982). The usual definition is extended in Uhlig (1992c) to allow for values $q = n/2$ for any integer $n \geq 1$, see appendix A for the details pertinent to this paper. The process (14) is described in terms of the precision matrix H_t rather than the more familiar error variance-covariance matrix $\Sigma_t = H_t^{-1}$ to streamline the notation and the analysis. Equation (14) is one of two rather natural generalization of the multiplication of two real numbers in equation (3) in order to guarantee the symmetry of the resulting matrix H_{t+1}. The other natural generalization switches H_t and Q_t in (14): it turns out that only (14) works for the proof of the updating formulas.

Stack the system as follows. Let $X_t = \left[C_t \, Y'_{t-1} \, Y'_{t-2} \, \ldots \, Y'_{t-k} \right]'$ and $B = \left[B_{(0)} \, B_{(1)} \, B_{(2)} \, \ldots \, B_{(k)} \right]$. Rewrite the first equation (12) more concisely as

$$Y_t = BX_t + \epsilon_t \tag{15}$$

Define $\lambda = 1/(\nu + 1)$. Consider the following algorithm, which generalizes the Kalman Filter to the nonlinear filtering problem at hand by including an exact updating formula for the mean precision matrix.

Method 1 (General BVARTEC)

1. *Choose a $m \times (c+km)$-matrix $\bar{B}_0$, a positive definite $(c+km) \times (c+km)$-matrix N_0 and a positive definite $m \times m$-matrix S_0. These parameters fix the prior.*

2. *For each $t = 1, \ldots, T$ calculate*

$$N_t = N_{t-1} + X_t X'_t \tag{16}$$

introduce them, however.

$$\bar{B}_t = \left(\bar{B}_{t-1}N_{t-1} + Y_t X_t'\right) N_t^{-1} \tag{17}$$

$$\begin{aligned} S_t = (1-\lambda)S_{t-1} + \\ \lambda\left(Y_t - \bar{B}_{t-1}X_t\right)\left(1 - X_t'N_t^{-1}X_t\right)\left(Y_t - \bar{B}_{t-1}X_t\right)'. \end{aligned} \tag{18}$$

3. Keep B_T, N_T and S_T as result.

(Note that N_t^{-1} can be computed numerically cheaply via

$$N_t^{-1} = N_{t-1}^{-1} - N_{t-1}^{-1}X_t X_t' N_{t-1}^{-1}/(X_t'N_{t-1}^{-1}X_t + 1), \tag{19}$$

as can be verified directly or with rule (T8), p. 324 in Leamer (1978).) Treating ν as a hyperparameter (see section 5) and taking the initial observations $Y_t, t = 1-k, \ldots, 0$ as given (see the discussion in Uhlig (1992a)), this method finds the Normal-Wishart posterior[3]

$$f_{NW}(B, H_{T+1} \mid \bar{B}_T, N_T, S_T, \nu)$$

in B and H_{T+1} for the Normal-Wishart prior

$$f_{NW}(B, H_1 \mid \bar{B}_0, N_0, S_0, \nu).$$

The proof for this assertion follows directly from the two theorems below with $\mu \equiv 1$. Key to the proof is a conjugacy between the Wishart distribution and the multivariate beta-distribution. This conjugacy, which is well known for beta distributions defined for the usual parameter range, has recently been established as well for the singular beta distributions $\mathcal{B}_m(\nu/2, 1/2)$ needed here in Uhlig (1992c), see appendix A.

Given a prior π_0 with a density of the form $\Phi(B)f_{NW}(B, H_1 \mid \bar{B}_0, N_0, S_0, \nu)$, the posterior is proportional to $\Phi(B)f_{NW}(B, H_{T+1} \mid \bar{B}_T, N_T, S_T, \nu)$, so that treating explosive roots or roots near unity differently from small roots can be done in this framework as well, cmp. Uhlig (1992a).

[3]See Appendix A for the definition of a Normal-Wishart distribution.

Theorem 1 *Let a prior for the $m \times (c + km)$ coefficient matrix B and the $m(m+1)/2$ distinct elements of the precision matrix H be given by a Normal-Wishart distribution with density $f_{NW}(B, H \mid \bar{B}, N, S, \nu)$, where N and S are positive definite. Suppose additionally, that there is one observation of data X and Y (where X is $(c + km) \times 1$ and Y is $m \times 1$), obeying the equation*

$$Y = BX + \epsilon,\ \epsilon \sim \mathcal{N}\left(0, H^{-1}\right) \tag{20}$$

Then the posterior for B and H is given by a Normal-Wishart distribution with density $f_{NW}(B, H \mid \tilde{\bar{B}}, \tilde{N}, \tilde{S}, \tilde{\nu})$, where $\tilde{N}$ and $\tilde{S}$ are positive definite and where

$$\tilde{N} = N + XX' \tag{21}$$

$$\tilde{\bar{B}} = \left(\bar{B}N + YX'\right)\tilde{N}^{-1} \tag{22}$$

$$\tilde{\nu} = \nu + 1 \tag{23}$$

$$\tilde{S} = \frac{\nu}{\nu+1}S + \frac{1}{\nu+1}\left(Y - \bar{B}X\right)\left(1 - X'\tilde{N}^{-1}X\right)\left(Y - \bar{B}X\right)' \tag{24}$$

The proof is in appendix B.

Theorem 2 *Let a prior for the $m \times (c + km)$ coefficient matrix B and the $m(m+1)/2$ distinct elements of the precision matrix H be given by a Normal-Wishart distribution with density $f_{NW}(B, H \mid \bar{B}, N, S, \nu + \mu)$, $\nu > m - 1$, μ a positive integer or $\mu > m - 1$. Suppose additionally, that there is one unobserved shock to the precision matrix obeying the equation*

$$\tilde{H} = \frac{\nu + \mu}{\nu}\mathcal{U}(H)'Q\mathcal{U}(H),\ Q \sim \mathcal{B}_m\left(\nu/2, \mu/2\right), \tag{25}$$

where $\mathcal{B}_m(p, q)$ is the multivariate beta-distribution, see appendix A. Then the posterior for B and $\tilde{H}$ is given by a Normal-Wishart distribution with density $f_{NW}(B, \tilde{H} \mid \bar{B}, N, S, \nu)$.

The proof is in appendix B.

4 Prior Selection

The selection of a prior is always an issue in Bayesian time series analysis and has recently been subject to much debate, see Phillips' (1991) critic of Sims and Uhlig (1991), his discussants and the summary in Uhlig (1992a). For the analysis of macroeconomic time series we suggest the following choice, modelled after the "Minnesota prior" in Doan,Litterman and Sims (1984) or Doan's RATS manual (1990). Use logarithms of the levels of the time series except for series expressed in percent (like interest rates), which are used in its raw form. Include a constant and a time trend, $C_t = [1\ t]'$ and $c = 2$. Let S_0 be the diagonal matrix of the MLEs of the residual variances for AR(1) univariate regressions for each included data series[4]. Include between one and two years of lags (e.g. 6 lags for quarterly data). Let $\bar{B}_0$ be the random walk prior

$$\bar{B}_0 = [0_{m,2}\ I_m\ 0_m\ \ldots\ 0_m]\,. \tag{26}$$

Choose N_0 to be block diagonal. The first block is of size 2×2 with

$$\begin{bmatrix} N_0(1,1) & N_0(1,2) \\ N_0(2,1) & N_0(2,2) \end{bmatrix} = \begin{bmatrix} \zeta_3 & -\zeta_3^2/2 \\ -\zeta_3^2/2 & \zeta_3^3/3 \end{bmatrix}. \tag{27}$$

The second block is of size $(km) \times (km)$ and diagonal with

$$N_0(2+m(l-1)+i, 2+m(l-1)+i) = Y_{0,i}^2 \zeta_1 l^{\zeta_2},\ i = 1,\ldots,m,\ l = 1,\ldots,k. \tag{28}$$

Here, $\zeta_1 \geq 0$, $\zeta_2 \geq 0$ and $\zeta_3 \geq 0$ are hyperparameters. The motivation for this particular form of the prior can be seen from the updating equation (16) for N_t and from the role N_t plays in equation (56) given in appendix A for the Normal-Wishart posterior: N_t is the precision "along the rows" of B. ζ_1 and ζ_2 determine the overall level of tightness and the increase in the prior precision for lagged coefficients. Heuristically, ζ_1 gives the number of earlier data observation one would need to have in order to arrive from a flat prior at a similarly precise prior before analyzing the rest of the data

[4]This, of course, amounts to a first pass through the data, which, strictly speaking, is not legitimate. However, this "loss of $m(m+1)/2$ degrees of freedom" should not be big for coming up with a reasonable starting point S_0 in most practical applications.

$Y_1, \ldots, Y_T$. $\zeta_2 > 0$ is a "soft" way to impose exclusion restrictions: the prior is increasingly tightened around a value of zero for coefficients on variables with increasingly higher lags. $\zeta_1 = 5$ and $\zeta_2 = 2$ seem to be reasonable choices[5]. ζ_3 corresponds roughly to the number of time periods of observing data regressed on the constant and the time trend for $t = -\zeta_3, \ldots, 0$ one would need to arrive at the prior given above from a flat prior. We chose $\zeta_3 = 8$. Our choices for the hyperparameters are quite weak except perhaps for ζ_3 (see the discussion below), since the prior imposed that way is swamped in the first few observations, see equation (56) in appendix A.

Rescaling of a time series is automatically taken care of with this prior in practice. Suppose, an interest rate of 3.1% is denoted by 0.031 rather than 3.1. Upon inspection of equations (12), (16) and the Normal-Wishart density, the appropriate rescaling is generally achieved by multiplying all $\bar{B}_0$-coefficients on the interest rate in the equations for the other variables by 100, dividing all $\bar{B}_0$-coefficients on other variables in the equation for the interest rate by 100, dividing the entries in N_0 which correspond to the interest rate coefficients by 10000 and dividing all rows and columns of S_0, which measure covariances with the interest rate innovation, by 100 (and therefore dividing the variance of the interest rate innovation itelf by 10000). The prior recommended above takes care of that automatically, unless the date-0 observation on the interest rate happens to be exactly zero.

It is important to note that e.g. a flat prior over the coefficients ($N \equiv 0$) does not represent ignorance as was pointed out by Phillips (1991) and that therefore the prior above imposes the prior view of downweighing explosive roots. While this is a desirable feature of the prior, a more problematic issue is the specification of the prior with regard to the constant and the trend term, see Schotman and Van Dijk (1991). Essentially the same time trend coefficient implies quite a different behaviour for the underlying time series depending on whether the largest root of the autoregressive part is small or large. In particular at a unit root, a nonzero coefficient on the constant

[5]Note, that we have $\zeta_2 = 2 > 0$ here in seeming contrast to standard BVAR methodology as in Doan, Litterman and Sims (1984), since the elements of N denote precision, not variance.

generates a linear time trend and a nonzero coefficient on the time trend generates a quadratic. To take care of this problem, Uhlig (1992a,1992b) suggests in a simple univariate context to modify the Normal-Wishart prior by premultiplying it with a factor driving the time trend coefficient to zero as the largest root approaches unity from below. The appropriate modification in the context of a VAR is still subject to research and not taken up in this paper. Another approach is to exclude the time-trend as a regressor: while this may lead to misspecification, it also seems to result in a more robust behaviour of the method. As a practical intermediate solution, we suggest to force the constant and the time trend close to zero a priori through the somewhat high value of $\zeta_3 = 8$ or an even higher choice.

5 Inference about the hyperparameter ν.

The degree of time variation $\nu > m - 1$ or, equivalently, $\lambda = 1/(\nu + 1)$ has been treated as a constant so far. To subject it to inference as well, one ought to specify a prior over ν and proceed to calculate a posterior. Practically this is difficult and a good prior for ν hard to come by. We suggest instead to treat ν as a hyperparameter (see e.g. Doan, Litterman and Sims (1984) or Leamer (1978)) and choose that value for ν which permits the best one-step ahead forecasts, given the prior, rather than integrating over or maximizing the joint posterior in all unobserved variables. This is a quasi-maximum posterior procedure. As we will point out, even this procedure yields some puzzling results, which merit further investigation.

Note that conditional on information up to date t, the predictive density, i.e. the density for the one-step ahead prediction error $e_{t+1} \equiv Y_{t+1} - \bar{B}_t X_{t+1}$ has a multivariate Student t-distribution[6] with parameters $\mu = 0$, $S_{t-1}\left(1 + X'_{t+1} N_t^{-1} X_{t+1}\right)$ and $\nu + 1 - m$: this follows easily from Leamers (1978) Theorem on p. 338 upon noting that in that theorem, the multivariate Student distribution parameter $\mathrm{S}/((\nu+1-k)T)$ has been misprinted as S/T. Rather than performing the difficult calculation of the exact joint

[6]For a definition of the multivariate Student t-distribution, see Leamer (1978), p. 336.

posterior in all variables and ν, we suggest to calculate the following log-quasi-posterior $\phi_T(\nu)$ for the benchmark prior $\pi(\nu) \equiv 1$. The idea is to use the predictive density for the one-step ahead prediction errors and the density of the multivariate beta distribution for that matrix $\bar{Q}_t$, which would explain the actual update of the mean precision. All the calculations are done in logs.

1. Select ν. Set $\delta = \nu + 1 - m$ and

$$\phi_{0,\text{data}} = 0 \tag{29}$$
$$\phi_{0,\text{beta}} = 0. \tag{30}$$

2. At each $t = 1, \ldots, T$,

   (a) Calculate the one-step ahead prediction error

$$e_t = Y_t - \bar{B}_{t-1} X_t, \tag{31}$$

   its log-likelihood, given the predictive density,

$$l_{t,\text{data}} = \log\left(\frac{\delta^{\delta/2}\Gamma((\delta+m)/2)}{\pi^{m/2}\Gamma(\delta/2)} \det\left(S_{t-1}\left(1 + X_t' N_{t-1}^{-1} X_t\right)\right)^{-1/2}\right) + \log\left(\left(\delta + e_t'\left(S_{t-1}\left(1 + X_t' N_{t-1}^{-1} X_t\right)\right)^{-1} e_t\right)^{(\delta+m)/2}\right) \tag{32}$$

   as well as the inverted upper Cholesky-Factor $U_{t-1} = \left(\mathcal{U}(S_{t-1}^{-1})\right)^{-1}$ of the mean precision S_{t-1}^{-1}.

   (b) Update to N_t, $\bar{B}_t$ and S_t,

   (c) Calculate the beta-distributed shock necessary to explain the update of the mean precisions as

$$\bar{Q}_{t-1} = \frac{\nu}{\nu+1} U_{t-1}' S_t^{-1} U_{t-1}. \tag{33}$$

   Calculate its log-likelihood via

$$l_{t,\text{beta}} = \log\left(\lambda^{m/2} f_{B,m,\nu/2,1/2}(Q_{t-1})\right), \tag{34}$$

where $f_{B,m,\nu/2,1/2}$ is the density of beta-distributed $m \times m$-matrices $Q \sim \mathcal{B}_m(\nu/2, 1/2)$, for which $I_m - Q$ is of rank 1 as calculated in Uhlig (1992c), see appendix A. For the extra factor $\lambda^{m/2}$, see the discussion below.

(d) Calculate

$$\phi_{t,\text{data}} = \phi_{t-1,\text{data}} + l_{t,\text{data}} \tag{35}$$

$$\phi_{t,\text{beta}} = \phi_{t-1,\text{beta}} + l_{t,\text{beta}}. \tag{36}$$

3. The result is $\phi_T(\nu) = \phi_{T,\text{data}} + \phi_{T,\text{beta}}$.

It then seems reasonable to choose ν which maximizes $\phi_T(\nu)$. However, complications in doing so seem to arise due to the singularity of the beta-distribution. The density (54) for $X \sim \mathcal{B}_m(\nu/2, 1/2)$ contains a factor $L^{-m/2}$, where L is the (only) positive eigenvalue of $I_m - X$. For the eigenvalue L_{t-1} of $I_m - \bar{Q}_{t-1}$, a short calculation, aided by rule (T8) in Leamer (1978), p. 324, shows that

$$L_{t-1} = \left(1 + \frac{1}{\left(1 - X_t' N_t^{-1} X_t\right)\left(e_t' S_{t-1}^{-1} e_t\right)\lambda/(1-\lambda)}\right)^{-1}. \tag{37}$$

Approximating $1 - X_t' N_t^{-1} X_t$ by 1, its limit as $t \to \infty$, and $S_{t-1}^{-1} e_t$ by 1, its expectation, yields that $L_{t-1} \approx \lambda$. Thus, the likelihood of $\bar{Q}_{t-1}$ will be of the order $\lambda^{-m/2}$, diverging to infinity as $\lambda \to 0$. Since this is an artifact of the likelihood function rather than a feature of the data, we have included the factor $\lambda^{m/2}$ in equation (34) as compensation. In application, this turns out not to be enough quite yet and $\phi_{T,\text{beta}}(\nu)$ often still diverges in applications as $\nu \to \infty$, see section 8. Avoiding the beta-density alltogether, a practical solution may be to select the maximizer of $\phi_{T,\text{data}}(\nu)$ or a value which is somewhat higher and therefore a somewhat more conservative choice. Clearly, further research on a reasonable way for determining ν is warranted.

6 Variations

Useful variations of the general method are easily obtained. We will show how to perform the updating, when the time variation of the variance-covariance matrix is given by a single factor in subsection 6.1, and when the only uncertainty is the uncertainty about the time-varying precision matrix itself in subsection 6.2. We compare our method to standard approaches for calculating Bayesian vectorautoregressions (BVARs) in subsection 6.3.

6.1 Time-Variation in the General Level of Volatility

Suppose the variance-covariance matrix is known a priori except for a multiplicative factor, which varies over time. Let Σ be of size $m \times m$ and positive definite. With the notation of section 3, the model is described by

$$Y_t = BX_t + \epsilon_t, \tag{38}$$

$$\epsilon_t \sim \mathcal{N}\left(0, \sigma_t^2 \Sigma\right), \tag{39}$$

$$\sigma_{t+1}^{-2} = \frac{\nu + m}{\nu} \sigma_t^{-2} u_t, \; u_t \sim \mathcal{B}_1\left(\nu/2, m/2\right), \tag{40}$$

where all ϵ_t and ν_t are independent.

Define $\lambda = 1/(\nu + m)$ (rather than $\lambda = 1/(\nu + 1)$ as before). The method for updating the prior takes the following form.

Method 2 (Time-Varying Level of Volatility)

1. *Choose a $m \times (c+km)$-matrix $\bar{B}_0$, a positive definite $(c+km) \times (c+km)$-matrix N_0 and a positive number s_0^2. These parameters fix the prior.*

2. *For each time $t = 1, \ldots, T$ calculate*

$$N_t = N_{t-1} + X_t X_t' \tag{41}$$

$$\bar{B}_t = \left(\bar{B}_{t-1} N_{t-1} + Y_t X_t'\right) N_t^{-1} \tag{42}$$

$$\begin{aligned} s_t^2 = {} & (1-\lambda) s_{t-1}^2 + \\ & \lambda tr\left(\Sigma^{-1}\left(Y_t - \bar{B}_{t-1} X_t\right)\left(1 - X_t' N_t^{-1} X_t\right)\left(Y_t - \bar{B}_{t-1} X_t\right)'\right). \end{aligned} \tag{43}$$

3. Keep B_T, N_T and s_T^2 as result..

With ν fixed, this method finds the Normal-Gamma posterior[7]

$$f_{NG}(B, \sigma_{T+1}^2 \mid \bar{B}_T, N_T, s_T^2, \nu)$$

in B and σ_{T+1}^2 for a Normal-Gamma prior

$$f_{NG}(B, \sigma_1^2 \mid \bar{B}_0, N_0, s_0^2, \nu).$$

This follows formally from theorem 2 applied for $\mu = m$ together with the following variation of theorem 1.

Theorem 3 *Let a prior for the $m \times (c + km)$ coefficient matrix B and the precision level σ^{-2} be given by a Normal-Gamma distribution with density $f_{NG}(B, \sigma^{-2} \mid \bar{B}, N, s^2, \nu)$, where N is positive define. Suppose additionally, that there is one observation of data X and Y (where X is $(c + km) \times 1$ and Y is $m \times 1$), obeying the equation*

$$Y = BX + \epsilon,\ \epsilon \sim \mathcal{N}\left(0, \sigma^2 \Sigma\right) \tag{44}$$

Then the posterior for B and σ^{-2} is given by a Normal-Gamma distribution with density $f_{NG}(B, \sigma^{-2} \mid \tilde{\bar{B}}, \tilde{N}, \tilde{s^2}, \tilde{\nu})$, where $\tilde{N}$ is positive definite and where

$$\tilde{N} = N + XX' \tag{45}$$

$$\tilde{\bar{B}} = \left(\bar{B}N + YX'\right)\tilde{N}^{-1} \tag{46}$$

$$\tilde{\nu} = \nu + m \tag{47}$$

$$\tilde{s^2} = \frac{\nu}{\nu + m} s^2 + \frac{1}{\nu + m}\, tr\left(\Sigma^{-1}\left(Y - \bar{B}X\right)\left(1 - X'\tilde{N}^{-1}X\right)\left(Y - \bar{B}X\right)'\right). \tag{48}$$

The proof is in appendix B.

[7] See Appendix A for the definition of a Normal-Gamma distribution.

6.2 Tracking a time-varying variance-covariance matrix

The efficient market hypothesis states that stock-returns follow a random walk, so that the coefficient matrix B in a VAR in logs of several stock-prices should be treated as known and equal to the right hand side of equation (26). With B fixed and just the time-variation of the variance covariance matrix at issue, the updating formulas for the model given by (15), (13) and (14) with $\lambda = 1/(\nu + 1)$ are as follows.

Method 3 (Tracking a Variance-Covariance Matrix)

1. *Choose a positive definite $m \times m$- matrix S_0, fixing the prior.*
2. *For each $t = 1, \ldots, T$ calculate*

$$S_t = (1 - \lambda)S_{t-1} + \lambda (Y_t - B_{t-1}X_t)(Y_t - B_{t-1}X_t)' \qquad (49)$$

3. *Keep S_T as result.*

With ν fixed, this method finds the Wishart posterior

$$H_{T+1} \sim \mathcal{W}_m \left(\nu, S_T^{-1}/\nu\right)$$

for a Wishart prior

$$H_1 \sim \mathcal{W}_m \left(\nu, S_0^{-1}/\nu\right).$$

The formal proof for this assertion is obtained from the general BVARTEC method for $N = n\bar{N}, n \to \infty$.

6.3 Relationship to Standard BVAR Methodology.

By eliminating the time-variation in the precision matrix H, the method reduces to the usual Kalman-Filter updating formulas for the parameters of a Normal-Wishart prior or, with H known a priori, a Normal prior. Consider the model

$$Y_t = BX_t + \epsilon_t \tag{50}$$

$$\epsilon_t \sim \mathcal{N}\left(0, H^{-1}\right) \text{ i.i.d..} \tag{51}$$

For the case where H is treated as unknown, choose some $\nu_0 > m-1$ in addition to the matrices chosen in step 1 of method 1. Equations (16) and (17) of step 2 of that method together with the equations

$$\nu_t = \nu_{t-1} + 1 \tag{52}$$

and

$$S_t = \frac{\nu_{t-1}}{\nu_t} S_{t-1} + \frac{1}{\nu_t}\left(Y_t - \bar{B}_{t-1}X_t\right)\left(1 - X_t' N_t^{-1} X_t\right)\left(Y_t - \bar{B}_{t-1}X_t\right)' \tag{53}$$

provide the recursive formulas for updating a Normal-Wishart prior $f_{NW}(B, H \mid \bar{B}_0, N_0, S_0, \nu_0)$ to obtain the final posterior $f_{NW}(B, H \mid \bar{B}_T, N_T, S_T, \nu_T)$. The formal justification is theorem 1. An improper, "flat" prior as in Koop (1990) or De Jong (1992) is obtained for $N_0 = 0$ and $\nu_0 = 0$ (see the discussion in appendix 2 in Uhlig (1992a)).

For the case where $\Sigma = H^{-1}$ is treated as known, choose only N_0 and $\bar{B}_0$ as in step 1 of method 1. Equations (16) and (17) of step 2 of that method deliver the recursive updating formulas to obtain the final Normal posterior given by

$$vec(B') \sim \mathcal{N}\left(vec(\bar{B}_T'), \Sigma \otimes N_T^{-1}\right)$$

for the Normal prior

$$vec(B') \sim \mathcal{N}\left(vec(\bar{B}_0'), \Sigma \otimes N_0^{-1}\right).$$

This version of a BVAR is a special case of the general form of BVARs typically used for example in Doan, Litterman and Sims (1984) or the RATS Manual by Doan (1990). It is special, since $vec(B')$ is assumed to have a variance-covariance matrix of the form $\Sigma \otimes N_0^{-1}$ in the prior. Once, this form is assumed, the posterior will have the same form as well. Via N_0, this prior

allows complete freedom in the choice of the variance-covariance structure for the coefficients of exactly one equation. The covariance structure across equations is then fixed by Σ and N_0. For a diagonal Σ, for example, coefficients of different equations are uncorrelated and the covariances between the coefficients in one equation are the covariances of the corresponding coefficients of another equation multiplied by the ratio of the two error variances of these equations as given by the diagonal elements in Σ. In particular, it is not possible to follow the procedure in Doan-Litterman-Sims (1984) and tighten the prior uncertainty about the coefficients in the stock market equation, say, to a random walk prior independently from the other equations: in fact, given the usual volatility of stock market data, the method here will probably generate higher uncertainty about the coefficients of that equation than most other equations of such a system. Further research should investigate the possibilities for equation-individual tightening. The restriction to the special priors here may be regarded as an advantage since fewer prior choices have to be made by the researcher.

7 An Experiment with Artificial Data

To get some feeling for how well this method behaves, artificial data has been generated and the estimation procedure applied. The data consists of $m = 2$ independent random walks for $t = -2, \ldots, 200$ with standard normal innovations, i.e the data variance does not change over time. The starting point is drawn from a standard normal distribution as well. We used $k = 3$ lags and the random walk prior specified in section 4 with $\zeta_1 = 5$, $\zeta_2 = 2$ and $\zeta_3 = 8$. We fixed $\lambda = .3/(m+1) = .1$. The initial variance-covariance matrix S_0 was estimated from the realized first differences of the data and should therefore be close to the identity matrix. Thus, the prior is actually (close to) the true data-generating process except for the fact that it assumes the variance to be time-varying. The results are in the four figures starting with the one labelled "Artificial Data". The matrices characterizing the posterior are available from the author upon request. Impulse response functions are produced by Cholesky-decomposing draws from the posterior Wishart distri-

bution, thereby assuming a lower triangular shape for the matrix translating orthogonal innovations into the innovations for each individual variable. This follows the usual BVAR-procedure, see e.g. Sims (1986), except that we do not compute the impulse responses directly at the posterior mean but rather from an average over 100 draws from the posterior distribution. For these 100 draws, we also calculated the variance and the skewness. The two dotted lines in each picture are mean + standard-deviation * (skewness +/- 1). The figures demonstrate that the asymmetry is quite substantial. It may possibly be due to the interaction of the time trend coefficient with the roots in the system. We proceeded similarly for the forecasts. In order to highlight just the posterior coefficient uncertainty, the error bands in the forecast figures reflect the coefficient uncertainty for calculating expected values of the future variables, not the randomness of the future variable itself.As can be seen there, the model reproduces well the random walk forecast for series 2, but does less well for series 1. Again, the skewness in the forecast distribution matters quite a lot: a topic that warrants more research.

8 An Application to a 4-variable forecasting model for the US

We have applied the general method to a 4-variable vector autoregression, using quarterly US time series data. The data series we used are obtained from CitiBase:

1. FPS6US is a monthly stock price index for 500 common stocks (1967 = 100),NSA
2. FYFF is the monthly federal funds rate in percent per annum,NSA
3. GNP82 is quarterly GNP in 1982 Dollars with 1982=100,
4. GDC is the implicit price deflator for personal consumption expenditures as quarterly data.

The choice of these variables results in a simple yet interesting system to study the interactions of monetary policy, characterized by the federal funds rate, aggregate output and inflation, cmp. SIms (1980,1986,1987). To obtain quarterly data, the monthly data has been time-averaged for the three month making up a quarter. The logarithm of all variables except the federal funds rate has been taken. That way, first differences can be interpreted as percent increases: this is what we have done in all the graphs. All the data has been cut to the time interval 1955:1 to 1991:2, since more data was not available on the federal funds rate. Thus, the length of the data is $T = 146$. We estimated S_0 and choose B_0 as well as N_0 as specified in section 4. We fixed the number of lags $k = 5$ and applied the general BVARTEC method.

Results can be found in the six figures starting with the one labelled "The Data". The matrices characterizing the posterior can be obtained from the author upon request. In addition to the figures produced for the experiment with artificial data, we have followed the procedure in section 5 to calculate the log-quasi-posterior $\phi(\nu)$, drawn as a function of λ rather than ν (without including the Jacobian of that transformation in order to keep the location of the maximum). Note that $\phi(\nu)$ diverges to infinity as $\lambda \to 0$. Inspection of the two parts $\phi_{T,\text{data}}$ and $\phi_{T,\text{beta}}$ in the next figure reveals, that based on the predictive density for the forecast errors alone, one would favor a value of $\lambda \approx 0.1$, and that it is the $\phi_{T,\text{beta}}$-piece which diverges as $\lambda \to 0$.

For all further calculations we selected the somewhat conservative "guess" $\lambda = 0.05$ corresponding to $\nu = 19$. Impulse response functions are again computed as averages over 100 draws from the final distribution, where we used each time the Cholesky-Decomposition to decompose the contemporaneous vector of innovations into orthogonal components in the manner common in standard BVAR research, see e.g. Sims (1986). The ordering of the variables used is the ordering given above. We plotted both the mean impulse response as well as error bands given by the mean + standard deviation * (skewness +/- 1). As in the experiment with artificial data, asymmetry is clearly important.

Similarly, forecasts have been obtained for the next 4 years, using draws from the posterior. Again the error bands reflect the coefficient uncertainty

for calculating expected values of the future variables, not the randomness of the future variable itself. The values in these forecast figures are differences between future and current values multiplied by 100 for the logarithmized variables: this can be interpreted as cumulative growth in percent. Thus, the graph for real GNP, for example, shows the forecasted cumulative growth over the current level in percent over the next 4 years, whereas the graph for the federal funds rate forecast shows the absolute level of that rate forecasted for this time span.

The last plot compares the calculated predicted time-varying forecast standard deviations (defined as the inverse square root of the appropriate mean precision diagonal element) with the first difference as well as the one-step ahead prediction error, which it is meant to track.

9 Conclusion

This paper introduced Bayesian vector autoregressions with time-varying error covariances (BVARTEC). More precisely, a nonlinear generalization of the Kalman-Filter was used to derive in closed form the Bayesian posterior for the coefficients as well as for the error precision matrix in a vector-autoregressive model, when that error precision matrix is stochastically time-varying. Along with current, advanced ARCH methods,this provides for a method to deal with time varying error variances and covariances in a multivariate context.

It turns out, that the Kalman Filter formula resemble closely those used to describe the evolution of heteroskedasticity in GARCH(1,1)-models, except that the one-step-ahead forecast error, rather than the true underlying, but unobservable disturbance is used. The method was compared to standard BVAR procedures. Two useful variations were derived: especially the method tracking a time-varying precision matrix when the regression coefficients are known should prove useful for applications in finance.

Appendix

A Some distributions and their properties

This appendix has the purpose to fix the notation and gather some useful facts. More information on the distributions used in this paper can be found in Zellner (1971), Leamer (1978), Muirhead (1982) and Uhlig (1992c). A Wishart distribution $\mathcal{W}_m(\nu, \Omega)$, $\nu \geq m$ specifies the following density for the $m(m+1)/2$ distinct elements of a positive definite $m \times m$ random matrix X:

$$f_W(X \mid \Omega, \nu) = \frac{\mid X \mid^{(\nu-m-1)/2}}{2^{m\nu/2}\Gamma_m\left(\frac{\nu}{2}\right) \mid \Omega \mid^{\nu/2}} \exp\left(-\frac{1}{2} tr\left(\Omega^{-1}X\right)\right),$$

where Γ_m is the multivariate gamma function, defined in Muirhead [28], Definition 2.1.10. If $X \sim \mathcal{W}_m(\nu, \Omega)$, then $E[X] = \nu\Omega$. Note that $\mathcal{W}_1(\nu, s^{-2}/\nu) = \Gamma(s^2, \nu)$ and $\mathcal{W}_1(\nu, 1) = \chi^2_\nu$, where $\Gamma(s^2, \nu)$ denotes the gamma distribution with mean s^{-2} and ν degrees of freedom and where χ^2_ν denotes the chi-squared distribution with ν degrees of freedom. The multivariate beta distribution $\mathcal{B}_m(p, q)$, $p > (m-1)/2$, $q > (m-1)/2$, for the $m(m+1)/2$ distinct elements of a positive definite $m \times m$ random matrix X, where $I_m - X$ is positive definite as well, has the density

$$f_{B,m,p,q}(X) = \frac{\Gamma_m(p+q)}{\Gamma_m(p)\Gamma_m(q)} \mid X \mid^{p-(m+1)/2} \mid I_m - X \mid^{q-(m+1)/2}.$$

In this paper, singular multivariate beta distributions $\mathcal{B}_m(p, n/2)$ are needed, where $0 < n < m$ is an integer. These distributions are defined in general in Uhlig (1992b). For the special case $n = 1$, which is particularly relevant here, the density for the positive definite $m \times m$ random matrix $X \sim \mathcal{B}_m(p, 1/2)$, where $I_m - X$ is positive semidefinite and of rank 1 is given in Uhlig (1992b) as

$$f_{B,m,p,1/2}(X) = \pi^{(-m+1)/2} \frac{\Gamma_m(p+1/2)}{\Gamma(1/2)\Gamma_m(p)} L^{-m/2} \det(X)^{p-(m+1)/2} \qquad (54)$$

with respect to the volume element of the m distinct entries in $I_m - X$, where $L \in \mathbf{R}$ is given by the decomposition $I_m - X = H_1 L H_1'$, $\|H_1\| = 1$. If $X \sim \mathcal{B}_m(p,q)$, then $E[X] = p/(p+q)I_m$: this is shown in Corollary 1 below.

As priors and posteriors, Normal-Gamma distributions and Normal-Wishart distributions are used. Let $l \geq 1$ be an integer, let N be $l \times l$ and positive definite, let Σ be $m \times m$ and positive definite, let $\bar{B}$ be a $m \times l$-dimensional matrix and let $\nu > 0$. The Normal-Gamma distribution for a $m \times l$-dimensional coefficient matrix B and a one-dimensional precision σ^{-2} is given by the density

$$
\begin{aligned}
&f_{NG}(B, \sigma^{-2} \mid \bar{B}, N, s^2, \nu) = \qquad (55)\\
&\quad \kappa(N, s^2, \nu, m, l, \Sigma) \mid \Sigma \mid^{-km/2} \left(\sigma^{-2}\right)^{km^2/2}\\
&\quad \exp\left(-\frac{1}{2}\left(vec(B') - vec(\bar{B}')\right)' \left[\sigma^{-2}\Sigma^{-1} \otimes N\right] \left(vec(B') - vec(\bar{B}')\right)\right)\\
&\quad \sigma^{\nu/2-1} \exp\left(-\frac{1}{2}\nu\, s^2 \sigma^{-2}\right)
\end{aligned}
$$

where $\kappa(N, s^2, \nu, m, l, \Sigma)$ is the appropriate integrating constant and $vec(\cdot)$ denotes columnwise vectorization. The Normal-Gamma distribution specifies, that the precision σ^{-2} follows a Gamma distribution $\Gamma(s^2, \nu)$ and that, conditional on σ^{-2}, the coefficient vector $vec(B')$ follows a Normal distribution $\mathcal{N}\left(vec(\bar{B}'), \sigma^2 \Sigma \otimes N^{-1}\right)$.

In the Normal-Wishart distribution, all entries of Σ are random individually. Let $l \geq 1$ be an integer: in the main text we usually have $l = c + km$. Let N be $l \times l$ and positive definite, let S be $m \times m$ and positive definite, let $\bar{B}$ be a $m \times l$-dimensional matrix and let $\nu \geq m$. The Normal-Wishart distribution for a $m \times l$-dimensional coefficient matrix B and a $m \times m$ precision matrix H is given by the density

$$
\begin{aligned}
&f_{NW}(B, H \mid \bar{B}, N, S, \nu) = \qquad (56)\\
&\quad \kappa(N, S, \nu, m) \mid H \mid^{l/2}\\
&\quad \exp\left(-\frac{1}{2}\left(vec(B') - vec(\bar{B}')\right)' [H \otimes N] \left(vec(B') - vec(\bar{B}')\right)\right)
\end{aligned}
$$

$$| H |^{(\nu-m-1)/2} \exp\left(-\frac{1}{2}\nu \; tr HS\right)$$

where $\kappa(N, S, \nu, m)$ is the appropriate integrating constant. The Normal-Wishart distribution specifies, that the precision matrix $H = \Sigma^{-1}$ follows a Wishart distribution $\mathcal{W}_m(\nu, S^{-1}/\nu)$ with $(E[\Sigma^{-1}])^{-1} = S$, and that, conditional on H, the coefficient matrix B in its vectorized form $vec(B')$ follows a Normal Distribution $\mathcal{N}\left(vec(\bar{B}'), H^{-1} \otimes N^{-1}\right)$. This definition is a slight generalization of Leamer (1978). The Normal-Wishart distribution is popular in traditional Bayesian multivariate regression analysis, see Zellner (1971).

The key ingredient for our method is based on the following theorem about the conjugacy between Wishart distributions and multivariate beta distributions. For nonsingular beta distributions, this theorem is a restatement of Muirhead [28], theorem 3.3.1. For singular beta distributions (the case needed in this paper, except in subsection 6.1), this theorem is shown in Uhlig (1992b).

Theorem 4 (The Conjugacy of Wishart- and Beta Distributions.) *Let $m > 0$ be an integer and either $n > m - 1$ or $n > 0$ an integer. Let $\nu > m - 1$. Let $H \sim \mathcal{W}_m(\nu + n, \Sigma)$ and $Q \sim \mathcal{B}_m(p/2, n/2)$ be independent. Then*

$$G \equiv \mathcal{U}(H)'Q\mathcal{U}(H) \sim \mathcal{W}_m(\nu, \Sigma).$$

As a corollary, we can compute the mean of multivariate beta distributions.

Corollary 1 *Let $X \sim \mathcal{B}_m(p, q)$. Then*

$$E[X] = \frac{p}{p+q} I_m.$$

Proof: *Choose some positive definite Σ of size $m \times m$ and obtain its Cholesky-decomposition $\Sigma = U'U$. Let $\nu = 2p, n = 2q$ and choose H, Q and G as in the previous theorem. Let $Z = E[Q]$ and $T = E[\mathcal{U}(H)]$. Then*

$\nu U'U = E[G] = T'ZT$, since H and Q are independent, and $(\nu+n)U'U = E[H] = T'T$. Since the Cholesky decomposition is unique, $\sqrt{\nu+n}U = T$ and thus $Z = \nu T'^{-1}U'UT^{-1} = \nu/(\nu+n)I_m$. •

It is easy to draw from these distributions, as long as the most common univariate distributions are available, by exploiting theorems 3.2.5 and 3.3.1 in Muirhead (1982) to generate draws from a Wishart distribution, and the definition in Uhlig (1992c) to draw from a multivariate singular beta distribution. A program in *MATHEMATICA* [27], version 2.0, is available from the author upon request.

B Proofs

Proof: Proof of Theorem 1. *The proof proceeds by directly calculating the densities. Note first, that $\tilde{N}$ and $\tilde{S}$ are indeed positive definite. Further, recall that the likelihood function for the observation is given by*

$$L(B,H \mid X,Y) \sim \mid H \mid^{1/2} \exp\left(-\frac{1}{2}(Y-BX)'H(Y-BX)\right). \tag{57}$$

Rewrite the Normal-Wishart density more conveniently as

$$\begin{aligned} f_{NW}(B,H \mid \bar{B},N,S,\nu) = & \qquad (58)\\ \kappa(N,S,\nu,m) \mid H \mid^{l/2} & \qquad (59)\\ \exp\left(-\frac{1}{2}tr\,H\left(B-\bar{B}\right)N\left(B-\bar{B}\right)'\right) & \\ \mid H \mid^{(\nu-m-1)/2} \exp\left(-\frac{1}{2}\nu\ tr\,HS\right), & \end{aligned}$$

Thus, using the properties of the trace-operator, we have for the posterior $f(B,H)$

$$
\begin{aligned}
f(B,H) \sim & \mid H \mid^{1/2} \exp\left(-\tfrac{1}{2} tr\, H(Y-BX)(Y-BX)'\right) \\
& \mid H \mid^{(c+km)/2} \exp\left(-\tfrac{1}{2} tr\, H\left(B-\bar{B}\right) N \left(B-\bar{B}\right)'\right) \\
& \mid H \mid^{(\nu-m-1)/2} \exp\left(-\tfrac{1}{2}\nu\ tr\, HS\right) \\
= & \mid H \mid^{(c+km)/2} \exp\left(-\tfrac{1}{2} tr\, H\left(B-\tilde{\bar{B}}\right) \tilde{N} \left(B-\tilde{\bar{B}}\right)'\right) \\
& \mid H \mid^{(\tilde{\nu}-m-1)/2} \exp\left(-\tfrac{1}{2}\tilde{\nu}\ tr\, H\tilde{S}\right) \\
& \exp\left(-\tfrac{1}{2}\ tr\, H\phi(B)\right),
\end{aligned}
\tag{60}
$$

where

$$
\begin{aligned}
\phi(B) = & \left(B-\bar{B}\right) N \left(B-\bar{B}\right)' \\
& +(Y-BX)(Y-BX)' \\
& -\left(B-\tilde{\bar{B}}\right) \tilde{N} \left(B-\tilde{\bar{B}}\right)' \\
& -\left(Y-\bar{B}X\right)\left(1-X'\tilde{N}^{-1}X\right)\left(Y-\bar{B}X\right)'
\end{aligned}
\tag{61}
$$

It remains to show that $\phi(B) \equiv 0$ *or, equivalently, the*
Claim:

1. $\phi(\bar{B}) = 0$ and
2. $\frac{\partial \phi}{\partial B_{ij}} \equiv 0$ for all i,j.

Calculate, that

$$
\begin{aligned}
& \left(\bar{B}-\tilde{\bar{B}}\right) \tilde{N} \left(\bar{B}-\tilde{\bar{B}}\right)' \\
& \qquad = \left(\bar{B}\tilde{N}-\tilde{\bar{B}}\tilde{N}\right) \tilde{N}^{-1} \left(\bar{B}\tilde{N}-\tilde{\bar{B}}\tilde{N}\right)' \\
& \qquad = \left(\bar{B}XX'-YX'\right) \tilde{N}^{-1} \left(\bar{B}XX'-YX'\right)' \\
& \qquad = \left(Y-\bar{B}X\right) X'\tilde{N}^{-1}X \left(Y-\bar{B}X\right)'.
\end{aligned}
\tag{62}
$$

It follows that

$$\begin{aligned}\phi(\bar{B}) = \; & (Y-\bar{B}X)(Y-\bar{B}X)' \\ & -\left(\bar{B}-\tilde{\bar{B}}\right)\tilde{N}\left(\bar{B}-\tilde{\bar{B}}\right)' \\ & -\left(Y-\bar{B}X\right)\left(1-X'\tilde{N}^{-1}X\right)\left(Y-\bar{B}X\right)' \\ = \; & 0,\end{aligned} \tag{63}$$

proving the first part of the claim. For the second part, let E_{ij} be the $m \times (c+km)$-matrix with a 1 as its (i,j)-entry and zeros everywhere else. Using the definition of $\tilde{N}$ as well as the fact, that $\tilde{\bar{B}}\tilde{N} = \bar{B}N + YX'$, we have

$$\begin{aligned}\frac{\partial\phi}{\partial\bar{B}_{ij}}(B) = \; & E_{ij}N\left(B-\bar{B}\right)' + \left(B-\bar{B}\right)NE_{ij}' \\ & -E_{ij}X\left(Y-BX\right)' - \left(Y-BX\right)X'E_{ij}' \\ & -E_{ij}\tilde{N}\left(B-\tilde{\bar{B}}\right)' + \left(B-\tilde{\bar{B}}\right)\tilde{N}E_{ij}' \\ = \; & -E_{ij}N\bar{B}' - \bar{B}NE_{ij}' \\ & -E_{ij}XY' - YX'E_{ij}' \\ & +E_{ij}\tilde{N}\tilde{\bar{B}}' + \tilde{\bar{B}}\tilde{N}E_{ij}' \\ = \; & 0.\end{aligned} \tag{64}$$

This finishes the proof. •

Proof: Proof of Theorem 2. *For notation see appendix A. All that is needed to show is that if H follows a Wishart $\mathcal{W}_m(\nu+\mu, S^{-1}/(\nu+\mu))$, then $\tilde{H}$ follows a Wishart $\mathcal{W}_m(\nu, S^{-1}/\nu)$ or, equivalently, that*

$$G \equiv \nu/(\nu+\mu)\tilde{H} = \mathcal{U}(H)'Q\mathcal{U}(H)$$

follows a Wishart $\mathcal{W}_m(\nu, S^{-1}/(\nu+\mu))$ distribution. But this is just theorem 4. •

Proof: Proof of Theorem 3. *The theorem follows directly from equation (60) and $\phi(B) \equiv 0$ in the proof above for theorem 1, when H is replaced by $\sigma^{-2}\Sigma^{-1}$ etc., using the usual rules for calculating with determinants.* •

References

[1] Baba, Yoshihisa, Robert F. Engle, Dennis F. Kraft and Kenneth F. Kroner,"Multivariate Simultaneous Generalized ARCH," draft, University of San Diego (1987)

[2] Baillie, Richard T. and Tim Bollerslev,"A multivariate generalized ARCH approach to modeling risk premia in forward foreign exchange rate markets," Journal of International Money and Finance (1990), 9, 309-324

[3] Bollerslev, Tim, Robert F. Engle and Jeffrey M. Wooldridge, "A Capital Asset Pricing Model with Time-Varying Covariances," Journal of Political Economy (1988), vol. 96, No. 1, 116-131

[4] Bollerslev, T., Ray Y. Chou and Kenneth F. Kroner, "ARCH Modeling in Finance: A Review of the Theory and Empirical Evidence," Journal of Econometrics 52 (1992), 5-59

[5] Canova, Fabio, "Vector Autoregressive Models: Specification, Estimation, Inference and Forecasting," draft, Department of Economics, European University Institute (1991)

[6] DeJong, David N., "Co-Integration and Trend-Stationarity in Macroeconomic Time Series - Evidence from the Likelihood Function," Journal of Econometrics, 52 (1992), 347-370

[7] Diebold, Francis X., and Mark Nerlove,"The Dynamics of Exchange Rate Volatility: A Multivariate Latent Factor ARCH Model," Journal of Applied Econometrics 4 (1990), 1-21

[8] Doan, Thomas, Robert Litterman, and Christopher Sims, "Forecasting and conditional projection using realistic prior distributions," Econometric Reviews, Vol. 3, No. 1 (1984), p. 1-100,

[9] Doan, Thomas, *RATS Manual*, VAR Econometrics, 1800 Sherman Ave. Suite 612, Evanston, IL 60201 (1990)

[10] Engel, Charles and Anthony P. Rodrigues, "Tests of International CAPM with Time-Varying Covariances," Journal of Applied Econometrics, Vol. 4, 119-138 (1989)

[11] Engle, Robert F., "Autoregressive conditional heteroskedasticity with estimates of the variance of United Kingdom inflation," Econometrica, 50 (1982), 987-1008

[12] Engle, Robert F.,"Multivariate ARCH with Factor Structures - Cointegration in Variance," draft, University of San Diego (1987)

[13] Engle, Robert F., Takatoshi Ito, and Wen-Ling Lin,"Meteor Showers or Heat Waves? Heteroskedastic Intra-Daily Volatility in the Foreign Exchange Markets," Econometrica, Vol. 58, No. 3, (1990), 525-542

[14] Geweke, John, "Exact inference in models with autoregressive conditional heteroskedasticity," Chapter 5 in "Dynamic econometric modelling," W.A. Barnett, E.R. Berndt, H. White, eds., Cambridge University Press (1988), p. 73-105

[15] Geweke, John, "Exact predictive densities for linear models with ARCH disturbances," Journal of Econometrics, Vol. 40 (1989), p. 63-86

[16] Harvey, Andrew, Esther Ruiz and Neil Shephard, "Multivariate Stochastic Variance Models," LSE Financial Markets Group Discussion Paper No. 132 (1992)

[17] Harvey, Andrew and Neil Shephard, "Structural Time Series Models," in G.S. Maddala, C.R. Rao and H.D. Vinod, eds., *Handbook of Statistics, Vol. 11: Econometrics,* North-Holland, Amsterdam (1992),

[18] Kleibergen, Frank and Herman K. van Kijk, "On the Shape of the Likelihood/Posterior in Cointegration Models," draft, econometric Institute and Tinbergen Institute, Erasmus University, Rotterdam (1992)

[19] Koop, Gary,"Aggregate shocks and macroeconomic fluctuations: a predictive approach," working paper, Boston University (1990)

[20] Leamer, Edward E., *Specification searches - ad hoc inference with non-experimental data,* John Wiley & Sons, Inc., New York (1978)

[21] Litterman, Robert," Techniques of forecasting using vector autoregressions", Federal Reserve Bank of Minneapolis, working paper 115 (1979),

[22] Litterman, Robert,"A Bayesian procedure for forecasting with vector autoregressions," working paper, MIT (1980)

[23] Litterman, Robert,"Forecasting and policy analysis with Bayesian vector autoregression models," Federal Reserve Bank of Minneapolis, Quarterly Review (1984), p. 30-41

[24] Litterman, Robert,"Specifying vector autoregressions for macroeconomic forecasting," Federal Reserve Bank of Minneapolis, Staff Report 92 (1984)

[25] Litterman, Robert,"Forecasting with Bayesian vector autoregressions - five years of experience," Federal Reserve Bank of Minneapolis, working paper 274 (1985)

[26] Lütkepohl, Helmut, *Introduction to Multiple Time Series Analysis,* Springer Verlag, Berlin (1990)

[27] *MATHEMATICA* - a System for Doing Mathematics, version 2.0, Wolfram Research, Champaign, Illinois, publ. Addison-Wesley, Redwood City (1991)

[28] Muirhead, Robb J., *Aspects of multivariate statistical theory,* John Wiley & Sons, Inc., New York (1982)

[29] Nelson, Daniel B., "Stationarity and persistence in the GARCH(1,1) model," Econometric Theory 6 (1990), 318-334

[30] Nelson, Daniel B.,"Conditional heteroskedasticity in asset returns: a new approach," Econometrica, Vol. 59, No. 2 (1991), 347-370

[31] Phillips, P.C.B., "To Criticize the Critics: An Objective Bayesian Analysis of Stochastic Trends," Journal of Applied Econometrics, vol. 6, no. 4 (1991), 333-364,

[32] Shephard, N.G., "A Local Scale Model: An Unobserved Component Alternative to Integrated GARCH Processes," Discussion Paper No. EM/90/220, London School of Economics (1990), Journal of Econometrics, forthcoming

[33] Schotman, Peter C. and Herman K. van Dijk, "On Bayesian Routes to Unit Roots," Journal of Applied Econometrics, vol. 6, no. 4 (1991), 387-402

[34] Sims, Christopher A., "Macroeconomics and reality," Econometrica, Vol. 48, No. 1 (1980), p.1-48,

[35] Sims, Christopher A.,"Are forecasting models usable for policy analysis?" Federal Reserve Bank of Minneapolis, Quarterly Review (1986), p. 2-15

[36] Sims, Christopher A.,"Identifying policy effects," Federal Reserve Bank of Minneapolis, working paper 351 (1987),

[37] Sims, Christopher A.,"A nine variable probabilistic macroeconomic forecasting model," Federal Reserve Bank of Minneapolis, Institute for Empirical Macroeconomics, discussion paper 14 (1989),

[38] Sims, Christopher A. and H. Uhlig,"Understanding unit rooters: a helicopter tour," Econometrica, Dec. 1991

[39] Todd, Richard M., "Implementing Bayesian vector autoregressions," Federal Reserve Bank of Minneapolis, working paper 384 (1988)

[40] Uhlig, H., "What Macroeconomists Should Know About Unit Roots as well: The Bayesian Perspective," draft, Princeton University (1992a)

[41] Uhlig, H., "A Note on Jeffreys' Prior When Using the Exact Likelihood Function," draft, Princeton University (1992b)

[42] Uhlig, H., "On Singular Wishart and Singular Multivariate Beta Distributions," draft, Princeton University (1992c)

[43] Watanabe, Toshiaki, "Alternative Approach to Conditional Heteroskedasticity in Stock Returns: Approximate Non-Gaussian Filtering," draft, Yale University (1992)

[44] West, Mike and Jeff Harrison, *Bayesian Forecasting and Dynamic Models*, Springer Series in Statistics (1989)

[45] Zellner, Arnold, *An introduction to Bayesian inference in econometrics*, John Wiley & Sons, Inc., New York (1971)

Figure 1

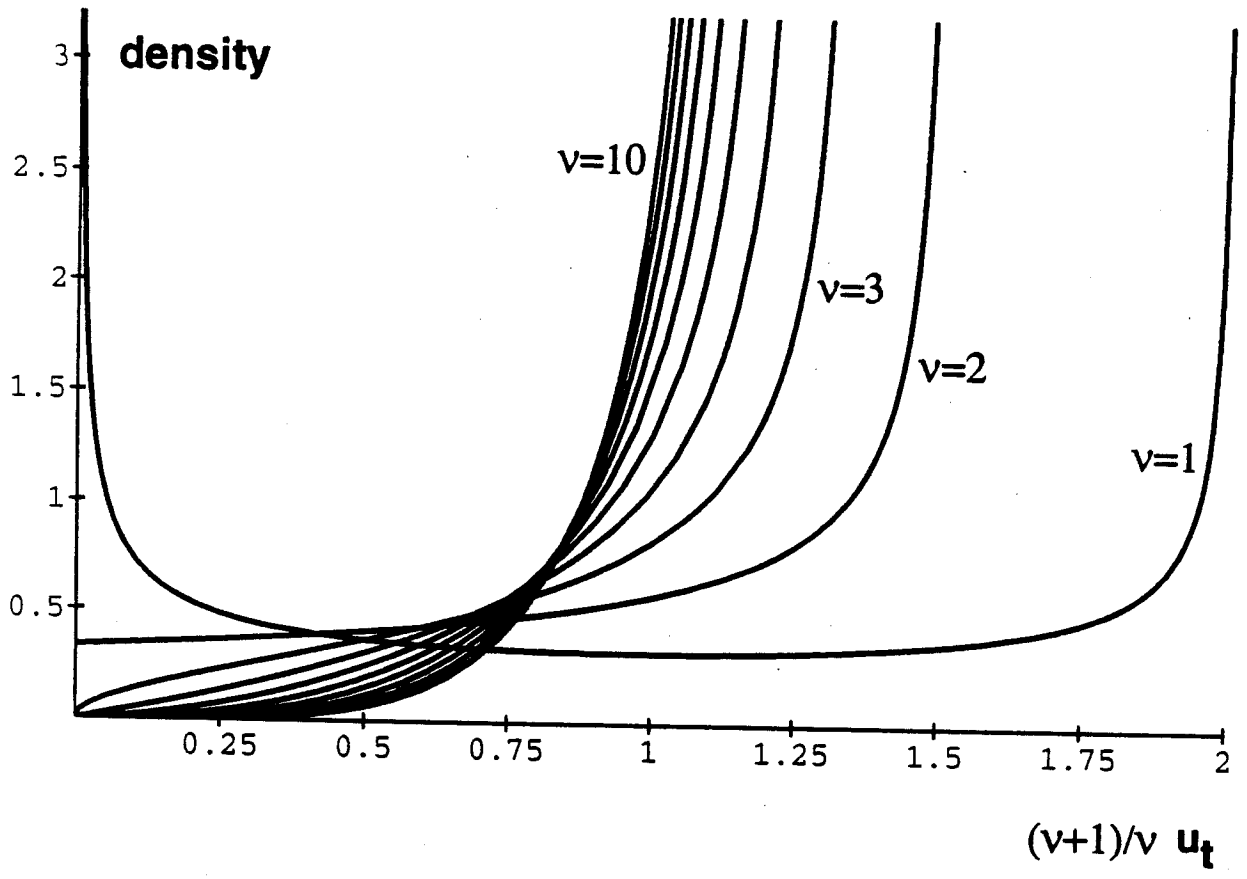


Figure 2

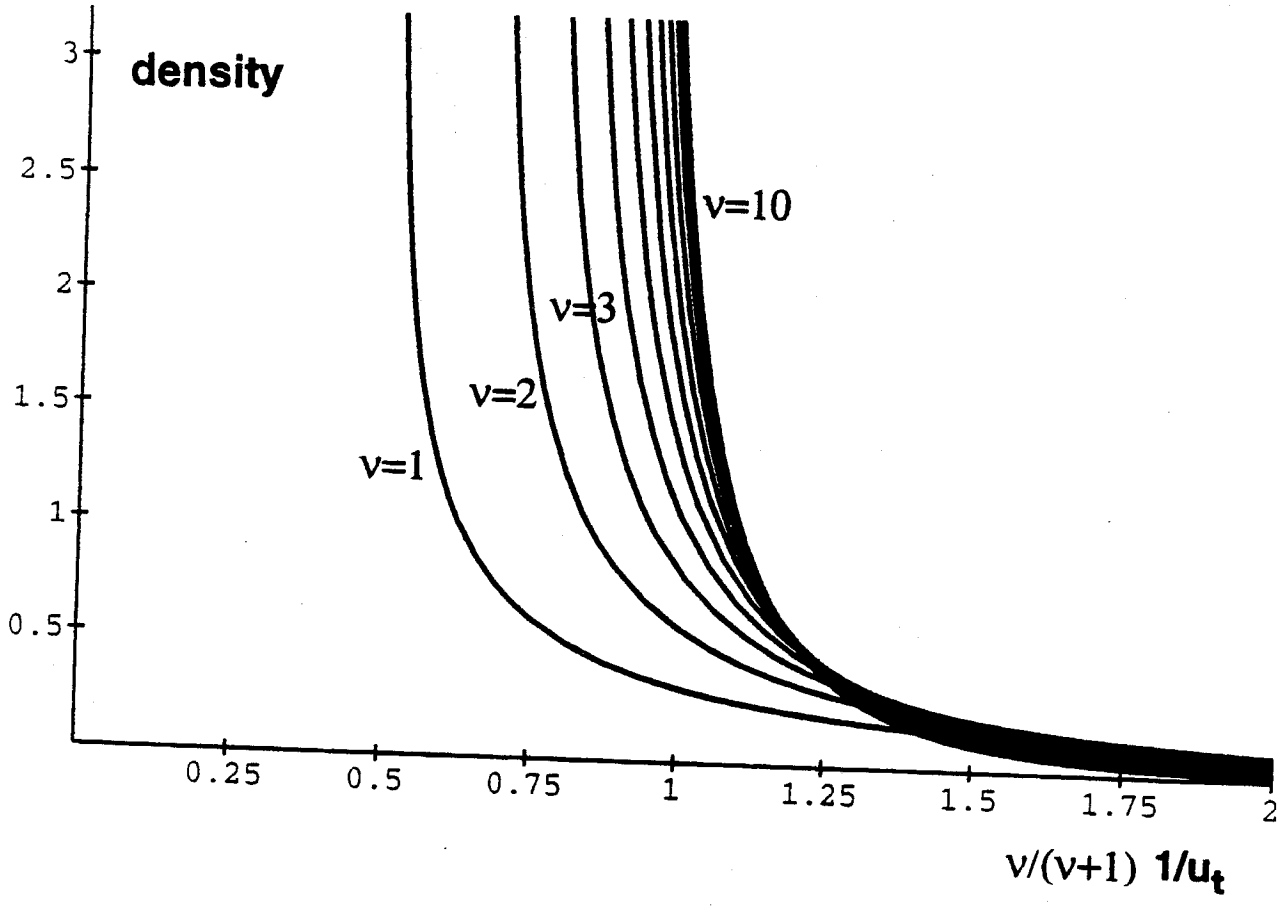


Figure 3: $\nu = 9$

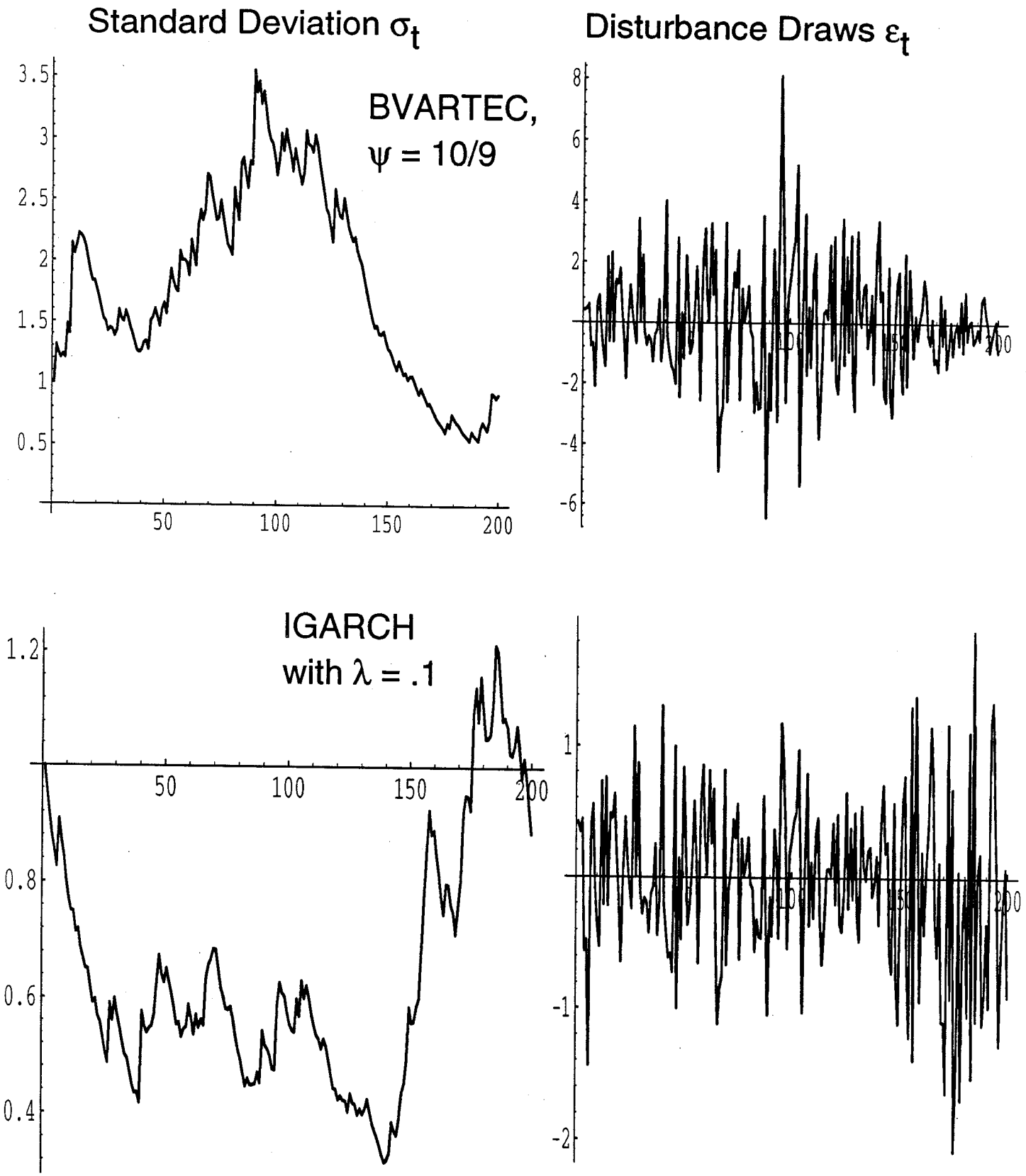
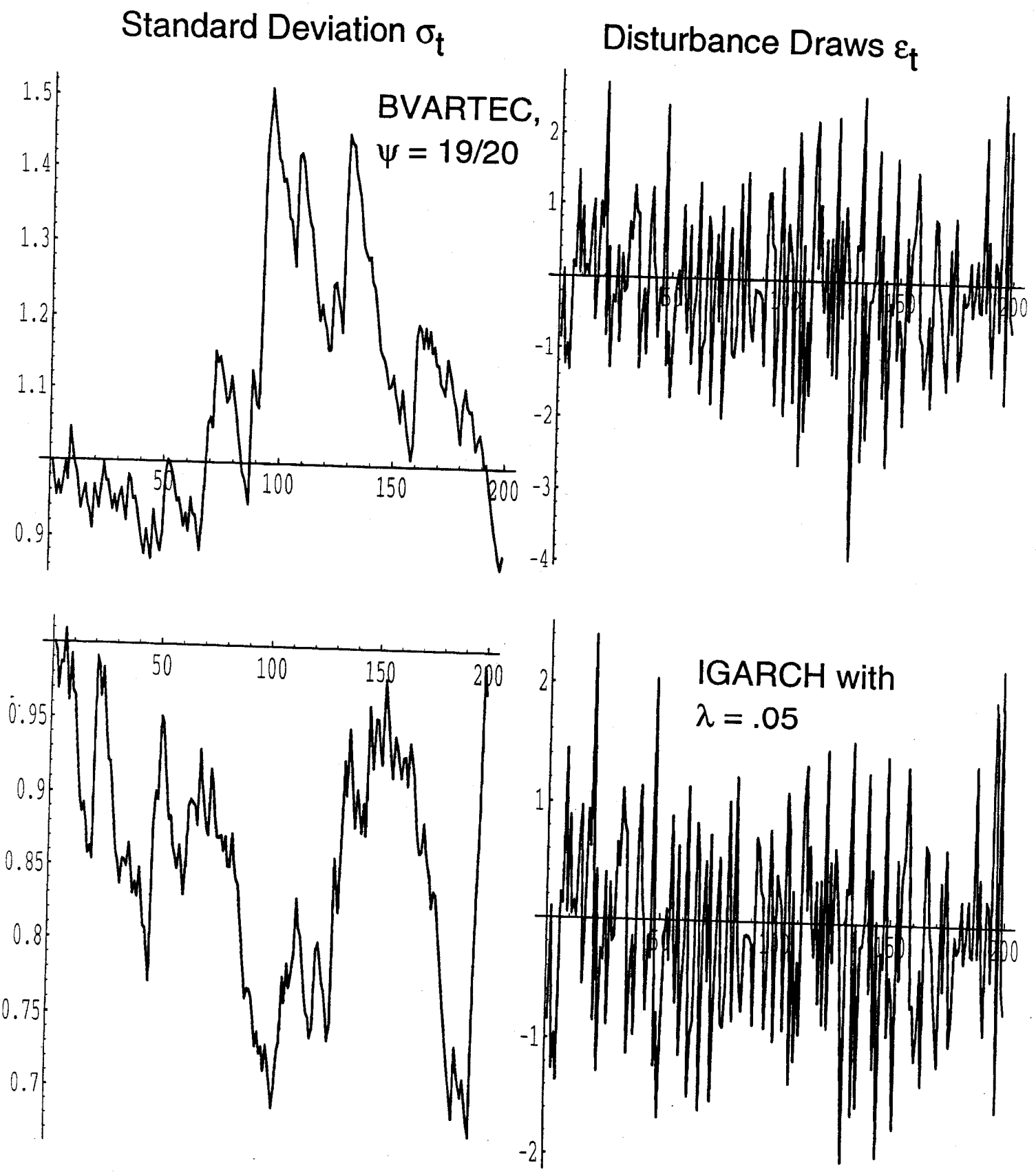
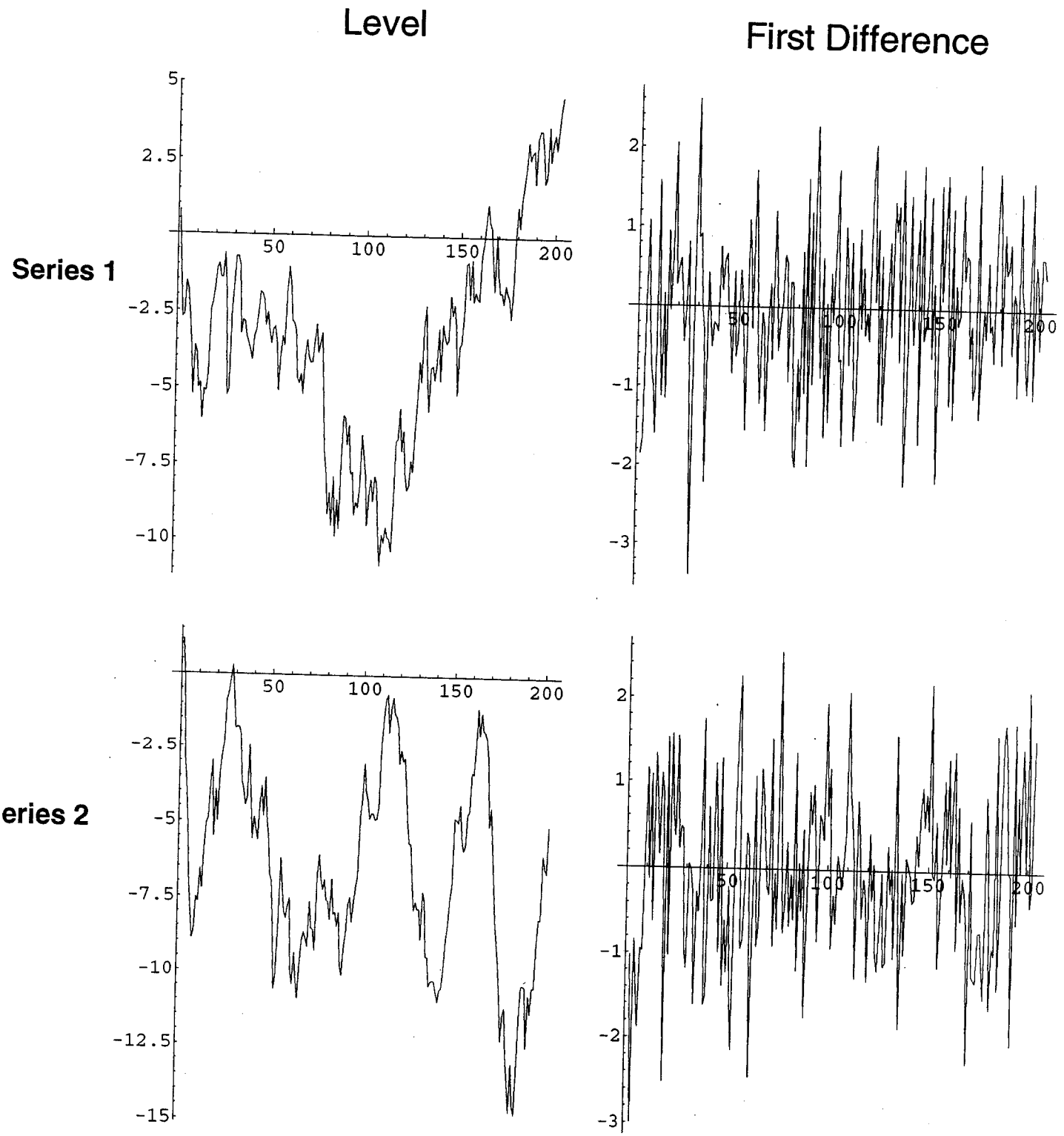

Standard Deviation σ_t
Disturbance Draws ε_t
BVARTEC,
$\psi = 10/9$
IGARCH
with $\lambda = .1$

Figure 4: $\nu = 19$



Artificial Data

(Two independent random walks with standard normal innovations.)



Impulse Response Functions

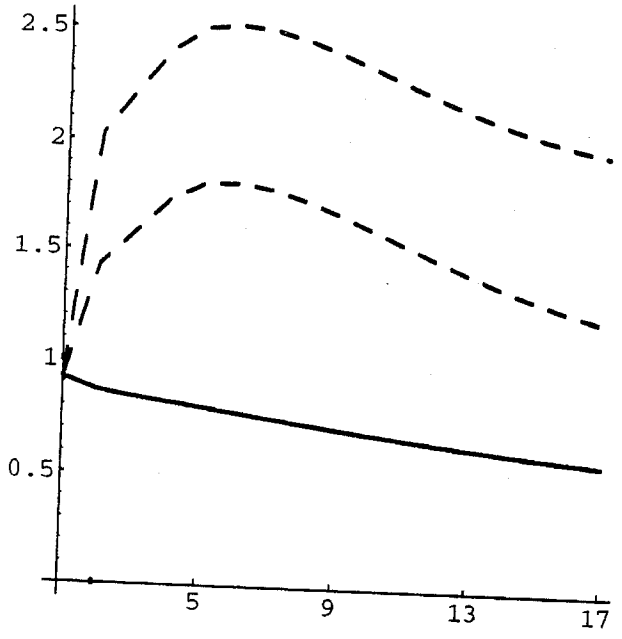
Response to innovation one standard deviation in size, using Cholesky-decomposition. Values in Level of Series.

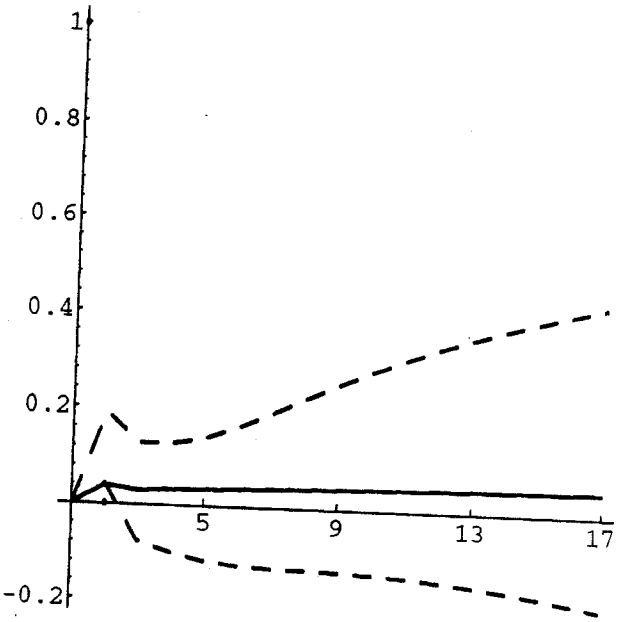
Innovation to / Response of

Series 1

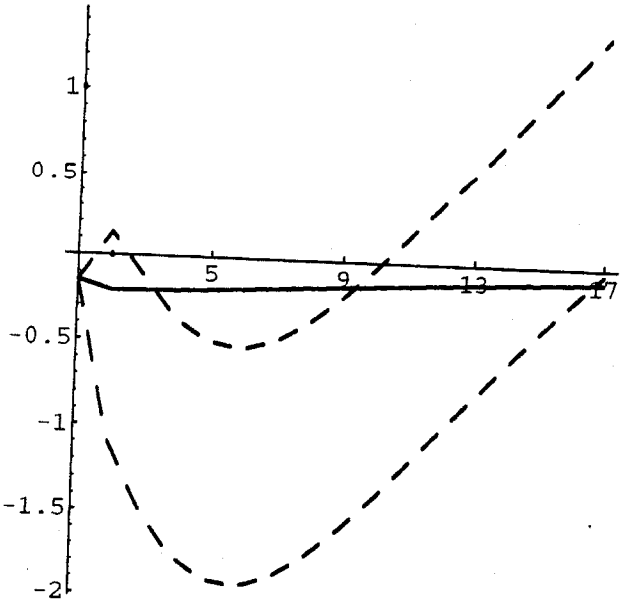
Series 2

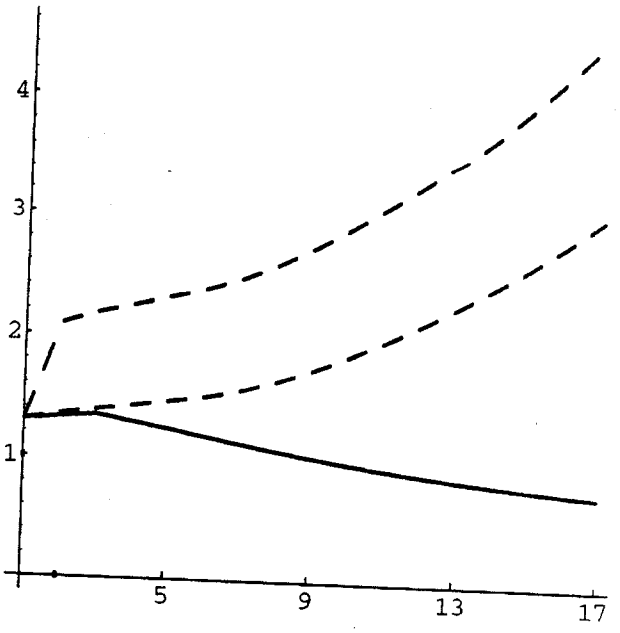
Series 1





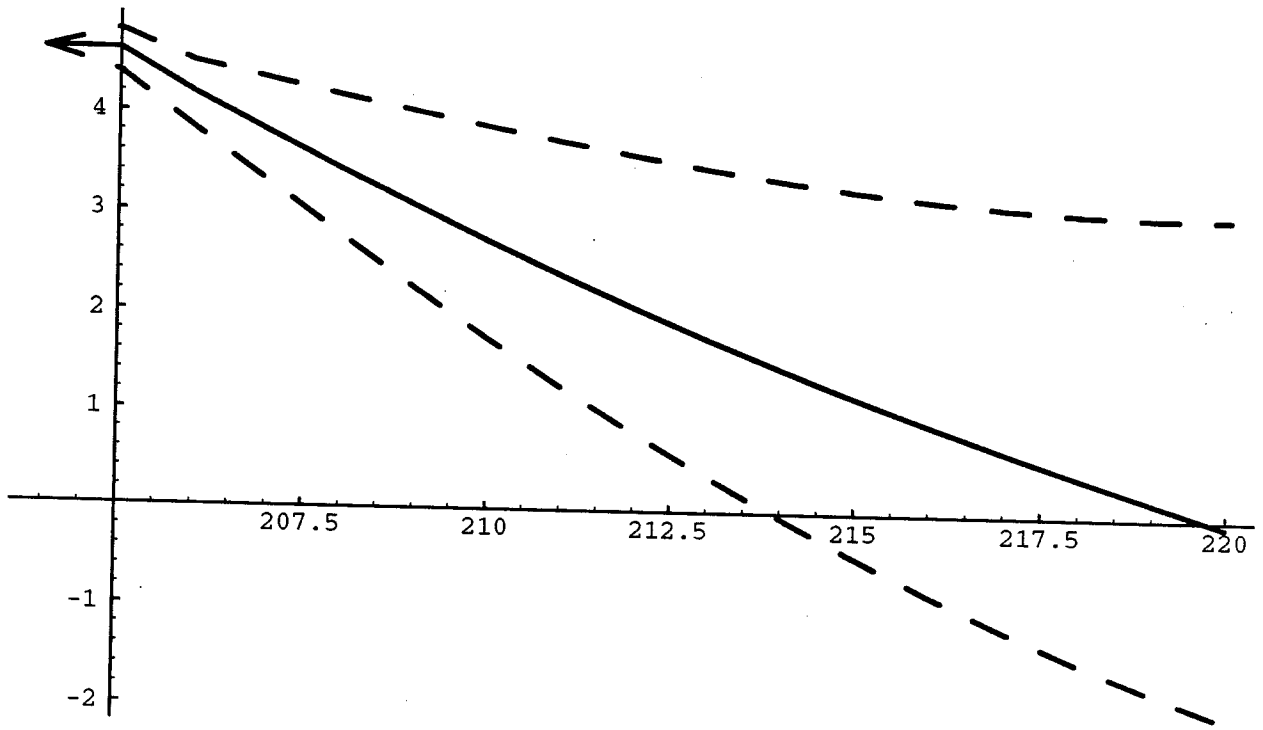
Series 2

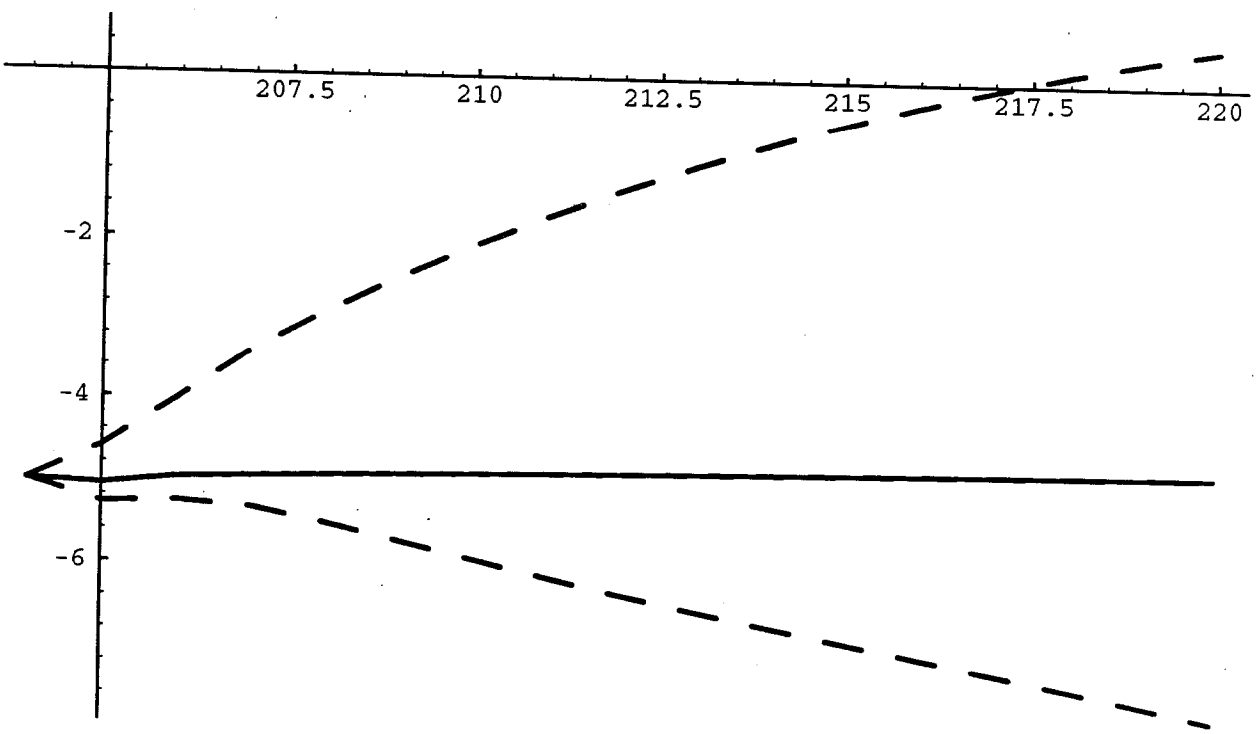




Artificial Data: Forecast

The dotted lines represent the posterior coefficient uncertainty.
Forecast is in level.

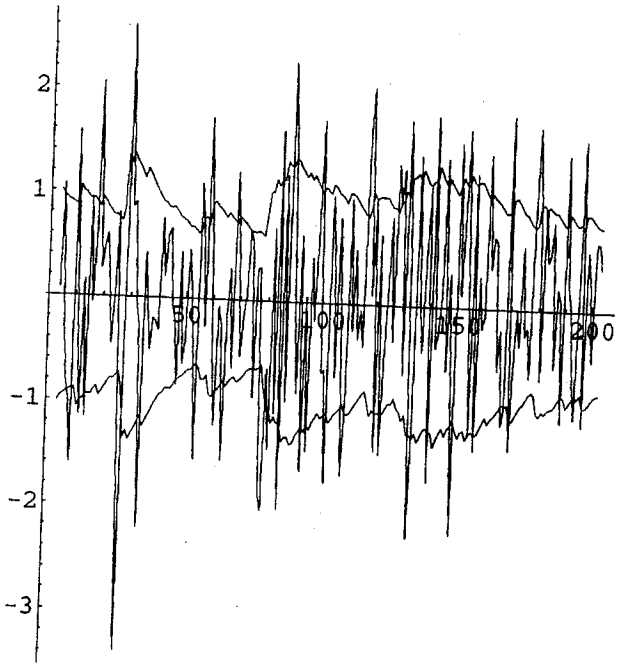


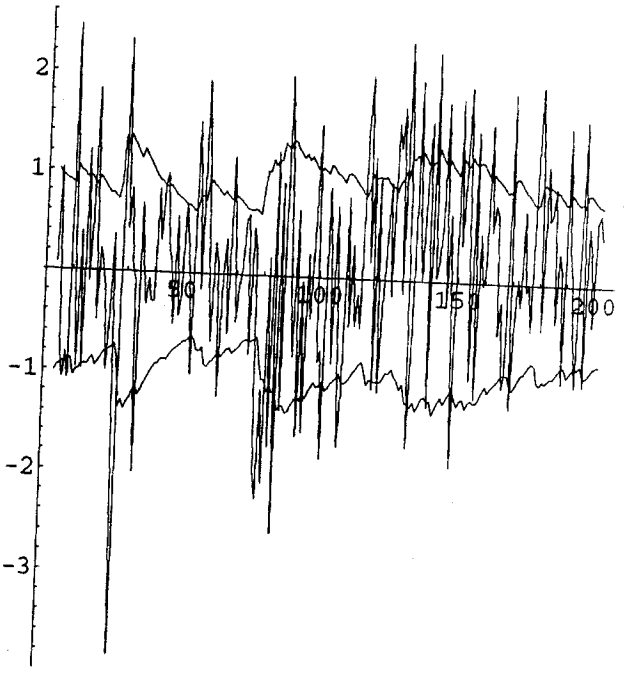


First Differences, Realized Forecast Errors and Predicted Standard Deviations.

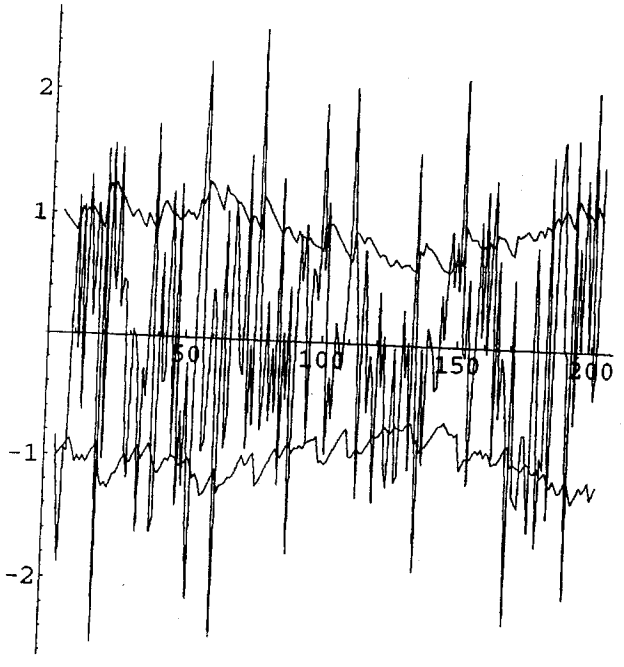
The rather smooth line in each graph is thte predicted standard deviation for the one-step ahead forecast error.

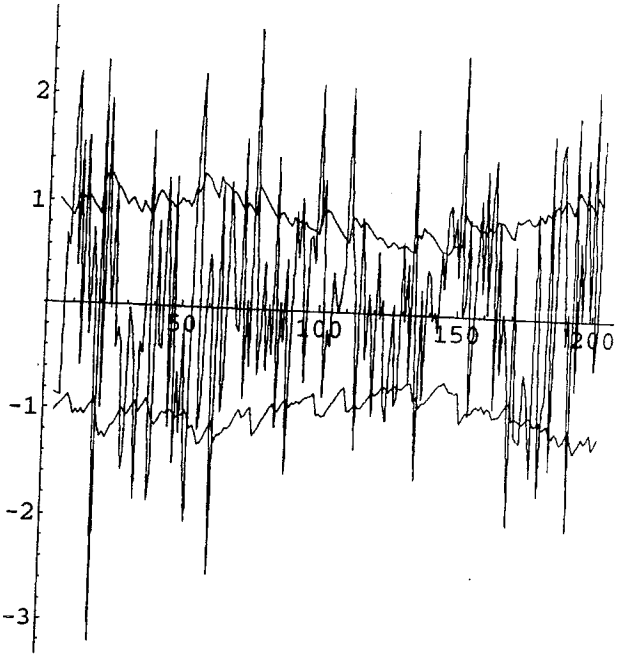
Series 1





Series 2

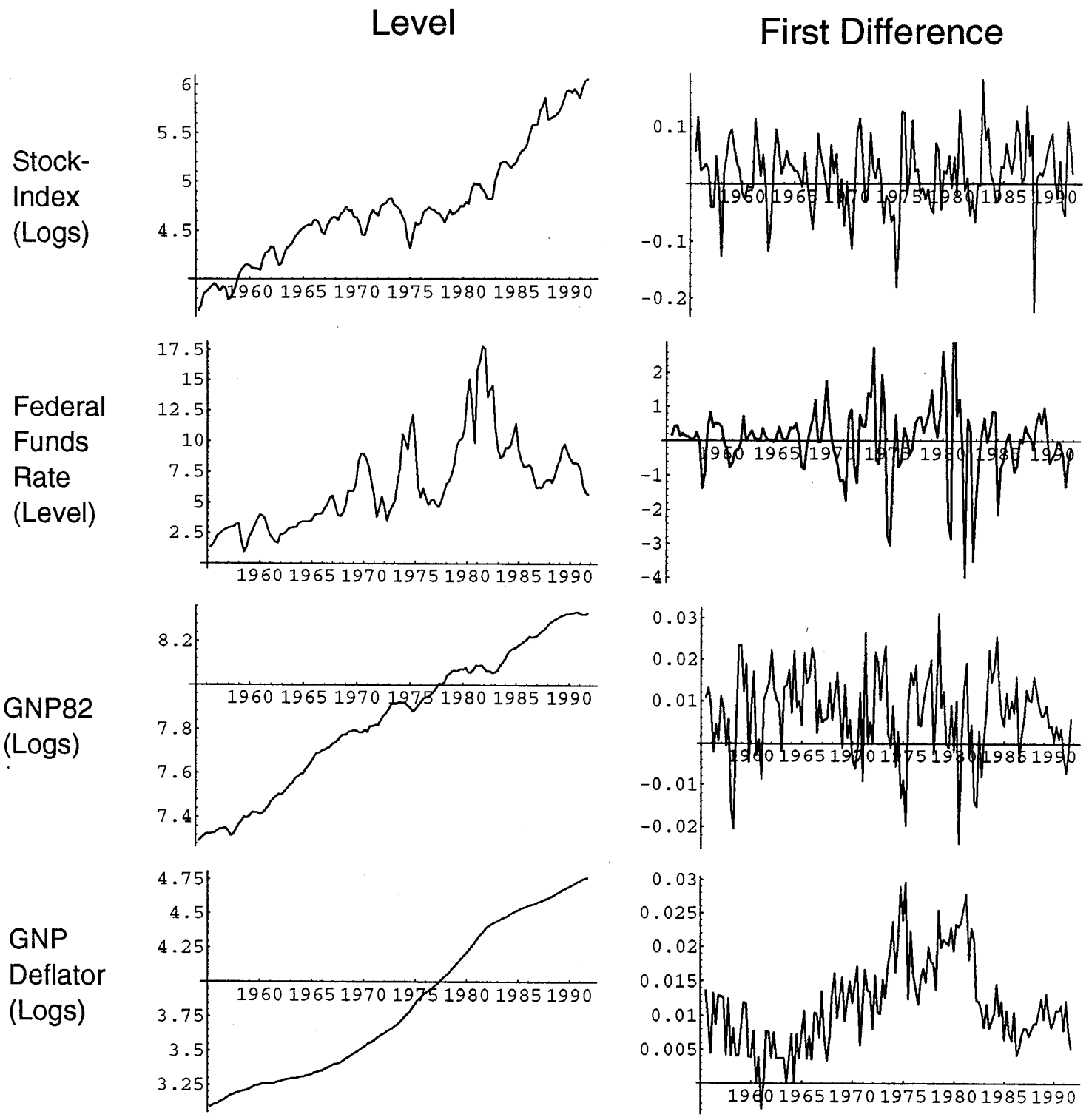




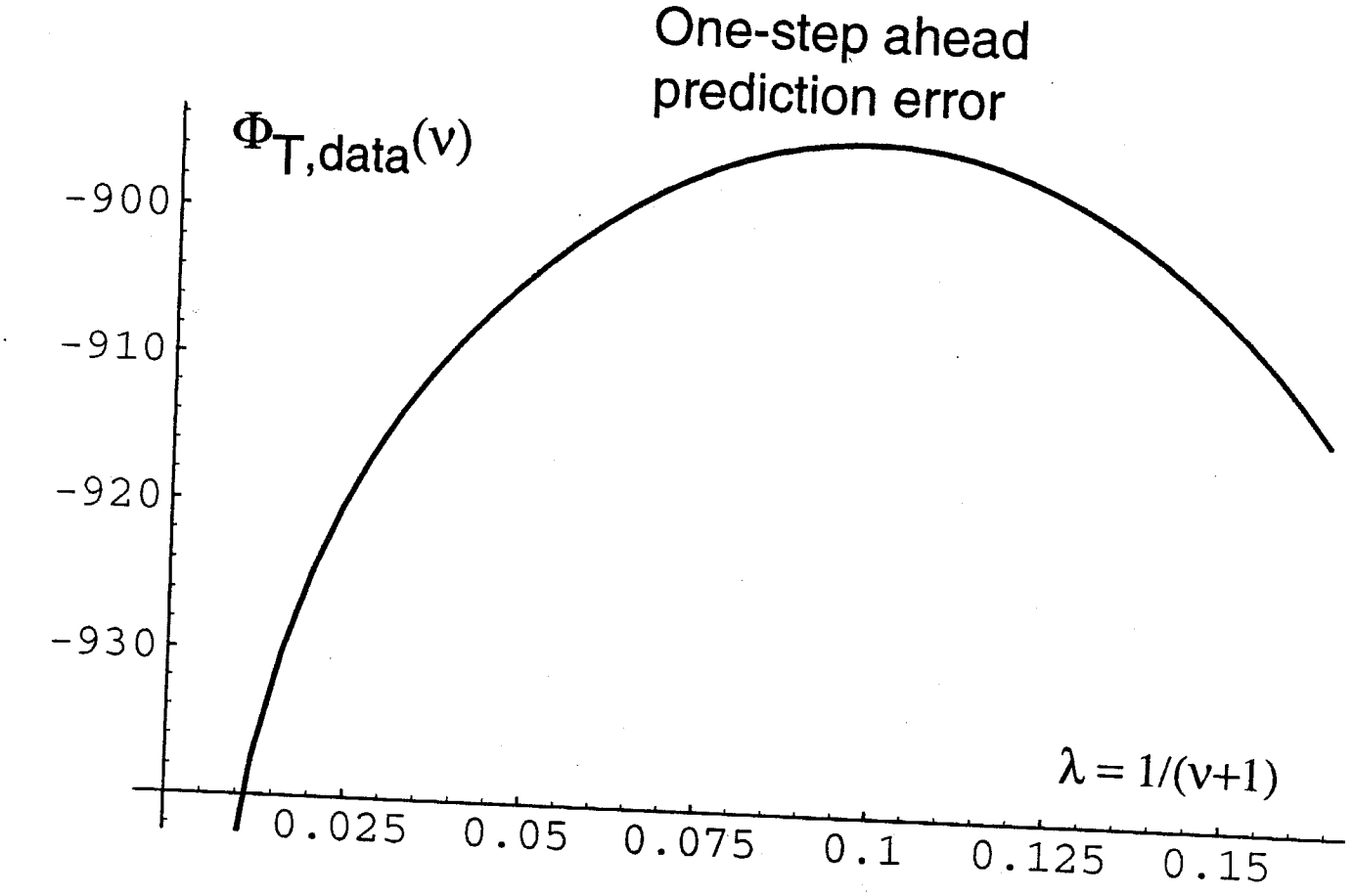
First Differences

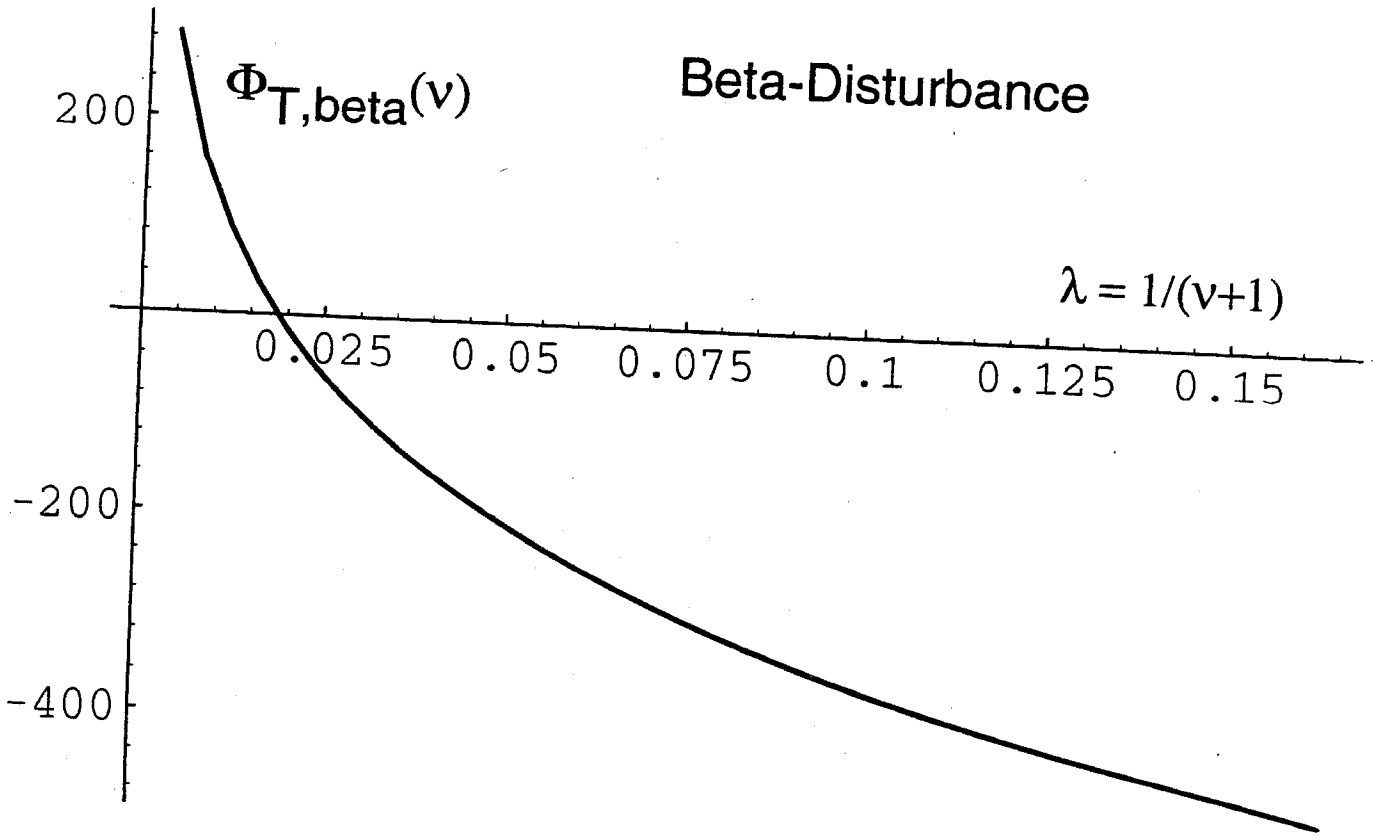
One-Step Ahead Prediction error

The Data

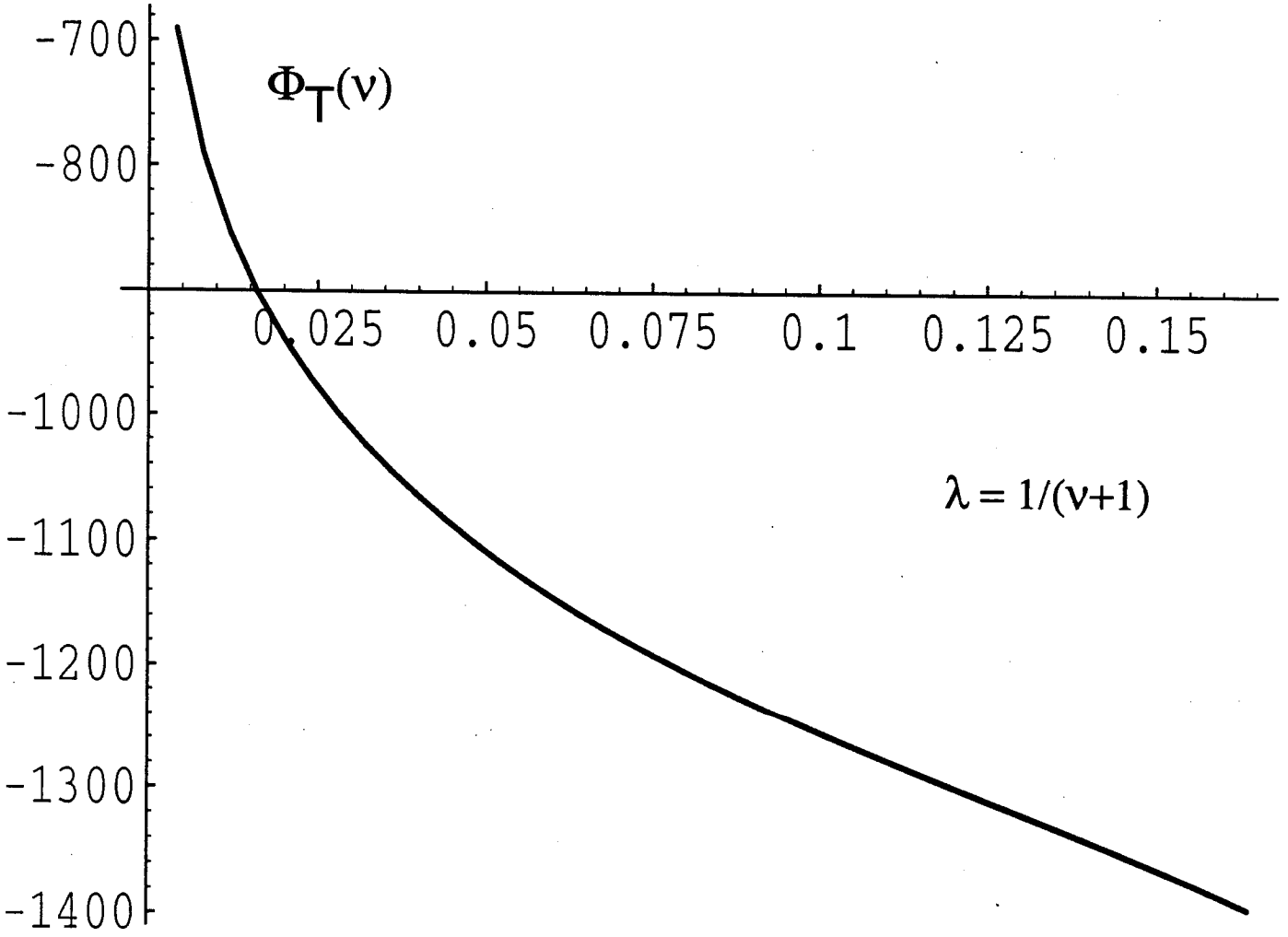

Level
First Difference
Stock-Index (Logs)
Federal Funds Rate (Level)
GNP82 (Logs)
GNP Deflator (Logs)
1960 1965 1970 1975 1980 1985 1990

The Two Pieces of $\Phi_T(\nu)$



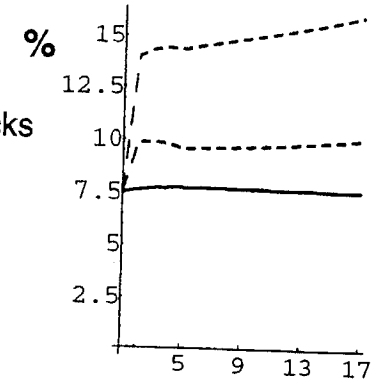
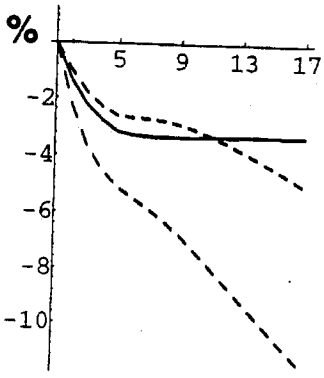
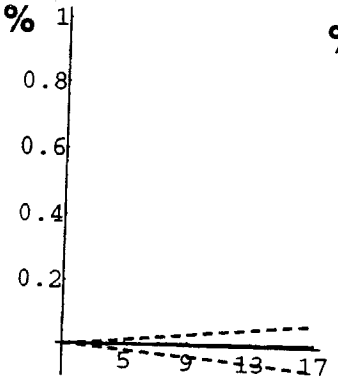
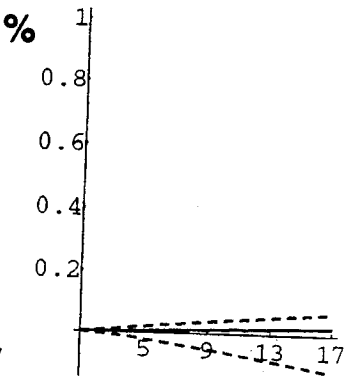
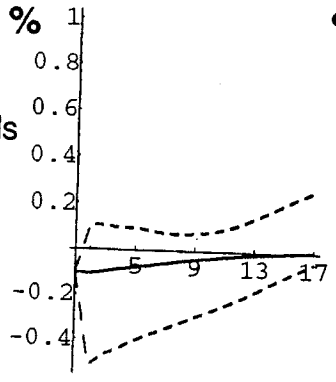
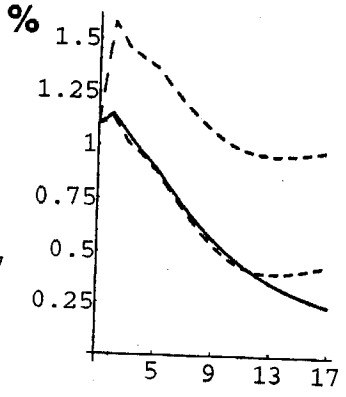
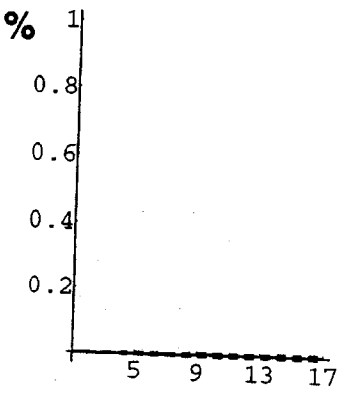
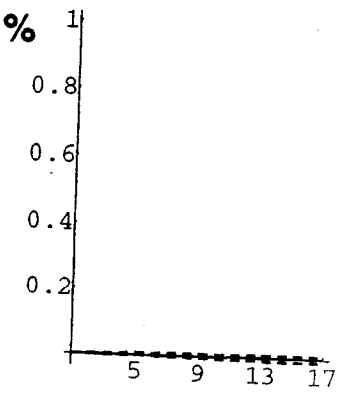
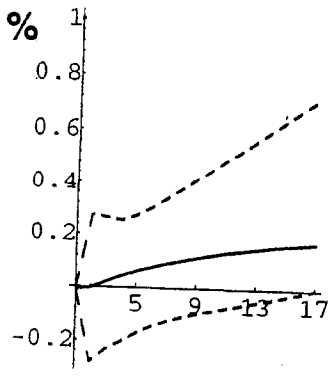
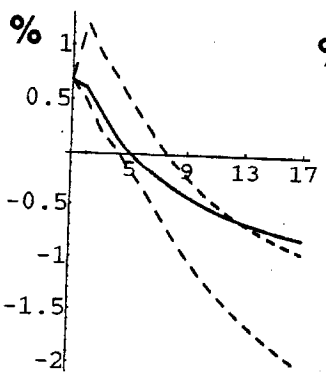
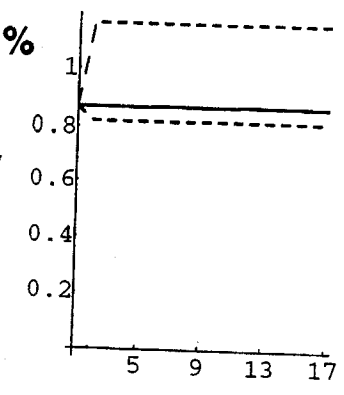
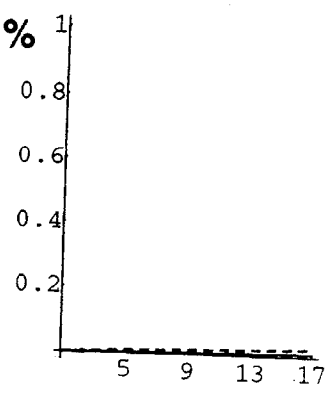
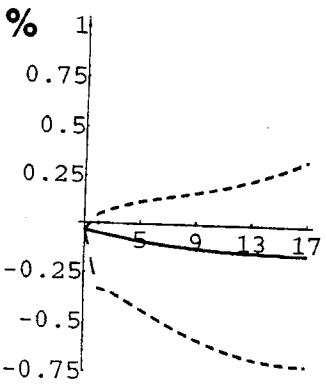
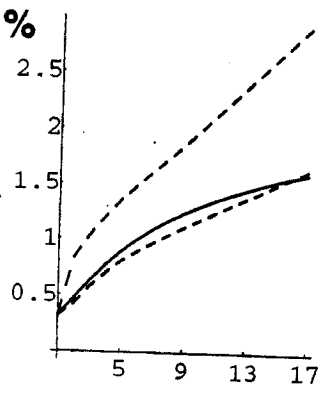
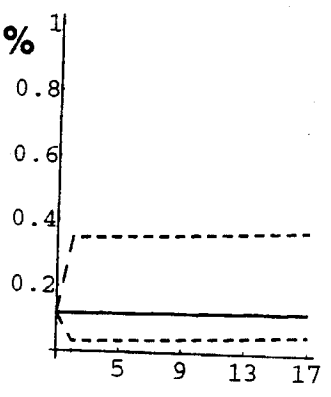
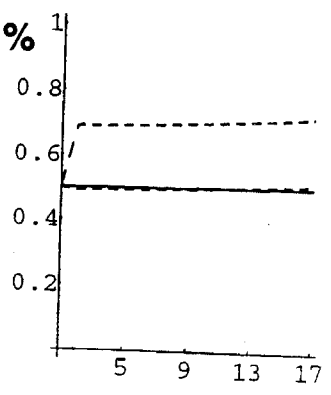


Function $\Phi_T(\nu)$ for the hyperparameter ν (four-variable US BVARTEC).



Impulse Response Functions

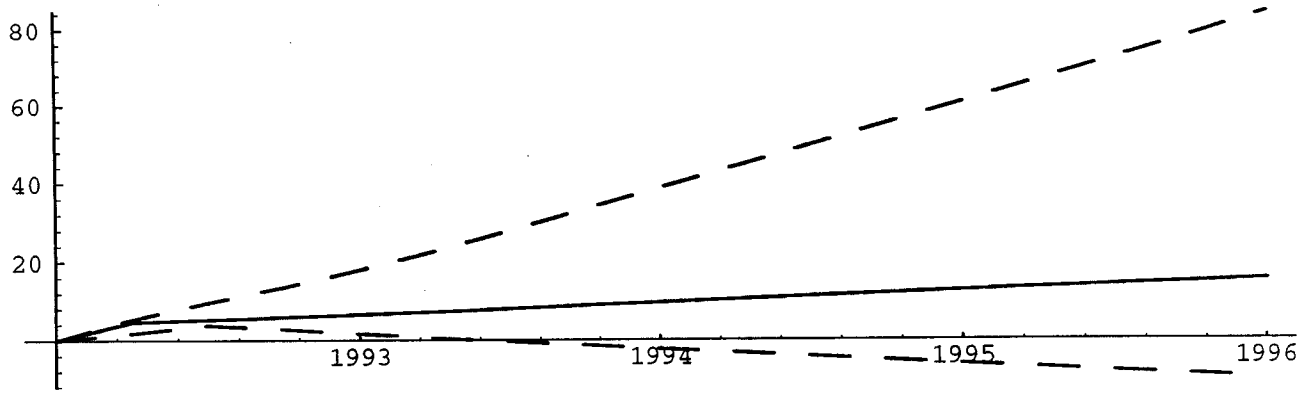
Response to innovation one standard deviation in size, using Cholesky-decomposition. Values in Percent of Level.

Response of \ Innovation to	Stocks	FedFunds	GNP82	GNPDefl
Stocks				
FedFunds				
GNP82				
GNPDefl				

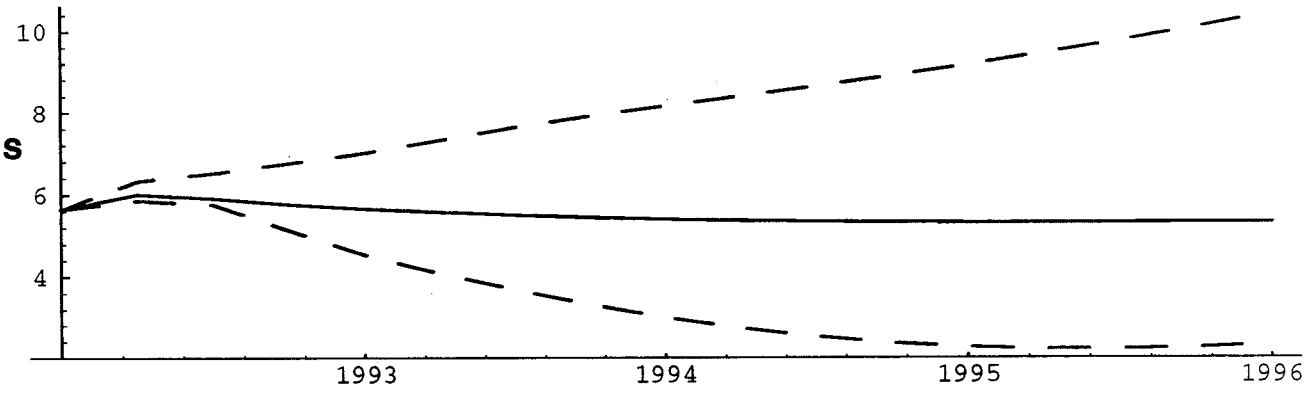
Forecast

The dotted lines represent the posterior coefficient uncertainty.
Everything is in cumulative Percent, except the Federal Funds Rate.

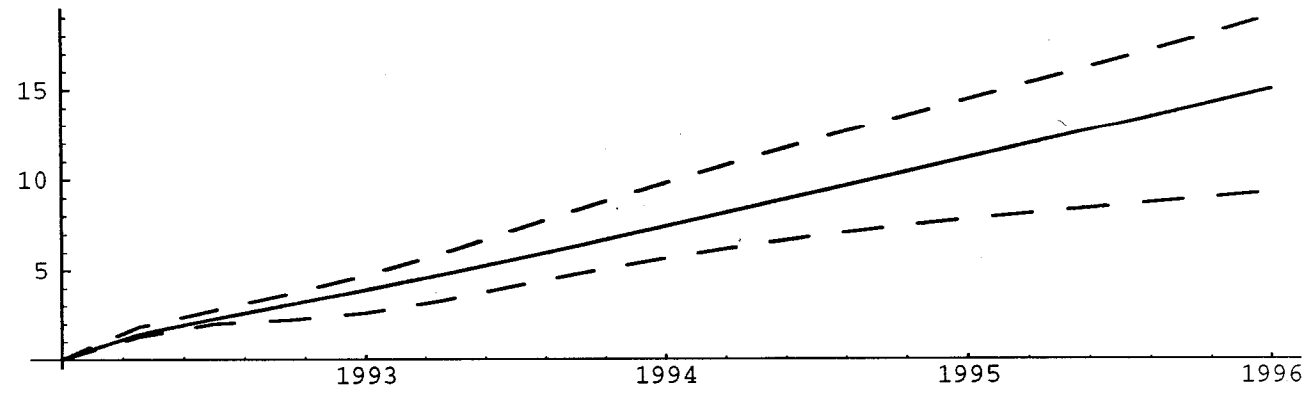
Stocks



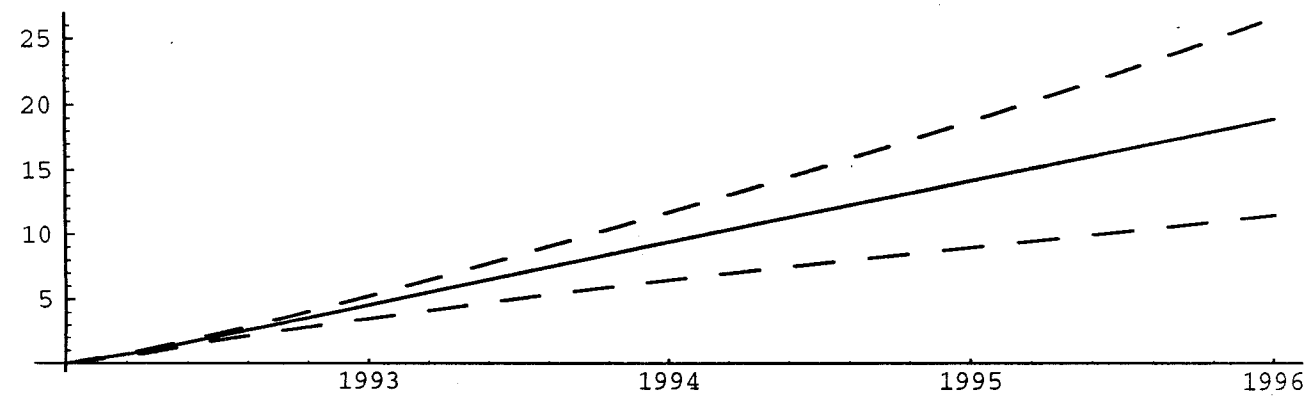
FedFunds



GNP82



GNP-Deflator



First Differences, Realized Forecast Errors and Predicted Standard Deviations.

The rather smooth lines in each graph are the predicted standard deviation for the one-step ahead prediction error.