15th Triennial World Congress, Barcelona, Spain

IDENTIFICATION OF STATE SPACE SYSTEMS WITH CONDITIONALLY HETEROSKEDASTIC INNOVATIONS

Dietmar Bauer *,[1]

* *Institute for Econometrics, Operations Research and System Theory, TU Wien, Argentinierstr. 8, A-1040 Vienna, Austria*

Abstract: In this paper consistency of estimates of linear dynamic systems obtained by using subspace algorithms under quite general assumptions on the innovations are derived. The assumptions include i.a. GARCH type of errors as well as E-GARCH. Also the consistent estimation of the model for the conditional variance is discussed. A small simulation study shows the potential of subspace algorithms in the context of GARCH modelling in comparison with the optimization based method implemented in MATLAB.

Keywords: subspace methods, GARCH models, finance, asymptotic properties

1. INTRODUCTION

The concept of heteroskedastic innovations has been introduced in the analysis of financial time series to explain the phenomenon of volatility clustering: Periods of high fluctuations alternate with periods of low fluctuation, which can be modelled via introducing a dependence of the conditional variances of the innovations. As a second property, GARCH models also helped to explain the 'fat tails' often observed in financial time series. The conditional first two moments build the basis of the most prominent portfolio selection methods, which are based on the assumption, that the investor measures his benefit using expected returns and his risk using the variance. Thus a model for the conditional first two moments is the core of any investment strategy building on these assumptions.

Since the introduction of ARCH models by (Engle, 1982) a number of different algorithms for the estimation have been proposed. Most of these procedures resort to optimization of some criterion function, such as the likelihood or the one step ahead prediction error. It is well known, that the prediction error approach neglecting the ARCH property of the errors leads to preliminary estimates, which are consistent but not efficient in the presence of ARCH effects (cf. e.g. Gourieroux, 1997). Also the asymptotic properties of maximum likelihood estimates in the ARMA case are known (cf. e.g. Gourieroux, 1997, for a discussion). However, in all situations, where the optimization of the criterion function is performed using standard numerical methods, the question of initial estimates is virulent. Especially in a multivariate context a good initial estimate is needed in order to achieve a low probability of being trapped in a local minimum. In the conventional homoskedastic case, where the conditional variance of the innovations is constant, it has been shown in (Bauer, 2000) that a particular subspace algorithm sometimes called CCA, which has been proposed by (Larimore, 1983), asymptotically is equivalent to a generalized pseudo maximum likelihood estimate, i.e. optimizing the Gaussian likelihood. Here equivalent means, that square root sample size times the difference of the two estimates converges to zero almost sure, so that the estimates tend to the same asymptotic distribution. In this paper it is shown, that the subspace

[1] Support by the Austrian FWF under the project number P14438-INF is gratefully acknowledged.

estimates possess some robustness properties with respect to the assumptions on the innovations.

2. MODEL SET AND ASSUMPTIONS

This paper deals with finite dimensional, discrete time, time invariant, linear, dynamical state space systems of the form

$$x_{t+1} = Ax_t + K\varepsilon_t, \quad y_t = Cx_t + \varepsilon_t \qquad (1)$$

where y_t denotes the s-dimensional observed output, x_t the n-dimensional state and ε_t the s-dimensional innovation sequence. $A \in \mathbb{R}^{n\times n}, K \in \mathbb{R}^{n\times s}, C \in \mathbb{R}^{s\times n}$ are real matrices. Note, that it is not assumed, that y_t is univariate. Throughout the paper it is assumed, that the system is stable, i.e. all the eigenvalues of A are assumed to lie within the open unit disc, and strictly minimum-phase, i.e. all the eigenvalues if $A - KC$ are assumed to lie within the unit circle.

It is well known (cf. e.g. Hannan and Deistler, 1988) that state space models and ARMA models are just two representations of the same mathematical object, namely the transfer function: It is easy to verify (using some mild assumptions on the noise sequence ε_t) that one solution to the difference equation given above is of the form

$$y_t = \varepsilon_t + \sum_{j=1}^{\infty} K(j)\varepsilon_{t-j}$$

where $K(j) = CA^{j-1}K, j > 0$ and the infinite sum corresponds to a.s. convergence (or limit in mean square, according to the assumptions imposed upon ε_t). The transfer function $k(z)$, where z denotes the backward shift operator, then is defined as $k(z) = I + zC(I - zA)^{-1}K$. Further let $M(n)$ denote the set of all transfer functions of McMillan degree equal to n fulfilling the stability and the strict minimumphase assumption. $k(z) \in M(n)$ is a rational function in z seen as a complex variable. Therefore the transfer function has a representation as an ARMA system according to $k(z) = a^{-1}(z)b(z)$. A more detailed discussion on the relation between ARMA and state space systems can be found in (Hannan and Deistler, 1988).

The solution y_t as given above is stationary, if the noise ε_t is a stationary sequence. This statement holds both in the weak sense and the strict stationary setting. Throughout this paper it will always be assumed, that ε_t is a martingale difference sequence with respect to the sequence of increasing sigma fields $\mathcal{F}_t$, i.e. $\mathbb{E}\{\varepsilon_t|\mathcal{F}_{t-1}\} = 0$. Furthermore it is assumed, that ε_t is ergodic and of finite fourth moment, i.e. $\mathbb{E}\varepsilon_{t,i}^4 < \infty$, where the notation indicates the i-th component of ε_t. It is also assumed, that $\lim_{k\to\infty} \mathbb{E}\{\varepsilon_t\varepsilon_t'|\mathcal{F}_{t-k}\} = \Sigma = \mathbb{E}\varepsilon_t\varepsilon_t'$ a.s. This property is sometimes referred to as *linear regularity.*

3. SUBSPACE ALGORITHMS

The subspace algorithm investigated in this paper is the CCA method proposed by (Larimore, 1983). Up to now a great number of results exist only for the case, where the innovations also fulfill $\mathbb{E}\{\varepsilon_t\varepsilon_t'|\mathcal{F}_{t-1}\} = \Sigma$, i.e. where no heteroskedasty is present. For the CCA case consistency has been shown in (Deistler *et al.*, 1995), asymptotic normality in (Bauer *et al.*, 1999) and asymptotic equivalence to pseudo maximum likelihood estimation in (Bauer, 2000). Especially the last result seems to be valuable, since it shows, that the computationally advantageous subspace algorithms are a very good substitute for pseudo maximum likelihood estimation. We also note, that in (Bauer and Wagner, 2001) it is shown, that an adaptation of the algorithm is able to produce (weakly) consistent estimates also in the case of cointegrated processes, where also some tests for the number of cointegrating relations are presented. It is the aim of the present paper to show, that the consistency property of the subspace algorithm holds for an extended range of innovation sequences. The asymptotic distribution is a matter of future research.

The CCA algorithm builds on the properties of the state. In the following we will only give a brief outline. For a more detailed description, also of different subspace algorithms cf. e.g. (Bauer, 1998). Fix two integers f and p. Denoting $Y_{t,p}^- = [y_{t-1}', y_{t-2}', \cdots, y_{t-p}']'$ and $Y_{t,f}^+ = [y_t', y_{t+1}', \cdots, y_{t+f-1}']'$ we obtain the following equation:

$$Y_{t,f}^+ = \mathcal{O}_f\mathcal{K}_p Y_{t,p}^- + \mathcal{O}_f(A - KC)^p x_{t-p} + N_{t,f}^+ \qquad (2)$$

where $N_{t,f}^+$ summarizes the effects of the future of the noise, which is orthogonal to the two other terms due to the assumptions on ε_t. Further $\mathcal{O}_f = [C', A'C', \cdots, (A^{f-1})'C']'$, $\mathcal{K}_p = [K, (A-KC)K, \cdots, (A-KC)^{p-1}K]$. Finally let $\langle a_t, b_t\rangle = T^{-1}\sum_{t=p+1}^{T-f} a_t b_t'$. Neglecting the second term in (2), since $(A - KC)^p$ tends to zero for $p \to \infty$, CCA obtains estimates of the system in the following three steps:

- Estimate $\mathcal{O}_f\mathcal{K}_p$ by LS regression in (2) as $\hat{\beta}_{f,p} = \langle Y_{t,f}^+, Y_{t,p}^-\rangle\langle Y_{t,p}^+, Y_{t,p}^-\rangle^{-1}$.
- $\hat{\beta}_{f,p}$ will be of full rank in general, whereas $\mathcal{O}_f\mathcal{K}_p$ is of rank n, where n denotes the system order. Thus approximate

$$\langle Y_{t,f}^+, Y_{t,f}^+\rangle^{-1/2}\hat{\beta}_{f,p}\langle Y_{t,p}^-, Y_{t,p}^-\rangle^{1/2} = \hat{U}\hat{\Sigma}\hat{V}' = \hat{U}_n\hat{\Sigma}_n\hat{V}_n' + \hat{R}_n$$

to obtain estimates $\hat{\mathcal{O}}_f = \langle Y^+_{t,f}, Y^+_{t,f}\rangle^{1/2}\hat{U}_n\hat{\Sigma}_n$ and $\hat{\mathcal{K}}_p = \hat{V}_n'\langle Y^-_{t,p}, Y^-_{t,p}\rangle^{-1/2}$. Here $\hat{U}\hat{\Sigma}\hat{V}'$ denotes the SVD of

$$\langle Y^+_{t,f}, Y^+_{t,f}\rangle^{-1/2}\hat{\beta}_{f,p}\langle Y^-_{t,p}, Y^-_{t,p}\rangle^{1/2}.$$

Thus e.g. $\hat{\Sigma}$ is the diagonal matrix containing the singular values ordered in decreasing size as diagonal entries. $\hat{U}_n \in \mathbb{R}^{fs\times n}, \hat{V}_n \in \mathbb{R}^{ps\times n}$ and $\hat{\Sigma}_n \in \mathbb{R}^{n\times n}$ correspond to the submatrices obtained by neglecting the singular values numbered $n+1$ and higher. Therefore in this step the order is specified.

- Given the estimate $\hat{\mathcal{K}}_p$ from the second step the state is estimated as $\hat{x}_t = \hat{\mathcal{K}}_p Y^-_{t,p}$ and the system matrices are obtained using least squares regressions in the system equations (1), where the estimated state takes the place of the state.

Estimation of the order can be performed using the information contained in the estimated singular values in a number of different ways (for a discussion see Bauer, 1998, Chapter 5). Here we will deal with the criterion SVC. Let

$$SVC(n) = \hat{\sigma}^2_{n+1} + \frac{C_T d(n)}{T}$$

where $d(n) = 2ns$ denotes the number of parameters and $C_T > 0, C_T/T \to 0$ denotes a penalty term. Here $\hat{\sigma}_i$ denotes the estimated singular values ordered decreasing in size. In the homoskedastic case it is known, that a penalty such that $C_T/(fp\log T) \to \infty$ leads to almost sure (a.s.) consistent estimates of the order $\hat{n} = \arg\min SVC(n), 0 \le n \le H_T, H_T = O((\log T)^a), a < \infty$.

4. RESULTS

The key to the results in this section lies in the uniform convergence of the estimated covariance sequence. The conditions in Theorem 5.3.2. of (Hannan and Deistler, 1988) require, that in order for the sequence of covariance estimates to converge uniformly of order $O(Q_T)$ the noise has to be homoskedastic. Here $g_T = O(f_T)$ means that there exists a constant M, such that $g_T/f_T < M$ a.s. and $Q_T = \sqrt{\log\log T/T}$. However, equation (5.3.7.) in the same book provides the result, that if the limiting covariance sequence is replaced with a sequence, where the innovation variance Σ is replaced with $T^{-1}\sum_{t=1}^T \varepsilon_t\varepsilon_t'$ the same results holds under weaker assumptions. This enables the results in the next theorem:

Theorem 1. Let the process $\{y_t\}$ be generated by a stable, strictly minimumphase system $k(z) \in M(n)$, where the innovation process is an ergodic, strictly stationary martingale difference sequence satisfying $\mathbb{E}\{\varepsilon_t|\mathcal{F}_{t-1}\} = 0, \mathbb{E}\varepsilon^4_{t,i} < \infty$ and $\lim_{k\to\infty}\mathbb{E}\{\varepsilon_t\varepsilon_t'|\mathcal{F}_{t-k}\} = \Sigma = \mathbb{E}\varepsilon_t\varepsilon_t'$. a.s. Let $(\hat{A}, \hat{K}, \hat{C})$ denote the estimates obtained via the CCA subspace algorithm using the true order n for the estimation, which have been transformed to the corresponding echelon canonical form. Then the following statements hold:

i) $I + z\hat{C}(I - z\hat{A})^{-1}\hat{K} \to k(z)$ a.s. for each fixed $z = \exp(i\omega)$, if $f \ge n, p = p(T) \to \infty, \max\{f,p\} = O((\log T)^a), a < \infty$. That is, the transfer function is estimated consistently.

ii) Let (A_0, K_0, C_0) denote the representation of the system in the echelon canonical form. Then for $k(z)$ in the generic neighbourhood of the echelon canonical form and if $p \ge -d\log T/(2\log\rho_0), d > 1$

$$\max\{\|\hat{A}-A_0\|, \|\hat{K}-K_0\|, \|\hat{C}-C_0\|\} = O(Q_T)$$

Here $0 < \rho_0 < 1$ denotes the maximal modulus of the eigenvalues of $A_0 - K_0C_0$.

iii) The order estimate $\hat{n}$ obtained by minimizing the SVC criterion is strongly consistent, i.e. $\hat{n} \to n$ a.s., for $C_T/(fp\log T) \to \infty$.

The three parts of the theorem state that with regard to consistency there is no major difference between the homoskedastic and the heteroskedastic case, as long as stationarity is preserved: The subspace estimates still are consistent, the estimation error can be bounded as in the homoskedastic case. Note that the result ii) has the form of a law of the iterated logarithm, except that the constant is not evaluated exactly. This result is only given for the generic neighbourhood of the echelon form, however, using overlapping forms (see e.g. Hannan and Deistler, 1988, Chapter 2) one can show, that an equivalent error bound is indeed valid for all $k \in M(n)$. The last result shows, that also the order estimation can be performed as in the homoskedastic case. This essentially means, that one can use the same code as in the homoskedastic case for the identification irrespective if the system is homo- or heteroskedastic. The derivation of the asymptotic distribution and the investigation of the comparison with prediction error methods is left as a topic of future research.

The theorem imposes an order of convergence for the integer p as a function of the sample size, which is only needed for the derivation of the error bound. This order of convergence includes system dependent quantities and thus might be seen as useless in practice. However, Theorem 6.6.3 in (Hannan and Deistler, 1988) shows, that if p is chosen as $\lfloor d\hat{p}_{AIC}\rfloor$ for $d > 1$, where $\lfloor x\rfloor$ denotes the largest integer smaller than x and where $\hat{p}_{AIC}$ is chosen as the order estimate of a long autoregression for approximating y_t using AIC, then p fulfills the assumption of part ii) a.s.

for large T.[2] Thus an algorithm using this choice of the integer p will lead to consistent estimates, where also the error bound on the estimation error holds.

In comparison to the homoskedastic case the theorem leaves out two important results: The asymptotic distribution of the estimates is not analyzed and secondly the consistency result should also be extended to the unit root case. Both questions are topics of future research.

4.1 *ARCH(p) innovations*

(Engle, 1982) introduced the class of ARCH(p) models, where the conditional variance h_t of the univariate innovations ε_t is modelled as a linear function of the last p squares of the innovations:

$$h_t = c + \sum_{j=1}^{p} a_j \varepsilon_{t-j}^2$$

where ε_t conditional on $\mathcal{F}_{t-1}$, the sigma algebra spanned by $\{\varepsilon_{t-1}, \varepsilon_{t-2}, \cdots\}$, is identically Gaussian distributed with mean zero and variance h_t. Here $0 \leq a_j, c > 0$ is assumed. In order for the process to be strictly stationary with finite variance it is assumed that $\sum_{j=1}^{p} a_j < 1$. It follows from (Bougerol and Picard, 1992) that in this situation the process ε_t is ergodic. Furthermore it is assumed for ψ being equal to the matrix with typical element $\psi_{i,j} = a_{i+j} + a_{i-j}$, where $a_j = 0, j \notin \{1, \cdots, p\}$ that $3(a_1, \cdots, a_p)(I - \psi)^{-1}(a_1, \cdots, a_p)' < 1$. This condition is sufficient for the existence of fourth moments (see e.g. Gourieroux, 1997, Exercise 3.4). Thus the system estimates obtained using subspace methods are consistent. The assumption on Gaussianity of $\varepsilon_t|\mathcal{F}_{t-1}$ is not necessary and can be replaced by other assumptions, which imply the existence of the fourth moment of the process ε_t.

¿From the discussion given above it follows, that a regression of $\hat{\varepsilon}_t^2$ onto $[1, \hat{\varepsilon}_{t-1}^2, \cdots, \hat{\varepsilon}_{t-p}^2]$ results in consistent estimates of the model for the conditional variance. This follows from the finiteness of the fourth moment, the strict stationarity and ergodicity of ε_t and the consistency of $\hat{\varepsilon}_t$ for ε_t.

4.2 *GARCH(p,q) innovations*

(Bollerslev, 1986) extended the ARCH(p) specification to also include MA terms, leading to GARCH(p,q) systems: Let the conditional variance be denoted as $h_t = \mathbb{E}\{\varepsilon_t^2|\mathcal{F}_{t-1}\}$, then the model assumes that

$$h_t = c + \sum_{j=1}^{p} a_j \varepsilon_{t-j}^2 + \sum_{j=1}^{q} b_j h_{t-j}$$

where again $c > 0, a_j \geq 0, b_j \geq 0$. (Bougerol and Picard, 1992) show, that the process ε_t is strictly stationary and ergodic, if $h_t^{-1/2}\varepsilon_t$ is identically standard normally distributed and if $\sum_{j=1}^{p} a_j + \sum_{j=1}^{q} b_j < 1$. In this case also the second moments exist and the process is also wide sense stationary. It remains to find a bound for the fourth moment: Conditions for this to hold are fairly complicated and can be found in (He and Teräsvirta, 1999). Thus in this case the result above shows the consistency of the transfer function estimates. Therefore also the estimated residuals are consistent. The estimation of the model for the innovations leads again to an ARMA model with heteroskedastic innovations. Thus in order to apply the results in this paper, the existence of an eighth moment has to be assumed: Although it follows from (Hannan and Deistler, 1988) that also in this case finite fourth moments are sufficient to achieve a uniform convergence of the sample covariances, no bound on the order of convergence can be given and thus the arguments given above fail for $p \to \infty$. Holding f and p fixed leads to consistent estimates in the sense, that the estimated system matrices converge to some constants a.s., but the estimated system will be asymptotically biased, where the bias depends on the magnitude of ρ_0^p.

4.3 *E-GARCH processes*

As a final example consider the exponential GARCH models considered in (Nelson, 1991): In order to guarantee positivity of the conditional variances the following model has been introduced:

$$\log h_t = \alpha_t + \sum_{j=1}^{\infty} \beta_k g(z_{t-k})$$

Here $\varepsilon_t = z_t h_t^{1/2}$, where z_t is assumed to be i.i.d. with mean zero and variance unity and α_t is a deterministic sequence e.g. constant. The function g is assumed to be of the lin-lin type: $g(z) = \theta z + \gamma(|z| - \mathbb{E}|z|)$ Further the distribution of z_t is assumed to be of the GED type with tail thickness parameter $\nu > 1$. Under these assumptions it follows that $\exp(-\alpha_t)\varepsilon_t$ is strictly stationary and ergodic with finite moments of all orders. Furthermore $\mathbb{E}\{\varepsilon_t^2|\mathcal{F}_{t-k}\} \to \sigma^2$ a.s. for $k \to \infty$. Thus the assumptions of Theorem 1 are fulfilled and the subspace estimates are a.s. consistent.

5. SIMULATIONS

In this section a simple simulation study compares the properties of the subspace estimates to the

[2] This does not hold for AR(p) systems. In this case $\rho_0 = 0$ and $\hat{p}_{AIC}$ stays bounded. However, all results remain true.

estimates obtained by using a likelihood approach. The procedure, which serves as a benchmark, is the one provided in the MATLAB toolbox. The investigated properties are the accuracy of the estimates and the computation times as measured by the MATLAB function `profile`. It should be noted, that both the ML procedure as well as the subspace algorithm have not been trimmed to have minimum computations and there seems to be much potential of improving the subspace algorithms, but on the other hand also the ML approach uses some consistency checks on the data, which increase the computations as well.

The system we will use is an ARMA model with GARCH(1,1) innovations and thus very simple. The specification in full detail is as follows:

$$y_t = 0.8y_{t-1} + \varepsilon_t + 0.3\varepsilon_{t-1}$$
$$h_t = 0.3h_{t-1} + \varepsilon_t^2 + 0.2\varepsilon_{t-1}^2 + 1$$

The conditional distribution of the innovations is Gaussian. The processes are generated using the MATLAB function `garchsim`. For each sample size $T = 200, T = 500, T = 1000$ and $T = 2000$ a total of 1000 time series have been generated and the system estimated using the function garchfit and the correct specification. Also the subspace procedure is used with $f = p = 2\hat{p}_{AIC}$, where $\hat{p}_{AIC}$ denotes the lag length selected by the AIC criterion.

The summary statistics of the estimates can be seen in Table 2 for the ML procedure and in Table 3 for the subspace procedure: The better accuracy of the ML method is clearly visible, however the difference does not seem to be striking for the ARMA model for the output series. Especially for $T = 2000$ the difference in accuracy is minor, except for the occurrence of some outliers in the subspace case. The estimates for the variance model achieved using subspace procedures however, are not very reliable, and this is in particular true for the estimated zeros of the variance model. Even at sample size $T = 2000$ there seems to be a downward bias in the estimates. These facts are also visible in Figure 1: The upper plot here shows a scatter plot of the estimated autoregressive parameters, the lower plot shows the scatter plot for the zero of the estimated variance models, both for sample size $T = 2000$. The upper plot shows a high correlation between the estimates, whereas the lower plot indicates a number of aberrant estimates for the subspace algorithms.

Also in a number of cases some outliers occur, which inflate the estimated variability. This is the reason for using robust estimates of the root mean square and the mean. It should also be mentioned, that in a number of cases the MATLAB routine `garchfit` crashed, giving no resulting system at all. These cases have been taken out of the simulations, leading to some bias in the comparison.

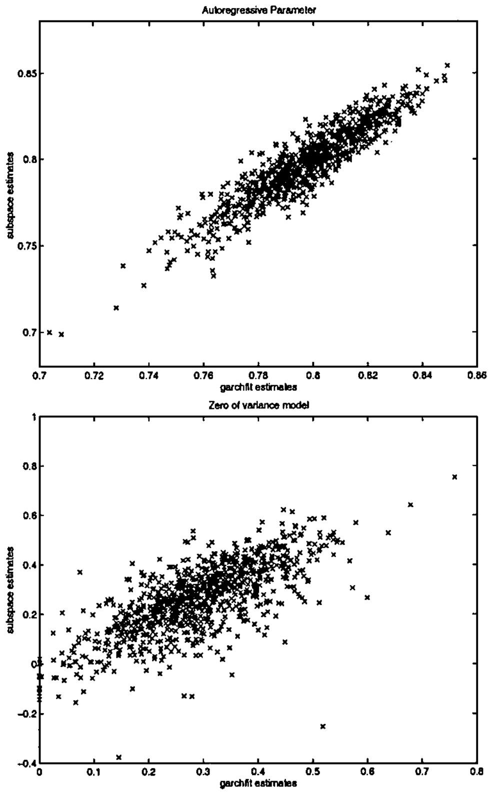

Fig. 1. Upper plot: estimates of the autoregressive parameter of the conditional mean model estimated using `garchfit` (x-axis) versus the estimates obtained using the subspace method (y-axis) for sample size $T = 2000$ in 1000 trials. Lower plot: analogous picture for the estimated zero of the variance model.

Method	$T = 200$	$T = 500$	$T = 1000$	$T = 2000$
garchfit	3.52	3.77	5.17	6.38
subspace	0.61	0.65	0.74	0.89
Quotient	5.77	5.8	7.0	7.2

Table 1. Mean computation time per identification experiment in seconds for the various sample sizes.

Finally the computational time can be analyzed, which clearly shows a huge advantage for the (not even optimized) subspace methods (see Table 1). It is clearly visible, that the subspace method requires only a fraction of computations, while still providing reasonable estimates. The main conclusion of the small simulation study is that the subspace algorithms provide relatively good initial estimates for a subsequent pseudo ML approach in terms of the asymptotic statistical properties, while still keeping the amount of computations required at a low level.

6. CONCLUSIONS

In this paper the asymptotic properties of estimates of state space models using subspace methods with heteroskedastic innovations are investigated. Consistency is shown and a bound on the obtainable order of consistency is provided. The result is stated in a general fashion such that it applies for a wide range of models for the heteroskedasticity, including ARCH(p), GARCH(p,q) and E-GARCH(p,q) models. This shows, that the standard subspace algorithms provide consistent estimates of the system also in situations, where the model for the conditional variance might be doubted. This of course is due to the fact, that the subspace algorithms are based mainly on regression techniques, which are robust with respect to the variance structure of the innovations. With respect to the estimation of the model for the conditional variances consistency can be achieved in the ARCH(p) case, whereas no comparable results are given for the general case. A simulation study compares the estimates with the estimates obtained using the GARCH toolbox implemented in `MATLAB` both with respect to accuracy and computation time. The loss of efficiency in the estimation of the model for the heteroskedasticity is clearly visible, however, the accuracy of the model for the conditional mean seems to be acceptable. Finally the main power of subspace algorithms, namely their low computational load is demonstrated in comparison with a GARCH routine implemented in the `MATLAB` GARCH toolbox.

T	Meas.	a	b	a_v	b_v	σ^2
	True	0.8	-0.3	0.2	0.3	2.0
200	Mean	0.776	-0.283	0.188	0.293	1.97
	RMSE	0.070	0.111	0.100	0.261	0.28
500	Mean	0.788	-0.288	0.194	0.290	1.99
	RMSE	0.041	0.068	0.064	0.201	0.18
1000	Mean	0.796	-0.297	0.198	0.277	1.99
	RMSE	0.026	0.045	0.045	0.147	0.12
2000	Mean	0.798	-0.298	0.199	0.291	1.99
	RMSE	0.019	0.032	0.032	0.104	0.09

Table 2. Summary of estimation results for the ARMA model for the conditional mean (parameters a and b) and the ARMA model for the conditional variance (parameters a_v and b_v) and implied stationary variance σ^2 for various sample sizes and for `garchfit`. For each sample size the trimmed mean and the trimmed root mean square error (RMSE) neglecting the extreme 5%, are calculated.

T	Meas.	a	b	a_v	b_v	σ^2
	True	0.8	-0.3	0.2	0.3	2.0
200	Mean	0.778	-0.286	0.155	0.095	1.97
	RMSE	0.074	0.121	0.111	0.391	0.27
500	Mean	0.787	-0.286	0.177	0.196	1.99
	RMSE	0.044	0.075	0.078	0.264	0.18
1000	Mean	0.795	-0.296	0.187	0.233	1.99
	RMSE	0.031	0.054	0.065	0.219	0.13
2000	Mean	0.798	-0.298	0.192	0.265	1.99
	RMSE	0.020	0.036	0.042	0.132	0.09

Table 3. Summary of estimation results for the subspace procedure.

7. REFERENCES

Bauer, D. (1998). Some Asymptotic Theory for the Estimation of Linear Systems Using Maximum Likelihood Methods or Subspace Algorithms. PhD thesis. TU Wien.

Bauer, D. (2000). Asymptotic efficiency of the CCA subspace method in the case of no exogenous inputs. *Submitted to Journal of Time Series Analysis.*

Bauer, D. and M. Wagner (2001). Estimating cointegrated systems using subspace algorithms. *to appear in Journal of Econometrics.*

Bauer, D., M. Deistler and W. Scherrer (1999). Consistency and asymptotic normality of some subspace algorithms for systems without observed inputs. *Automatica* **35**, 1243–1254.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* **31**, 307–327.

Bougerol, P. and N. Picard (1992). Stationarity of GARCH processes and of some nonnegative time series. *Journal of Econometrics* **52**, 115–127.

Deistler, M., K. Peternell and W. Scherrer (1995). Consistency and relative efficiency of subspace methods. *Automatica* **31**, 1865-1875.

Engle, R. (1982). Autoregressive conditional heteroskedasticity with estimates of the variance of u.k. inflation. *Econometrica* **50**, 987–1008.

Gourieroux, Ch. (1997). *ARCH Models and Financial Applications.* Springer Series in Statistics.

Hannan, E. J. and M. Deistler (1988). *The Statistical Theory of Linear Systems.* John Wiley. New York.

He, C. and T. Teräsvirta (1999). Fourth moment structure of the GARCH(p,q) process. *Econometric Theory* **15**(6), 824-846.

Larimore, W. E. (1983). System identification, reduced order filters and modelling via canonical variate analysis. In: *Proc. 1983 Amer. Control Conference 2.*, Piscataway, NJ. pp. 445–451.

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

Peternell, K., W. Scherrer and M. Deistler (1996). Statistical analysis of novel subspace identification methods. *Signal Processing* **52**, 161–177.