SFB-Discussion Paper No. B-406

A Simple Regime-Switching Model for Stochastic Volatilities

Norbert Christopeit[1] and Axel Cron[2]

Juli 1997

Abstract: In this paper, a simple Markov switching model for the volatility of financial asset returns is presented. We discuss a moment estimation procedure and develop forecasts for future squared volatilities.

Keywords: Stochastic volatilities, Markov regime switching, Moment estimator.

JEL classification: C51, C53.

Financial support by Deutsche Forschungsgemeinschaft, Sonderforschungsbereich 303, is gratefully acknowledged.

[1]Institute for Econometrics and Operations Research, University of Bonn. Adenauerallee 24-42, D-53113 Bonn, Germany.
e-mail: christop@argos.ect.uni-bonn.de

[2]Institute for Econometrics and Operations Research, University of Bonn. Adenauerallee 24-42, D-53113 Bonn, Germany.
e-mail: cron@eos.ect.uni-bonn.de

1 Introduction

In recent years, stochastic volatility models have been a major issue in the econometrics of financial markets. The purpose of all these models is to account for empirical findings such as *fat tailed distributions* and *volatility clustering.*

The most prominent class of such models is provided by the ARCH-type models (cf. Engle(1995), Bollerslev et al.(1992)). In these models, the conditional variance of a financial time series is specified as a function of past observations of the series as well as of its own past. At least in the parametric case, this amounts to specifying a certain class of parametric functions (e. g. linear or logarithmic).

In this paper, we propose a completely different approach which is based on probabilistic assumptions instead of functional specifications. In particular, we pick up a simple model used in Duffie and Gray[1995] for the prediction of oil price returns. In this model, the conditional variance is driven by a two-state homogeneous Markov process. As a compensation for the lack of a functional specification, which is essential for the calculation of the likelihood function, a certain type of noncausality requirement for the return process in relation to the conditional variance has to made.

The main purposes of this paper are to investigate the probabilistic properties of the resulting return process, to present and to analyze an easy-to-handle estimation procedure for the parameters of interest, and, finally, to construct forecasts for future squared volatilities.

The paper is organized as follows. In section 2, we present the model and derive some basic properties of the resulting return process. Section 3 introduces the moment estimator the asymptotics of which are analyzed in section 4. The forecasts are constructed in section 5. In the final section, we illustrate the finite-sample properties of the estimators and the predictors by applying them to artificial data in Monte-Carlo experiments as well as to some historical exchange rate data.

2 The model and some asymptotic properties

We assume that the returns are generated by a model of the form

$$R_t = \mu + \sigma_t \epsilon_t, \tag{1}$$

where the random shocks ϵ_t to the returns R_t are normalized white noise. Under the assumptions to be made below, it will make no difference whether we work on a double sided time horizon ($t \in \mathbf{Z}$) or on a one sided horizon ($t \in \mathbf{N}$) with an appropriate initial condition. In either case, however, we shall assume that observations are available at times $t = 1, 2, \ldots, T$. In applications, we may take either the relative price changes

$$R_t = \frac{P_t - P_{t-1}}{P_{t-1}}$$

or the logarithmic price changes

$$R_t = \log P_t - \log P_{t-1}$$

as returns. The σ_t may be interpreted as random shocks to the variance, since

$$E\left\{(R_t - \mu)^2 | \mathcal{F}_{t-1}\right\} = E\{\sigma_t^2 | \mathcal{F}_{t-1}\} \tag{2}$$

if $\mathcal{F}_t$ denotes the $\sigma-$field containing the information available at time t and the ϵ_t satisfy the following assumption.

Assumption 1.
(i) The ϵ_t are i. i. d. , symmetric, normalized to variance one and possess an everywhere positive density w. r. to Lebesgue measure.
(ii) The ϵ_t possess moments up to order 12.
(iii) For each t, ϵ_t is independent of $\mathcal{F}_{t-1} \vee \sigma\{\sigma_t\}$.

Let us remark right here that, for the estimation procedure used below, the first three even moments of ϵ_t are assumed to be *known*. Hence, as is common practice in ARCH-related models, the standard normal distribution would be the canonical choice.

Assumption 1 leaves open a wide range of specifications for σ_t. In the majority of models treated in the literature, a functonal specification is made, which, in its most general form, may be written as

$$\sigma_t = F_t(R_{t-1}, R_{t-2}, \ldots; \sigma_{t-1}, \sigma_{t-2}, \ldots; \boldsymbol{\xi}_{t-1}, \boldsymbol{\xi}_{t-2}, \ldots; \boldsymbol{\eta}_t, \boldsymbol{\eta}_{t-1}, \ldots), \tag{3}$$

i. e. σ_t is a function of the past returns, its own history, as well as the past of some additional *observable* exogenous variables (e. g. prices of other assets), gathered in a random vector $\boldsymbol{\xi}_t$ plus *unobservable* random influences $\boldsymbol{\eta}_t$. The information $\sigma-$field $\mathcal{F}_t$ would then be given by either $\mathcal{F}_t = \mathcal{F}_t^{R,\sigma,\boldsymbol{\xi}} = \sigma(R_t, R_{t-1}, \ldots; \sigma_t, \sigma_{t-1}, \ldots; \boldsymbol{\xi}_t, \boldsymbol{\xi}_{t-1}, \ldots)$ if σ_t is itself observable or $\mathcal{F}_t = \mathcal{F}_t^{R,\boldsymbol{\xi}}$ in case σ_t cannot be directly observed (which should be the more realistic case). Note that, if the $\boldsymbol{\eta}_t, \boldsymbol{\eta}_{t-1}, \ldots,$ are lacking in (3), σ_t may be obtained as a function of the (infinite) past of the observed processes (R_t) and $(\boldsymbol{\xi}_t)$ under certain technical assumptions about the functions F_t which ensure that infinite substitution for $\sigma_{t-1}, \ldots,$ is possible (cf. Cron(1997)). In this case, σ_t is $\mathcal{F}_{t-1} = \mathcal{F}_{t-1}^{R,\boldsymbol{\xi}}-$measurable and (2) becomes

$$E\left\{(R_t - \mu)^2 | \mathcal{F}_{t-1}\right\} = \sigma_t^2,$$

i. e. σ_t is the *conditional standard deviation* of the return given the information up to time $t-1$. This approach is in the spirit of the (parametric or nonparametric) (G)ARCH-models (cf. Engle(1995)).

Instead of specifying a class of functions F_t, we shall assume the following simple regime-switching model considered in Duffie and Gray(1995).

Assumption 2.
σ_t may take two different levels v_a and v_b, and evolves as a homogeneous Markov process w. r. to its own history with transition probabilities given by the matrix

$$P = \begin{pmatrix} p_{aa} & p_{ab} \\ p_{ba} & p_{bb} \end{pmatrix}.$$

Here, e. g. , p_{ab} is the conditional probability for $\sigma_t = v_b$, given that $\sigma_{t-1} = v_a$. Note that $p_{ab} = 1 - p_{aa}$, $p_{ba} = 1 - p_{bb}$.

This approach of modelling regime switches by means of homogeneous Markov chains has been made popular by Hamilton (e.g. Hamilton (1989) and Hamilton (1994)). There, however, it is used to describe changes in the regime of a trend component or of parameters of an autoregression, whereas the noise is not subject to regime changes.

In order to analyze the asymptotic behavior of this Markov chain, denote $\pi = p_{aa} + p_{bb} - 1$ and assume that $0 < p_{aa} < 1, 0 < p_{bb} < 1$, $\pi \neq 1$. Then it is easily calculated (by induction) that the $t-$step transition matrix is given by

$$P^t = \frac{1}{1-\pi} \begin{pmatrix} 1 - p_{bb} + (1-p_{aa})\pi^t & (1-p_{aa})(1-\pi^t) \\ (1-p_{bb})(1-\pi^t) & 1 - p_{aa} + (1-p_{bb})\pi^t \end{pmatrix}.$$

Consequently, as $t \to \infty$,

$$P^t \longrightarrow \frac{1}{1-\pi} \begin{pmatrix} 1-p_{bb} & 1-p_{aa} \\ 1-p_{bb} & 1-p_{aa} \end{pmatrix},$$

and

$$p_a = \frac{1-p_{bb}}{1-\pi}, \quad p_b = \frac{1-p_{aa}}{1-\pi}$$

is the (unique) invariant distribution. In particular, the Markov chain (σ_t) is ergodic. Henceforth, we shall assume that the initial distribution of σ_t at $t = 0$ is given by the invariant distribution (p_a, p_b), i. e. that the process has been running long enough to attain stationarity. Then (σ_t) is a (strictly) stationary ergodic process. Unfortunately, with respect to the observable process (R_t), which is of main concern for our estimation procedure, assumptions 1 and 2 do not imply very much, except that it is a martingale difference sequence w. r. to the observation $\sigma-$field. In order to get a tractable form of the first order autocorrelation of the squared returns or of the likelihood function (cf. section 3), we impose the following strengthened version of assumption 2 , which captures the notion that dependence of σ_t on the whole past should only be through σ_{t-1}.

Assumption 3.
σ_t is conditionally independent of $\mathcal{F}_{t-1}$, given σ_{t-1}.

As pointed out in the introduction, assumptions 2 and 3 are meant as an alternative to popular functional specifications of the conditional variance. It is therefore not surprising that the relationship between both kinds of models is rather intricate. While assumption 2 in itself is rather harmless (apart form the binary structure of σ_t) and is satisfied by a large class of specifications (3), assumption 3 is compatible with (3) only under some rather restrictive additional conditions on the functional specification, ruling out in particular the case of completely predictible σ_t. Therefore, the functional and the probabilistic models should be considered as basically different approaches, and it seems to make little sense to start out with a functional model and try to fit it into the above framework.

Assumption 3 — or rather the somewhat stronger version in which conditional independence of σ_t of $\mathcal{F}_{t-1}$, given $\underline{\sigma}_{t-1} = (\sigma_t, \sigma_{t-1}, \ldots)$ is required — may be thought of as a nonlinear version of Granger-noncausality of (R_t) for (σ_t). Actually, assumption 3 implies that (R_t) and (σ_t) are uncorrelated and hence (R_t) is Granger-noncausal for (σ_t) in the linear sense.

Assumptions 1-3 allow us to calculate the following moments of $x_t = R_t - \mu$.

$$\begin{aligned} m_k &= E(x_t^k) = E(\sigma_t^k \epsilon_t^k) = (p_a v_a^k + p_b v_b^k)\alpha_k, && (4)\\ m_{kl} &= E\left\{x_{t-1}^k x_t^l\right\} = E(x_{t-1}^k \sigma_t^l \epsilon_t^l) && (5)\\ &= \left[p_a v_a^k (p_{aa} v_a^l + p_{ab} v_b^l) + p_b v_b^k (p_{ba} v_a^l + p_{bb} v_b^l)\right] \alpha_k \alpha_l, \end{aligned}$$

with $\alpha_k = E(\epsilon_t^k)$.

Lemma 1.
Under assumptions 1-3, the (unconditional) likelihood function of the $(x_t)-$ process is given by

$$L(\boldsymbol{\theta}|x_1, \ldots, x_T) = f(x_1, \ldots, x_T|\boldsymbol{\theta}) = tr\left\{\Pi \prod_{t=1}^{T} P\Phi(x_t)\right\} \qquad (6)$$

with

$$\Phi(x) := \begin{pmatrix} \phi(x, v_a) & 0 \\ 0 & \phi(x, v_b) \end{pmatrix}, \quad \Pi = (p_a \boldsymbol{\iota}, (1-p_a)\boldsymbol{\iota}), \quad \boldsymbol{\iota}' = (1,1),$$

$\phi(x, v) = v^{-1} g(x/v)$, g being the density of ϵ_t and $\boldsymbol{\theta} = (\mu, v_a^2, v_b^2, p_{aa}, p_{bb})$ the vector of parameters to be determined.

Proof.
Let $g_1, \ldots, g_T$ be bounded Borel functions, and put $I_a = 1_{\{v_a\}}, I_b = 1_{\{v_b\}}$ (indicator functions on $V = \{v_a, v_b\}$). Then, since the conditional distribution of x_T given $\mathcal{F}_{T-1} \vee \sigma(\sigma_T)$ is $\phi(\xi_T, \sigma_T)$,

$$\begin{aligned}
& E\{g_1(x_1)\cdots g_T(x_T)I_a(\sigma_T)\} \\
= \; & E\{g_1(x_1)\cdots g_{T-1}(x_{T-1})I_a(\sigma_T)E\{g_T(x_T)|\mathcal{F}_{T-1}, \sigma_T\}\} \\
= \; & E\{g_1(x_1)\cdots g_{T-1}(x_{T-1})I_a(\sigma_T)\int g_T(\xi_T)\phi(\xi_T, v_a)d\xi_T\} \\
= \; & \int g_T(\xi_T)E\{g_1(x_1)\cdots g_{T-1}(x_{T-1})P(\sigma_T = v_a|\mathcal{F}_{T-1}, \sigma_{T-1})\}\phi(\xi_T, v_a)d\xi_T \\
= \; & \int g_T(\xi_T)E\{g_1(x_1)\cdots g_{T-1}(x_{T-1})[p_{aa}I_a(\sigma_{T-1}) + p_{ba}I_b(\sigma_{T-1})]\}\phi(\xi_T, v_a)d\xi_T.
\end{aligned}$$

In the fourth equality we have made use of assumption 3. Similarly,

$$\begin{aligned}
& E\{g_1(x_1)\cdots g_T(x_T)I_b(\sigma_T)\} \\
= \; & \int g_T(\xi_T)E\{g_1(x_1)\cdots g_{T-1}(x_{T-1})[p_{ab}I_a(\sigma_{T-1}) + p_{bb}I_b(\sigma_{T-1})]\}\phi(\xi_T, v_b)d\xi_T.
\end{aligned}$$

Hence

$$\begin{aligned}
& E\{g_1(x_1)\cdots g_T(x_T)(I_a(\sigma_T), I_b(\sigma_T))\} \\
= \; & \int g_T(\xi_T)E\{g_1(x_1)\cdots g_{T-1}(x_{T-1})(I_a(\sigma_{T-1}), I_b(\sigma_{T-1}))\}P\Phi(\xi_T)d\xi_T.
\end{aligned}$$

Proceeding inductively, we find that

$$\begin{aligned}
& E\{g_1(x_1)\cdots g_T(x_T)\} \\
= \; & \int g_T(\xi_T)\int g_{T-1}(\xi_{T-1})\int\cdots \\
& \int g_2(\xi_2))E\{g_1(x_1)(I_a(\sigma_1, I_b(\sigma_1))\}\prod_{t=2}^{T}P\Phi(\xi_t)\boldsymbol{\iota}d\xi_2\cdots d\xi_T.
\end{aligned} \tag{7}$$

But

$$\begin{aligned}
E\{g_1(x_1)I_a(\sigma_1)\} &= \int g_1(\xi_1)p_a\phi(\xi_1, v_a)d\xi_1, \\
E\{g_1(x_1)I_b(\sigma_1)\} &= \int g_1(\xi_1)p_b\phi(\xi_1, v_b)d\xi_1,
\end{aligned}$$

hence

$$E\{g_1(x_1)(I_a(\sigma_1), I_b(\sigma_1))\} = \int g_1(\xi_1)(p_a, p_b)\Phi(\xi_1)d\xi_1.$$

In view of (7), the unconditional density of $(x_T, \ldots, x_1)$ is therefore given by

$$(p_a, p_b)\Phi(\xi_1)\prod_{t=2}^{T}P\Phi(\xi_t)\boldsymbol{\iota} = \operatorname{tr}\left\{\Pi\Phi(\xi_1)\prod_{t=2}^{T}P\Phi(\xi_t)\right\},$$

which coincides with (6) since $\Pi P = \Pi$. ■

From (6) it is easily calculated that the (unconditional) distribution of x_t is a mixture with density

$$f(x_t|\boldsymbol{\theta}) = p_a\phi(x_t, v_a) + p_b\phi(x_t, v_b).$$

Proposition 1.
(x_t) is stationary and $\phi-$mixing.

Proof.
Stationarity is immediately clear from the time invariance of the likelihood function (6). As for the the mixing property, note first that, for every bounded Borel function $g(v, x)$,

$$\begin{aligned} E\{g(\sigma_{t+1}, x_{t+1})|\mathcal{F}_t\} &= E\{E\{g(\sigma_{t+1}, \sigma_{t+1}\epsilon_{t+1})|\mathcal{F}_t, \sigma_{t+1}\}|\mathcal{F}_t\} \\ &= E\{\overline{g}(\sigma_{t+1})|\mathcal{F}_t\} \\ &= E\{\{\overline{g}(\sigma_{t+1})|\mathcal{F}_t, \sigma_t\}|\mathcal{F}_t\} \\ &= E\{\overline{\overline{g}}(\sigma_t)|\mathcal{F}_t\}, \end{aligned} \tag{8}$$

where for the second equality we have used assumption 1(iii) and for the fourth assumption 3. Note that the Borel functions $\overline{g}, \overline{\overline{g}}$ satisfy the same bound as g. By induction it follows that, for every bounded Borel function $g(x_1, \ldots, x_m)$,

$$E\{g(x_{t+n+1}, \ldots, x_{t+n+m})|\mathcal{F}_{t+n}\} = E\{\overline{g}(\sigma_{t+n})|\mathcal{F}_{t+n}\}, \tag{9}$$

with the function $\overline{g}$ satisfying the same bound as g. (Do it first for products of indicator functions $g = 1_{B_1}\ldots 1_{B_m}$, using (8), and extend to all measurable bounded functions, using a monotone class argument.) We shall show that

$$|E\{\overline{g}(\sigma_{t+n})|\mathcal{F}_t\} - E\{\overline{g}(\sigma_{t+n})\}| \leq \phi(n) \tag{10}$$

for all Borel functions $g(x_1, \ldots, x_m)$ bounded by 1, from which, by virtue of (9),

$$|E\{g(x_{t+n+1}, \ldots, x_{t+n+m})|\mathcal{F}_t\} - E\{g(x_{t+n+1}, \ldots, x_{t+n+m})\}| \leq \phi(n),$$

implying the mixing property. As for (10), note that, by induction (using assumption 3)

$$E\{\overline{g}(\sigma_{t+n})|\mathcal{F}_t\} = \mathbf{p}'_{t|t}P^n\overline{\mathbf{g}},$$

where we have put

$$\begin{aligned} \overline{\mathbf{g}} &= (\overline{g}(v_a), \overline{g}(v_b))', \\ \mathbf{p}_{t|t} &= (p_{t|t}(v_a), p_{t|t}(v_b))', \\ p_{t|t}(v) &= P(\sigma_t = v|\mathcal{F}_t), \end{aligned}$$

and

$$E\{\overline{g}(\sigma_{t+n}\} = \mathbf{p}' P^n \overline{\mathbf{g}},$$

with $\mathbf{p} = (p_a, 1-p_a)'$. Hence the left hand side of (10) can be estimated by

$$|(\mathbf{p}_{t|t} - \mathbf{p})' P^n \overline{\mathbf{g}}| \leq 2\pi^n =: \phi(n),$$

since, for every initial distribution $\tilde{\mathbf{p}} = (\tilde{p}_a, 1-\tilde{p}_a)'$

$$(\tilde{\mathbf{p}} - \mathbf{p})' P^n = (\tilde{p}_a - p_a)\pi^n(1, -1). \quad \blacksquare$$

As a consequence, the processes (x_t^k) and $(x_{t-1}^k x_t^k)$ are ergodic and we have the following asymptotics.

Proposition 2.

$$\frac{1}{T}\sum_{t=1}^{T} x_t^k \to m_k \quad a.\ s.\ ,$$
$$\frac{1}{T}\sum_{t=1}^{T} x_{t-1}^k x_t^l \to m_{kl} \quad a.\ s.$$

(whenever the theoretical moments exist).

Define

$$X_t(k) = \begin{cases} x_t & \text{if} \quad k = 0 \\ x_t^{2k} - m_{2k} & \text{if} \quad k = 1, 2, 3, \\ x_{t-1}^2 x_t^2 - m_{22} & \text{if} \quad k = 4, \end{cases} \tag{11}$$

$\mathbf{X}_t = (X_t(0), \ldots, X_t(4))'$. Then $(\mathbf{X}_t)$ is $\phi-$mixing with the $\phi(n)-$sequence from the proof of Proposition 1, and standard CLT's for $\phi-$mixing processes show (cf. Billingsley(1968))

Proposition 3.

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T} \mathbf{X}_t \xrightarrow{D} \mathcal{N}(\mathbf{0}, \Sigma), \tag{12}$$

where Σ *is given by*

$$\Sigma = E(\mathbf{X}_1 \mathbf{X}_1') + \sum_{t=2}^{\infty}[E(\mathbf{X}_1 \mathbf{X}_t') + E(\mathbf{X}_t' \mathbf{X}_1)].$$

Actual, a functional central limit theorem is true, but we shall not need it. The matrix Σ is calculated in appendix B.

3 The estimator

The model presented above contains five unknown parameters $\mu, v_a^2, v_b^2, p_{aa}$, p_{bb}. For the estimation procedure to be chosen it turns out more convenient to work with the equivalent parametrization $\boldsymbol{\theta} = (\mu, v_a^2, v_b^2, p_a, p_{aa})$. Since $(v_a^2, v_b^2, p_{aa}, p_{bb})$ and $(v_b^2, v_a^2, p_{bb}, p_{aa})$ lead to observationally equivalent structures, we impose the restriction $v_a^2 > v_b^2$. As estimator of $\boldsymbol{\theta}$, we take the *moment estimator* based on the first three even central moments and the mean of the observed process (R_t) plus the first order autocorrelation of the squared deviations from the mean. I. e. , $\hat{\boldsymbol{\theta}} = (\hat{\mu}, \hat{v}_a^2, \hat{v}_b^2, \hat{p}_a, \hat{p}_{aa})$ is determined by solving the following system of nonlinear equations

$$(p_a v_a^{2k} + p_b v_b^{2k})\alpha_{2k} = M_{2k}, \quad k = 1, 2, 3, \tag{13}$$

$$\left[p_a v_a^2 (p_{aa} v_a^2 + p_{ab} v_b^2) + p_b v_b^2 (p_{ba} v_a^2 + p_{bb} v_b^2)\right] \alpha_2^2 = M_{2,2}, \tag{14}$$

with the empirical moments

$$\begin{aligned}
\hat{\mu} &= \overline{R} = \frac{1}{T}\Sigma_{t=1}^T R_t, \\
M_{2k} &= \frac{1}{T}\Sigma_{t=1}^T (R_t - \overline{R})^{2k}, \\
M_{2,2} &= \frac{1}{T-1}\Sigma_{t=2}^T (R_t - \overline{R})^2 (R_{t-1} - \overline{R})^2.
\end{aligned}$$

The left hand sides of (13) and (14) are the theoretical moments $m_{2k} = E(R_t - \mu)^{2k}$ and $m_{2,2} = E\{(R_t - \mu)^2 (R_{t-1} - \mu)^2\}$, respectively. (13)-(14) has to be solved for $v_a^2, v_b^2, p_a, p_{aa}$ (under the restriction $v_a^2 > v_b^2$), using the definitorial equations of section 2 to eliminate p_b, etc. :

$$p_b = 1 - p_a, \quad p_{ab} = 1 - p_{aa}, \quad p_{ba} = \frac{p_a(1 - p_{aa})}{1 - p_a},$$

$$p_{bb} = \frac{1 - (2 - p_{aa})p_a}{1 - p_a}. \tag{15}$$

With these substitutions, (14) may be written

$$p_{aa} p_a (v_a^2 - v_b^2)^2 + 2 p_a v_a^2 v_b^2 + (1 - 2p_a) v_b^4 = M_{2,2}.$$

As usual, the motivation behind this choice of estimation procedure is that, in view Proposition 2, the empirical moments approach the theoretical ones as the sample size increases and therefore the the solutions of (13) and (14) should be expected to tend to the true parameter values. Indeed, if the system (13)-(14) possesses a unique solution with probability one and if its Jacobian is nonzero for all admissible parameter values, then strong consistency of the moment estimators can be shown.

Actually, as is shown in appendix A, we may obtain explicit formulas for the solution of (13)-(14), namely

$$\hat{v}_a^2 = \Gamma_1 + \sqrt{\frac{\sqrt{4+C^2}+C}{\sqrt{4+C^2}-C}}\sqrt{\Gamma_2-\Gamma_1^2}, \tag{16}$$

$$\hat{v}_b^2 = \Gamma_1 - \sqrt{\frac{\sqrt{4+C^2}-C}{\sqrt{4+C^2}+C}}\sqrt{\Gamma_2-\Gamma_1^2}, \tag{17}$$

$$\hat{p}_a = \frac{1}{2}\left[1-\frac{C}{\sqrt{4+C^2}}\right], \tag{18}$$

$$\hat{p}_{aa} = \frac{\Gamma_{2,2}-2\hat{p}_a\hat{v}_a^2\hat{v}_b^2-(1-2\hat{p}_a)\hat{v}_b^2}{\hat{p}_a(\hat{v}_b^2-\hat{v}_b^2)^2}, \tag{19}$$

with $\Gamma_k = M_{2k}/\alpha_{2k}, \Gamma_{2,2} = M_{2,2}/\alpha_2^2, C = [\Gamma_3 - \Gamma_1^3 - 3\Gamma_1(\Gamma_2 - \Gamma_1^2)]/(\sqrt{\Gamma_2-\Gamma_1^2})^3$.

4 Asymptotics of the moment estimator

The asymptotic behavior of the $(x_t)-$process transfers in a straightforward manner to the empirical moments M_{2k} and $M_{2,2}$.

Proposition 4.
(i) With probability one, $\overline{R} \to \mu, \quad M_{2k} \to m_{2k}, \quad k = 1,2,3, \quad M_{2,2} \to m_{2,2}$.
(ii) $\sqrt{T}(\overline{R}-\mu), \sqrt{T}(M_{2k}-m_{2k}), k=1,2,3,$ *and* $\sqrt{T}(M_{2,2}-m_{2,2})$ *are asymptotically jointly normal with mean* $\mathbf{0}$ *and covariance matrix* Σ *(from Prop. 3).*

This is an immediate consequence of the following (not surprising) lemma.

Lemma 2.
(i) $M_{2k} = \frac{1}{T}\sum_{t=1}^T x_t^{2k} + R_T(k)$ *with* $R_T(k) \to 0$ *a. s. and* $\sqrt{T}R_T(k) \xrightarrow{P} 0$.
(ii) $M_{2,2} = \frac{1}{T}\sum_{t=2}^T x_{t-1}^2 x_t^2 + S_T$ *with* $S_T \to 0$ *a. s. and* $\sqrt{T}S_T \xrightarrow{P} 0$.

Proof.
(i) Since

$$M_{2k} = \frac{1}{T}\sum_{t=1}^T (x_t+\mu-\overline{R})^{2k},$$

binomial expansion yields

$$\begin{aligned} R_T(k) &= \sum_{\nu=1}^{2k}\binom{2k}{\nu}(\mu-\overline{R})^\nu \frac{1}{T}\sum_{t=1}^T x_t^{2k-\nu} \\ &= \sum_{\nu=1}^{2k}\binom{2k}{\nu}\left(-\frac{1}{T}\sum_{t=1}^T x_t\right)^\nu \frac{1}{T}\sum_{t=1}^T x_t^{2k-\nu}. \end{aligned}$$

Then the first assertion follows from the ergodic behavior of (x_t) (cf. Prop. 2). Moreover, for $\nu > 1$,

$$\sqrt{T}\left(\frac{1}{T}\Sigma_{t=1}^{T}x_t\right)^{\nu} = T^{-(\nu-1)/2}\left(\frac{1}{\sqrt{T}}\Sigma_{t=1}^{T}x_t\right)^{\nu} \xrightarrow{P} 0 \tag{20}$$

by Prop. 3, while for $\nu = 1$

$$\frac{1}{\sqrt{T}}\Sigma_{t=1}^{T}x_t \cdot \frac{1}{T}\Sigma_{t=1}^{T}x_t^{2k-1} \xrightarrow{P} 0$$

(since $E(x_t^{2k-1}) = 0$). This shows the second part of (i).
(ii) Note that

$$\begin{aligned} M_{2,2} &= \frac{1}{T}\Sigma_{t=1}^{T}[x_{t-1} + (\mu - \overline{R})]^2[x_t + (\mu - \overline{R})]^2 \\ &= \frac{1}{T}\Sigma_{t=1}^{T}x_{t-1}^2x_t^2 + S_T \end{aligned}$$

with

$$\begin{aligned} S_T &= (\mu - \overline{R})\frac{2}{T}\Sigma_{t=1}^{T}(x_{t-1}^2x_t + x_{t-1}x_t^2) \\ &\quad +(\mu - \overline{R})^2\frac{1}{T}\Sigma_{t=1}^{T}(x_{t-1}^2 + 4x_{t-1}x_t + x_t^2) \\ &\quad +(\mu - \overline{R})^3\frac{2}{T}\Sigma_{t=1}^{T}(x_{t-1} + x_t) + (\mu - \overline{R})^4. \end{aligned}$$

Ergodicity of (x_t) yields $S_T \to 0$ a. s. $\sqrt{T}S_T \xrightarrow{P} 0$ follows from $\sqrt{T}(\overline{R} - \mu)^{\nu} \xrightarrow{P} 0$ for $\nu > 1$ (cf. (20)) and

$$\lim_{T\to\infty}\frac{1}{T}\Sigma_{t=1}^{T}x_{t-1}^2x_t = \lim_{T\to\infty}\frac{1}{T}\Sigma_{t=1}^{T}x_{t-1}x_t^2 = 0 \quad a.\ s. \quad \blacksquare$$

The moment estimator $\hat{\boldsymbol{\theta}}_T$ has been determined as the (unique) solution of a nonlinear equation of the form

$$\mathbf{f}(\hat{\boldsymbol{\theta}}_T) = \boldsymbol{\Gamma}, \tag{21}$$

where $\boldsymbol{\Gamma} = \boldsymbol{\Gamma}(T) = (\overline{R}, \Gamma_1, \Gamma_2, \Gamma_3, \Gamma_{2,2})'$ with $\Gamma_k = M_{2k}/\alpha_{2k}$, $\Gamma_{2,2} = M_{2,2}$ (cf. section 3) and $\mathbf{f}(\boldsymbol{\theta}) = (f_0(\boldsymbol{\theta}), \ldots, f_4(\boldsymbol{\theta})'$ with $f_0(\boldsymbol{\theta}) = \mu$ and $f_k(\boldsymbol{\theta})$ given by the left hand side of (13)-(14) for $k = 1, 2, 3, 4$, (without the $\alpha-$factors) in conjunction with (15). Since $\mathbf{f}$ has a continuous inverse on the whole parameter space (cf. (16)-(19)) and the true parameter value $\boldsymbol{\theta}$ satisfies the same equation (21) with $\boldsymbol{\Gamma}$ replaced by its theoretical counterpart

$$\boldsymbol{\gamma} = (\mu, \gamma_1, \gamma_2, \gamma_3, \gamma_{2,2})',$$

$\gamma_k = m_{2k}/\alpha_{2k}$, $\gamma_{2,2} = m_{2,2}$ (cf. (4),(5)), it follows from Proposition 2 that

$$\hat{\boldsymbol{\theta}}_T \to \boldsymbol{\theta} \quad a.\ s. \tag{22}$$

Next, the first derivatives of $\mathbf{f}$ w. r. to $\boldsymbol{\theta}$ are easily calculated to be given by

$$\begin{array}{ll}
\frac{\partial f_0}{\partial \mu} = 1, & \frac{\partial f_0}{\partial \theta_j} = 0 \quad \text{for} \quad j = 1,2,3,4, \\
\frac{\partial f_k}{\partial \mu} = 0, & \frac{\partial f_k}{\partial v_a^2} = k p_a v_a^{2(k-1)}, \quad \frac{\partial f_k}{\partial v_b^2} = k p_b v_b^{2(k-1)}, \\
 & \frac{\partial f_k}{\partial p_a} = v_a^{2k} - v_b^{2k}, \quad \frac{\partial f_k}{\partial p_{aa}} = 0 \quad \text{for} \quad k = 1,2,3, \\
\frac{\partial f_4}{\partial \mu} = 0, & \frac{\partial f_4}{\partial v_a^2} = 2p_a[p_{aa} v_a^2 + (1 - p_{aa}) v_b^2], \\
 & \frac{\partial f_4}{\partial v_b^2} = 2[p_a(1 - p_{aa}) v_a^2 + (1 - (2 - p_{aa}) p_a) v_b^2], \\
 & \frac{\partial f_4}{\partial p_a} = p_{aa} v_a^4 + v_b^2[2(1 - p_{aa}) v_a^2 + (2 - p_{aa}) v_b^2], \\
 & \frac{\partial f_4}{\partial p_{aa}} = p_a (v_a^2 - v_b^2)^2.
\end{array}$$

Hence the matrix $F(\boldsymbol{\theta}) = D\mathbf{f}(\boldsymbol{\theta})$ of first partial derivatives is of the form

$$F = F(\boldsymbol{\theta}) = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & & & & 0 \\ 0 & & \tilde{F} & & 0 \\ 0 & & & & 0 \\ 0 & * & * & * & f_{44} \end{pmatrix}$$

with $f_{44} > 0$. The determinant of $\tilde{F}$ is easily calculated as

$$|\tilde{F}| = 3 p_a p_b (v_a^4 + v_b^4)(v_a^2 - v_b^2) > 0$$

on the whole parameter space (note that we have imposed the constraint $v_a^2 > v_b^2$). Hence F^{-1} exists and is continuous on the whole parameter space. By first order Taylor expansion

$$\begin{aligned} \mathbf{f}(\hat{\boldsymbol{\theta}}_T) - \mathbf{f}(\hat{\boldsymbol{\theta}}) &= F(\boldsymbol{\theta} + \lambda_T(\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}))(\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}) \\ &= \boldsymbol{\Gamma}(T) - \boldsymbol{\gamma}, \quad 0 < \lambda_T < 1. \end{aligned}$$

This together with (22) and Prop. 3 implies that

$$\sqrt{T}(\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \Sigma_\theta)$$

with

$$\begin{aligned} \Sigma_\theta &= F(\boldsymbol{\theta})^{-1} \Lambda \Sigma \Lambda F(\boldsymbol{\theta})'^{-1}, & (23) \\ \Lambda^{-1} &= \text{diag}(1, 1, \alpha_4, \alpha_6, 1). & (24) \end{aligned}$$

Proposition 5.
(i) The moments estimator $\hat{\boldsymbol{\theta}}_T$ is strongly consistent.
(ii) $\sqrt{T}(\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta})$ is asymptotically normal with asymptotic covariance Σ_θ given by (23)-(24).

5 Forecasting

If the return process (R_t) (or the centered process (x_t)) is observed up to time T, we are interested in forecasting the behavior of R at some future time $T+\tau$ or over the time interval $[T+1,\ldots,T+\tau]$. This requires knowledge of the conditional probabilities

$$p_{s|t}(v|\underline{x}_t) = P(\sigma_s = v|\underline{x}_t), \quad s \geq t,$$

with $\underline{x}_t = (x_t, x_{t-1},\ldots)$. Maintaining our assumption of conditional independence of σ_t and $\underline{x}_{t-1}$, given σ_{t-1} and assuming in addition that $\mathcal{F}_t = \mathcal{F}_t^x$, i. e. (R_t) is the only observable process (this is not essential, but makes the formulas look more appealing), it is easy to see that

$$\mathbf{p}_{s|t}(\underline{x}_t) = \mathbf{p}_{t|t}(\underline{x}_t)P^{s-t} \quad \text{for} \quad s > t,$$

where we have put

$$\begin{aligned}
\mathbf{p}_{s|t}(\underline{x}_t) &= (p_{s|t}(v_a|\underline{x}_t), p_{s|t}(v_b|\underline{x}_t)) \\
&= (p_{s|t}(v_a|\underline{x}_t), 1 - p_{s|t}(v_a|\underline{x}_t)).
\end{aligned}$$

(For the sake of lean notation, we shall treat the $\mathbf{p}_{s|t}$ as row vectors in this section.) So everything is reduced to the *filter problem* of calculating $\mathbf{p}_{t|t}(\underline{x}_t)$. As usual, this is solved by deriving a recursive formula for $\mathbf{p}_{t|t}$. To this end, let $f_{t|t-1}(x,v|\underline{x}_{t-1})$ and $f_{t|t-1}(x|\underline{x}_{t-1})$ denote the conditional densities of (x_t,σ_t) and x_t, resp. , given $\underline{x}_{t-1}$ (the first one with respect to $\lambda \otimes \nu$). Then, by Bayes formula,

$$\begin{aligned}
f_{t|t-1}(x,v|\underline{x}_{t-1}) &= \phi(x,v)p_{t|t-1}(v|\underline{x}_{t-1}), \qquad (25)\\
f_{t|t-1}(x|\underline{x}_{t-1}) &= \int_V \phi(x,v)p_{t|t-1}(v|\underline{x}_{t-1})\nu(dv) \\
&= \phi(x,v_a)p_{t|t-1}(v_a|\underline{x}_{t-1}) + \phi(x,v_b)p_{t|t-1}(v_b|\underline{x}_{t-1}) \\
&= \mathbf{p}_{t|t-1}(\underline{x}_{t-1})\Phi(x)\boldsymbol{\iota}
\end{aligned}$$

(with $\Phi(x)$ as in section 3). Since (again by Bayes formula)

$$p_{t|t}(v|\underline{x}_t) = \frac{f_{t|t-1}(x_t,v|\underline{x}_{t-1})}{f_{t|t-1}(x_t|\underline{x}_{t-1})},$$

some elementary calculations lead to the fundamental recursion

$$\mathbf{p}_{t|t}(\underline{x}_t) = \frac{\mathbf{p}_{t-1|t-1}(\underline{x}_{t-1})P\Phi(x_t)}{\mathbf{p}_{t-1|t-1}(\underline{x}_{t-1})P\Phi(x_t)\boldsymbol{\iota}}.$$

Hence, by induction,

$$\mathbf{p}_{T|T}(\underline{x}_T) = \frac{\mathbf{p}_{t|t}(\underline{x}_t)P\Phi(x_{t+1})\ldots P\Phi(x_T)}{\mathbf{p}_{t|t}(\underline{x}_t)P\Phi(x_{t+1})\ldots P\Phi(x_T)\boldsymbol{\iota}}.$$

In particular, for $t = 0$ and with the stationary initial distribution $\mathbf{p}_{0|0} = (p_a, 1 - p_a) = \mathbf{e}_1'\Pi \quad (\mathbf{e}_1' = (1,0))$,

$$\mathbf{p}_{T|T}(\underline{x}_T) = \frac{\mathbf{e}_1'\Pi \prod_{t=1}^{T} P\Phi(x_t)}{\operatorname{tr}\left\{\Pi \prod_{t=1}^{T} P\Phi(x_t)\right\}}.$$

(Note, however, that the latter expression tends to be numerically instable. For actual calculations, it is therefore advisable to use the recursive formula, starting at $t = 0$.) We are now in the position to calculate forecasts of future squared volatilities.

$$\begin{aligned} v_{t|T}^2 = E(x_t^2|\underline{x}_T) = E(\sigma_t^2|\underline{x}_T) &= \mathbf{p}_{t|T}(\underline{x}_T)\begin{pmatrix} v_a^2 \\ v_b^2 \end{pmatrix} \\ &= \mathbf{p}_{T|T}(\underline{x}_T)P^{t-T}\begin{pmatrix} v_a^2 \\ v_b^2 \end{pmatrix}, \quad t > T, \qquad (26) \end{aligned}$$

or forecasts of average future squared volatilities

$$V_{T+\tau|T}^2 = \frac{1}{\tau}\sum_{t=T+1}^{T+\tau} v_{t|T}^2.$$

Of course, in calculating the right hand side, the estimated parameter values $\hat{v}_a$, etc. have to be substituted for v_a, etc., and $R_t - \hat{\mu}$ for x_t. Alternatively, one may be interested in the most probable regime active in some period, i.e. in calculating $\max\{p_{t|T}(v_a|\underline{x}_T), p_{t|T}(v_b|\underline{x}_T)\}$. Also, in the same way as in the proof of Lemma 1, one may obtain the joint conditional density

$$f(x_t, \ldots, x_{T+1}|\underline{x}_T, \underline{\sigma}_T; \boldsymbol{\theta}) = (I_a(\sigma_T) I_b(\sigma_T))\left[\prod_{s=T+1}^{t} P\Phi(x_s)\right]\boldsymbol{\iota}.$$

Integrating out σ_T,

$$f(x_t, \ldots, x_{T+1}|\underline{x}_T; \boldsymbol{\theta}) = \mathbf{p}_{T|T}\left[\prod_{s=T+1}^{t} P\Phi(x_s)\right]\boldsymbol{\iota}$$

is the joint conditional density of $x_t, \ldots, x_{T+1}$, given the observations up to time T. In particular, since

$$\int P\Phi(x_s)dx_s = P,$$

the marginal conditional distribution of x_t is

$$f(x_t|\underline{x}_T; \boldsymbol{\theta}) = \mathbf{p}_{T|T}P^{t-T}\Phi(x_t)\boldsymbol{\iota},$$

a mixture. This may be used to construct prediction intervals for x_t.

6 Finite-sample properties

In this section we demonstrate the finite-sample performance of the moment estimator and the predictor by applying them both to artificial data sets and to some historical exchange rate data.

The artificial data sets of size $T = 1000$ (which corresponds with about the size of daily observations available in financial markets) were generated (for different parameter values $\boldsymbol{\theta}$) by the following recursion.

$$
\begin{aligned}
R_t &= \sigma_t \epsilon_t, \\
\sigma_t &= I_a(\sigma_{t-1}) \left[v_a 1_{[0,p_{aa}]}(u_t) + v_b 1_{[p_{aa},1]}(u_t) \right] \\
&\qquad + I_b(\sigma_{t-1}) \left[v_a 1_{[p_{bb},1]}(u_t) + v_b 1_{[0,p_{bb}]}(u_t) \right], \\
u_t &\overset{i.\ i.\ d.}{\sim} U[0,1], \\
\epsilon_t &\overset{i.\ i.\ d.}{\sim} \mathcal{N}(0,1).
\end{aligned}
\tag{27}
$$

In order to minimize the effect of initial values on the trajectories, we eliminated the first 200 observations. Any solution of (27) evidently satisfies assumptions 1-3 of section 2. The number of Monte-Carlo runs was 500. For each of the 500 data sets, we tried (see below) to (a) estimate the parameters θ_i using (16)-(19) (in combination with (15)), (b) estimate the standard deviations $\hat{\sigma}_{\hat{\theta}_i}$ of the resulting estimates $\hat{\theta}_i$ using the estimated main diagonal elements of (23), and (c) predict the squared volatility one period ahead applying (26) (replacing the true parameter values by its non-updated estimates) to each of the 50 observations following the sample. However, for some artificial data sets the estimation method proposed here failed (either because the nonnegativity constraint on the variances was violated or since the procedure was numerically instable). In order to determine parameter sets where the method is not operational, we report the share of runs where the method failed (denoted S_f). We report the mean of the estimates, their standard deviation, the mean of their estimated standard deviation, the standard deviation of their estimated standard deviation and the mean of the average (over 50 periods) squared prediction error (ASPE_c) — all determined from the succesfully operated runs — for different parameters in Tables 2-11. Additionally, we report the theoretical kurtosis (κ), the mean of the estimated kurtosis'($\hat{\kappa}$), and, for sake of comparison with the ASPE_c, the mean of the average squared prediction errors (ASPE_u) of the unconditional predictions (i. e. the estimated stationary squared volatilities).

A first glance at the time series plots of the real data used here (see below) indicated that regimes of high volatility are characterized by relatively short clusters of deflections of either sign that are *some* times higher than the deflections in "normal times". On the other hand, Duffie and Gray(1995) found high transition probabilities for returns on oil prices. In order to reproduce both findings, we chose the parameters as exhibited in Table 1 in our experiments (ordered according to increasing relative persistence, see below).

Nr. Exp.	v_a	v_b	p_{aa}	p_{bb}
1	2	.5	.5	.5
2	2	.5	.8	.95
3	2	.5	.5	.95
4	2	.5	.2	.9
5	2	.5	.1	.95
6	3	.5	.5	.5
7	3	.5	.8	.95
8	3	.5	.5	.95
9	3	.5	.2	.9
10	3	.5	.1	.95

Table 1: List of Experiments.

The main findings of the Monte-Carlo experiments can be summarized as follows.

- The operationality of the estimation method proposed here seems to (be partially) positively correlated with the ratio p_{bb}/p_{aa} (which may be interpreted as *relative persistence* of the lower volatility).

- The relative biases are reasonably small (clearly below 10% apart from few exceptions), but ambiguous in sign.

- With the exception of the estimates for p_{aa} (and, less definite, the estimates of μ), the means of the estimated (assuming limiting distribution) standard deviations are higher than means of the actual (provided that the Monte-Carlo approximation of the latter is sufficiently accurate) standard deviations. The problem with the estimates for p_{aa} is presumably caused by the small number of observed regime transitions (with $p_{aa} = .5$, the problem did not arise) and by numerical instability (see (19)).

- The relative improvement of using conditional predictions over the use of unconditional ones is (with one exception) is recognizable, but small in either case (the maximal relative improvement amounts roughly 13% in Experiment 9). However, in financial market applications, these seemingly insignificant improvements might mean considerable gains (or losses).

Eventually, in order to examplify the performance of the moment estimator in practice, we applied it to daily compound returns (scaled to variance one by dividing the original time series through their empirical standard deviations σ_D) of four foreign exchange rates (FRF,GBP,ITL,USD). For each currency, we estimated the parameters from samples of size 1200,1250 and 1300 and predicted the future squared volatility from one day to another as before (again without updating the estimates, however) for the 50 days following the respective samples. The samples start on the second of January 1990 in either case and end, depending on their lenght, on 17.10.1994, 28.12.1994 or 8.3.1995. The results are reported in Tables 11-14. For two selected cases, we additionally provide plots of the actual squared returns and their conditional predictions in Figures 1 and 2. Some findings are

- The estimates indicate a high relative persistence (see above) for FRF,GBP and ITL. Apart from the instability of the estimates $\hat{\sigma}_{\hat{p}_a a}$, the estimation procedure performed stable for all currencies.
- Due to the high estimated standard deviations $\hat{\sigma}_{\hat{v}_a}$, the estimates v_a are highly insignificant(ly different form zero) for FRF,GBP and ITL, whereas this insignificance is less pronounced for USD. Though it did not come out in the experiments, we suppose that the high relative persistence (which implies very small stationary probabilities p_a and consequently very few observations of the high volatility regime) of FRF,GBP and ITL is to be made responsible for the insignificance.
- The improvement in forecasting is more ambiguous than in the experiments, but again small.

The economic interpretation of the results is left to the willing reader.

Appendix A

Here we show how to obtain explicit formulas for the solution of (13), (14). Denote $\Gamma_k = M_{2k}/\alpha_{2k}, \Gamma_{2,2} = M_{2,2}, v_a^2 = x, v_b^2 = y, p_a = p, p_{aa} = r$. We then have to solve the four equations

$$\begin{aligned} px + (1-p)y &= \Gamma_1, \quad &\text{(A-1)}\\ px^2 + (1-p)y^2 &= \Gamma_2, \quad &\text{(A-2)}\\ px^3 + (1-p)y^3 &= \Gamma_3, \quad &\text{(A-3)}\\ rp(x-y)^2 + 2pxy + (1-2p)y^2 &= \Gamma_{2,2}. \quad &\text{(A-4)} \end{aligned}$$

The first step consists in solving (A-1),(A-2) for x, y, given p.

$$y = \frac{\Gamma_1 - px}{q} \quad (\text{with} \quad p = q) \tag{A-5}$$

$$px^2 + \frac{1}{q}(\Gamma_1 - px)^2 = \Gamma_2. \tag{A-6}$$

The last equation has the two roots

$$x_{1/2} = \Gamma_1 \pm \sqrt{\frac{q}{p}}\sqrt{\Gamma_2 - \Gamma_1^2},$$

with corresponding y values given by

$$y_{1/2} = \Gamma_1 \mp \sqrt{\frac{p}{q}}\sqrt{\Gamma_2 - \Gamma_1^2}.$$

Note that there is no guarantee that the roots are real, since the discriminant may become negative. If, however, the theoretical moments $m_{2k} = E(R_t - \mu)^{2k} = \alpha_{2k}E(\sigma_t^{2k})$ are substituted for the empirical moments M_{2k}, then, with $\gamma_k = m_{2k}/\alpha_{2k} = E(\sigma_t)^{2k}$, it is trivially true that $\gamma_2 - \gamma_1^2 > 0$ so that two real roots do exist. Since, in view of Proposition 2, the empirical moments approach the theoretical ones and hence $\Gamma_k \to \gamma_k$ for increasing sample size, it can be expected that in most cases $\Gamma_2 - \Gamma_1^2 > 0$ for large sample size. The correct choice (satisfying the restriction $x > y$) is then

$$x = \Gamma_1 + \sqrt{\frac{q}{p}}\sqrt{\Gamma_2 - \Gamma_1^2}, \quad y = \Gamma_1 - \sqrt{\frac{p}{q}}\sqrt{\Gamma_2 - \Gamma_1^2}.$$

These values are now inserted in (A-3).

$$p\left(\Gamma_1 + \sqrt{\frac{q}{p}}d\right)^3 + q\left(\Gamma_1 - \sqrt{\frac{p}{q}}d\right)^3 = \Gamma_3$$

with $d = \sqrt{\Gamma_2 - \Gamma_1^2}$, leading to

$$p\left(\sqrt{\frac{q}{p}}\right)^3 - q\left(\sqrt{\frac{p}{q}}\right)^3 = \frac{\Gamma_3 - \Gamma_1^3 - 3\Gamma_1 d^2}{d^3} = C$$

and, after some straightforward algebraic manipulations, to

$$1 - 2p = \sqrt{p(1-p)}C.$$

The solution is

$$p = \frac{1}{2}\left[1 - \frac{C}{\sqrt{4+C^2}}\right].$$

Since $\frac{q}{p} = \frac{\sqrt{4+C^2}+C}{\sqrt{4+C^2}-C}$, the formulas for x and y take the form

$$\begin{aligned} x &= \Gamma_1 + \sqrt{\frac{\sqrt{4+C^2}+C}{\sqrt{4+C^2}-C}}\sqrt{\Gamma_2 - \Gamma_1^2}, \\ y &= \Gamma_1 - \sqrt{\frac{\sqrt{4+C^2}-C}{\sqrt{4+C^2}+C}}\sqrt{\Gamma_2 - \Gamma_1^2}. \end{aligned}$$

Finally, r is calculated from (A-4) in a straightforward way.

Appendix B

The entries of the asymptotic covariance matrix Σ are as follows.

$$\begin{aligned} \Sigma_{00} &= m_2; \\ \Sigma_{0k} &= \Sigma_{k0} = 0, \quad k = 1,2,3,4; \\ \Sigma_{kl} &= \Sigma_{lk} = \mathbf{v}(k)'\left[\alpha_{2(k+l)}P_0 - \alpha_{2k}\alpha_{2l}\mathbf{pp}'\right]\mathbf{v}(l) \\ &\qquad +2\alpha_{2k}\alpha_{2l}p_a p_b \frac{\pi}{1-\pi}\mathbf{v}(k)'\begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}\mathbf{v}(l), \quad k,l = 1,2,3; \\ \Sigma_{k4} &= \Sigma_{4k} = 2\mathbf{v}(1)'P'V_1[\alpha_{2(k+l)}P_0 - \alpha_{2k}\alpha_{2l}\mathbf{pp}' \\ &\qquad +\alpha_{2k}p_a p_b \frac{\pi}{1-\pi}\begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}]\mathbf{v}(l), \quad k = 1,2,3; \\ \Sigma_{44} &= \mathbf{v}(1)'\left[\alpha_4^2 V_1 P_0 P V_1 - P'V_1\mathbf{pp}'V_1 P\right]\mathbf{v}(1) \\ &\qquad +2\mathbf{v}(1)'P'V_1\left[\alpha_4 P_0 - \mathbf{pp}' + 2p_a p_b \frac{\pi}{1-\pi}\begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}\right]V_1 P\mathbf{v}(1). \end{aligned}$$

Here

$$\mathbf{p} = (p_a, p_b)', \quad P_0 = \mathrm{diag}(p_a, p_b),$$
$$\mathbf{v}(k) = (v_a^{2k}, v_b^{2k})', \quad V_k = \mathrm{diag}(v_a^{2k}, v_b^{2k}).$$

The calculations are somewhat tedious, but straightforward, making repeated use of the following formulas.
(i) If ϵ_t is symmetric,

$$E\{g(\underline{x}_{t-1})x_t^{2k+1}\} = 0.$$

(ii) $E\{x_1 g(x_t)\} = 0 \quad$ (for $\quad t \geq 2$).
(iii) $E\{x_0^{2r} x_1^{2s} x_{t-1}^{2m} x_t^{2n}\}$

$$= \begin{cases} \alpha_{2r}\alpha_{2s}\alpha_{2m}\alpha_{2n}\mathbf{v}(r)' P_0 P V_s P^{t-2} V_m P \mathbf{v}(n) & \text{for} \quad t \geq 3, \\ \alpha_{2r}\alpha_{2(s+m)}\alpha_{2n}\mathbf{v}(r)' P_0 P V_s V_m P \mathbf{v}(n) & \text{for} \quad t = 2. \end{cases}$$

(iv) $m_{2n} = E(x_t^{2n}) = \alpha_{2n}\mathbf{p}'\mathbf{v}(n)$.
(i)-(iv) can be shown by some cumbersome calculations making repeated use of assumptions 1-3 as well as the telescoping property of conditional expectations.
(v) $P_0 P^{t-1} - \mathbf{p}\mathbf{p}' = p_a p_b \pi^{t-1} \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}$.
(vi) $\begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} P = \pi \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}$.

References

[1] Billingsley, P., 1968, Convergence of probability measures (John Wiley & Sons, New York).

[2] Bollerslev, T., R. Y. Chou and K. F. Kroner, 1992, ARCH modeling in finance, Journal of Econometrics 52, 5-59.

[3] Cron, A., 1997, Robust nonparametric estimation and prediction in ARCH-related models, Technical report (Department of Econometrics, University of Bonn).

[4] Duffie, D. and St. Gray, 1995, Volatility in energy prices, in: R. Jameson, ed., Managing Energy Price Risk (Risk Publications, London) 39-55.

[5] Engle, R. F. (ed.), 1995, ARCH selected readings (Oxford University Press, New York).

[6] Hamilton, J. D., 1989, A new approach to the economic analysis of nonstationary time series and the business cycle, Econometrica 57, 357-384.

[7] Hamilton, J. D., 1994, Time Series analysis (Princeton University Press, Princeton).

θ_i		$\hat{\theta}_i$		$\hat{\sigma}_{\hat{\theta}_i}$	
Name	True	Mean	SD	Mean	SD
μ	0.000	0.000	(0.045)	0.046	(0.002)
v_a	2.000	2.043	(0.157)	0.880	(0.367)
v_b	0.500	0.499	(0.229)	0.471	(0.052)
p_{aa}	0.500	0.472	(0.136)	0.147	(0.023)
p_{bb}	0.500	0.523	(0.127)	0.953	(0.205)
p_a	0.500	0.475	(0.094)	0.155	(0.011)
p_b	0.500	0.525	(0.094)	0.155	(0.011)
S_f	38.6%	κ	5.336	$\hat{\kappa}$	5.435
ASPE_c	0.163	ASPE_u	0.163		

Table 2: Experiment 1.

θ_i		$\hat{\theta}_i$		$\hat{\sigma}_{\hat{\theta}_i}$	
Name	True	Mean	SD	Mean	SD
μ	0.000	-0.000	(0.031)	0.032	(0.002)
v_a	2.000	2.048	(0.000)	1.243	(0.726)
v_b	0.500	0.451	(0.103)	0.192	(0.029)
p_{aa}	0.800	0.789	(0.036)	0.052	(0.018)
p_{bb}	0.950	0.972	(0.000)	0.492	(0.286)
p_a	0.200	0.232	(0.058)	0.109	(0.021)
p_b	0.800	0.805	(0.000)	0.109	(0.021)
S_f	20.0%	κ	9.750	$\hat{\kappa}$	11.723
ASPE_c	0.016	ASPE_u	0.017		

Table 3: Experiment 2.

θ_i		$\hat{\theta}_i$		$\hat{\sigma}_{\hat{\theta}_i}$	
Name	True	Mean	SD	Mean	SD
μ	0.000	0.000	(0.024)	0.024	(0.002)
v_a	2.000	1.894	(0.288)	1.591	(1.003)
v_b	0.500	0.449	(0.060)	0.115	(0.018)
p_{aa}	0.500	0.523	(0.199)	0.082	(0.025)
p_{bb}	0.950	0.939	(0.000)	0.160	(0.062)
p_a	0.091	0.123	(0.047)	0.079	(0.017)
p_b	0.909	0.879	(0.026)	0.079	(0.017)
S_f	2.2%	κ	12.985	$\hat{\kappa}$	12.952
ASPE_c	0.009	ASPE_u	0.008		

Table 4: Experiment 3.

θ_i		$\hat{\theta}_i$		$\hat{\sigma}_{\hat{\theta}_i}$	
Name	True	Mean	SD	Mean	SD
μ	0.000	-0.000	(0.025)	0.026	(0.001)
v_a	2.000	1.920	(0.295)	1.522	(0.967)
v_b	0.500	0.449	(0.072)	0.129	(0.017)
p_{aa}	0.200	0.221	(0.129)	0.091	(0.016)
p_{bb}	0.900	0.873	(0.048)	0.145	(0.041)
p_a	0.111	0.140	(0.049)	0.082	(0.015)
p_b	0.889	0.860	(0.049)	0.082	(0.015)
S_f	0.0%	κ	12.375	$\hat{\kappa}$	12.221
ASPE_c	0.005	ASPE_u	0.005		

Table 5: Experiment 4.

θ_i		$\hat{\theta}_i$		$\hat{\sigma}_{\hat{\theta}_i}$	
Name	True	Mean	SD	Mean	SD
μ	0.000	-0.001	(0.021)	0.021	(0.001)
v_a	2.000	1.829	(0.333)	1.866	(1.138)
v_b	0.500	0.462	(0.039)	0.085	(0.015)
p_{aa}	0.100	0.135	(0.151)	0.113	(0.035)
p_{bb}	0.950	0.923	(0.034)	0.085	(0.027)
p_a	0.053	0.082	(0.034)	0.065	(0.016)
p_b	0.947	0.918	(0.034)	0.065	(0.016)
S_f	0.0%	κ	13.520	$\hat{\kappa}$	12.936
ASPE_c	0.000	ASPE_u	0.000		

Table 6: Experiment 5.

θ_i		$\hat{\theta}_i$		$\hat{\sigma}_{\hat{\theta}_i}$	
Name	True	Mean	SD	Mean	SD
μ	0.000	0.005	(0.069)	0.068	(0.003)
v_a	3.000	3.126	(0.208)	2.055	(0.650)
v_b	0.500	0.704	(0.311)	0.988	(0.103)
p_{aa}	0.500	0.438	(0.118)	0.136	(0.019)
p_{bb}	0.500	0.552	(0.102)	3.859	(0.907)
p_a	0.500	0.445	(0.077)	0.147	(0.009)
p_b	0.500	0.555	(0.077)	0.147	(0.009)
S_f	53.0%	κ	5.684	$\hat{\kappa}$	5.863
ASPE_c	0.135	ASPE_u	0.135		

Table 7: Experiment 6.

θ_i		$\hat{\theta}_i$		$\hat{\sigma}_{\hat{\theta}_i}$	
Name	True	Mean	SD	Mean	SD
μ	0.000	0.003	(0.046)	0.044	(0.004)
v_a	3.000	3.083	(0.000)	2.916	(1.323)
v_b	0.500	0.465	(0.176)	0.401	(0.063)
p_{aa}	0.800	0.765	(0.116)	0.033	(0.014)
p_{bb}	0.950	0.956	(0.000)	2.431	(1.886)
p_a	0.200	0.198	(0.045)	0.098	(0.024)
p_b	0.800	0.814	(0.000)	0.098	(0.024)
S_f	35.6%	κ	12.187	$\hat{\kappa}$	14.138
ASPE_c	0.124	ASPE_u	0.130		

Table 8: Experiment 7.

θ_i		$\hat{\theta}_i$		$\hat{\sigma}_{\hat{\theta}_i}$	
Name	True	Mean	SD	Mean	SD
μ	0.000	0.000	(0.031)	0.032	(0.003)
v_a	3.000	2.840	(0.431)	3.490	(2.143)
v_b	0.500	0.394	(0.124)	0.234	(0.044)
p_{aa}	0.500	0.504	(0.179)	0.045	(0.015)
p_{bb}	0.950	0.940	(0.000)	0.678	(0.307)
p_a	0.091	0.115	(0.036)	0.070	(0.012)
p_b	0.909	0.887	(0.000)	0.070	(0.012)
S_f	7.2%	κ	20.368	$\hat{\kappa}$	20.279
ASPE_c	0.081	ASPE_u	0.089		

Table 9: Experiment 8.

θ_i		$\hat{\theta}_i$		$\hat{\sigma}_{\hat{\theta}_i}$	
Name	True	Mean	SD	Mean	SD
μ	0.000	-0.001	(0.035)	0.035	(0.002)
v_a	3.000	2.922	(0.429)	3.445	(2.213)
v_b	0.500	0.405	(0.148)	0.270	(0.041)
p_{aa}	0.200	0.208	(0.121)	0.065	(0.013)
p_{bb}	0.900	0.883	(0.038)	0.506	(0.227)
p_a	0.111	0.130	(0.039)	0.072	(0.010)
p_b	0.889	0.870	(0.039)	0.072	(0.010)
S_f	9.2%	κ	18.186	$\hat{\kappa}$	18.185
ASPE_c	0.006	ASPE_u	0.008		

Table 10: Experiment 9.

θ_i		$\hat{\theta}_i$		$\hat{\sigma}_{\hat{\theta}_i}$	
Name	True	Mean	SD	Mean	SD
μ	0.000	-0.002	(0.028)	0.027	(0.002)
v_a	3.000	2.775	(0.515)	4.236	(3.019)
v_b	0.500	0.417	(0.084)	0.176	(0.038)
p_{aa}	0.100	0.118	(0.118)	0.057	(0.012)
p_{bb}	0.950	0.927	(0.028)	0.220	(0.160)
p_a	0.053	0.077	(0.027)	0.056	(0.011)
p_b	0.947	0.923	(0.027)	0.056	(0.011)
S_f	1.2%	κ	25.685	$\hat{\kappa}$	24.510
ASPE_c	0.006	ASPE_u	0.006		

Table 11: Experiment 10.

A	$T = 1200$		$T = 1250$		$T = 1300$	
θ_i	$\hat{\theta}_i$	$\hat{\sigma}_{\hat{\theta}_i}$	$\hat{\theta}_i$	$\hat{\sigma}_{\hat{\theta}_i}$	$\hat{\theta}_i$	$\hat{\sigma}_{\hat{\theta}_i}$
μ	-0.002	0.029	-0.006	0.027	-0.018	0.024
v_a	5.988	36.708	5.873	35.284	5.229	27.039
v_b	0.728	0.297	0.706	0.275	0.658	0.217
p_{aa}	0.406	NA	0.406	NA	0.405	NA
p_{bb}	0.992	2.915	0.992	2.591	0.992	1.631
p_a	0.013	0.021	0.013	0.020	0.013	0.020
p_b	0.987	0.021	0.987	0.020	0.987	0.020
σ_D	0.000002	-	0.000002	-	0.000003	-
$\hat{\kappa}$	52.149	-	53.151	-	48.522	-
ASPE$_c$	0.431	-	31.846	-	79.466	-
ASPE$_u$	0.645	-	30.031	-	85.019	-

Table 12: Results A.

B	$T = 1200$		$T = 1250$		$T = 1300$	
θ_i	$\hat{\theta}_i$	$\hat{\sigma}_{\hat{\theta}_i}$	$\hat{\theta}_i$	$\hat{\sigma}_{\hat{\theta}_i}$	$\hat{\theta}_i$	$\hat{\sigma}_{\hat{\theta}_i}$
μ	-0.021	0.029	-0.020	0.028	-0.030	0.026
v_a	3.688	10.188	3.670	10.067	3.544	9.216
v_b	0.803	0.173	0.792	0.164	0.779	0.151
p_{aa}	0.589	NA	0.590	NA	0.589	NA
p_{bb}	0.989	0.691	0.989	0.650	0.989	0.557
p_a	0.026	0.033	0.025	0.032	0.025	0.031
p_b	0.974	0.033	0.975	0.032	0.975	0.031
σ_D	0.000023	-	0.000024	-	0.000025	-
$\hat{\kappa}$	16.265	-	16.461	-	15.790	-
ASPE$_c$	0.466	-	8.884	-	17.405	-
ASPE$_u$	0.647	-	6.449	-	13.999	-

Table 13: Results B.

C	$T = 1200$		$T = 1250$		$T = 1300$	
θ_i	$\hat{\theta}_i$	$\hat{\sigma}_{\hat{\theta}_i}$	$\hat{\theta}_i$	$\hat{\sigma}_{\hat{\theta}_i}$	$\hat{\theta}_i$	$\hat{\sigma}_{\hat{\theta}_i}$
μ	-0.047	0.028	-0.047	0.026	-0.056	0.025
v_a	3.210	5.919	3.019	5.215	2.714	3.541
v_b	0.721	0.175	0.688	0.149	0.645	0.142
p_{aa}	0.323	NA	0.327	NA	0.366	NA
p_{bb}	0.968	0.485	0.969	0.375	0.960	0.310
p_a	0.045	0.043	0.044	0.041	0.060	0.048
p_b	0.955	0.043	0.956	0.041	0.940	0.048
σ_D	0.000030	-	0.000033	-	0.000038	-
$\hat{\kappa}$	16.381	-	15.906	-	14.781	-
ASPE$_c$	2.324	-	57.273	-	50.147	-
ASPE$_u$	2.484	-	62.058	-	53.310	-

Table 14: Results C.

D	$T = 1200$		$T = 1250$		$T = 1300$	
θ_i	$\hat{\theta}_i$	$\hat{\sigma}_{\hat{\theta}_i}$	$\hat{\theta}_i$	$\hat{\sigma}_{\hat{\theta}_i}$	$\hat{\theta}_i$	$\hat{\sigma}_{\hat{\theta}_i}$
μ	-0.012	0.029	-0.008	0.028	-0.018	0.027
v_a	1.495	0.858	1.490	0.837	1.521	0.842
v_b	0.811	0.163	0.801	0.154	0.789	0.145
p_{aa}	0.386	0.222	0.395	0.215	0.340	0.196
p_{bb}	0.840	0.140	0.848	0.131	0.842	0.135
p_a	0.206	0.188	0.201	0.178	0.193	0.159
p_b	0.794	0.188	0.799	0.178	0.807	0.159
σ_D	0.000065	-	0.000065	-	0.000066	-
$\hat{\kappa}$	4.265	-	4.305	-	4.487	-
ASPE$_c$	0.594	-	7.700	-	5.146	-
ASPE$_u$	0.618	-	7.688	-	5.160	-

Table 15: Results D.