Bayesian Analysis of Switching ARCH Models

Sylvia Kaufmann

University of Vienna, Department of Economics
Hohenstaufengasse 9, 1010 Vienna
Austria
(email: sylvia.kaufmann@univie.ac.at)

and

Sylvia Frühwirth-Schnatter [1]

Vienna University of Economics and Business Administration
Department of Statistics
Augasse 2-6, 1090 Vienna
Austria
(email: sfruehwi@isis.wu-wien.ac.at)

---

[1]Corresponding author

Bayesian Analysis of Switching ARCH Models

Sylvia Kaufmann [2] and Sylvia Frühwirth-Schnatter [3]

August 28, 2000

Abstract

We consider a time series model with autoregressive conditional heteroskedasticity that is subject to changes in regime. The regimes evolve according to a multistate latent Markov switching process with unknown transition probabilities, and it is the constant in the variance process of the innovations that is subject to regime shifts. The joint estimation of the latent process and all model parameters is performed within a Bayesian framework using the method of Markov Chain Monte Carlo simulation. We perform model selection with respect to the number of states and the number of autoregressive parameters in the variance process using Bayes factors and model likelihoods. To this aim, the model likelihood is estimated by the method of bridge sampling. The usefulness of the sampler is demonstrated by applying it to the data set previously used by Hamilton and Susmel who investigated models with switching autoregressive conditional heteroskedasticity using maximum likelihood methods. The paper concludes with some issues related to maximum likelihood methods, to classical model selection, and to potential straightforward extensions of the model presented here.

Keywords: Bayesian analysis, bridge sampling, MCMC estimation, model selection, switching ARCH-models

1 Introduction

The basic switching ARCH-model has been introduced independently in Cai (1994) and Hamilton and Susmel (1994) and has been generalized in Gray (1996). In these papers the basic switching ARCH-model is derived from the classical

[2]University of Vienna, Department of Economics, Hohenstaufengasse 9, 1010 Vienna, Tel. +43 1 4277-37430, email sylvia.kaufmann@univie.ac.at

[3]Vienna University of Economics and Business Administration, Department of Statistics, Augasse 2-6, 1090 Vienna, Tel. 43 1 31336-5053, email sfruehwi@isis.wu-wien.ac.at

ARCH-model

$$h_t = \gamma + \alpha_1 u_{t-1}^2 + \cdots + \alpha_m u_{t-m}^2, \tag{1}$$

where the parameter γ is time invariant, by assuming that γ changes over time. From now on we will use the notation γ_t rather than γ to emphasize its time varying nature:

$$h_t = \gamma_t + \alpha_1 u_{t-1}^2 + \cdots + \alpha_m u_{t-m}^2. \tag{2}$$

The classical ARCH-model (1) obviously is that special case of (2), where $\gamma_t \equiv \gamma$. Note that there exists an alternative way to parameterize the ARCH-model (1), namely:

$$u_t = \sqrt{\gamma} \cdot \tilde{u}_t, \qquad \tilde{u}_t = \sqrt{h_t} \cdot v_t,$$
$$h_t = 1 + \alpha_1 \tilde{u}_{t-1}^2 + \cdots + \alpha_m \tilde{u}_{t-m}^2. \tag{3}$$

Again we could introduce a time varying parameter γ_t rather than a constant parameter γ. It is easy to verify that such a model could be rewritten as:

$$u_t = \sqrt{\gamma_t h_t} \cdot v_t,$$
$$h_t = 1 + \alpha_1 \frac{u_{t-1}^2}{\gamma_{t-1}} + \cdots + \alpha_m \frac{u_{t-m}^2}{\gamma_{t-m}}. \tag{4}$$

For a constant parameter $\gamma_t \equiv \gamma$ the alternative parameterization is equivalent to parameterization (1). However, obviously model (2) and (4) are different, if γ_t is time dependent.

To get an identified model one needs further assumptions on the way how γ_t changes over time. In Cai (1994), Hamilton and Susmel (1994) and Gray (1996) changes of γ_t are described by the framework of Markov switching models introduced by Hamilton (1989). One assumes that γ_t takes one out of K different values according to a hidden Markov chain I_t taking values between 1 and K: $\gamma_t = \gamma_{I_t}$.

The switching ARCH-model of Hamilton and Susmel (1994) is obtained by introducing such a switching parameter into parameterization (4):

$$u_t = \sqrt{\gamma_{I_t} \cdot h_t} \cdot v_t,$$
$$h_t = 1 + \alpha_1 \frac{u_{t-1}^2}{\gamma_{I_{t-1}}} + \cdots + \alpha_m \frac{u_{t-m}^2}{\gamma_{I_{t-m}}}. \tag{5}$$

In the present paper we study the slightly different switching ARCH-model where the switching parameter is introduced into parameterization (2):

$$u_t = \sqrt{h_t} \cdot v_t,$$
$$h_t = \gamma_{I_t} + \alpha_1 u_{t-1}^2 + \cdots + \alpha_m u_{t-m}^2. \tag{6}$$

It is easy to verify that the switching ARCH-model of Cai (1994) which is restricted to $K = 2$ and uses the state dependent formulation $\gamma_t = \gamma_0 + S_t \gamma_1$, with S_t either 0 or 1, is a special case of (6). Gray (1996) also uses this parameterization, and introduces additionally switching into the coefficients of the ARCH-process.

In the present paper we carry out a fully Bayesian analysis of the basic switching ARCH-model (6). The joint estimation of the latent Markov switching process and all model parameters for fixed orders K and m of the switching ARCH-model is performed using the method of Markov Chain Monte Carlo simulation (see e.g. Smith and Roberts, 1993 for a general introduction to MCMC methods). The design of suitable MCMC methods to generate a sample from the posterior of a switching ARCH-model has not been studied before, previous papers used maximum likelihood methods for parameter estimation. We combine well-known results for multi move sampling of a hidden Markov process (Carter and Kohn, 1994; Shephard, 1994; Chib, 1996) with recent results available for MCMC estimation of ARCH-models (Nakatsuma, 2000). One iteration of the sampler involves first a multi-move step to simulate the latent process out of its conditional distribution. The Gibbs sampler can then be used to simulate the parameters, in particular the transition probabilities, for which the full conditional posterior distribution is known. For most parameters, however, the full conditionals do not belong to any well-known family of distributions. The simulations are then based on the Metropolis-Hastings algorithm with carefully chosen proposal densities.

For practical volatility modeling the order parameters, however, will be unknown. We perform model selection with respect to the number of states K and the number of autoregressive parameters m in the variance process using Bayes factors and model likelihoods. We derive an estimate of the model likelihood for a given order K and m from the MCMC output by combining the candidate's formula (Chib, 1995) with importance sampling, where the importance density is constructed from the MCMC sample. An alternative Bayesian approach, not pursued in the present paper, would be the inclusion of K and m into the MCMC scheme along the lines of the jump-diffusion-approach outlined by Richardson and Green (1997).

The next section presents the basic switching ARCH model along with the prior specification. MCMC Estimation is discussed in detail in section 3, followed by issues on Bayesian model selection in section 4. The usefulness of the Bayesian approach is illustrated in section 5 by reanalyzing the data set used in Hamilton and Susmel (1994). The paper concludes in section 6 with some comments on maximum likelihood versus Bayesian methods, some issues related to classical model selection, and with some potential straightforward extensions of the methods presented here.

2 The Basic Switching ARCH Model

2.1 Model formulation

Let $y^N = (y_1, \dots, y_N)$ denote a sequence of N observations y_t. We assume that the observations y_t are generated by the following model:

$$y_t = z_t^{'}\beta + u_t, \tag{7}$$

where the predictor z_t may contain exogenous variables as well as lagged values of y_t. u_t has the following normal distribution:

$$u_t = \sqrt{h_t} \cdot v_t, \tag{8}$$

where v_t is an iid normal sequence with zero mean and unit variance: $E(v_t) = 0, E(v_t^2) = 1$. We assume that h_t follows the switching ARCH model defined in (6):

$$\begin{aligned} u_t &= \sqrt{h_t} \cdot v_t, \\ h_t &= \gamma_{I_t} + \alpha_1 u_{t-1}^2 + \cdots + \alpha_m u_{t-m}^2. \end{aligned} \tag{9}$$

As in Cai (1994) and Hamilton and Susmel (1994), I_t is a latent discrete variable modeled as a stationary, irreducible Markov process with discrete state space $\{1, \dots, K\}$ and unknown transition probabilities $\eta_{ij} = Pr\{I_t = j | I_{t-1} = i\}$. I_0 is assumed to have some starting distribution $Pr\{I_0 = i_0\} = \pi(i_0)$. For convenience, we refer to the K-state, m order Markov switching ARCH model defined by (9) as SWARCH(K, m).

Subsequently we will use the following notations: $\eta = (\eta_{1\cdot}, \dots, \eta_{K\cdot})$, where $\eta_{i\cdot} = (\eta_{i1}, \dots, \eta_{iK})$ is the conditional transition distribution, collects all probabilities of the transition matrix of I_t; $\phi = (\alpha, \beta, \gamma, \eta)$ summarizes all unknown model parameters, with $\alpha = (\alpha_1, \dots, \alpha_m)$ and $\gamma = (\gamma_1, \dots, \gamma_K)$; and finally $I^N = (I_0, I_1, \dots, I_N)$ denotes the whole sequence of switching variables and correspondingly i^N denotes a realization of I^N. The SWARCH(K, m)-model has the structure of a hierarchical model including latent variables:

1. Conditionally on known realizations i^N of the switching process I^N and on a known model parameter ϕ the conditional distribution of $y_1, \dots, y_N$ factorizes in the following way:

$$f(y_1, \dots, y_N | i^N, \phi) = \prod_{t=1}^{N} f(y_t | y^{t-1}, \alpha, \beta, \gamma_{i_t}), \tag{10}$$

6

where the one-step ahead predictive densities are Gaussian:

$$f(y_t|y^{t-1}, \alpha, \beta, \gamma_{i_t}) = \sqrt{\frac{1}{2\pi h_t(\alpha, \beta, \gamma_{i_t}, y^{t-1})}} \exp\left(\frac{-(y_t - z_t'\beta)^2}{2h_t(\alpha, \beta, \gamma_{i_t}, y^{t-1})}\right). \quad (11)$$

2. For each ϕ, the latent switching process I^N is a Markov chain with transition matrix depending on $\eta = (\eta_{1.}, \ldots, \eta_{K.})$. The density $\pi(i^N|\eta)$ of the prior distribution of I^N w.r.t. the counting measure is equal to

$$\pi(i^N|\phi) = \pi(i^N|\eta) \propto \prod_{t=1}^{N} \eta_{i_{t-1}, i_t} \pi(i_0|\phi) = \prod_{j=1}^{K}\prod_{i=1}^{K} \eta_{ij}^{N_{ij}} \pi(i_0),$$

$$N_{ij} = \#\{I_t = j|I_{t-1} = i\}.$$

3. ϕ has a prior distribution with density $\pi(\phi)$.

The factorization (10) shows that the observation density $f(y_t|y^{t-1}, i^N, \phi)$ depends only on the present value i_t of the switching process I^N. This property results from introducing the switching parameter directly into the parameterization of h_t given by (2). For the parameterization (4) which has been used by Hamilton and Susmel (1994) the conditional density $f(y_t|y^{t-1}, i^N, \phi)$ depends on i_t as well as on the lagged values $i_{t-1}, \ldots, i_{t-m}$.

Note that the first two layers of the model are sufficient to compute the marginal likelihood $L(y^N|\phi)$ which could be maximized such as in Hamilton and Susmel (1994) to obtain ML-estimates of the model parameters ϕ. For a complete Bayesian analysis of the model the third layer specifying the prior $\pi(\phi)$ on the model parameters ϕ has to be added (see subsection 2.3 for further details).

We conclude this section with discussing the marginal model implied by the basic SWARCH(K, m)-model. Under the well-known condition that the absolute values of all eigenvalues of the matrix A, where A is defined from the transition matrix η of I_t by $A_{jl} = \eta_{lj} - \eta_{Kj}$, $j, l = 1, \ldots, K-1$, are smaller than one, the switching process I_t has got a stationary distribution $\eta^\star = (\eta_1^\star, \ldots, \eta_K^\star)$, where $(\eta_1^\star, \ldots, \eta_{K-1}^\star)' = (I - A)^{-1}(\eta_{K1}, \ldots, \eta_{K,K-1})'$ and $\eta_K^\star = 1 - \sum_{j=1}^{K-1} \eta_j^\star$. Then the marginal distribution of y_t, where the latent process I^N is integrated out, is a mixture of normal distributions with weights $\eta_j^\star$:

$$F(y_t|\mathcal{F}_{t-1}) = \sum_{j=1}^{K} \eta_j^\star \Phi\left(\frac{y_t - z_t'\beta}{h_t}\right),$$

where $F(y_t|\mathcal{F}_{t-1})$ denotes the distribution function of Y_t given information up to $t-1$, and $\Phi(\cdot)$ is the standard Gaussian distribution function. Thus the basic

7

SWARCH(K, m)-model could be regarded as a special case of the MAR-ARCH-model discussed in Wong and Li (1999). From Theorem 2 of that paper we obtain covariance stationarity of the process $u_t = y_t - z_t' \beta$ under the necessary and sufficient condition that the roots of the equation

$$1 - \sum_{l=1}^{m} \alpha_l x^l = 0 \tag{12}$$

are all outside the unit circle. The second unconditional moment of u_t is given by:

$$E(u_t^2) = \frac{\sum_{j=1}^{K} \gamma_j \eta_j^\star}{1 - \sum_{l=1}^{m} \alpha_l}. \tag{13}$$

For the SWARCH$(2, m)$-process this result has been proven already by Cai (1994). Further applications of the results derived in Wong and Li (1999) lead to the conclusion that the third order moment of u_t is zero, impling symmetry of the unconditional distribution, and that for a SWARCH$(K, 1)$-model u_t^2 is covariance stationary if $\alpha_1^2 < 1/3$ and the autocorrelation function of u_t^2 is given by $E(u_t^2 u_{t-l}^2) = \alpha_1^l$.

2.2 Identifiability

It is well known that for any model including a latent, discrete structure, an identifiability problem is present, since the labeling of the states of the switching variable I_t can be permuted without changing the (marginal) likelihood: $\exists \phi \neq \tilde{\phi}$ such that $L(y^N | \phi) = L(y^N | \tilde{\phi})$ (see e.g. Frühwirth-Schnatter (2000) for a recent discussion of this issue). Therefore the unconstrained SWARCH(K, m)-model is not identifiable in a strict sense. It is possible to estimate quantities from the unconstrained SWARCH(K, m)-model which are invariant to relabeling, the most important examples being the time-varying parameter γ_{I_t} and the volatility h_t defined in (9). Invariance of γ_{I_t} follows from:

$$\gamma_{I_t} = \gamma_1 S_t^1 + \ldots + \gamma_K S_t^K = \gamma_{\rho(1)} S_t^{\rho(1)} + \ldots + \gamma_{\rho(K)} S_t^{\rho(K)},$$

where $\rho(1), \ldots, \rho(K)$ is an arbitrary permutation of $1, \ldots, K$. Thus extracting volatility estimates from a SWARCH(K, m)-model is possible without identifying a unique labeling.

If focus lies on estimating the variance γ_i of each state i individually, and to estimate the probability of being in a certain state at a certain time, however, it is necessary to introduce a unique labeling. To render the model identified usually an identifiability constraint is put on the state specific parameters. For a

SWARCH(K, m)-model a standard constraint based on the notion that the first state is the state of lowest volatility whereas the last state is the state of highest volatility would be a constraint on the state specific variances $\gamma_1, \ldots, \gamma_K$:

$$\gamma_1 < \ldots < \gamma_K. \tag{14}$$

Such a standard constraint, however, may turn out to be a poor constraint especially for a higher number of classes. We found in our case study in section 5 within the 4-state model two states of about medium volatility which were difficult to separate through γ_j. We found that the main difference between these states is primarily not the level of volatility but the persistence of remaining in the current state. For such a model a suitable identifiability constraint would be:

$$\gamma_1 < \min(\gamma_2, \ldots, \gamma_4), \quad \gamma_4 > \max(\gamma_2, \gamma_3), \qquad \eta_{22} > \eta_{33}. \tag{15}$$

2.3   Choice of the prior

In this paper the focus lies on Bayesian estimation in situations where we in general do not have strong prior information. From a theoretical point of view, being completely non-informative about ϕ is possible only for state independent parameters such as α and β. Theoretically, being non-informative about $\gamma = (\gamma_1, \ldots, \gamma_K)$ and η is not possible, as improper priors on γ and η result in improper posteriors. There is always the possibility that no observation is allocated to a certain state, say j. Improper priors on γ_j and η_j. will then lead to improper posterior distributions.

In what follows we will make use of the following independence assumptions concerning the prior distribution of ϕ: $\pi(\phi) = \pi(\beta)\pi(\alpha)\pi(\gamma, \eta)$. As common with regression types models such as (7) we assume a normal prior $\mathcal{N}(b_0, B_0)$ for the regression parameter β. The choice of the prior on α should reflect important constraints such as that all α_l are positive and $\sum_{l=1}^{m} \alpha_l$, which is the persistence parameter, is smaller than 1.[4] Therefore we assume that (α, α_{m+1}), where $\alpha_{m+1} = 1 - \sum_{l=1}^{m} \alpha_l$, follows a Dirichlet distribution $D(a_1, \ldots, a_{m+1})$. Note that the Dirichlet distribution ensures that the constraints mentioned above are fulfilled with probability 1. Furthermore this choice implies that the marginal distribution of the persistence parameter $\sum_{l=1}^{m} \alpha_l$ is a Beta distribution and could therefore be regarded as an extension of the Beta prior imposed in Kim *et al.* (1998) on the persistence parameter of a GARCH(1,1)-model.

Concerning the priors for the state dependent parameter η and γ we discuss two possible choices. First we could assume that the parameters corresponding to

---

[4] We want to thank the referee for drawing our attention to the fact that the normal prior $N(0, \kappa I)$ we assumed in a previous version of the paper has been a particularly poor choice in this respect.

the various states are independent apriori and the hyperparameters of the state specific priors are the same for all states. A natural choice for the distribution of each γ_i is an inverse gamma prior $IG(g_0, G_0)$, with g_0 and G_0 being state independent, whereas it is standard to assume that each conditional transition distributions $\eta_{i\cdot}$ follows a Dirichlet distribution $D(e_1, \ldots, e_K)$, with $e_1, \ldots, e_K$ being state independent. This choice leads to a symmetric prior in the sense that the resulting prior is invariant to relabeling the states and gives equal probability to each labeling pattern. The prior integrates to 1 over the unconstrained parameter space and to $1/K!$ over each subspaces corresponding to a unique labeling.

This prior may be unsatisfactory, if we have vague prior ideas concerning the differences between the states. With a state independent prior we could not include, for instance, the prior belief, that one state correspond to a persistent, low volatility state, whereas another state corresponds to a non-persistent, high volatility state. To include such information we could take the priors used above and make the hyperparameters state dependent: $IG(g_{0,i}, G_{0,i}), D(e_{i1}, \ldots, e_{iK})$, $i = 1, \ldots, K$. The problem is now, how to associate these state dependent priors with the various components of γ and η. Based on some apriori labeling we could connect each state dependent prior with a certain component through $\gamma_i \sim IG(g_{0,i}, G_{0,i})$ and $\eta_{i\cdot} \sim D(e_{i1}, \ldots, e_{iK})$. This strategy, however, leads to a prior which is no longer invariant to relabeling. To preserve invariance of the prior, on the one hand, and to include state specific information on the other hand, we use the following mixture prior:

$$\pi(\gamma, \eta) = 1/K! \sum_{m=1}^{K!} \prod_{i=1}^{K} IG(\gamma_i, g_{0,\rho_m(i)}, G_{0,\rho_m(i)}) \prod_{i=1}^{K} D(\eta_{i\cdot}, e_{\rho_m(i),\rho_m(1)}, \ldots, e_{\rho_m(i),\rho_m(K)}), (16)$$

where $\rho_m(1), \ldots, \rho_m(K)$, $m = 1, \ldots, K!$ correspond to the $K!$ possible ways of relabeling the states with $\rho_1(\cdot)$ being the identity. With such a prior is possible to express, for instance, the belief that one state has higher variance and lower persistence than another state without being specific which state this will be. Note that this prior implies apriori dependence between γ and η whenever the hyperparameters are actually different. If all hyperparameters are the same, it collapses to the state independent prior discussed above. The mixture prior (16) is by definition invariant to relabeling the states, gives equal probability to each labeling pattern, integrates to 1 over the unconstrained parameter space and to $1/K!$ over each subspaces corresponding to a unique labeling.

3 MCMC Estimation of the Basic Switching ARCH-Model

In what follows we are going to estimate the augmented parameter vector $\psi = (\phi, I^N)$ by sampling from the posterior density $\pi(\psi|y^N)$ by means of MCMC methods. We combine well-known results for multi move sampling of a hidden Markov process (Carter and Kohn, 1994; Shephard, 1994; Chib, 1996) with recent results available for MCMC estimation of ARCH-models (Nakatsuma, 2000). Doing so we end up with the following sampling scheme.

Blocked MCMC Sampling for the SWARCH(K, m)-model. Generate the various blocks of ψ from the following conditional densities:

(i) $\pi(i^N|\alpha, \beta, \gamma, \eta, y^N)$
(ii) $\pi(\eta|i^N, \gamma)$
(iii) $\pi(\beta|\alpha, \gamma, i^N, y^N)$
(iv) $\pi(\gamma, \alpha|\beta, \eta, i^N, y^N)$

Sampling I^N and η. Step (i) and (ii) are standard steps occurring in the MCMC estimation of models including a latent Markov switching variable and are easily adapted to the framework of a SWARCH-model. Step (i) is carried out in a multi move manner as in Chib (1996). First we run a (forward) filter to compute $\pi(i_t|\phi, y^t)$ starting for $t = 1$ from the prior distribution $\pi(i_0)$. This step is straightforward as the observation density $f(y_t|y^{t-1}, i^N, \phi)$ depends on the past i^N through i_t, only:

$$\pi(i_t|\phi, y^t) \propto f(y_t|y^{t-1}, \gamma_{i_t}, \alpha, \beta)\pi(i_t|\phi, y^{t-1}) \tag{17}$$

where the predictive density $f(y_t|y^{t-1}, \gamma_{i_t}, \alpha, \beta)$ is equal to the observation density appearing in (11) and $\pi(i_t|\phi, y^{t-1})$ is given by extrapolation:

$$\pi(i_t|\phi, y^{t-1}) = \sum_{i_{t-1}=1}^{K} \pi(i_{t-1}|\phi, y^{t-1})\eta_{i_{t-1}, i_t}.$$

Given the filter probabilities we run a backward sampler starting from $t = N$ with sampling i_N from $\pi(i_N|\phi, y^N)$. For $t = N - 1, \dots, 0$ we sample from $\pi(i_t|i_{t+1}, \dots, i_N, \phi, y^N)$ which is given by:

$$\pi(i_t|i_{t+1}, \dots, i_N, \phi, y^N) = \pi(i_t|i_{t+1}, \phi, y^t) \propto \pi(i_t|\phi, y^t)\eta_{i_t, i_{t+1}}$$

One way to sample i_t from some (eventually non-normalized) distribution $p_{i_t} = \pi(i_t|\cdot)$ is to sample from the conditional probabilities $q_j = Pr\{I_t = j|I_t \geq j, \cdot\}$,

which are given by

$$q_j = Pr\{I_t = j | I_t \geq j, \cdot\} = p_j / \sum_{l=j}^{K} p_l.$$

This is carried out by starting from $j = 1$ by sampling a uniform number U. If $U \leq q_j$, then $i_t = j$ and stop sampling. Otherwise increase j to $j + 1$ and continue sampling another uniform number U.

Sampling η is completely standard, if the hyperparameters of the Dirichlet prior on η are state independent. In this case the conditional posterior $\pi(\eta | \psi, y^N)$ depends only on the prior structure of I^N and the conditional transition distributions $\eta_{1\cdot}, \ldots, \eta_{K\cdot}$ are independent *a posteriori* given I^N with each $\eta_{i\cdot}$ following a Dirichlet distribution $D(e_{i1} + N_{i1}, \ldots, e_{iK} + N_{iK})$:

$$\pi(\eta_{i\cdot} | i^N) \propto \prod_{j=1}^{K} \eta_{ij}^{e_{ij} + N_{ij} - 1}, \qquad N_{ij} = \#\{I_t = j | I_{t-1} = i\}.$$

The easiest way to sample from this Dirichlet posterior distribution $D(e_{i1} + N_{i1}, \ldots, e_{iK} + N_{iK})$ is to sample K independent non-normalized values $\tilde{\eta}_{i1}, \ldots, \tilde{\eta}_{iK}$ from the Gamma distributions $G(e_{i1} + N_{i1}, 1), \ldots, G(e_{iK} + N_{iK}, 1)$ and to normalize: $\eta_{ij} = \tilde{\eta}_{ij} / \sum_{j=1}^{K} \tilde{\eta}_{ij}$. Sampling is slightly more envolved for the mixture prior (16) with state dependent hyperparameters. In this case the posterior $\pi(\eta | \psi, y^N)$ is a mixture of Dirichlet distributions and we use the Metropolis-Hastings algorithm (Hastings, 1970) with the proposal density

$$q(\eta) = \prod_{i=1}^{K} D(\eta_{i\cdot}, e_{i1} + N_{i1}, \ldots, e_{iK} + N_{iK}) \tag{18}$$

to sample from this distribution. A proposal value $\eta^\star$ from $q(\eta)$ is accepted with probability

$$\alpha(\eta^\star | \eta^{(old)}) = \min\left\{ \frac{\prod_{i=1}^{K} D(\eta_{i\cdot}^\star, e_{i1}, \ldots, e_{iK}) \pi(\gamma, \eta^{(old)})}{\prod_{i=1}^{K} D(\eta_{i\cdot}^{(old)}, e_{i1}, \ldots, e_{iK}) \pi(\gamma, \eta^\star)}, 1 \right\},$$

where $\pi(\gamma, \eta)$ is equal to the mixture prior (16). If $\eta^\star$ is rejected, the new sampled value is taken to be equal to the old value $\eta^{(old)}$.

Sampling the regression parameters and the ARCH parameters. MCMC sampling of the parameters (α, β) of a regression model with ARCH errors has been studied before only for the classical ARCH model without switching. Bauwens and Lubrano (1998) use the griddy Gibbs sampler (Ritter and Tanner, 1992) for sampling the parameters of a GARCH(1,1) in a single move manner. Single

12

move samplers, however, might exhibit slow convergence if the parameters are highly correlated. This problem is avoided in Kim *et al.* (1998) who use single move ARMS rejection sampling (Gilks *et al.*, 1995) by reparameterizing the GARCH(1,1)-model in terms of a persistence parameter and a moving average type parameter. Multi move sampling of all ARCH-parameters using a multivariate Metropolis-Hastings algorithm is discussed in Nakatsuma (2000). In the present paper we also use a multivariate Metropolis-Hastings algorithm within step (iii) and (iv) which for $K = 1$ is equivalent to the sampler suggested in Nakatsuma (2000).

No direct method of sampling the regression parameters β exists, even if the ARCH parameters (α, γ) are known. Recall that we started from model (7) with heteroskedastic error term: $V(u_t) = h_t$. The crucial point for inference on the regression parameter β is the fact that h_t itself depends on β through the lagged residuals $u_{t-1}, \ldots, u_{t-m}$. If h_t were independent of β, we would be faced with a model with heteroskedastic, but independent errors and the conditional posterior of β would be a normal density. In the present paper we use a Metropolis-Hastings-algorithm with a proposal density which is derived by substituting $h_t(\beta)$ by the estimate $\hat{h}_t = h_t(\beta^{(old)})$. $\beta^{(old)}$ is the previously sampled value. This leads to the following proposal density $q(\beta|\beta^{(old)})$:

$$q(\beta|\beta^{(old)}) \sim \mathcal{N}(\beta, b_N(\beta^{(old)}), B_N(\beta^{(old)})), \tag{19}$$

$$b_N(\beta) = B_N(\beta)(Z^{'}W(\beta)y + B_0^{-1}b_0), \quad B_N(\beta) = (Z^{'}W(\beta)Z + B_0^{-1})^{-1},$$

$$W(\beta) = \text{Diag}\left(\ w_1(\beta)\ \cdots\ w_N(\beta)\ \right), \quad w_t(\beta)^{-1} = h_t(\beta) = \gamma_{i_t} + \sum_{l=1}^{m} \alpha_l \cdot (y_{t-l} - z^{'}_{t-l}\beta)^2.$$

y and Z are the observation vector and the predictor matrix of model (7), respectively. A value $\beta^{\star}$ sampled from $q(\beta|\beta^{(old)})$ is accepted with probability

$$\alpha(\beta^{\star}|\beta^{(old)}) = \min\left\{ \frac{\pi(\beta^{\star}|\alpha, \gamma, i^N, y^N)}{\pi(\beta^{(old)}|\alpha, \gamma, i^N, y^N)} \cdot \frac{q(\beta^{(old)}|\beta^{\star})}{q(\beta^{\star}|\beta^{(old)})}, 1 \right\},$$

where $\pi(\beta|\alpha, \gamma, i^N, y^N)$ is proportional to $f(y_1, \ldots, y_N|\beta, \alpha, \gamma, i^N)\pi(\beta)$ (see (10)). If $\beta^{\star}$ is rejected, the new sampled value is taken to be equal to the old value $\beta^{(old)}$.

For joint sampling of γ and α we reformulate the switching ARCH-models as generalized linear model. From $u_t = \sqrt{h_t} \cdot v_t$, where v_t is standard normal, we obtain:

$$\frac{u_t^2}{h_t} = v_t^2, \tag{20}$$

where v_t^2 is a χ_1^2 random variable with one degree of freedom. v_t^2 may be expressed as $v_t^2 = 1 + \varepsilon_t$, where $E(\varepsilon_t) = 0$ and $V(\varepsilon_t) = 2$. Therefore (20) can be rewritten

as:

$$u_t^2 = h_t + h_t \varepsilon_t.$$

As h_t is linear in α and γ, we obtain the following linear regression model with independent, however non-normal and heteroskedastic errors:

$$u_t^2 = \gamma_1 S_t^1 + \ldots + \gamma_K S_t^K + u_{t-1}^2 \alpha_1 + \ldots + u_{t-m}^2 \alpha_m + h_t \varepsilon_t, \qquad (21)$$

where we used a $0/1$-coding for I_t by introducing dummy variables $S_t^1, \ldots, S_t^K$ which for each $j = 1, \ldots, K$ and each t are defined by $S_t^j = 1$ iff $I_t = j$ and $S_t^j = 0$ otherwise. Again we are faced with a regression model with heteroskedastic, however non-normal errors. A simple multivariate proposal for (γ, α) may be derived from model (21), if the variance of the error term, which is equal to $2 \cdot h_t^2$, is estimated from the previous draw of α and γ by $\hat{h}_t = h_t((\gamma, \alpha)^{(old)})$ and non-normal errors are substituted by normal ones:

$$\begin{aligned}
q(\gamma, \alpha | (\gamma, \alpha)^{(old)}) &\sim \mathcal{N}(c_N((\gamma, \alpha)^{(old)}), C_N((\gamma, \alpha)^{(old)})), \qquad (22) \\
c_N(\gamma, \alpha) &= C_N(\gamma, \alpha)(Z' W(\gamma, \alpha)\tilde{u} + C_0^{-1} c_0), \\
C_N(\gamma, \alpha) &= (Z' W(\gamma, \alpha) Z + C_0^{-1})^{-1}, \\
W(\gamma, \alpha) &= \operatorname{diag}(w_1(\gamma, \alpha) \cdots w_N(\gamma, \alpha)), \\
w_t(\gamma, \alpha)^{-1} &= 2 \cdot h_t(\gamma, \alpha)^2 = 2 \cdot (\gamma_{i_t} + \sum_{l=1}^{m} \alpha_l u_{t-l}^2)^2.
\end{aligned}$$

$\tilde{u}$ and Z are the observation vector and the predictor matrix of model (21): $\tilde{u}_t = u_t^2$ and $Z_{t\cdot} = (S_t^1 \cdots S_t^K \; u_{t-1}^2 \cdots u_{t-m}^2)$. For each MCMC move it may happen that none of the sampled indicators $(I_1, \ldots, I_N)$ takes a certain value, say j. Then at the next move the approximate model (21) will not contain any information on γ_j. In order to avoid problems with sampling γ_j in such a case the prior on γ has been included into the construction of the proposal density through the parameters c_0 and C_0^{-1} containing the mode and the information matrix of each inverted gamma prior $IG(g_{0,i}, G_{0,i})$:

$$c_0 = \begin{pmatrix} \frac{G_{0,1}}{1+g_{0,1}} \\ \vdots \\ \frac{G_{0,K}}{1+g_{0,K}} \\ 0_{m \times 1} \end{pmatrix}, \qquad C_0^{-1} = \begin{pmatrix} \operatorname{diag}\left(\frac{(1+g_{0,1})^3}{G_{0,1}^2}, \ldots, \frac{(1+g_{0,K})^3}{G_{0,K}^2}\right) & 0_{K \times m} \\ 0_{m \times K} & 0_{m \times m} \end{pmatrix}.$$

As all observations contribute to the information on α we did not include information from the prior $\pi(\alpha)$ into the construction of the proposal. A proposal value $(\gamma, \alpha)^\star$ sampled from $q(\gamma, \alpha | (\gamma, \alpha)^{(old)})$ is accepted with probability

$$\alpha((\gamma, \alpha)^\star | (\gamma, \alpha)^{(old)}) = \min\left\{ \frac{\pi((\gamma, \alpha)^\star | \beta, \eta, i^N, y^N)}{\pi((\gamma, \alpha)^{(old)} | \beta, \eta, i^N, y^N)} \frac{q((\gamma, \alpha)^{(old)} | (\gamma, \alpha)^\star)}{q((\gamma, \alpha)^\star | (\gamma, \alpha)^{(old)})}, 1 \right\}, (23)$$

where $\pi(\gamma, \alpha | \beta, \eta, i^N, y^N)$ proportional to $f(y_1, \ldots, y_N | \beta, \alpha, \gamma, i^N) \pi(\gamma, \eta) \pi(\alpha)$ (see (11)). If $(\gamma, \alpha)^\star$ is rejected, the new sampled value is taken to be the old value $(\gamma, \alpha)^{(old)}$. Note that although parameters sampled from the normal proposal might violate important parameter restrictions such as $\gamma_j, \alpha_l > 0$, the inclusion of the prior into the acceptance probability (23) will lead to an automatic rejection of such values.

We may run through the steps (i)–(iv) in two ways, depending on whether we estimate the unconstrained or a constrained model, using the permutation sampler of Frühwirth-Schnatter (2000), where each iteration is concluded by a permutation ρ of the current labeling:

$$
\begin{aligned}
(\gamma_1, \ldots, \gamma_K) &:= (\gamma_{\rho(1)}, \ldots, \gamma_{\rho(K)}), \\
i^N &:= (\rho(i_0), \rho(i_1), \ldots, \rho(i_N)), \\
\eta &:= (\tilde{\eta}_{1\cdot}, \ldots, \tilde{\eta}_{K\cdot}), \quad \tilde{\eta}_{i\cdot} := (\eta_{\rho(i),\rho(1)}, \ldots, \eta_{\rho(i),\rho(K)}).
\end{aligned}
\tag{24}
$$

If we want to estimate the unconstrained model, we select this permutation randomly from all $K!$ possible ones. For a random permutation sampler under a prior with state dependent parameters, the best acceptance rates in the Metropolis-Hastings algorithm are obtained, if the state dependent parameters are permuted as well. If we want to estimate the model under an identifiability constraint such as (14) a permutation is applied only, if the identifiability constraint is violated and in this case the permutation is selected in such a way that the identifiability constraint is fulfilled (for more details see Frühwirth-Schnatter, 2000).

4 Issues in Model Selection

For practical volatility modeling the two order parameters K and m will be unknown and some comments are in order concerning model selection. In the present paper we use Bayes factors and model likelihoods to infer on K and m. For fixed K and m the model likelihood is defined by:

$$
L(y^N | K, m) = \int f(y_1, \ldots, y_N | \psi, K, m) \pi(\psi) \nu(d\psi),
\tag{25}
$$

with r.t. an appropriate measure ν. Analytical integration of (25) with respect to the whole parameter $\psi = (\phi, I^N)$ is not possible, however, the dimension of integration can be reduced substantially, as it is possible to integrate analytically with respect to I^N:

$$
L(y^N | K, m) = \int L(y_1, \ldots, y_N | \phi, K, m) \pi(\phi) d\phi,
\tag{26}
$$

where an explicit formula for the marginal likelihood $L(y_1, \ldots, y_N | \phi, K, m)$ is available. The computation of the marginal likelihood follows along the lines of Hamilton and Susmel (1994), however due to the alternative parameterization of the model the marginal likelihood is easier to compute than for the SWARCH model defined in Hamilton and Susmel (1994). For more details see the appendix.

Model likelihoods may be estimated from the MCMC output using methods such as the candidate's formula (Chib, 1995), importance sampling based on mixture approximations (Frühwirth-Schnatter, 1995), combining MCMC simulations and asymptotic approximation (Gelfand and Dey, 1994; DiCiccio *et al.*, 1997) and bridge sampling (Meng and Wong, 1998). The application of these methods to computing the model likelihood for switching and mixture models has been discussed in detail in Frühwirth-Schnatter (1999) with the following main results: first, estimation of the model likelihood turns out to be sensitive to the problem of label switching. Especially the candidate's formula (Chib, 1995) should not be applied, if label switching is present, and is a suitable estimation method only for identified models. Second, it is not necessary to identify the model in order to compute the model likelihood. Third, the best result with the lowest standard error is obtained by using the method of bridge sampling where a MCMC sample obtained by random permutation sampling is combined with an iid sample from an importance density $q(\phi)$. In the present paper we will apply bridge sampling to compute the model likelihood of a SWARCH-model. The importance density $q(\phi)$ is constructed in an unsupervised manner from the MCMC output $(\psi^{(1)}, \ldots, \psi^{(M)})$ of the random permutation sampler using a mixture of the following conditional densities:

$$q(\phi) = 1/M_L \sum_{m=1}^{M_L} q(\eta | \psi^{(m)}) q(\beta | \psi^{(m)}, y^N) q(\gamma, \alpha | \psi^{(m)}, y^N), \qquad (27)$$

where $q(\eta | \psi^{(m)})$ is equal to the Dirichlet proposal density in (18), whereas $q(\beta | \psi^{(m)}, y^N)$ and $q(\gamma, \alpha | \psi^{(m)}, y^N)$ are equal to the normal proposal densities in (19) and (22), respectively. Note that the components $q(\gamma, \alpha | \psi^{(m)}, y^N)$ are not normalized over the range of admissible parameters (γ, α). The normalising constants are computed simultaniously with sampling from each component. For further details the reader is referred to Frühwirth-Schnatter (1999).

5 A Case Study

5.1 Data and model specification

For illustration we will apply the constrained permutation sampler for the SWARCH(K, m) model presented in the previous section to the data set investigated in Hamilton and Susmel (1994).[5] The series originates from the CRISP data tapes and consists of a value-weighted portfolio of stocks traded on the New York Stock Exchange. Figure 1 shows weekly returns starting with the week ending Tuesday, July 3, 1962 and ending with the week ending Tuesday, December 29, 1987. We use the whole sample so that each model estimated in the following uses 1330 observations.

The models will be extended to include a leverage effect in the ARCH specification. This will account for the fact that price decreases tend to increase volatility subsequently by more than would a stock price increase. Hamilton and Susmel (1994) also used this extension that has been suggested by several authors before (see e.g. Black, 1976 and Nelson, 1991):

$$y_t = \beta_0 + \beta_1 y_{t-1} + u_t \tag{28}$$

$$u_t = \sqrt{h_t} v_t \qquad v_t \sim \mathcal{N}(0, 1)$$

$$h_t = \gamma_{I_t} + \alpha_1 u_{t-1}^2 + \cdots + \alpha_m u_{t-m}^2 + \xi d_{t-1} u_{t-1}^2 \qquad I_t = 1, \dots, K \tag{29}$$

where $d_t = 1$ if $u_t \leq 0$, $d_t = 0$ if $u_t > 0$ and $\xi > 0$. In the following, we will denote the specification of SWARCH(K, m) model with leverage effect as SWARCH-L(K, m).

As the additional parameters ξ influences the persistence of the model (see below), we assume that $(\alpha, \xi/2, \alpha_{m+1})$, where $\alpha_{m+1} = 1 - \sum_{l=1}^{m} \alpha_l - \xi/2$, follows a Dirichlet distribution $D(a_1, \dots, a_{m+1}, a_{m+2})$. Again the marginal distribution of the persistence parameter $\sum_{l=1}^{m} \alpha_l + \xi/2$ is a Beta distribution.

Sampling ξ is done within block (iv) described previously, but now out of the conditional distribution $\pi(\gamma, \alpha, \xi | \beta, i^N, y^N)$. To this aim, the Metropolis-Hastings step can be adjusted in a straightforward manner where the multivariate normal proposal for (γ, α, ξ) is now based on the following equation:

$$u_t^2 = \gamma_1 S_t^1 + \ldots + \gamma_K S_t^K + u_{t-1}^2 \alpha_1 + \ldots + u_{t-m}^2 \alpha_m + \xi d_{t-1} u_{t-1}^2 + h_t \varepsilon_t, \tag{30}$$

rather than equation (21).

---

[5]We kindly thank James Hamilton and Raul Susmel for making available their data set.

5.2 Prior specification and permutation

Below we investigate SWARCH-L(K, m) specifications for various combinations of $K = 3, 4$ and $m = 2, 3, 4$. The priors used in particular are:

- $\beta = (\beta_0, \beta_1)' \sim \mathcal{N}(0, B_0)$, where $B_0^{-1} = \mathrm{diag}(1, 4)$.

- $(\alpha, \xi/2, \alpha_{m+1})$, where $\alpha_{m+1} = 1 - \sum_{l=1}^m \alpha_l - \xi/2, \sim D(1, \ldots, 1)$.

- Prior on (γ, η) as in (16) with

    - $g_{0,i} = 1, \forall i = 1, \ldots, K, K = 3, 4$;
      $G_{0,1} = 3, G_{0,2} = 10, G_{0,3} = 800$ for $K = 3$;
      $G_{0,1} = 2, G_{0,2} = 10, G_{0,3} = 30, G_{0,4} = 2000$ for $K = 4$;
    - $e_{1.} = (2, 1, 1)$, $e_{2.} = (1, 2, 1)$ and $e_{3.} = (1, 1, 1)$ for $K = 3$;
      $e_{1.} = (2, 1, 1, 1)$, $e_{2.} = (1, 2, 1, 1)$, and $e_{4.} = (1, 1, 1, 1)$, for $K = 4$,
      with $e_{3.} = (1, 2, 1, 1)$ for $m = 4$ and $e_{3.} = (2, 1, 1, 1)$ for $m = 2, 3$,
      respectively.

A few words are in order concerning the prior specifications of γ and η. The choice of the hyperparameters for the prior distribution of γ reflects the belief that the states differ in the variances and that there is one state with a considerably higher variance than the others. The inclusion of a state with extremely high variance turned out to be crucial, as normally, only a few observations pertain to this state. Therefore, the proposal distribution within the Metropolis-Hastings is driven by the prior distribution (see also section 2.2), and excess rejection rates resulted if the normal proposal with mean and variance nearly equal to the mode and the inverse of the information matrix of the prior, respectively, did not allow for high enough sampled values for the highest-variance state. The choice of the hyperparameters for the prior distribution of η reflects apriori dependence between persistence in a state and the associated variance. The hyperparameters $e_{1.}$ and $e_{2.}$ account for our belief that for a low-variance state the persistence to remain in this state is higher than the switching probability. With the prior specification on $e_{3.}$ we stress the belief that for a higher-variance state switching from this state might be more probable than not switching, and if switching, the probability to switch to a specific state is higher than the probability to switch to another one.

Throughout, we use the constrained permutation sampler, sampling first out of the unrestricted posterior of the parameters. The permutation step puts identification restrictions on γ for the 4- and 3-states models, and additionally on η for the 4-states models. Within the 4-states models we find 2 states of medium volatility that are difficult to separate through γ_j. However, they differ in the

18

persistence to remain in the respective state (see figure 5, panel a) and b), below). For $K = 4$ we set the following identification restrictions:

$$\gamma_1 < \min(\gamma_2, \ldots, \gamma_4), \quad \gamma_4 > \max(\gamma_2, \gamma_3), \quad \eta_{22} > \eta_{33}. \tag{31}$$

The restriction

$$\gamma_1 < \gamma_2 < \gamma_3 \tag{32}$$

identifies the 3-states models. The MCMC output for $m = 4$, however, reveals a potential fourth state of medium volatility but lower persistence (see again figure 5, panel c), below).

5.3   Output analysis

The most general specification with $K = 4$ and $m = 4$ will serve as a baseline case. We iterated the permutation sampler based on the constraint (31) over 40'000 times and discarded the first 10'000 for the posterior analysis to remove the dependence on initial conditions. The sampler converges rather quickly so the initial burn-in phase of 10'000 seems adequate.

The inference about the posterior distributions of the parameters presented in table 1 is readily available from the MCMC output. Mean and standard error are estimated by the mean and the standard deviation of the sampled values, respectively. The 95% confidence interval is determined using the 2.5th and the 97.5th percentiles of the simulated values. Figure 2a plots the estimated posterior distribution of the transition probabilities, figure 2b and 2c reproduce the simulated 50'000 values of γ, α and ξ, respectively, the respective posterior distribution inferred from the last 30'000 iterations and the autocorrelation of the MCMC sample up to lag 500. Whereas some of the posteriors are not too far from normality, the posteriors of some transition probabilities are extremely skewed. We find a low and a medium variance state with high persistence to remain in the respective state, and two states with low persistence (see again figure 5, panel a). The switching probability from state 3 to state 2 is higher than the switching probability to state 1.

Figure 3 displays the inferred posterior probabilities as averaged over the 30'000 sampled paths for I_t along with lines dating business cycle turning points according to the NBER. Using smoothed probabilities gives us a notion about the state variables' ability to capture specific time series features or to identify specific time periods. Here, the "usual" state seems to be state 2, relative short periods of lower variance (equal to state 1) interrupting it. The last state clearly captures "crash"

or crisis periods. There is only one observation that pertains to this regime with a probability nearly one, the 1987 October crash of the stock market. All in all however, no clear association between stock market volatility and some specific economic period can be made.

Restricting the model to $m = 3$, we obtain a slight, interesting change in the results (see table 2).[6] First, the posterior distributions of γ report higher means for the first two states than in the case $m = 4$. The estimates rise from 0.76 to 1.18 and from 2.65 to 6.19, respectively. Volatility captured previously through a persistence parameter, α_4 in particular, is now captured by the switching constant in the ARCH equation. Note also, that the probability to switch from state 3 is now highest for the switch to state 1. These differences alter the inference about the smoothed posterior state probabilities (figure 4). Now the first state seems to be the "usual" state and usually coincides with periods of economic recovery, the second state can be associated with recession periods. The third state seems to capture short moments of uncertainty. Generally, the probability of this state peaks shortly before a switch from state 1 to state 2 (or vice versa) takes place, and especially several times before the 1987 October crash.

Next, we turn to the analysis of the more restrictive model with just $K = 3$ states. We will again restrict the discussion to the specification $m = 4, 3$ as the results for $m = 2$ are very similar to those for $m = 3$. The posterior distributions of the parameters for the SWARCH-L(3,4) and the SWARCH-L(3,3) models are found in table 3 and 4, respectively. For both specifications, we find the first two states to be persistent, and the third seems to have higher switching probability (from state 3 to state 2) than remaining in the same state. Interestingly, post-processing the MCMC output for $m = 4$ reveals a potential fourth state with similar γ_2 but lower persistence than state 2 (see figure 5, panel c), hence pointing towards a misspecification of the model. This fourth state does not show up when we restrict the model to $m = 3$ and $m = 2$ (see panel d) of figure 5). Therefore, in figure 6 we display the smoothed posterior state probabilities of the SWARCH-L(3,3) specification. The second and third state of the 4-states model seem to have been subsumed in state 2 of the 3-states model. State 3 still captures unusual events of very high volatility, the October 1987 crash being one of these and the oil price shock being another one in this specification.

---

[6]Further restricting $m = 2$ does not change the interpretation given in the following, so we skip an extensive discussion of the results for the SWARCH-L(4,2) specification in order to save space.

5.4 Inefficiency

The inefficiency factors reported in table 5 are based on 30'000 iterations of the sampler after an initial burn-in phase of 10'000 iterations. Their values are computed as the ratio of the squared numerical standard error of the MCMC simulation and the variance estimate divided by the number of iterations (the variance of the sample mean from a hypothetical iid sampler). The squared numerical standard error is estimated using time series methods to account for serial dependence in the sampled values (Geweke, 1992):

$$SE^2 = \Omega_0 + 2 \sum_{j=1}^{J} \left(1 - \frac{j}{J+1}\right) \Omega_j,$$

where the bandwidth chosen is $J = 1000$ and Ω_j represents the autocovariance at lag j of the sampled parameter values. Using 30'000 simulations out of the posterior distribution seems appropriate if we require that the Monte Carlo error in estimating the mean is smaller than one percentage of the variation of the error due to the data. The very high inefficiency factors reported for η_1. and η_2. in the case of $K = 3$, $m = 4$ are reflected in a persistent autocorrelation in the simulated values even at very high lags. As already mentioned before, plotting the simulated values for $\log \gamma$ against persistence reveals that these high inefficiency factors might be due to misspecification rather than an inefficient sampler (figure 5). Note finally, that the acceptance probabilities of the Metropolis-Hastings simulation step for the parameters in the ARCH equation lie in an acceptable range between 0.13 and 0.26 for the various specifications.

5.5 Comparison to previous study

The evidence presented so far deviates from the findings documented in Hamilton and Susmel (1994). For both the 4-states and the 3-states models they find a highly persistent third state, while our results report a third state with low persistence, the probability to switch to another state rather than remaining in the state being much higher. The estimates for the parameters γ_3 and γ_4 reflect this difference. In the SWARCH-L(4,4) specification we infer a mean of 12.60 and 1487.02 for γ_3 and γ_4, respectively, while the point estimates for the SWARCH-L(4,2) model reported in Hamilton and Susmel are 7.59 and 92.95, respectively. Also, the mean estimates of γ_3 in our SWARCH-L(3,3) and SWARCH-L(3,2) specifications are 391.45 and 375.95, respectively, while the point estimate of the SWARCH-L(3,2) specification is 7.47 in Hamilton and Susmel.

Moreover, our inference subsumes the third state into the second one when we restrict the model to 3 states, while in the investigation of Hamilton and Susmel

the last, the fourth, state is subsumed into state 3. Briefly, the constrained permutation sampler associates the highest-variance state with crash periods, while the maximum likelihood inference suppresses this state when K is restricted to 3.

Last, we would like to point out that standard errors of the parameters, in particular of the transition probabilities are readily available from the MCMC output, even if the posteriors are very skewed towards the boundary space. In contrast, numerical maximum likelihood estimation often encounters this boundary space problem, and therefore, Hamilton and Susmel have to put zero restrictions on seven transition probabilities to get standard errors for the parameters of the 4-state model.

5.6 Bayes factors and persistence

So far, the discussion of the various specifications leaves some ambiguity with respect to the number of states present in the data and the number of lags that are appropriate in the ARCH process of the variance. Table 6 summarizes model likelihoods $L(y^N|K, m)$ and the Schwarz critrion for various SWARCH-L(K, m)-models. The model likelihoods are estimated using bridge sampling as described in section 4 by setting $M_L = 1000$.

The computation of the Schwarz criterion is based on the maximum value $\mathcal{L}$ of the log marginal likelihood $\log L(y_1, \ldots, y_N|\phi^{(l)}, K, m)$ which is obtained as a by-product at each iteration of the sampler (see appendix):

$$\mathcal{L} - k/2 \ln N, \tag{33}$$

where k is the number of parameters, and $N = 1330$. The Schwarz criterion favors the SWARCH-L(3,2) specification. In section 6.2 we will briefly argue why the Schwarz criterion has to be used with care for model selection concerning the number of states.

Testing between $K = 3$ and $K = 4$ using model likelihoods, however, turns out to be problematic in this case, too. From the MCMC output of the four state models visualized in figure 5, panel a) and b), as well as from the evidence given in figure 5, panel c), which reveals a potential fourth state within a 3-states specification, we would expect the model likelihood to select a model with four states. Interestingly, the model likelihoods reported in table 6 point towards a SWARCH-L(3,4) specification. An extensive sensitivity study, however, revealed high sensitivity of the model selection procedure with respect to the prior on γ. Among various other priors we tested, a prior with $G_{0,1} = 2, G_{0,2} = 12, G_{0,3} = 600$ for $K = 3$ and $G_{0,1} = 1.2, G_{0,2} = 4, G_{0,3} = 20, G_{0,4} = 1600$ for $K = 4$ which is

only a slight modification of the prior used so far (see Subsection 5.2), results in favouring a SWARCH-L(4,4) specification rather than the SWARCH-L(3,4) – see the additional model likelihoods reported in table 6. This sensitivity may be explained by the fact that for a 4-state model one of the states corresponds to a single outlier. Thus little information on the parameters of the fourth state is available from the likelihood and the prior dominates the posterior distribution. Selecting the number m of lags within a specification with fixed number of states, however, is unaffected by the prior on γ. Both for the 3-states as well as for the 4-states specifications $m = 4$ performs better than $m = 3$ and $m = 2$. Overall, from the exploratory Bayesian analysis in figure 5 as well as from the model likelihoods we tend to favor the SWARCH-L(4,4) model.

We proceed with a discussion of the persistence implied by each model. In analogy to Hamilton and Susmel (1994), our measure of persistence for the SWARCH-L(K, m) models will be the largest eigenvalue of the matrix:

$$\begin{bmatrix} (\alpha_1 + \xi/2) & \alpha_2 & \cdots & \alpha_m \\ I_{m-1} & & & 0 \end{bmatrix},$$

where I_{m-1} is an identity matrix of dimension $m - 1$. The last column of table 6 summarizes this measure of persistence for each model. As expected, our results document a considerable reduction in persistence relative to the high persistence usually found in ARCH and GARCH models. The reported persistence measures range from 0.40 to 0.62 and are well below values above 0.7 for Gaussian and Student-t ARCH and GARCH specifications reported elsewhere in the literature and in particular in Hamilton and Susmel (1994).

5.7   Model diagnostics

We are going to conclude our case study by discussing issues in model diagnostics based on recursive residuals which are an important tool for model diagnostics especially for time series models of continuous observations (see Smith, 1985; Frühwirth-Schnatter, 1996; Gerlach et al., 1999 for a detailed discussion).

Let $y_1, \ldots, y_N$ be observed time series observations generated by a model $\mathcal{M}$ and let $Y_1, \ldots, Y_N$ be the corresponding random variables. We use the notation y^{t-1} to denote all observations $\{y_1, \ldots, y_{t-1}\}$ up to $t - 1$. Smith (1985), following Dawid (1984), defined recursive residuals for model $\mathcal{M}$ for time series of continuous observations by

$$u_t = Pr(Y_t \leq y_t | y^{t-1}),$$

where $Pr(Y_t \leq y_t | y^{t-1})$ is the one-step ahead predictive distribution of a future value y_t of a time series given observations up to $t - 1$. $u_1, \ldots, u_N$ will be called

P-scores in the following. What makes the P-scores so useful is the property that they are iid uniform on $[0, 1]$ under a correct model and model diagnostics turns out to be testing for independence and uniformness of the sequence of P-scores. It is often more convenient to work with the transformed P-scores $v_t = \Phi^{-1}(u_t)$, where Φ is the standard normal distribution. If the model is correct, the sequence of the transformed P-scores is iid standard normal.

It is in general impossible to compute u_t explicitly, because the predictive distribution of y_t depends on unknown parameters ϕ which cannot be integrated out analytically. If ϕ is an unknown, but fixed parameter one may proceed in two ways. One strategy which will, however, not be pursued in the present paper is to assume that ϕ is a random variable with prior $\pi(\phi)$. The exact P-scores u_t are then given by the infinite mixture

$$u_t = \int Pr(Y_t \leq y_t | y^{t-1}, \phi) \pi(\phi | y^{t-1}) \, d\phi$$

where $\pi(\phi | y^{t-1})$ is the posterior density of ϕ given data up to $t - 1$. Approximate P-scores are obtained by substituting the infinite mixture by a finite mixture approximation using methods from multi-process-filtering (Frühwirth-Schnatter, 1996) or MCMC sampling combined with importance sampling (Gerlach *et al.*, 1999). The second strategy is to estimate $\hat{\phi}$ from the data, e.g. by taking the posterior mean estimated from the MCMC simulations, and to derive the P-scores conditional on $\phi = \hat{\phi}$. Given a consistent estimate $\hat{\phi}$ of ϕ the P-scores approximated from $Pr(Y_t \leq y_t | y^{t-1}, \hat{\phi})$ are iid uniform at least asymptotically (Dawid, 1984).

Following Kim *et al.* (1998) and Kaufmann (2000) we apply the second approach to perform model diagnostics of the SWARCH-model using to the following approximate P-scores:

$$u_t \approx \sum_{j=1}^{K} Pr(Y_t \leq y_t | y^{t-1}, I_t = j, \hat{\phi}) Pr(I_t = j | y^{t-1}, \hat{\phi}),$$

where $Pr(Y_t \leq y_t | y^{t-1}, I_t = j, \hat{\phi})$ is the distribution of the one-step ahead normal predictive density given by (11) and $Pr(I_t = j | y^{t-1}, \hat{\phi})$ is computed by one run of the discrete filter described in appendix.

Various exploratory and formal tools of model diagnostics may be applied to these approximate P-scores. A simple graphical device is a plot of the ordered P-scores $\{u_t\}$ and of the transformed P-scores $\{v_t\}, t = 1, \dots, N$ against normal order statistic. Under the hypothesis of a correct model the P-scores as well as the transformed P-scores should lie on a straight line. To check for remaining serial correlation we can use the empirical correlogram of the transformed P-scores,

while a visual inspection of the correlogram of the transformed reflected P-scores, $2|u_t - 0.5|$, yields an assessment about neglected volatility clustering. Figure 7 and 8 gather the respective graphics for the SWARCH-L(4,4) and SWARCH-L(3,3) specification. Obviously, both specifications are able to fit the data quite well.

Table 7 reports some quantitative diagnostic measures to test for departure from the assumed model. The indices (see Frühwirth-Schnatter, 1996) are derived from the first four central moments of the transformed P-scores, $V_i = \sum_{t=1}^{N} (v_t - \bar{V})^i / N$, where $\bar{V} = \sum_{t=1}^{N} v_t / N$, $i = 2, \ldots, 4$. The computed statistics can be used to check for bias (B), for over- or underdispersion (D), for skewness (S) and for the tail properties (T) of the predictive density. Accordingly, they are defined:

$$
\begin{array}{ll}
B = \sqrt{N}\bar{V} & D_N = \dfrac{NV_2 - N + 1}{\sqrt{2(N-1)}} \\[2ex]
S_N = \sqrt{\dfrac{(N+1)(N+3)}{6(N-2)}} \dfrac{V_3}{V_2^{3/2}} & L_N = \dfrac{(N+1)\sqrt{(N+3)(N+5)}}{\sqrt{24(N-2)(N-3)N}} \left(\dfrac{V_4}{V_2^2} - \dfrac{3(N-1)}{(N+1)} \right) \\[2ex]
J_N = S_N^2 + L_N^2 & A_N = \sqrt{N}\left(\rho_1 + \dfrac{1}{N-1} \right)
\end{array}
\tag{34}
$$

where ρ_1 is the first order empirical autocorrelation coefficient of the transformed P-scores. The $AC(1)$-index A_N tests whether there is some remaining autocorrelation in the transformed residuals that has not been accounted for by the model. The quantiles of the exact distribution have been tabulated in Frühwirth-Schnatter (1996). Given the number of observations, however, we will compare the statistics against standard normal statistics and against χ^2 statistic with two degrees of freedom. Table 7 summarizes our results for the various SWARCH-L(K, m) specifications. All models pass the diagnostic tests except for skewness and for normality in the case for the SWARCH-L(4,4) model. However, the excess skewness might be an artefact of the data, as the crash of October 1987 produces several big, very rarely observed, negative observations.

6 Discussion

6.1 ML estimation versus Bayesian estimation

In the present paper we use Bayesian estimation rather than ML estimation. We see several advantages of the Bayesian approach compared to ML-estimation. First, ML estimation of the SWARCH model appears to be rather difficult especially in situations where we are close to a boundary space problem. Hamilton and Susmel (1994), for instance, report extreme difficulties to maximize the marginal likelihood of a SWARCH-L(4,2)-model. These difficulties appear to be caused by the fact that some transition probabilities are extremely small and close to

the boundary of the parameter space. To report standard errors for their model, Hamilton and Susmel (1994) imposed a zero constraint on seven(!) transition probabilities.

Second, we can explore the MCMC output in a similar way as one would explore data. Standard errors of all parameters involved, for instance, are directly available and may be estimated from the MCMC output simply by the standard deviation of the sampled values. Marginal densities and scatter plots of the sampled parameters can be explored from the MCMC output and may lead to hypotheses concerning order selection.

Finally, the smoothed probabilities $P(I_t = j|y^N)$ of being in state j at time t may be estimated from the MCMC output $I_t^{(1)}, \dots, I_t^{(L)}$ simply by $1/L \#\{I_t^{(l)} = j\}$ and no conditioning on the model parameters is necessary to obtain these estimates.

6.2   Testing for the presence of Markov switching

In the present paper we are testing for the presence of Markov switching and select the number of states within a Bayesian framework. Testing for the presence of Markov switching and selecting the number of states is not possible within the classical framework of maximum likelihood. Although an ARCH(m)-model could be viewed as a special case of a SWARCH(K, m)-model with $K = 1$, the regularity conditions for justifying the χ^2 approximation to the likelihood ratio statistics do not hold, as γ_2 is unidentified under the hypothesis that there is really one state. For similar reasons, Schwarz's criterion should not be applied for selecting the number of states. Schwarz's criterion is an approximation to the model likelihood which is based on asymptotic normality of the posterior. Although they are aware of these problems, Hamilton and Susmel (1994), Cai (1994) and Gray (1996) based inference concerning the presence of Markov switching on likelihood ratio statistics as well as on the Schwarz criterion.

We tried to overcome these problems by computing model likelihood from the MCMC output using the method of bridge sampling and selecting the model with the largest model likelihood. This method of selecting the number of states has been applied sucessfully in the context of classical Markov switching time series models (Frühwirth-Schnatter, 1999) and switching Gaussian state space models (Frühwirth-Schnatter, 2001). For the case study of the present paper, however, we encountered problems insofar, as testing between $K = 3$ and $K = 4$ turned out to be sensitive to the prior on the switching parameters $\gamma_1, \dots, \gamma_K$, leaving some ambiguity with respect to the appropriate number of states also within the Bayesian approach.

6.3 Extensions

In the present paper we studied Bayesian analysis of SWARCH models through MCMC methods. Details were given for the basic switching ARCH-model, where only a switching intercept is included into the volatility equation. Extension of the proposed MCMC method to various more general SWARCH models is possible. As outlined in our case study, the inclusion of a „leverage" effect when modeling stock returns with a SWARCH-model (Hamilton and Susmel, 1994) is straightforward. Similarly, for the totally switching ARCH-model (Gray, 1996), where all parameters appearing in the volatility equation are switching,

$$h_t = \gamma_{I_t} + \alpha_{1,I_t} u_{t-1}^2 + \cdots + \alpha_{m,I_t} u_{t-m}^2,$$

joint sampling of all parameters appearing in the volatility equation is possible. In general, the proposed method to multi move sampling of all parameters in the volatility equation can be applied, whenever the volatility equation remains linear in the unknown parameters. It is also possible to include switching means into the regression equation such as in Cai (1994) and Gray (1996). In general, extension of the proposed method of multi move sampling of all regression parameters is possible, whenever the regression equation is linear in the unknown regression parameters. Note however that for a switching mean the conditional densities $f(y_t|y^{t-1}, I^N, \phi)$ will not only depend on I_t, but also on the history $I_{t-1}, \ldots, I_{t-m}$ through the lagged residuals $u_{t-1}^2, \ldots, u_{t-m}^2$. Multi move sampling of I^N has to be modified accordingly (see e.g. Kaufmann, 2000). Our method needs some obvious modifications if we want to include t-errors such as in Hamilton and Susmel (1994).

Unfortunately, the MCMC algorithm discussed in this paper can not be applied to GARCH-models with Markov-switching behaviour in a straightforward manner. It has been noted earlier (Hamilton and Susmel, 1994; Cai, 1994; Gray, 1996) that the combination of GARCH-models with Markov-switching models introduces tremendous complications in estimation. The conditional observation distribution $f(y_t|y^{t-1}, I^N, \phi)$ appearing in equation (10) depends on the *whole* history $I_0, \ldots, I_t$. Therefore no exact fixed memory filter for I_t is available. Approximate ML-estimation by an approximate filter has been studied in Gray (1996). The MCMC algorithm discussed in this paper, however, can not be applied to switching GARCH-models, for the same difficulties as in ML-estimation are encountered in the Bayesian analysis: multi move sampling of I^N along the lines of Chib (1996) is based on the existence of an exact finite dimensional filter for I_t.

Appendix - Computing the marginal likelihood

The marginal likelihood $L(y_1, \ldots, y_N | \phi)$ is given by:

$$L(y_1, \ldots, y_N | \phi) = \prod_{t=1}^{N} f(y_t | \phi, y^{t-1}),$$

$$f(y_t | \phi, y^{t-1}) = \sum_{j=1}^{K} f(y_t | y^{t-1}, \alpha, \beta, \gamma_j) Pr\{I_t = j | y^{t-1}, \phi\} \quad (35)$$

The probability $Pr\{I_1 = j | y^0, \phi\}$ is given by:

$$Pr\{I_1 = j | y^0, \phi\} = \sum_{i=1}^{K} \eta_{ij} Pr\{I_0 = i | \phi\},$$

For $t > 1$ the probability $Pr\{I_t = j | y^{t-1}, \phi\}$ is given by the following recursion:

$$Pr\{I_t = j | y^{t-1}, \phi\} = \sum_{i=1}^{K} \eta_{ij} Pr\{I_{t-1} = i | y^{t-1}, \phi\}$$

$$Pr\{I_{t-1} = i | y^{t-1}, \phi\} = \frac{f(y_{t-1} | y^{t-2}, \alpha, \beta, \gamma_i) Pr\{I_{t-1} = i | y^{t-2}, \phi\}}{\sum_{j=1}^{K} f(y_{t-1} | y^{t-2}, \alpha, \beta, \gamma_j) Pr\{I_{t-1} = j | y^{t-2}, \phi\}}.$$

Note that for each t, $f(y_t | \phi, y^{t-1})$ is the normalizing constant of the filtering distribution $\pi(i_t | y^t, \phi)$. Therefore the marginal likelihood is easily computed from (35) after one run of a discrete filter for $i_0, \ldots, i_N$.

References

Bauwens, L. and Lubrano M. (1998), Bayesian inference on GARCH models using the Gibbs sampler. *Econometrics Journal*, 1, C23-C46.

Besag, J. (1989), A Candidate's Formula: A Curious Result in Bayesian Prediction. *Biometrika*, 76, 183.

Black, F. (1976), Studies of stock market volatility changes. *1976 Proceedings of the American Statistical Association*, Business and Economic Statistics Section, 177-181.

Cai, J. (1994), A Markov model of switching-regime ARCH. *Journal of Business & Economic Statistics*, 12, 309-316.

Carter, C.K. and Kohn, R. (1994), On Gibbs sampling for state space models, *Biometrika*, 81, 541-53.

Chib, S. (1995), Marginal Likelihoods from the Gibbs Output. *Journal of the American Statistical Assosiation*, 90, 1313-21.

Chib, S. (1996), Calculating posterior distributions and modal estimates in Markov mixture models, *Journal of Econometrics*, 75, 79-97.

Dawid, A.P. (1984) Statistical Theory – The Prequential Approach. *Journal of the Royal Statistical Society* A, **147**, 278-292.

DiCiccio, T.J., Kass, R., Raftery, A. and L. Wasserman (1997): Computing Bayes Factors By Combining Simulations and Asymptotic Approximations. *Journal of the American Statistical Assosiation*, 92, 903-15.

Frühwirth-Schnatter, S. (1995), Bayesian Model Discrimination and Bayes Factors for Linear Gaussian State Space Models. *Journal of the Royal Statistical Society*, B, 57, 237-246.

Frühwirth-Schnatter, S. (1996), Recursive Residuals and Modeldiagnostics for Normal and Non-normal State Space Models. *Environmental and Ecological Statistics*, **3**, 291-309.

Frühwirth-Schnatter, S. (1999), Model Likelihoods and Bayes Factors for Switching and Mixture Models. Submitted.

Frühwirth-Schnatter, S. (2000). MCMC Estimation of Classical and Dynamic Switching and Mixture Models, Final revision submitted to the *Journal of the American Statistical Assosiation*.

Frühwirth-Schnatter, S. (2001). Fully Bayesian Analysis of Switching Gaussian State Space Models. *Annals of the Institute of Mathematical Statistics*, forthcoming.

Gelfand, A.E., and D.K. Dey (1994), Bayesian Model Choice: Asymptotics and Exact Calculations. *Journal of the Royal Statistical Society*, Ser. B, 501-514.

Gerlach, R., Carter, C., and R. Kohn (1999): Diagnostics for time series. *Journal of Time Series Analysis*, **20**, 309-330.

Geweke, J. (1992), Evaluating the Accuracy of Sampling-Based Approaches to the Calculation of Posterior Moments. In *Bayesian Statistics 4*, eds. J.M. Bernardo, J.O. Berger, A.P. Dawid and A.F.M. Smith, Oxford: Oxford University Press, pp. 169-193.

Gilks, W.R., N.G. Best and K.K.C. Tan (1995), Adaptive rejection Metropolis sampling within Gibbs sampling. *Applied Statistics*, 44, 155-73.

Gray, S.F. (1996), Modeling the conditional distribution of interest rates as a regime switching process. *Journal of Financial Economics*, 42, 27-62.

Hamilton, J.D. (1989), A new approach to the economic analysis of nonstationary time series and the business cycle, *Econometrica*, 57, 357-84.

Hamilton, J.D. and Susmel, R. (1994), Autoregressive conditional heteroscedasticity and changes in regime. *Journal of Econometrics*, 45, 39-70.

Kaufmann, S. (2000), Measuring business cycles with a dynamic Markov switching factor model: An assessment using Bayesian simulation methods. Forthcoming in *Econometrics Journal*, 3, issue 1.

Kim, S., N. Shephard, and S. Chib (1998): Stochastic volatility: Likelihood inference and comparison with ARCH models, *Review of Economics Studies*, 65, 361-393.

Nakatsuma, T. (2000), Bayesian analysis of ARMA-GARCH models: a Markov chain sampling approach. *Journal of Econometrics*, 95, 57-69.

Meng, X.L. and W.H. Wong (1996), Simulating Ratios of Normalising Constants via a Simple Identity. *Statistica Sinica*, 6, 831-60.

Nelson, D. (1991), Conditional heteroskedasticity in asset returns: A new approach. *Econometrica*, 59, 347-370.

Richardson, S. and Green, P. (1997), On Bayesian analysis of mixtures with an unknown number of components (with discussion). *Journal of the Royal Statistical Society*, Ser. B, 59, 731-92.

Ritter, C. and M. Tanner (1992), Facilitating the Gibbs sampler: the Gibbs stopper and the Griddy-Gibbs sampler. *Journal of the American Statistical Assosiation*, 87, 861-868.

Shephard, N. (1994), Partial Non-Gaussian state space models, *Biometrika*, 81, 115-31.

Smith, A.F.M. and Roberts, G. (1993), Bayesian computation via the Gibbs sampler and related Markov Chain Monte Carlo methods, *Journal of the Royal Statistical Society*, Ser. B, 55, 3-23.

Smith, J.Q. (1985), Diagnostic check of non-standard time series models. *Journal of Forecasting*, **4**, 283-291.

Wong, C.S. and Li, W.K. (1999): On a Mixture Autoregressive Conditonal Heteroscedastic Model. Research Report Department of Statistics, University of Hong Kong.

7 Tables

Table 1: SWARCH-L(4,4). Posterior distribution of the model parameters. The mean, standard deviation, and the median of 30'000 sampled values out of the posterior are reported. The 95% confidence intervals are estimated as the shortest interval containing 95% of the MCMC simulations.

parms	mean	st.dev.	median	95% conf. int.	
β_0	0.3187	0.05067	0.3186	0.2245	0.4241
β_1	0.2664	0.02809	0.2662	0.21	0.3196
γ_1	0.758689	0.190436	0.745121	0.451633	1.18577
γ_2	2.52497	1.02607	2.28651	1.13217	4.46657
γ_3	12.2664	4.03985	11.4756	5.92434	19.8874
γ_4	1481.03	915.217	1204.73	429.545	3508.39
α_1	0.04063	0.02506	0.03638	4.025e-005	0.085
α_2	0.1033	0.03949	0.1013	0.03089	0.1781
α_3	0.02841	0.01896	0.02352	3.71e-005	0.0634
α_4	0.04158	0.02345	0.04058	0.001522	0.08414
ξ	0.3394	0.0784	0.3355	0.1826	0.4753
η_{11}	0.9437	0.04447	0.9514	0.8944	0.9868
η_{12}	0.02461	0.02015	0.0206	2.578e-006	0.06034
η_{13}	0.02718	0.0423	0.01833	3.993e-008	0.07089
η_{14}	0.004552	0.004806	0.003045	6.442e-008	0.01404
η_{21}	0.01155	0.02291	0.006267	1.286e-007	0.03578
η_{22}	0.7959	0.1268	0.8221	0.5395	0.9821
η_{23}	0.1902	0.1254	0.1666	0.0004257	0.4369
η_{24}	0.002391	0.002416	0.001664	3.775e-008	0.007058
η_{31}	0.09395	0.1559	0.03877	1.845e-006	0.4713
η_{32}	0.7428	0.2075	0.8054	0.2596	0.9904
η_{33}	0.1488	0.1402	0.1057	5.08e-006	0.4307
η_{34}	0.01451	0.01898	0.009174	2.296e-007	0.04494
η_{41}	0.2437	0.1905	0.2007	7.557e-007	0.6203
η_{42}	0.2922	0.2069	0.2573	1.129e-005	0.6797
η_{43}	0.2763	0.2037	0.2354	1.471e-005	0.6664
η_{44}	0.1878	0.1544	0.1482	2.689e-006	0.4951

Table 2: SWARCH-L(4,3). Posterior distribution of the model parameters. The mean, standard deviation, and the median of 30'000 sampled values out of the posterior are reported. The 95% confidence intervals are estimated as the shortest interval containing 95% of the MCMC simulations.

parms	mean	st.dev.	median	95% conf. int.	
β_0	0.3391	0.05214	0.3388	0.2343	0.4369
β_1	0.2669	0.02899	0.267	0.2114	0.3245
γ_1	1.15799	0.179338	1.16387	0.790343	1.49018
γ_2	6.00349	1.30544	6.09247	3.0302	8.44209
γ_3	11.5185	4.20424	10.6423	5.18288	20.0709
γ_4	1252.26	717.395	1066.92	267.23	2767.39
α_1	0.03459	0.02498	0.0298	7.022e-007	0.08435
α_2	0.09085	0.03805	0.08797	0.01948	0.1642
α_3	0.02378	0.01706	0.02046	0.0001106	0.05561
ξ	0.3097	0.07946	0.3056	0.1426	0.4586
η_{11}	0.8959	0.04729	0.9048	0.8025	0.9695
η_{12}	0.01547	0.01311	0.01244	1.283e-007	0.04066
η_{13}	0.08648	0.05138	0.07949	5.89e-006	0.1787
η_{14}	0.002124	0.002228	0.001461	8.248e-008	0.006344
η_{21}	0.02072	0.01857	0.01688	4.769e-007	0.05192
η_{22}	0.9395	0.05304	0.9532	0.8553	0.9903
η_{23}	0.03558	0.05087	0.02112	1.432e-007	0.1188
η_{24}	0.004232	0.004212	0.002962	1.9e-007	0.01252
η_{31}	0.6727	0.2292	0.7365	0.1348	0.9764
η_{32}	0.1727	0.1971	0.1011	1.182e-006	0.6628
η_{33}	0.1283	0.1174	0.09293	1.586e-006	0.3639
η_{34}	0.02634	0.02809	0.0183	3.365e-006	0.07702
η_{41}	0.2152	0.1776	0.1702	6.487e-006	0.5738
η_{42}	0.369	0.2176	0.3532	6.565e-005	0.7496
η_{43}	0.2283	0.1849	0.1824	4.908e-006	0.5976
η_{44}	0.1875	0.1536	0.1486	7.497e-007	0.496

Table 3: SWARCH-L(3,4). Posterior distribution of the model parameters. The mean, standard deviation, and the median of 30'000 sampled values out of the posterior are reported. The 95% confidence intervals are estimated as the shortest interval containing 95% of the MCMC simulations.

parms	mean	st.dev.	median	95% conf. int.	
β_0	0.3281	0.05196	0.3283	0.2254	0.4277
β_1	0.2697	0.02914	0.2699	0.2114	0.3254
γ_1	1.18075	0.195749	1.18443	0.778599	1.53868
γ_2	6.18438	1.28628	6.05651	3.6534	8.76535
γ_3	505.942	290.137	428.168	97.8024	1114.79
α_1	0.03083	0.02635	0.02372	0.0001928	0.08418
α_2	0.1097	0.04815	0.1055	0.02443	0.2045
α_3	0.0275	0.02254	0.02244	3.853e-005	0.07047
α_4	0.03362	0.02462	0.02833	4.47e-005	0.08117
ξ	0.2965	0.08175	0.2918	0.1329	0.454
η_{11}	0.9233	0.06033	0.9422	0.7783	0.9818
η_{12}	0.07344	0.06075	0.05448	0.01083	0.2177
η_{13}	0.003273	0.003246	0.002328	5.842e-007	0.009552
η_{21}	0.1356	0.1979	0.06379	0.001447	0.6868
η_{22}	0.8576	0.1992	0.9299	0.3011	0.9928
η_{23}	0.006809	0.005756	0.005423	4.044e-007	0.01773
η_{31}	0.3233	0.2352	0.2791	1.375e-005	0.7714
η_{32}	0.456	0.2435	0.4586	0.0001746	0.8541
η_{33}	0.2207	0.1733	0.1806	1.103e-006	0.5634

Table 4: SWARCH-L(3,3). Posterior distribution of the model parameters. The mean, standard deviation, and the median of 30'000 sampled values out of the posterior are reported. The 95% confidence intervals are estimated as the shortest interval containing 95% of the MCMC simulations.

parms	mean	st.dev.	median	95% conf. int.	
β_0	0.3324	0.05199	0.3324	0.2287	0.4318
β_1	0.2687	0.02947	0.2685	0.2091	0.3247
γ_1	1.2755	0.170814	1.27125	0.943472	1.61533
γ_2	6.23739	0.949275	6.10795	4.37096	8.07763
γ_3	531.642	349.48	426.366	164.735	1419.29
α_1	0.02516	0.02078	0.02064	1.51e-005	0.06632
α_2	0.09174	0.03946	0.09209	0.01846	0.1694
α_3	0.02287	0.01847	0.01781	7.372e-006	0.06079
ξ	0.2851	0.08046	0.2771	0.1369	0.4447
η_{11}	0.9458	0.02062	0.9492	0.9088	0.9777
η_{12}	0.05085	0.02081	0.04756	0.01842	0.08824
η_{13}	0.003315	0.003189	0.002415	2.882e-009	0.009479
η_{21}	0.06191	0.05084	0.05388	0.01488	0.114
η_{22}	0.9314	0.05158	0.9394	0.8787	0.9797
η_{23}	0.006706	0.005262	0.005506	2.698e-007	0.01677
η_{31}	0.265	0.2124	0.2117	3.601e-007	0.6927
η_{32}	0.5001	0.2316	0.5079	0.06054	0.9071
η_{33}	0.2349	0.1764	0.1985	8.96e-006	0.579

Table 5: Inefficiency factors and mean acceptance probability of the Metropolis-Hastings step for (α, ξ) for various model specifications. The first line refers to a SWARCH-L(K, m) specification where K is the number of states and m is the lag length of the ARCH process.

parameters	(4,4)	(4,3)	(4,2)	(3,4)	(3,3)	(3,2)
β_0	13.843	10.557	11.735	6.9378	8.349	4.165
β_1	6.443	6.044	3.228	5.8186	3.472	3.176
γ_1	307.724	137.745	40.553	158.667	86.791	42.504
γ_2	434.248	236.163	101.577	236.259	84.454	43.514
γ_3	171.694	226.599	105.633	79.704	185.650	70.775
γ_4	299.036	262.599	117.600			
α_1	46.840	63.517	27.403	185.239	59.490	12.695
α_2	81.256	34.720	24.302	123.957	71.268	32.502
α_3	62.646	24.379		131.234	41.522	
α_4	83.311			131.289		
ξ	120.382	41.485	26.917	30.632	62.809	31.399
η_{11}	76.484	142.537	98.129	612.007	21.741	12.401
η_{12}	24.038	45.919	65.162	612.858	22.674	11.615
η_{13}	136.126	146.630	109.779	28.043	5.510	3.887
η_{14}	18.730	4.920	6.122			
η_{21}	153.866	68.934	22.622	702.897	41.069	15.912
η_{22}	230.325	121.992	178.093	701.746	37.840	15.224
η_{23}	233.589	126.828	197.220	38.267	16.891	10.034
η_{24}	33.345	7.360	5.868			
η_{31}	351.481	111.233	131.773	22.922	6.766	3.906
η_{32}	238.969	122.137	163.780	23.090	5.362	2.079
η_{33}	61.751	28.205	22.824	3.853	2.653	4.737
η_{34}	55.790	19.693	11.120			
η_{41}	3.748	4.111	2.214			
η_{42}	4.824	5.423	5.802			
η_{43}	3.566	3.481	5.940			
η_{44}	0.922	1.241	1.211			
acc. prob.	0.19	0.18	0.26	0.16	0.13	0.22

Table 6: Comparison of some model selection criterions for various SWARCH-L(K, m) specifications; model likelihoods are computed using the priors of Subsection 5.2 ("prior 1") and Subsection 5.6 ("prior 2").

Model/No.of parameters	$L(y^N|\hat{\phi}, K, m)$	$L(y^N|K, m)$ ("prior 1")	$L(y^N|K, m)$ ("prior 2")	Schwarz criterion	Persistence
SWARCH-L(4,4)/27	-2808.7	-2859.1	-2855.9	-2905.8	0.62
SWARCH-L(4,3)/26	-2815.8	-2860.7	-2859.4	-2909.3	0.47
SWARCH-L(4,2)/25	-2817.7	-2861.0	-2859.7	-2907.7	0.40
SWARCH-L(3,4)/19	-2822.1	-2857.1	-2856.4	-2890.4	0.59
SWARCH-L(3,3)/18	-2824.1	-2858.2	-2857.7	-2888.8	0.47
SWARCH-L(3,2)/17	-2825.8	-2858.5	-2858.0	-2886.9	0.40

Table 7: Diagnostic statistics on the transformed P-scores for the various model specification. The statistics are computed from the first four centered moments of the transformed P-scores. * denotes significance at the 5% level. The bottom panel reproduces selected quantiles of the standard normal distribution and the χ^2 distribution with two degrees of freedom.

	B_N	D_N	S_N	T_N	J_N	A_N
SWARCH-L(4,4)	-1.28	-0.86	-1.97*	-1.59	6.40^a	1.07
SWARCH-L(4,3)	-1.48	-1.69	-1.89*	-1.12	4.82	0.92
SWARCH-L(4,2)	-1.54	-1.72	-1.77*	-1.19	4.53	0.76
SWARCH-L(3,4)	-1.44	0.31	-1.89*	-0.48	3.81	0.87
SWARCH-L(3,3)	-1.59	-1.03	-2.23*	0.03	4.99	0.87
SWARCH-L(3,2)	-1.61	-0.94	-2.17*	-0.04	4.70	0.79
	Quantiles					
	0.01	0.025	0.05	0.95	0.975	0.99
Standard Normal	-2.33	-1.96	-1.65	1.65	1.96	2.33
χ^2_2				5.99	7.38	9.21

8 Figures

Figure 1: Weekly returns of a value-weighted portfolio of the New York Stock Exchange. The observation period runs from the week ending Tuesday, July 3, 1962 through the week ending Tuesday, December 29 1987.
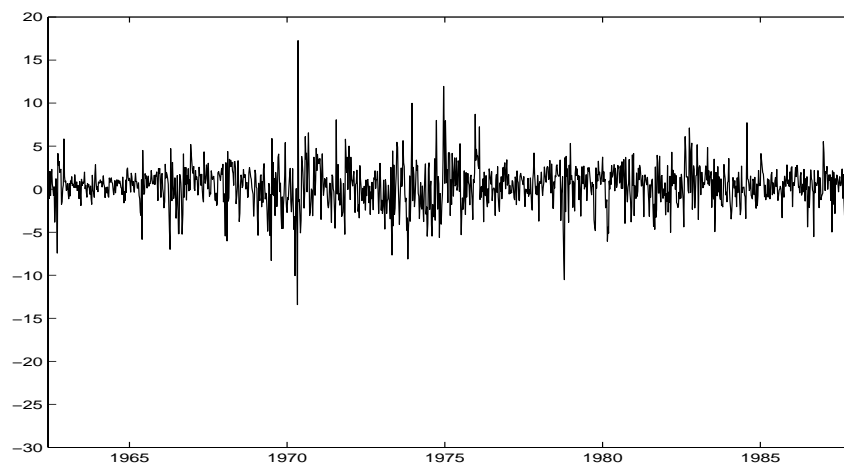
Figure 2: (a) SWARCH-L(4,4). Posterior density of the transition probabilities
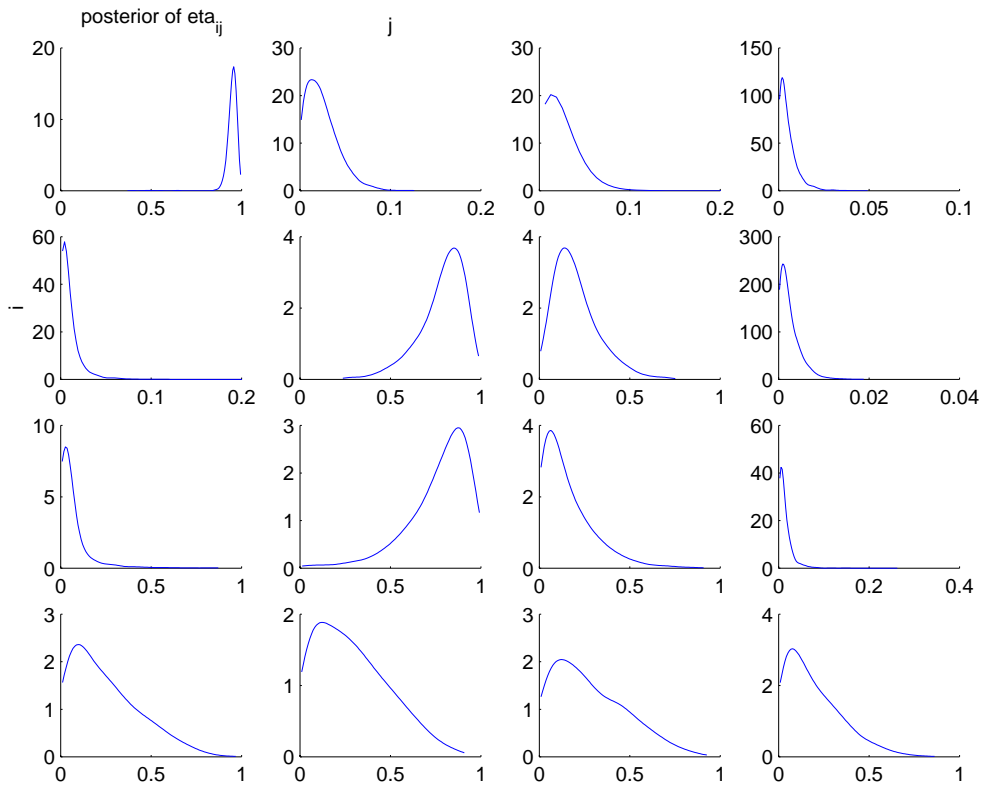
Figure 2: (b) SWARCH-L(4,4). Simulated values, the posterior, and the auto-correlation of the MCMC output for γ
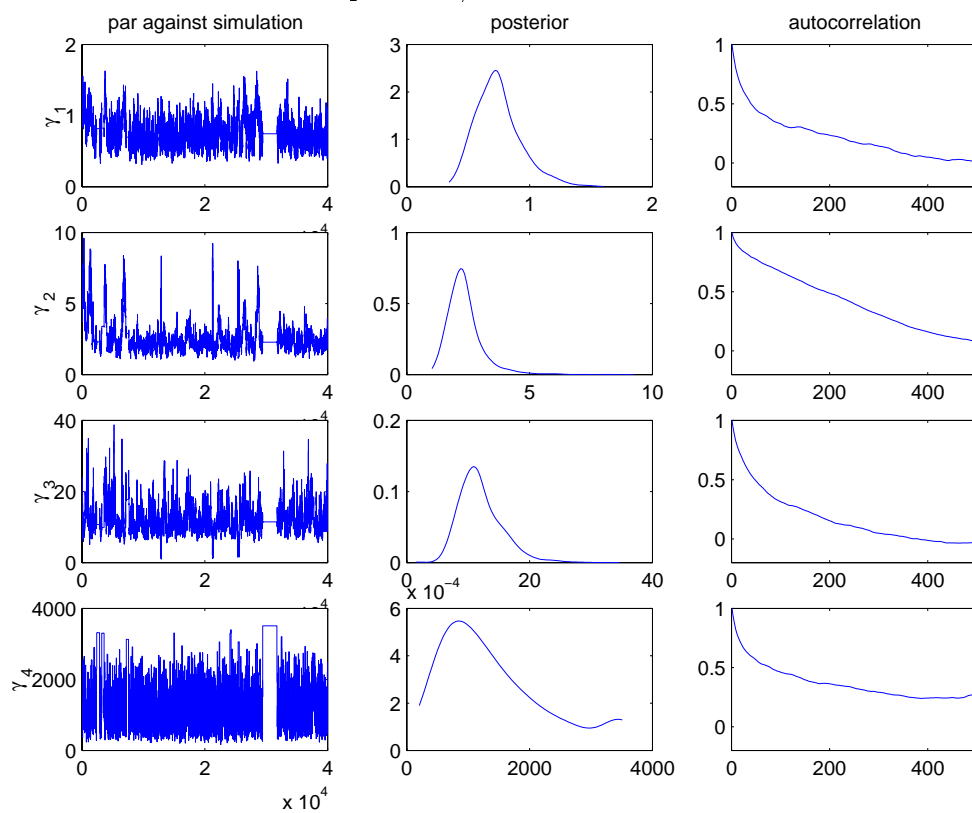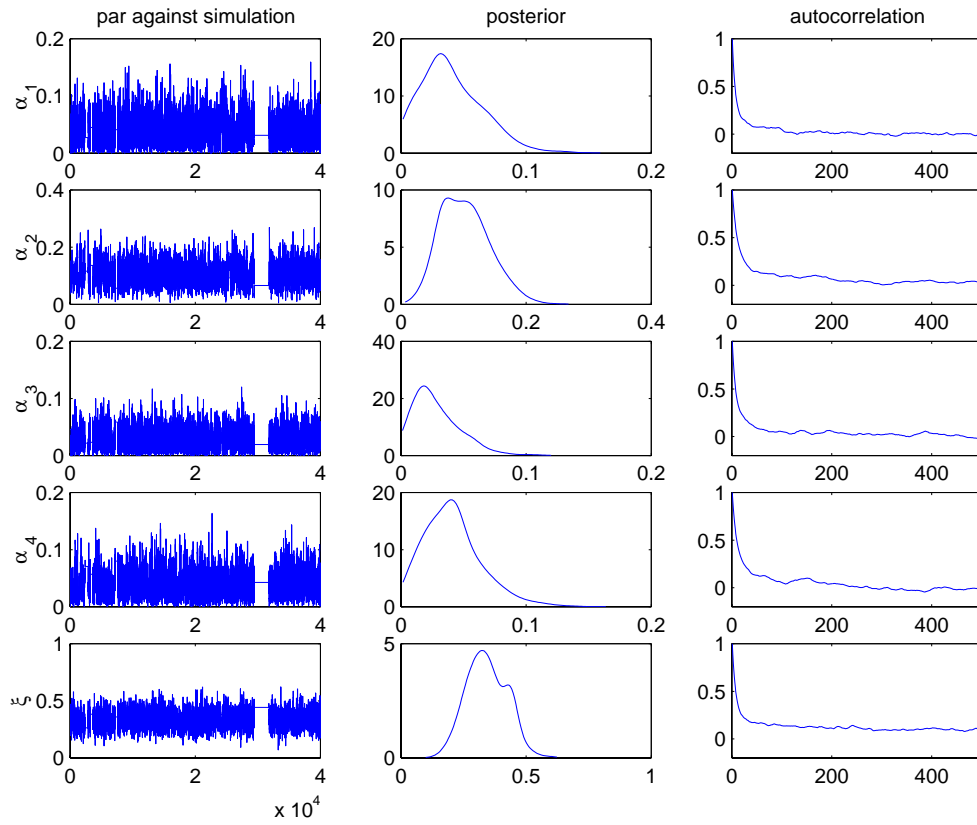
Figure 2: (c) SWARCH-L(4,4). Simulated values, the posterior, and the auto-correlation of the MCMC output for $(\alpha'\ \xi)$

Figure 3: SWARCH-L(4,4). Posterior smoothed probabilities of being in state j, $j = 1, \ldots, K$. The lines refer to business cycle turning points dated by the NBER.
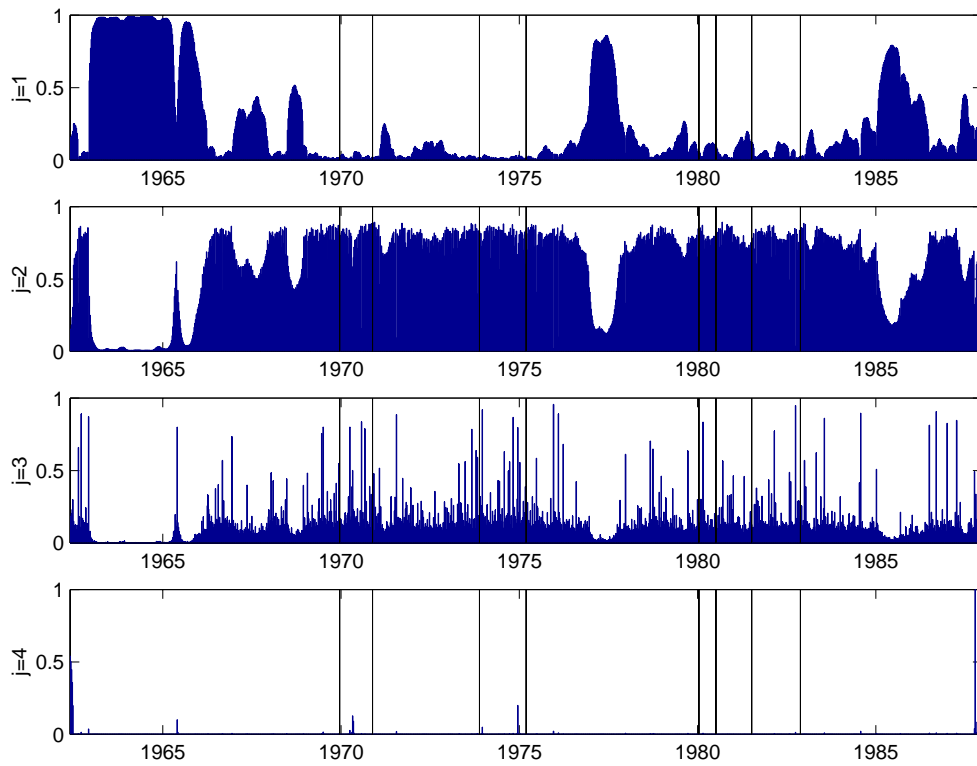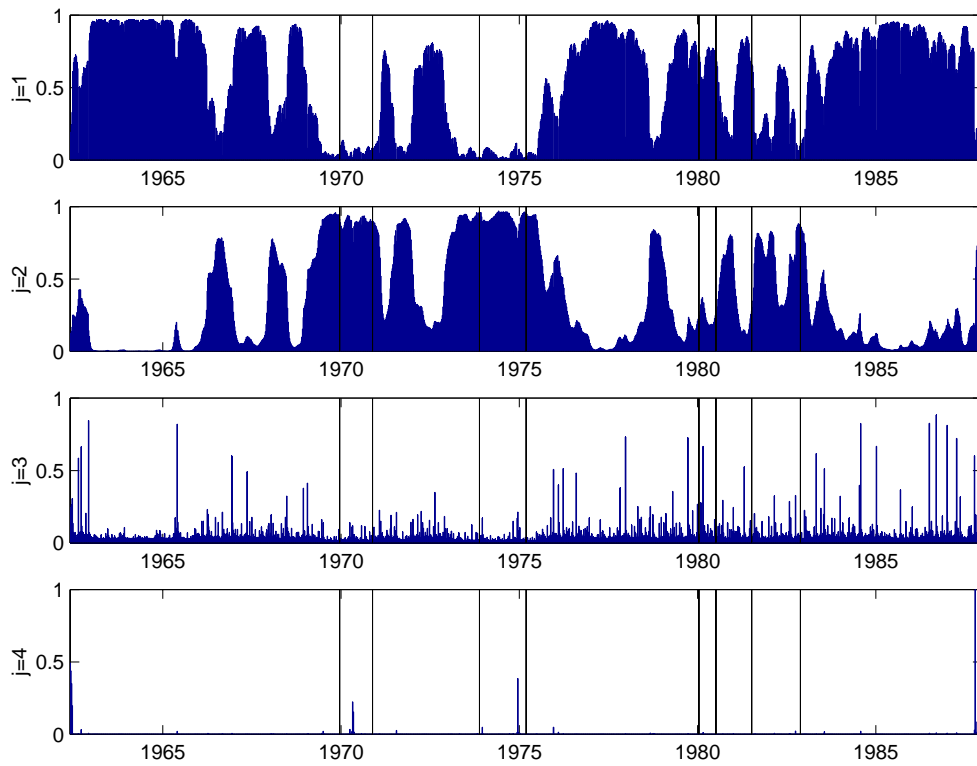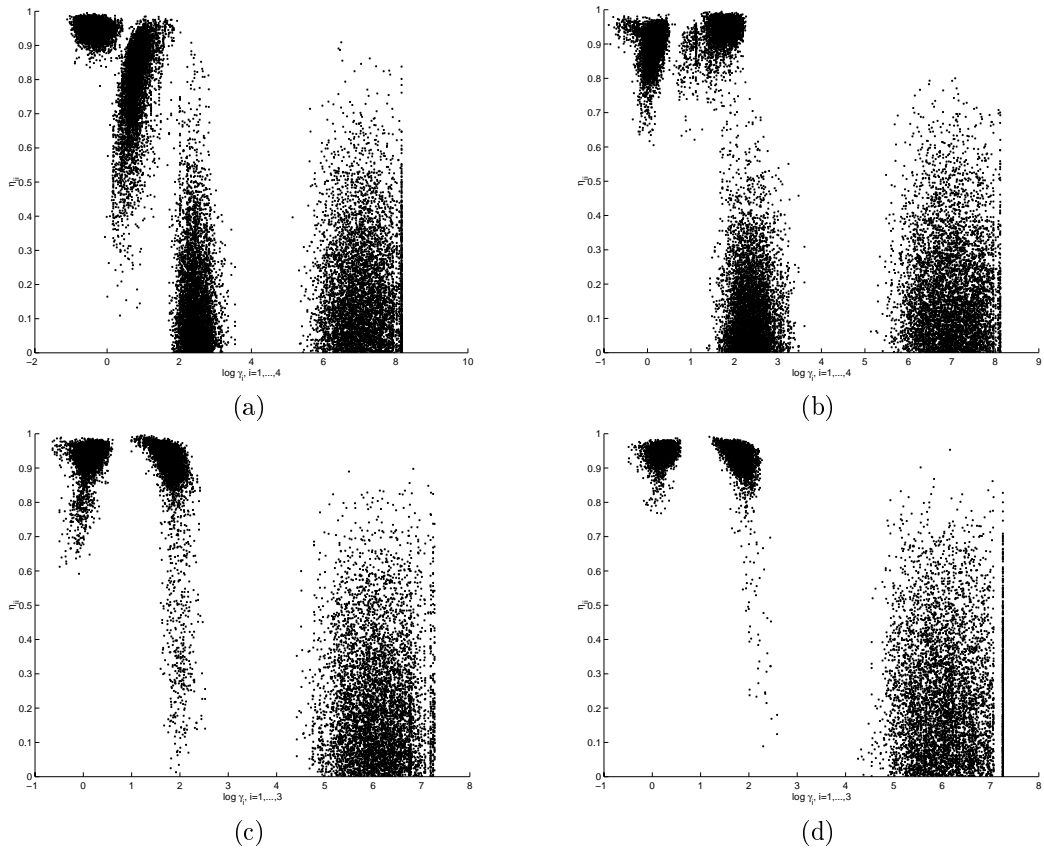
Figure 4: SWARCH-L(4,3). Posterior smoothed probabilities of being in state j, $j = 1, \ldots, K$. The lines refer to business cycle turning points dated by the NBER.

Figure 5: MCMC output of $\log \gamma$ against persistence. Top panels: $K = 4$, $m = 4, 3$, bottom panels: $K = 3$, $m = 4, 3$.

(a)

(b)

(c)

(d)

Figure 6: SWARCH-L(3,3). Posterior smoothed probabilities of being in state j, $j = 1, \ldots, K$. The lines refer to business cycle turning points dated by the NBER.
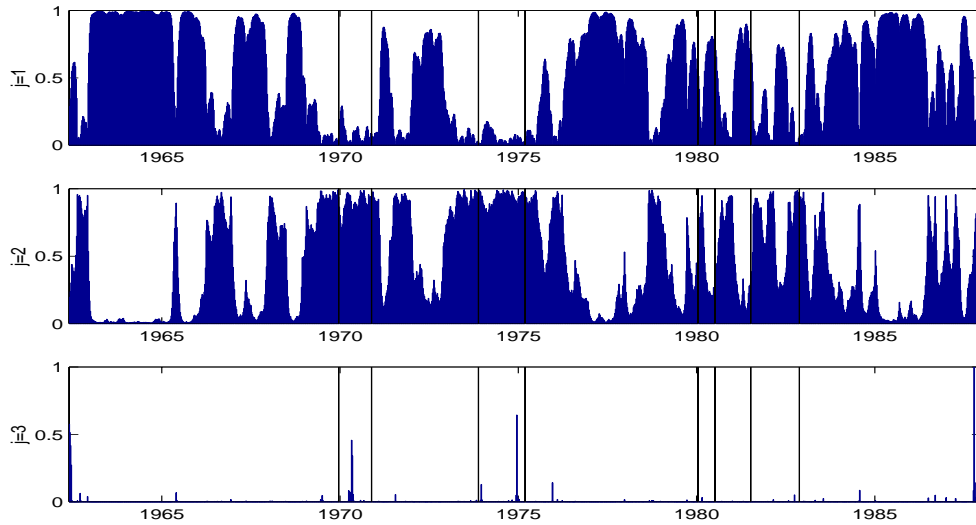

Figure 7: SWARCH-L(4,4). Plot of the empirical distribution function of the P-scores (edf), of the transformed P-scores against normal order statistics (normal plot), and the correlogram of the transformed P-scores and the transformed reflected P-scores (depicted along with a confidence band around zero of $[-1/(N-1) - 2/\sqrt{N}, -1/(N-1) + 2/\sqrt{N}]$).
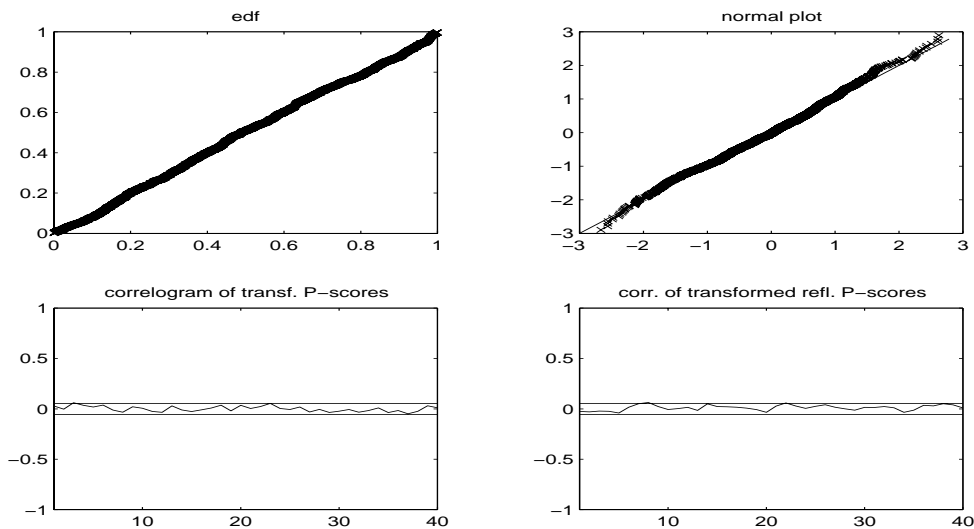
Figure 8: SWARCH-L(3,3). Plot of the empirical distribution function of the P-scores (edf), of the transformed P-scores against normal order statistics (normal plot), and the correlogram of the transformed P-scores and the transformed reflected P-scores (depicted along with a confidence band around zero of $[-1/(N-1) - 2/\sqrt{N}, -1/(N-1) + 2/\sqrt{N}]$).