Modelling economic high-frequency time series with STAR-STGARCH models

Stefan Lundbergh and Timo Teräsvirta

Department of Economic Statistics,

Stockholm School of Economics,

P.O.Box 6501, SE-113 83 Stockholm, Sweden

December 22, 1998

Abstract

In this paper we introduce the STAR-STGARCH model that can characterize nonlinear behaviour both in the conditional mean and the conditional variance. A modelling cycle for this family of models, consisting of specification, estimation, and evaluation stages is constructed. Misspecification tests for the estimated model are obtained using standard asymptotic distribution theory. We illustrate the actual modelling by applying the STAR-STGARCH model family to two series of daily observations, the Swedish OMX index and the exchange rate JPY-USD.

Keywords: Financial time series, model misspecification test, nonlinear time series, smooth transition autoregressive model, smooth transition GARCH, time series model specification.

JEL Classification Code: C51, C52, F31.

Acknowledgements: The first author acknowledges financial support from the Tore Browaldh's Foundation. The research of the second author has been supported by the Swedish Council for Research in Humanities and Social Sciences. Material from the paper has been presented at the Fifth Workshop on Finance and Econometrics, Brussels, December 1997, Conference on Financial Markets and the Macroeconomy, Bielefeld, December 1997, Sixth Annual Symposium of the Society for Nonlinear Dynamics and Econometrics, New York, NY, March 1998, and Stochastics '98, Prague, August 1998. Comments from participants, Marno Verbeek and Bruce Mizrach in particular, are gratefully acknowledged. We also wish to thank Changli He and Clive Granger for their comments. Any errors and shortcomings in the paper remain our own responsibility.

1. Introduction

Modelling financial time series has recently received considerable attention. Many of these series are "high frequency", that is, they consist of weekly, daily or even intradaily observations. In this paper, "high frequency" simply means that the conditional variance of the process is not constant over time. As many high frequency series show little or no linear dependence, the focus has been on modelling the conditional variance. The idea of conditional autoregressive heteroskedasticity (ARCH; Engle, 1982) has led to a large number of extensions of the original model and applications. Bollerslev, Engle and Nelson (1994) and Palm (1996) are examples of recent surveys of the area. Another variant of ARCH, the so-called Stochastic Volatility model (Taylor, 1986, p. 73ff.) has gained popularity recently, and the latest developments were surveyed in Ghysels, Harvey and Renault (1996). It has been argued, however, that despite the absence of linear dependence there may be nonlinear dependence in the conditional mean. This should then be appropriately modelled in order to avoid misspecification of the conditional variance. Tong (1990, p. 116) suggested combining the Self Exciting Threshold Autoregressive (SETAR) model for the conditional mean with an ARCH model for the conditional variance. Li and Lam (1995) followed this suggestion and also devised a specification strategy for building SETAR-ARCH models. They applied their model to the daily Hong Kong Hang Seng stock index and reported nonlinearities in the conditional mean.

The original ARCH model and its most important extension, the generalized ARCH (GARCH), are symmetric: while the size of the shock matters, the sign does not. Many authors have argued that shocks may have asymmetric effects to volatility: the dynamic response to a positive shock is not necessarily the mirror image of the response to a negative shock of the same size. Pagan (1996), in his survey of developments in financial econometrics, provided a useful review of models that can handle this type of asymmetry. A natural idea would be to combine such a parameterization of the conditional variance with a nonlinear model for the conditional mean. Li and Li (1996) did exactly that by defining a double threshold autoregressive heteroskedastic (DTARCH) time series model. A DTARCH model has a SETAR-type conditional mean. The conditional variance is parameterized similarly, and the authors called their specification the threshold ARCH (TARCH) model. Note, however, that it differs from the TARCH model of Zakoïan (1994) in that the latter is a parameterization of the conditional standard deviation. Li and Li (1996) also provided a comprehensive modelling strategy for DTARCH models.

It was based on the idea of ordered autoregressions which Tsay (1989) successfully applied to the specification of SETAR models. The authors fitted their DTARCH specification to the daily Hong Kong Hang Seng stock index.

Recently, Lee and Li (1998) generalized the DTARCH model by allowing the transition of the first and second regime to be smooth. They called this model the Smooth Transition Double Threshold model. In this paper we follow Lee and Li (1998) by adopting the idea of smooth transition in the conditional mean which first appeared in Bacon and Watts (1971). Our paper may be seen as an extension of Lee and Li's work in the sense that we simultaneously allow a rather flexible specification for the conditional variance as well. Besides, misspecification testing will receive plenty of attention in this paper. The conditional mean is thus specified as a Smooth Transition Autoregressive (STAR) model; see, for example, Chan and Tong (1986), Granger and Teräsvirta (1993) and Teräsvirta (1994). The conditional variance is specified as a Smooth Transition GARCH (STGARCH) model; see Hagerud (1997) and González-Rivera (1998). Our STGARCH model is a generalization of the GJR-GARCH model (Glosten, Jagannathan and Runkle, 1993) and the Generalized Quadratic ARCH model of Sentana (1995). It allows plenty of scope for explaining asymmetries in volatility. Our aim is to construct a complete modelling cycle for our STAR-STGARCH family of models, consisting of three stages: specification, estimation and evaluation.

The plan of the paper is as follows. We define the model in Section 2 and discuss its specification, estimation and evaluation in Sections 3 and 4. In Section 5 we apply the STAR-STGARCH model to daily returns of the Swedish OMX index and the daily JPY/USD exchange rate and study properties of one-step-ahead out-of-sample forecasts. Section 6 concludes.

2. The model

The logistic smooth transition version of the AR(m)-GARCH(p,q) parameterization is a special case of the following additive nonlinear model in which the conditional mean has the following structure:

$$y_t = \boldsymbol{\varphi}'\mathbf{w}_t + f\left(\mathbf{w}_t; \boldsymbol{\theta}\right) + u_t \tag{2.1}$$

where $\boldsymbol{\varphi} = (\varphi_0, \varphi_1, ..., \varphi_m)'$ is the parameter vector for the autoregressive part of the model and $\mathbf{w}_t = (1, y_{t-1}, ..., y_{t-m})'$ is the corresponding lag vector. Function $f\left(\mathbf{w}_t; \boldsymbol{\theta}\right)$ is nonlinear and assumed to be at least twice continuously differentiable for $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ everywhere in $\mathbf{w}_t \in \Re^m$. The error process of this model is parametrized as

$$u_t = \varepsilon_t \sqrt{h(\mathbf{w}_t, \boldsymbol{\varphi}, \boldsymbol{\theta}, \boldsymbol{\eta}, \boldsymbol{\zeta})} \tag{2.2}$$

where $\{\varepsilon_t\} \sim \text{nid}(0,1)$ and $h_t = h(\mathbf{w}_t, \boldsymbol{\varphi}, \boldsymbol{\theta}, \boldsymbol{\eta}, \boldsymbol{\zeta}) = \boldsymbol{\eta}'\mathbf{z}_t + g(\mathbf{z}_t; \boldsymbol{\zeta})$ is the conditional variance not dependent on ε_t and positive for every t with probability one. Definition (2.2) precludes linear dependence in $\{u_t\}$. Setting $\boldsymbol{\eta} = (\alpha_0, \alpha_1, ..., \alpha_q, \beta_1, ..., \beta_p)'$ and $\mathbf{z}_t = (1, u_{t-1}^2, ..., u_{t-q}^2, h_{t-1}, ..., h_{t-p})'$ where $h_t > 0$ for all t with probability one, makes the linear part of (2.2) a GARCH(p,q) model. Furthermore, $u_t = y_t - \boldsymbol{\varphi}'\mathbf{w}_t - f(\mathbf{w}_t; \boldsymbol{\theta})$ such that neither $\boldsymbol{\varphi}$ nor $\boldsymbol{\theta}$ is assumed to depend on either $\boldsymbol{\eta}$ or $\boldsymbol{\zeta}$. Function $g(\mathbf{z}_t; \boldsymbol{\zeta})$ is nonlinear and at least twice continuously differentiable for $\boldsymbol{\zeta} \in \boldsymbol{\Xi}$ everywhere in $\mathbf{z}_t \in \Re^{p+q+1}$. The normality assumption of errors $\{\varepsilon_t\}$ is not necessary but is retained for inference. It is also assumed that the moments necessary for the inference exist and that the parameters are subject to restrictions such that the model is stationary and ergodic. The usual restrictions imposed on $\boldsymbol{\eta}$ to ensure nonnegative conditional variance have been $\alpha_0 > 0, \alpha_j \geq 0, j = 1, ..., q-1, \alpha_q > 0, \beta_j \geq 0, j, ..., p$. They can be relaxed as in Nelson and Cao (1992); see also He and Teräsvirta (1999b).

In order to define $f(\mathbf{w}_t; \boldsymbol{\theta})$ and $g(\mathbf{z}_t; \boldsymbol{\zeta})$, let

$$H_n(s_t; \gamma, \mathbf{c}) = \left(1 + \exp(-\gamma \prod_{l=1}^{n} (s_t - c_l))\right)^{-1}, \gamma > 0, c_1 \leq ... \leq c_n \tag{2.3}$$

where s_t is the transition variable, γ is a slope parameter and $\mathbf{c} = (c_1, ..., c_n)'$ a location vector. The parameter restrictions $\gamma > 0$ and $c_1 \leq ... \leq c_n$ are identifying restrictions. The value of the logistic function (2.3), which is bounded between a and 1, $0 \leq a \leq 1/2$ depends on the transition variable s_t. Note that for $\gamma = 0$, $H_n(s_t; \gamma, \mathbf{c}) \equiv 1/2$ and when $\gamma \to \infty$, and $n = 1$, $H_n(s_t; \gamma, \mathbf{c})$ becomes a step-function. It becomes a "multistep" function as $\gamma \to \infty$, if $n > 1$.

In this paper, $f(\mathbf{w}_t; \boldsymbol{\theta})$ is defined as a product of the logistic function (2.3) of order n and another linear combination including lags of y_t. Setting $\boldsymbol{\theta} = (\boldsymbol{\phi}', \gamma, \mathbf{c}')'$, the function can be written as

$$f(\mathbf{w}_t; \boldsymbol{\theta}) = \boldsymbol{\phi}'\mathbf{w}_t H_n(s_t; \gamma, \mathbf{c}) \tag{2.4}$$

Function (2.4) is bounded only in probability. In this paper the transition variable is $s_t = y_{t-d}$ most of the time but $s_t = t$ is another important case. Other definitions such as s_t being a linear combination of variables are possible as well. By inserting

(2.4) into (2.1) we obtain the LSTAR(n) model which by definition becomes linear if $\gamma = 0$. Setting $n = 2$ and $c_1 = c_2$ yields a model that closely approximates the exponential STAR or ESTAR model; see Teräsvirta (1994). STAR models are capable of characterizing asymmetric series and series with sudden upswings and downturns. Chappell and Peel (1998) recently showed that they can also generate realizations that appear chaotic. Bacon and Watts (1971) were the first ones to apply the idea of smooth transition to statistical modelling.

Our conditional variance specification is a generalization of the GJR-GARCH model of Glosten, Jagannathan and Runkle (1993). We make the transition between the extreme regimes smooth by assuming that $g(\mathbf{z}_t;\boldsymbol{\zeta})$ has the same logistic structure as $f(\mathbf{w}_t;\boldsymbol{\theta})$. This is a natural extension of the idea of smooth transition to modelling conditional variance. Thus by setting $\boldsymbol{\zeta} = (\boldsymbol{\alpha}^*, \delta, \mathbf{k}')'$ where $\boldsymbol{\alpha}^* = (\alpha_{01}, ..., \alpha_{0q}, \alpha_{21}, ..., \alpha_{2q})'$, the nonlinear function $g(\mathbf{z}_t;\boldsymbol{\zeta})$ may be written as

$$g(\mathbf{z}_t;\boldsymbol{\zeta}) = \sum_{j=1}^{q} \alpha_{0j} H_{n^*}(u_{t-j};\delta,\mathbf{k}) + \sum_{j=1}^{q} \alpha_{2j} H_{n^*}(u_{t-j};\delta,\mathbf{k}) u_{t-j}^2 \tag{2.5}$$

In practice we restrict ourselves to cases $n^* = 1, 2$. No nonlinear structure is imposed on h_{t-j}, $j = 1, ..., p$, since the model is very flexible even without such an extension. The model can characterize processes with an asymmetric response to shocks with the same magnitude but opposite signs. With $n = 2$ and $k_1 = k_2$, a symmetric but nonlinear response may be characterized as well. The first sum on the right-hand side is a nonlinear variant of the corresponding structure in the GQARCH model of Sentana (1995). Conditions for positivity of the conditional variance are simpler than in the GQARCH model: $\alpha_0 > 0$, $\alpha_0 + \sum_{j=1}^{q} \alpha_{0j} > 0$, $\alpha_j \geq 0$, $\alpha_j + \alpha_{2j} \geq 0$, $j = 1, ..., q$; $\beta_j \geq 0$, $j = 1, ..., p$, form a set of sufficient conditions. We denote (2.5) inserted into (2.2) as the Logistic Smooth Transition GARCH (LSTGARCH(n^*)) model which by definition collapses into the standard GARCH model if $\delta = 0$. The nonlinear GARCH model in Hagerud (1997) may be viewed as a special case of this parametrization with $\alpha_{01} = ... = \alpha_{0q} = 0$.

Assuming normally distributed errors, Engle (1982) showed that the information matrix of the conditional mean-ARCH model is block-diagonal if some regularity conditions hold and if the parameterization of the conditional variance is symmetric in the sense that the model responds similarly to positive and negative inputs of the same size. This in turn implies that if the conditional mean is estimated with a consistent estimator, the conditional variance can be estimated from the residuals of the conditional mean model without loss of asymptotic efficiency. The classical

GARCH parameterization is symmetric and satisfies the regularity conditions, see Bollerslev (1986), so that the LSTAR-linear GARCH model has this property. But then, the general smooth transition GARCH model may not be symmetric, in which case the usual two-stage estimation strategy leads to consistent but not asymptotically efficient estimates. However, if $n^* = 2$ and $c_1 = -c_2$ in (2.3), the STGARCH model is symmetric.

3. Specification and estimation of a STAR-STGARCH model

The nonlinear STAR-STGARCH model defined in (2.1-2.5) is the most general parameterization considered in this paper. It is nevertheless possible that the time series under consideration may be adequately characterized by a submodel of the general STAR-STGARCH one. For instance, the conditional mean may be linear or the conditional variance constant. Furthermore, even if we eventually select a general model there are still choices to be made that have to be based on the data. The delay d in the conditional mean usually has to specified from the data as well as the maximum lag length and the type of the transition function ($n = 1$ or 2). We also have to select the lag length and the type of transition function ($n^* = 1$ or 2) in the STGARCH specification of the conditional variance. All this requires a coherent specification strategy such as, for example, in Box and Jenkins (1970), Li and Li (1996), Tsay (1989, 1998) and Teräsvirta (1994).

Our general rule is to specify the conditional mean first, followed by the conditional variance. The reason is that we may estimate the parameters of the conditional mean consistently even if the conditional variance is not specified, that is, even if it is assumed constant. On the other hand, it is not possible to estimate the parameters of the conditional variance consistently if the conditional mean is misspecified. The specification of the STAR-STGARCH model consists of the following stages:

1. Test linearity of conditional mean and, if rejected, choose d and n.

2. Estimate the parameters of the conditional mean assuming that the conditional variance remains constant and test the null hypothesis of no linear ARCH against ARCH of a given order. If the hypothesis of no ARCH is rejected, tentatively assume that the conditional variance follows a low-order standard GARCH process.

3. Estimate the parameters of the STAR-GARCH model and test the adequacy of

the STAR (conditional mean) and the GARCH (conditional variance) specifications by various misspecification tests. If rejected, specify a STAR-STGARCH model.

4. Estimate the parameters of the STAR-STGARCH model and test the adequacy of both the conditional mean and the conditional variance of that specification by appropriate misspecification tests.

5. If the model passes the tests tentatively accept it. In the opposite case try another specification search or choose another family of models.

It should be noted that by following the above modelling scheme we proceed from restricted models to more general ones. This may be simpler than to start from the most general model and gradually reduce its size, but there is also a statistical rationale behind this choice of direction. If the conditional mean is linear then no STAR specification is identified. As for the conditional variance, the same is true for any STGARCH specification if the linear GARCH already is a valid parameterization. The lack of identification leads to lack of consistency in the parameter estimation, which, in turn, is likely to create numerical difficulties in estimation. See Hansen (1996) for a recent discussion of this problem. To avoid estimating unidentified models we have to proceed from specific to general. In the following we consider the specification stages in detail.

3.1. Testing linearity of the conditional mean

We begin the modelling cycle with the specification of the conditional mean. In order to carry out the linearity tests we have to determine the maximum lag, m, of the linear AR model.

Following Teräsvirta (1994), linearity against a logistic STAR model of order n is tested with an LM-type test, where the null hypothesis is a linear AR model and the alternative an LSTAR(n) model. The null hypothesis is $\gamma = 0$. As mentioned above, function $f(\mathbf{w_t}; \boldsymbol{\theta})$ is not identified under the null hypothesis. To circumvent this problem, (2.4) is Taylor-expanded around $\gamma = 0$. Setting $s_t = y_{t-d}$ in (2.4), assuming $d \leq m$ without loss of generality, leads to

$$y_t = \boldsymbol{\pi}_0'\mathbf{w}_t + \boldsymbol{\pi}_1'\widetilde{\mathbf{w}}_t y_{t-d} + \boldsymbol{\pi}_2'\widetilde{\mathbf{w}}_t y_{t-d}^2 +, ..., \boldsymbol{\pi}_n'\widetilde{\mathbf{w}}_t y_{t-d}^n + R_1(\mathbf{w}_t; \boldsymbol{\theta}) + u_t \qquad (3.1)$$

where $\widetilde{\mathbf{w}}_t = (y_{t-1}, ..., y_{t-m})'$, $\boldsymbol{\pi}_i$ is a function of γ such that $\boldsymbol{\pi}_i = \mathbf{0}$, $i = 1, ..., n$, when $\gamma = 0$ and $R_1(\mathbf{w_t}; \boldsymbol{\theta})$ is the remainder. The new null hypothesis thus is

$H_0 : \boldsymbol{\pi}_i = \mathbf{0}$, $i = 1, ..., n$. Note that under H_0, $R_1(\mathbf{w}_t; \boldsymbol{\theta}) \equiv 0$ so the remainder does not affect the distribution theory when the test is based on the LM-principle. When the conditional variance is constant the LM-type statistic with an asymptotic χ^2 distribution (we assume that the necessary moments exist) under the null hypothesis, can be computed by two auxiliary regressions. The F-version of the test is often recommended as it has better small sample properties than the χ^2-version; see, for example, Granger and Teräsvirta (1993, p. 66). The sample sizes in the analysis of high-frequency series are usually so large, however, that in the present context this recommendation has no practical value. Both tests may be carried out in stages as follows.

1. Regress y_t on $\mathbf{w}_t$ and compute the sum of squared residuals, SSR_0.
2. Regress y_t on $\mathbf{w}_t, \widetilde{\mathbf{w}}_t y_{t-d}, \widetilde{\mathbf{w}}_t y_{t-d}^2, ..., \widetilde{\mathbf{w}}_t y_{t-d}^n$ and compute the sum of squared residuals, SSR_1.
3. Compute the F-version of the test statistic $F = \frac{(SSR_0 - SSR_1)/mn}{SSR_1/(T-m(n+1)-1)}$, or the χ^2-version, $\chi^2 = T\frac{(SSR_0 - SSR_1)}{SSR_0}$

This linearity test assumes constant conditional variance, and is therefore not robust against conditional heteroskedasticity. If H_0 cannot be rejected then the conclusion is that the conditional mean is linear. The problem arises when H_0 is rejected because then we do not in principle know if that is because of nonlinearity in the conditional mean or because of conditional heteroskedasticity. However, when the heteroskedasticity is of GARCH type, the size of the test may in some cases be affected. This would suggest using a robust version of the linearity test such as the one in Granger and Teräsvirta (1993, p. 69). We consider this possibility by a small simulation study whose results can be found in Appendix A. The results do show some size distortion in situations where the GARCH-type error process has fat tails, i.e., the kurtosis is high. On the other hand, the power simulations indicate that in those cases the robustification may remove most of the power, so that existing nonlinearity may remain undetected by a robustified linearity test. As our objective is to find and model any existing nonlinearity also in the conditional mean, robustification therefore cannot be recommended. We expect to discover false rejections of the null hypothesis of linearity, due to heteroskedasticity of GARCH type, at the evaluation stage of model building.

In order to carry out the linearity test(s) we have to determine the order of the linear stationary AR model representing the conditional mean under the null hypothesis. Teräsvirta (1994) suggested that the order could be determined by an order selection criterion such as the AIC, see Akaike (1974). The problem is that for high-frequency economic series the usual order selection criterion would typically select a model with no lags because there is normally little or no linear dependence in the series. To avoid the problem the maximum lag ($m \geq 6$ is used for daily observations) is fixed in advance. If the null hypothesis is not rejected we assume that the conditional mean is linear.

3.2. Specification and estimation of the conditional mean model and testing for ARCH

The specification of the STAR model for the conditional mean is carried out as follows. First, the linearity test above is used to select the delay parameter, d. This is followed by choosing n, that is, selecting the type of the STAR(n) model. After that, the estimated model is tested for ARCH in the error process.

The linearity test (3.1) is computed for different values of d, and the one for which the null hypothesis, $\gamma = 0$, has the smallest p-value is selected, see Teräsvirta (1994). This requires the smallest p-value to be lower than a pre-determined value chosen by the researcher. After fixing the delay parameter the order, $n \leqslant 2$, of the LSTAR model is selected. We can choose between the LSTAR(1) and the LSTAR(2) model by testing a sequence of nested hypotheses. The sequence is defined within (3.1) assuming $n = 3$:

$$
\begin{aligned}
&\mathrm{H}_{04} : \pi_3' = 0 \\
&\mathrm{H}_{03} : \pi_2' = 0 | \pi_3' = 0 \\
&\mathrm{H}_{02} : \pi_1' = 0 | \pi_2' = \pi_3' = 0.
\end{aligned}
$$

If H_{03} is most strongly rejected the LSTAR(2) model is selected. In the other two cases the choice is the LSTAR(1) model; for the rationale of this rule, see Teräsvirta (1994). This selection rule is not balanced as it sometimes has a tendency to favour the LSTAR(1) model. In practice this happens when the true model is LSTAR(2) and there are no (or very few) observations in one of the tails of the transition function. In such cases, the LSTAR(1) model is a good approximation to the LSTAR(2) model in the relevant subset of the sample space so that an erroneous choice has little practical significance. Escribano and Jordá (1996) recently presented a rule that purports to remedy this problem, and it could be applied here.

The idea of these selection rules has been to avoid estimating a possibly large number of nonlinear models, but with the steady increase in computational power an average modeller has at his/her disposal this step is no longer crucial. If nonlinear estimation is not considered a time-consuming task one may simply estimate an LSTAR(n) model for $n = 1, 2$, and make the choice between the LSTAR(1) and LSTAR(2) families at the evaluation stage. This can be done by considering the estimation results themselves and the results of the misspecification tests to be discussed in Section 4.1. The logistic part (2.3) of $f(w_t; \theta)$ should also be examined when determining n. If an element in the estimated location vector $c = (c_1, c_2)'$ of the LSTAR(2) model in practice does not affect the values of $f(w_t; \theta)$ in the sample then n may be reduced from two to one.

The estimation of the parameters of the STAR model is carried out by maximum likelihood. Our algorithm also checks if the size of the model can be reduced. The autoregressive parameters whose estimates are insignificant according to a predetermined level are removed using a backward elimination algorithm. In practice this is done by repetitively removing the parameter corresponding to the least significant (if nonsignificant) parameter estimate and reestimating the reduced model. This algorithm also considers restrictions of the form $\varphi_i = -\phi_j$ which are exclusion restrictions comparable to $\varphi_i = 0$. They allow an autoregressive parameter to be cancelled out smoothly in the transition between the two regimes. The backward elimination terminates when all the remaining autoregressive parameter estimates are significant.

The assumption that the error sequence $\{u_t\}$ in (2.1) has a constant conditional variance is normally not realistic when modelling high-frequency financial series and has to be tested. We first test it against the alternative that $\{u_t\}$ follows an ARCH(s) process. As a GARCH(p, q) model can be adequately approximated by a long ARCH specification we choose a large s (≥ 8) for the alternative. Engle (1982) suggested an LM-test which is used here. The asymptotically equivalent test of McLeod and Li (1983) is another possibility. If the null hypothesis is not rejected and the estimated conditional mean model passes the appropriate evaluation tests the modelling sequence terminates. On the other hand, if the null is rejected, as we generally expect when dealing with high-frequency economic series, we continue by assuming that the conditional variance follows a low-order GARCH process.

3.3. Estimation of the STAR-GARCH model

If the conditional variance is not constant the next step is to fit a STAR-GARCH model to the data. The usual way of obtaining the estimates for the conditional mean and the conditional variance when the latter has a standard GARCH representation is to make use of the block-diagonality of the information matrix. The conditional mean model is estimated first. This is followed by the estimation of the conditional variance model using the residuals from estimating the conditional mean. This procedure yields consistent estimates, but in this paper all parameters are ultimately estimated simultaneously. One advantage with simultaneous estimation is that it may lead to more parsimonious models, at least if the series are not very long. The two-step estimation has a tendency to yield over-parameterized models because some effects due to the nonconstant conditional variance may at first be captured by the estimated conditional mean. The autoregressive parameters turning out to be redundant are eliminated during joint estimation by applying the previous backward elimination algorithm. The estimation is carried out using analytical second derivatives which gives numerically reliable estimates for the information matrix. This is needed at the evaluation stage when the estimated model is tested for misspecification.

However, the two-step estimation is useful for obtaining initial values for the joint estimation. We proceed as follows. First estimate the STAR model for the conditional mean and then estimate a GARCH$(1,1)$ model for the residuals. As a first-order GARCH model has very often been found to be adequate in practice, it is only expanded if necessary. The decision to do that is based on a misspecification test of the functional form which together with other evaluation procedures is discussed in Section 4. To enforce the conditional variance generated by any higher-order GARCH model to be nonnegative, the constraints in Nelson and Cao (1992) for parameters of such models are imposed. The validity of restrictions constraining linear combinations of estimated parameters is verified after the estimation.

3.4. Specification and estimation of the STAR-STGARCH model

The STAR-GARCH model has to be subjected to misspecification tests. We postpone the discussion of such tests to Section 4. At this stage we assume that the linear GARCH specification is rejected in favour of STGARCH and proceed to discuss STAR-STGARCH models. We have to consider specification, estimation and evaluation of these models and begin by specification.

For the smooth transition type alternative (2.5) the problem of selecting the transition variable is not present. We only have to select the order, $n^* \leqslant 2$, of the logistic part in (2.5). One way of doing that is to apply a decision rule similar to that suggested for the STAR model. But then, one may instead simply estimate the STGARCH(n^*) model for $n^* = 1, 2$ and make the choice on the basis of the results, including the results of the misspecification tests in Section 4. The logistic function in $g(\mathbf{z}_t;\boldsymbol{\zeta})$ should also be examined when determining n^*. If an element in the estimated location vector $\mathbf{k} = (k_1, k_2)'$ of the STGARCH(2) model in practice does not affect the values of $g(\mathbf{z}_t;\boldsymbol{\zeta})$ in the sample then n^* may be reduced from two to one. Furthermore, the same type of exclusion restrictions that were considered for the STAR model are relevant for the STGARCH model.

When estimating STAR-STGARCH models it is not certain that the conditional variance is symmetric with respect to the error terms (most often it is not) and therefore the assumption of block diagonality of the information matrix may not hold. This implies that the two-stage estimation algorithm does not yield asymptotically efficient estimates. We maintain our previous strategy and ultimately estimate the conditional mean and the conditional variance jointly.

4. Model evaluation

As discussed above, the validity of the assumptions used in the estimation of parameters must be investigated once the parameters of the STAR-STGARCH model (or a submodel) has been estimated. These assumptions include:

1. The errors and the squared (and standardized) errors of the model are not serially correlated.
2. The parameters of the model are constant.
3. The squared (and standardized) errors of the model are independent and identically distributed.

These assumptions are testable. Furthermore, it is useful to find out whether or not there are any nonlinearities left in the process after fitting a STAR-STGARCH model to the series under consideration. In this paper that possibility is investigated by testing the hypothesis of no additive nonlinearity against this type of nonlinearity.

As for the three testable assumptions, the first two may be tested following Eitrheim and Teräsvirta (1996). These authors have also suggested a test of no additive nonlinearity for the conditional mean. We only have to generalize these tests to the case where the conditional variance follows a STGARCH process. As to the independence hypothesis, the BDS test, see Brock, Dechert, Scheinkman and LeBaron (1996), is applicable if the number of observations is sufficiently large.

4.1. Misspecification tests

4.1.1. General

This section follows Eitrheim and Teräsvirta (1996) and the companion paper Lundbergh and Teräsvirta (1998). Consider the estimated additive STAR-GARCH model as defined in (2.1) and (2.2). An additive extension of the model may be written as

$$\begin{aligned} y_t &= A(\mathbf{w}_t; \boldsymbol{\pi}_a) + \boldsymbol{\varphi}'\mathbf{w}_t + f(\mathbf{w}_t; \boldsymbol{\theta}) + u_t \\ u_t &= \varepsilon_t \sqrt{\boldsymbol{\eta}'\mathbf{z}_t + g(\mathbf{z}_t; \boldsymbol{\zeta}) + B(\mathbf{z}_t; \boldsymbol{\pi}_b)} \end{aligned} \tag{4.1}$$

where $f(\mathbf{w}_t; \boldsymbol{\theta})$ and $g(\mathbf{z}_t; \boldsymbol{\zeta})$ are defined in (2.4) respectively (2.5) and $\{\varepsilon_t\}$ is a sequence of independent standard normal variables. Model (4.1) forms a unifying framework for our tests. Set $\boldsymbol{\omega} = (\boldsymbol{\varphi}', \boldsymbol{\theta}', \boldsymbol{\eta}', \boldsymbol{\zeta}')'$ which comprises all the parameters of the model. Functions $A(\mathbf{w}_t; \boldsymbol{\pi}_a)$ and $B(\mathbf{z}_t; \boldsymbol{\pi}_b)$ are assumed twice continuously differentiable for all $\boldsymbol{\pi}_a$ and $\boldsymbol{\pi}_b$ everywhere in the corresponding sample spaces. For notational simplicity and without loss of generality we assume $A(\mathbf{w}_t; \boldsymbol{\pi}_a) \equiv 0$ for $\boldsymbol{\pi}_a = \mathbf{0}$ and $B(\mathbf{z}_t; \boldsymbol{\pi}_b) \equiv 0$ for $\boldsymbol{\pi}_b = \mathbf{0}$. Tests for various types of misspecification are obtained by different parameterizations of A and B. It is assumed that the maximum likelihood estimator of $\boldsymbol{\omega}$ is consistent and asymptotic normal under any null hypothesis to be considered, which implies that $\{y_t\}$ satisfies the regularity conditions for stationary and ergodicity. Also, the necessary moments needed for $\{u_t\}$ that are required for the asymptotic distribution theory to work are assumed to exist. The null hypothesis of no additional structure is $\mathrm{H}_0 : \boldsymbol{\pi}_a = \mathbf{0}$ and $\boldsymbol{\pi}_b = \mathbf{0}$. The Lagrange multiplier (or score) test statistic is defined as

$$\mathrm{LM} = T \begin{pmatrix} \frac{1}{T}\sum \frac{\partial l_t}{\partial \boldsymbol{\pi}_a}|_{\mathrm{H}_0} \\ 0 \\ \frac{1}{T}\sum \frac{\partial l_t}{\partial \boldsymbol{\pi}_b}|_{\mathrm{H}_0} \end{pmatrix}' \widehat{\mathbf{I}}\,(\boldsymbol{\pi}_a, \widehat{\boldsymbol{\omega}}, \boldsymbol{\pi}_b)^{-1}|_{\mathrm{H}_0} \begin{pmatrix} \frac{1}{T}\sum \frac{\partial l_t}{\partial \boldsymbol{\pi}_a}|_{\mathrm{H}_0} \\ 0 \\ \frac{1}{T}\sum \frac{\partial l_t}{\partial \boldsymbol{\pi}_b}|_{\mathrm{H}_0} \end{pmatrix} \tag{4.2}$$

where $\widehat{\mathbf{I}}$ is a consistent estimator of the information matrix under the null hypothesis. We use the estimated negative expectation of the Hessian as our estimator the information matrix. The partial derivatives forming the Hessian may be found in Appendix B. The test statistic (4.2) is asymptotically χ^2- distributed with $\dim(\boldsymbol{\pi}_a) + \dim(\boldsymbol{\pi}_b)$ degrees of freedom under the null hypothesis. If the information matrix is block-diagonal the test statistic may be computed simply by two artificial regressions; see Lundbergh and Teräsvirta (1998). This approach does not always apply to the STAR-STGARCH model, because the GARCH component may be asymmetric. Note that by letting either $A(\mathbf{w}_t;\boldsymbol{\pi}_a)$ or $B(\mathbf{z}_t;\boldsymbol{\pi}_b)$ to be identically equal to zero also under the alternative amounts to testing the conditional mean and the conditional variance specifications separately. From the modeller's point of view, this is often the most practical alternative. The above structure may also be used for evaluating submodels within the STAR-STGARCH parameterization. For example, by setting $g(\mathbf{z}_t;\boldsymbol{\zeta}) \equiv 0$ in (4.1) we can, among other things, test a STAR-GARCH specification against a STAR-STGARCH model nesting the former.

4.1.2. Test against serial dependence

To test the joint null hypothesis of no serial dependence in either the conditional mean or in the conditional variance or in both, the alternative is stated as remaining serial dependence of order p in the ordinary error process and of order p^* in the squared (and standardized) errors. In the general case, this gives the extended model (4.1) with $A(\mathbf{w}_t;\boldsymbol{\pi}_a) = \boldsymbol{\pi}_a'\mathbf{v}_t$ and $B(\mathbf{w}_t;\boldsymbol{\pi}_b) = \boldsymbol{\pi}_b'\mathbf{v}_t^*$ where $\boldsymbol{\pi}_a = (\pi_{a,1},...,\pi_{a,p})'$, $\mathbf{v}_t = (u_{t-1},...,u_{t-p})'$, $\boldsymbol{\pi}_b = (\pi_{b,1},...,\pi_{b,p^*})'$, and $\mathbf{v}_t^* = (h_{t-1},...,h_{t-p^*})'$. The null hypothesis of no remaining serial dependence in neither the conditional mean nor in the conditional variance is equivalent to $\boldsymbol{\pi}_a = \mathbf{0}$ and $\boldsymbol{\pi}_b = \mathbf{0}$. Under the null hypothesis and assuming that the necessary moments exits, the LM-statistic (4.2) is asymptotically χ^2- distributed with $\dim(\boldsymbol{\pi}_a) + \dim(\boldsymbol{\pi}_b)$ degrees of freedom. The details of an LM-test of this hypothesis for the squared standardized errors are given in Lundbergh and Teräsvirta (1998).

4.1.3. Test against nonconstant parameters

We assume that the alternative to constant parameters in either the conditional mean or in the conditional variance or in both is that the parameters change smoothly over time, see Lin and Teräsvirta (1994) and Lundbergh and Teräsvirta (1998). This gives rise to the following model:

$$\begin{aligned} y_t &= \boldsymbol{\varphi}(t)'\mathbf{w}_t + f(\mathbf{w}_t;\boldsymbol{\theta}(t)) + u_t \\ u_t &= \varepsilon_t\sqrt{\boldsymbol{\eta}(t)'\mathbf{z}_t + g(\mathbf{z}_t;\boldsymbol{\zeta}(t))} \end{aligned} \tag{4.3}$$

where $\boldsymbol{\theta}(t) = (\boldsymbol{\phi}(t)',\gamma,\mathbf{c}')'$ and $\boldsymbol{\zeta}(t) = (\boldsymbol{\psi}(t)',\delta,\mathbf{k}')'$. All time-varying parameter vectors are assumed to have the same structure. For instance, $\boldsymbol{\varphi}(t) = \boldsymbol{\varphi}^* + \lambda_{\varphi}H_{n_\varphi}(t;\gamma_\varphi,\mathbf{c}_\varphi)$ where the transition function $H_{n_\varphi}(t;\gamma_\varphi,\mathbf{c}_\varphi)$ is a logistic function of order n_φ defined in (2.3) with $s_t \equiv t$. If $\gamma_\varphi \to \infty$ while $n_\varphi = 1$, function $H_1(t;\gamma_\varphi,\mathbf{c}_\varphi)$ becomes a step function and the alternative to parameter constancy a single structural break. The null hypothesis of parameter constancy can be stated as $\mathrm{H}_0 : \gamma_{\varphi} = \gamma_{\phi} = \gamma_{\eta} = \gamma_{\psi} = 0$. We circumvent the identification problem under the null hypothesis as before by expanding $H_{n_i}(t; \gamma_i, \mathbf{c}_i)$ into a Taylor series around $\gamma_i = 0$. A realistic assumption in a test situation is to assume that $n_{\varphi} = n_{\phi} = l_1$ and $n_{\eta} = n_{\psi} = l_2$. Using the first-order expansion we obtain, after reparameterization,

$$
\begin{aligned}
y_t &= \boldsymbol{\pi}_a' \mathbf{v}_t + \varkappa_{a,0}' \mathbf{w}_t + \varkappa_{a,l_1+1}' \mathbf{w}_t H_n(s_t; \gamma, \mathbf{c}) + R_1(\mathbf{w}_t; \boldsymbol{\pi}_a, \boldsymbol{\omega}) + u_t \\
u_t &= \varepsilon_t \sqrt{\varkappa_{b,0}' \mathbf{z}_t + \varkappa_{b,l_2+1}' \mathbf{z}_t H_{n^*}(\mathbf{s}_t; \delta, \mathbf{k}) + \boldsymbol{\pi}_b' \mathbf{v}_t^* + R_2(\mathbf{z}_t; \boldsymbol{\pi}_b, \boldsymbol{\omega})}
\end{aligned}
$$

where the logistic transition functions for the STAR-GARCH are estimated with $H_n(s_t; \widehat{\gamma}, \widehat{\mathbf{c}})$ and $H_{n^*}(\mathbf{s}_t; \widehat{\delta}, \widehat{\mathbf{k}})$. For notational simplicity we denote these functions by H_n and H_{n^*}. This gives the extended model (4.1) with $A(\mathbf{w}_t; \boldsymbol{\pi}_a) = \boldsymbol{\pi}_a' \mathbf{v}_t$ and $B(\mathbf{w}_t; \boldsymbol{\pi}_b) = \boldsymbol{\pi}_b' \mathbf{v}_t^*$ where $\boldsymbol{\pi}_a = (\varkappa_{a,1}', ..., \varkappa_{a,l_1}', \varkappa_{a,l_1+2}', ..., \varkappa_{a,2l_1+1}')'$, $\boldsymbol{\pi}_b = (\varkappa_{b,1}', ..., \varkappa_{b,l_2}', \varkappa_{b,l_2+2}', ..., \varkappa_{b,2l_2+1}')$, $\mathbf{v}_t = \left(\mathbf{w}_t t, ..., \mathbf{w}_t t^{l_1}, \mathbf{w}_t t H_n, ..., \mathbf{w}_t t^{l_1} H_n\right)'$ and $\mathbf{v}_t^* = \left(\mathbf{z}_t t, ..., \mathbf{z}_t t^{l_2}, \mathbf{z}_t t H_{n^*}, ..., \mathbf{z}_t t^{l_2} H_{n^*}\right)'$.

The joint null hypothesis of parameter constancy both in the conditional mean and variance consists of the restrictions $\boldsymbol{\pi}_a = \mathbf{0}$ and $\boldsymbol{\pi}_b = \mathbf{0}$. Note that if the null hypothesis holds, $\varkappa_{a,0}' = \boldsymbol{\varphi}'$, $\varkappa_{a,l_1+1}' = \boldsymbol{\phi}'$, $\varkappa_{b,0}' = \boldsymbol{\eta}'$, $\varkappa_{b,l_2+1}' = \boldsymbol{\psi}'$. Furthermore, then the two remainder terms $R_1(\mathbf{w}_t; \boldsymbol{\pi}_a, \boldsymbol{\omega}) \equiv R_2(\mathbf{z}_t; \boldsymbol{\pi}_b, \boldsymbol{\omega}) \equiv 0$ so that they do not affect the asymptotic distribution theory. Under the null hypothesis, the LM-statistic (4.2) is asymptotically χ^2- distributed with $\dim(\boldsymbol{\pi}_a) + \dim(\boldsymbol{\pi}_b)$ degrees of freedom. Again, it is useful to test the constancy of the parameters of the conditional mean and the conditional variance separately. More details about the test of the latter hypothesis and its finite-sample properties can be found in Lundbergh and Teräsvirta (1998). These tests may also be applied to a given subset of parameters by assuming that the remaining ones are constant even under the alternative. This often helps locate the nonconstancy if it exists.

4.1.4. Test against remaining nonlinearity

As we are searching for an adequate nonlinear specification for our time series it is of considerable interest to try and check whether or not our estimated parametrization adequately characterizes all nonlinearity in the series. To keep things simple, we focus on the null hypothesis of no remaining additive nonlinearity. The alternative to this null hypothesis is assumed to be an additive smooth transition component of the same type as before. This alternative may be written as

$$\begin{aligned} y_t &= f_a(\mathbf{w}_t;\boldsymbol{\theta}_a) + \boldsymbol{\varphi}'\mathbf{w}_t + f(\mathbf{w}_t;\boldsymbol{\theta}) + u_t \\ u_t &= \varepsilon_t\sqrt{\boldsymbol{\eta}'\mathbf{z}_t + g(\mathbf{z}_t;\boldsymbol{\zeta}) + g_b(\mathbf{z}_t;\boldsymbol{\zeta}_b)} \end{aligned} \tag{4.4}$$

where the functions $f_a(\mathbf{w}_t;\boldsymbol{\theta}_a)$ and $g_b(\mathbf{z}_t;\boldsymbol{\zeta}_b)$ represent any remaining nonlinearity and have the same parameter structure as (2.4) and (2.5). Set $\boldsymbol{\theta}_a = (\boldsymbol{\phi}_a', \gamma_a, \mathbf{c}_a')'$ and $\boldsymbol{\zeta}_b = (\boldsymbol{\psi}_b', \delta_b, \mathbf{k}_b')'$. The null hypothesis of no remaining nonlinearity can be written as $\mathrm{H}_0 : \gamma_a = \delta_b = 0$. Even here, functions $f_a(\mathbf{w}_t;\boldsymbol{\theta}_a)$ and $g_b(\mathbf{z}_t;\boldsymbol{\zeta}_b)$ are Taylor-expanded to circumvent the identification problem, and we assume that the orders of the smooth transitions are l_1 and l_2, respectively. This yields, after reparameterization, the transformed model

$$\begin{aligned} y_t &= \boldsymbol{\pi}_a'\mathbf{v}_t + \boldsymbol{\varphi}'\mathbf{w}_t + f(\mathbf{w}_t;\boldsymbol{\theta}) + R_3(\mathbf{w}_t;\boldsymbol{\pi}_a,\boldsymbol{\omega}) + u_t \\ u_t &= \varepsilon_t\sqrt{\boldsymbol{\eta}'\mathbf{z}_t + g(\mathbf{z}_t;\boldsymbol{\zeta}) + \boldsymbol{\pi}_b\mathbf{v}_t^* + R_4(\mathbf{z}_t;\boldsymbol{\pi}_b,\boldsymbol{\omega})}. \end{aligned} \tag{4.5}$$

Model (4.5) is a special case of model (4.1) with $A(\mathbf{w}_t;\boldsymbol{\pi}_a) = \boldsymbol{\pi}_a'\mathbf{v}_t$ where $\boldsymbol{\pi}_a = (\varkappa_{a,0}', \varkappa_{a,1}', ..., \varkappa_{a,l_1}')'$ and $\mathbf{v}_t = (\mathbf{w}_t^{lin}, \mathbf{w}_t^{nc}y_{t-e}, ..., \mathbf{w}_t^{nc}y_{t-e}^{l_1})'$. Any linear terms (lags) that are not included in the estimated smooth transition AR part of the model are denoted by $\mathbf{w}_t^{lin}$ whereas $\mathbf{w}_t = (1, (\mathbf{w}_t^{nc})')'$. We assume that the nonlinearity in the conditional variance part is of smooth transition type and therefore given by $B(\mathbf{w}_t;\boldsymbol{\pi}_b) = \boldsymbol{\pi}_b'\mathbf{v}_t^*$ where $\boldsymbol{\pi}_b = (\varkappa_{b,1}, ..., \varkappa_{b,(1+l_2)q})'$ and $\mathbf{v}_t^* = (u_{t-1}, u_{t-1}^3, ..., u_{t-1}^{l_2+2}, ..., u_{t-q}, u_{t-q}^3, ..., u_{t-q}^{l_2+2})'$. The two remainders $R_3(\mathbf{w}_t;\boldsymbol{\pi}_a,\boldsymbol{\omega})$ and $R_4(\mathbf{z}_t;\boldsymbol{\pi}_b,\boldsymbol{\omega})$ do not affect the distribution theory because both are identically equal to zero under H_0. The null hypothesis is $\boldsymbol{\pi}_a = \mathbf{0}$ and $\boldsymbol{\pi}_b = \mathbf{0}$, under which the LM-statistic (4.2) is asymptotically χ^2- distributed with $\dim(\boldsymbol{\pi}_a) + \dim(\boldsymbol{\pi}_b)$ degrees of freedom. As for the conditional mean the test statistic is the same as in Eitrheim and Teräsvirta (1996). The corresponding test for the adequacy of the conditional variance is discussed in detail in Lundbergh and Teräsvirta (1998).

If we assume $g_b(\mathbf{z}_t;\boldsymbol{\zeta}_b) = f(\mathbf{w}_t;\boldsymbol{\theta}) \equiv 0$ in (4.4) and only test $\mathrm{H}_0 : \gamma_a = 0$ then the test is a linearity test for the conditional mean under nonconstant GARCH-type conditional variance. Wong and Li (1997) recently presented a linearity test for a related situation where the conditional mean follows a threshold autoregressive process under the alternative.

4.2. Test of independent, identically distributed disturbances

The misspecification tests of the previous section are tests against a parametric alternative. This is useful in model-building because a rejection or a set of rejections normally contain information about the nature of the misspecification. Nevertheless, we may complete our set of tests by a general test of independence of the (standardized) errors. The BDS statistic (Brock et al., 1996), appears a suitable large-sample test for this purpose. It is based on the m^{th} correlation integral which represents the fraction of all possible pairs of m consecutive points of the series that are closer than $\varepsilon > 0$ to each other. Closeness ε is defined by the Euclidean norm. Let T denote the number of observations and $C_{m,T}(\varepsilon)$ the m^{th} correlation integral. The test statistic is

$$W_{m,T}(\varepsilon) = \sqrt{T}\frac{C_{m,T}(\varepsilon) - C_{1,T}(\varepsilon)^m}{\widehat{\sigma}_{m,T}(\varepsilon)} \tag{4.6}$$

where $\widehat{\sigma}_{m,T}(\varepsilon)$ is an estimate of the standard deviation under the null hypothesis. Statistic (4.6) has an asymptotic $N(0,1)$ distribution under this hypothesis. For details see Brock et al. (1996). In this paper, we have used the C-code of LeBaron (1997) to compute the values of the test statistic.

The null distribution of the statistic depends on the two nuisance parameters, m and ε. In this paper the BDS statistic is computed with $m = 2$ and ε is chosen as the standard deviation of the residual series. These nuisance parameters cause problems in small samples because then the size of the test is a function of these two parameters. In order to have the size of the test under control we perform a simple bootstrap to generate an empirical null distribution of the test statistic for the selected combination of m and ε. At the sample sizes (≥ 1000) we are primarily interested in, however, this precaution seems no longer necessary.

5. Application to two high frequency series

As an illustration we apply the STAR-STGARCH modelling strategy to two series consisting of daily observations. The first series is the Swedish OMX index which consists of the values of the 30 most traded stocks at the Stockholm Stock Exchange. The observation period is December 30, 1983 to September 30, 1998, with a total of 3693 observations. The period until October 4, 1994, is used for estimating the model, whereas the remaining period of 1000 observations is reserved for studying the predictive properties of the model. The second series is the exchange rate for the Japanese yen (JPY) against the US dollar (USD). The observations extend from December 28, 1978 to September 30, 1997, a total of 4756 observations. The period until December 31, 1991, is the estimation period, and the remaining period of 1460 observations is used for a study of the predictive properties of the estimated STAR-STGARCH model.

Both series are transformed to percentage changes in continuously compounded rates. This is done by differencing the logarithms of the original series. Plots of the closing-bid and the percentage changes for both series can be found in Figures C.1-C.4.

5.1. Swedish OMX index

If the errors of a linear model are IID, then the skewness and leptokurtosis may be directly characterized by a nonsymmetric fat-tailed distribution; see for instance Mittnik and Rachev (1993). In such a case there is no need for further modelling. To consider linear dependence in the OMX series, we specified a linear AR model. This AR model was augmented by weekday and holiday dummies to account for possible weekday and holiday effects. An LM-test of no autocorrelation of order at most m, Breusch and Pagan (1980), was then applied to the residuals and there were no evidence of remaining serial dependence. On the other hand, the errors were found to be leptokurtic. Characteristics of the errors of the linear model can be found in Table C.1. The BDS test applied to the residuals from the linear model heavily rejected the null hypothesis of IID errors; see Table C.1. This rejection indicates that there may be some nonlinear structure to be modelled in the OMX index.

5.1.1. Estimation and evaluation

Following the specification strategy outlined above a linearity test, assuming constant conditional variance, was performed where the maximum lag of the linear autoregressive part was 7 and the alternative an LSTAR(3) model. The results of this linearity test for delay parameter values $0 < d \leq 5$, appear in Table C.1. The tests strongly rejects linearity. The strongest rejection occurs at $d = 1$, which is the delay we select. The STAR-STGARCH model we consider has the following form:

$$\begin{aligned} y_t &= D_{Mo}\mathrm{Mo} + D_{Tu}\mathrm{Tu} + D_{We}\mathrm{We} + D_{Th}\mathrm{Th} + D_{Hol}\mathrm{Hol} \qquad (5.1) \\ &\quad +\varphi_0 + \varphi_1 y_{t-1} + ... + \varphi_4 y_{t-4} \end{aligned}$$

$$
\begin{aligned}
& +(\phi_0+\phi_1 y_{t-1}+...+\phi_4 y_{t-4})(1+\exp(-\gamma(y_{t-d}-c))/\widehat{\sigma}(y))^{-1}+\varepsilon_t\sqrt{h_t} \\
h_t &= \alpha_0+\alpha_{11}u_{t-1}^2+\beta_1 h_{t-1} \qquad (5.2)\\
& +(\alpha_{01}+\alpha_{21}u_{t-1}^2)(1+\exp(-\delta(u_{t-1}-k))/\widehat{\sigma}(u))^{-1}
\end{aligned}
$$

where $\widehat{\sigma}(y)$ and $\widehat{\sigma}(u)$ are the sample standard deviations of $\{y_t\}$ and $\{u_t\}$ making γ and δ scale-free.

We first estimated the STAR-STGARCH model ($\alpha_{01}=\alpha_{21}=0$ in (5.2)) for the returns and tested it against the STAR-STGARCH specification. The results in Table C.5 show that the standard GARCH(1,1) model for the conditional variance is inadequate so that an extension to smooth transition GARCH was necessary. Model (5.2) was thus re-estimated without the symmetry restrictions, and the maximum likelihood estimates (standard deviations in parentheses) of the parameters based on analytical first and second derivatives are reported in Table C.2. A missing value in Table C.2 means that the corresponding parameter in (5.1) or (5.2) has been set to zero.

A few characteristic features of the standardized residuals of the estimated model can be found in Table C.3. The standardized residuals are less leptokurtic than the original observations. The results of the misspecification tests for the conditional mean of the STAR-STGARCH model can be found in Table C.4. The null hypothesis of the LM-test of no remaining autocorrelation cannot be rejected and the results of the parameter constancy tests indicate that parameters are constant, except for the daily dummies. This rejection has not been followed up, however. As for the LM-test of no additional nonlinearity against additional nonlinearity of LSTAR(3) type, the hypothesis of no additional nonlinearity cannot be rejected. This is remarkable given the very low p-values of the linearity tests.

The results of the misspecification tests for the conditional variance of the model appear in Table C.6. The null hypothesis no remaining multiplicative ARCH structure, which is asymptotically equivalent to the test of Li and Mak (1994), in the squared and standardized errors cannot be rejected. The test of the functional form indicates no remaining serial dependence of GARCH type. There is still some evidence of nonlinearity in the conditional variance, see Table C.6, but it is now very weak compared to the previous results in Table C.5. The parameter constancy test does not indicate nonconstancy at the 1% level of significance. Finally, it can be seen from Table C.3 that the hypothesis of IID errors cannot be rejected for the standardized errors of the STAR-STGARCH model.

5.1.2. Interpretation

Having obtained a satisfactory model for the OMX index we proceed to interpret the estimation results. A conspicuous feature in the model is that both a Monday and a Tuesday effect seem to exist at a 5 percent significance level. In general there appears to be a weak tendency for the index to display growth towards the end of each week. No holiday effect is found. The conditional mean model is asymmetric. A sufficiently large negative shock causes a secondary negative effect after four days. For positive shocks there is no similar positive secondary effect because $\widehat{\varphi}_4 + \widehat{\theta}_4$ changes sign and becomes slightly negative as the value of the transition function increases towards unity. Figure C.5 shows that both extreme regimes are invoked quite often. As to the conditional variance, the STGARCH model responds asymmetrically to positive and negative errors from the conditional mean part of the model. For positive residuals the behaviour is locally represented by a standard GARCH model, but for negative residuals the local dynamics are close to the dynamic behaviour of an IGARCH process.

Under the assumption of constant conditional variance, a model spectrum for each value of the observed logistic transition function in $f(w_t; \widehat{\theta})$ is plotted in Figure C.6. This 'sliced' spectrum was introduced by Skalin and Teräsvirta (1999). It is a model spectrum conditional on the value of the transition function and describes change in the local dynamics of the conditional mean with the transition from one of the two extreme regimes to the other in the estimated STAR model. The slight peak at the frequency corresponding to four days reflects the situation at $\widehat{H}_1 = 0$. This peak fades away as $\widehat{H}_1 \rightarrow 1$, and thus we obtain another description of the asymmetry in the conditional mean discussed above. Note, however, that the sliced spectrum characterizes local behaviour and cannot be interpreted as representing the "global" dynamic behaviour of the series. If one wants to estimate the global spectrum of the conditional mean process it has to be done numerically; see Skalin and Teräsvirta (1999) for discussion.

It should also be noted that the conditional variance of the model is clearly asymmetric. Positive and negative shocks, i.e, news, of the same size do not have the same impact on the conditional variance.

5.1.3. Forecasting one business day ahead

We computed 1000 one-day-ahead predictions of the conditional mean. The root mean square error (RMSE) equals 0.32 for the conditional mean part of the estimated STAR-STGARCH model while the corresponding linear model has RMSE of 0.39. We evaluate the predictive properties of the model in two ways. First we compare the forecasts from our model with those from a linear AR(7) model by applying the test in Granger and Newbold (1986, pp. 278-280) which is based on the correlation coefficient, r, of the sums and differences of the forecast errors. If this correlation coefficient equals zero then both forecast processes have the same RMSE. The null hypothesis of no correlation is $\mathrm{H}_0 : r = 0$ and the alternative $\mathrm{H}_1 : r > 0$. The alternative corresponds to the situation where the forecasts from the nonlinear model have a smaller RMSE than those from the linear model. For inference we use the well-known transformation

$$w = \frac{1}{2}\ln(\frac{1+r}{1-r}) \tag{5.3}$$

where $w \sim N(0, \frac{1}{N-3})$ under the null hypothesis, N being equal to the number of one-step-ahead forecasts. The test provides very strong evidence of a positive correlation, see Table C.7. It seems that the nonlinear structure captured by the model is useful in forecasting one step ahead.

Another way of evaluating the one-step-ahead forecasts is to compare them with the true outcomes using ordinal data. Positive returns and forecasts that are greater than 0.002 are given value 1. If a forecast or realization is less than -0.002 it obtains value -1. The remaining observations have value zero. This classification may be useful if we think that an agent facing transaction costs only acts upon a forecast if it deviates sufficiently much from zero. Considering prediction accuracy in such a framework may then be of interest. The ordinal data obtained this way are cross-tabulated in Table C.8. Following Agresti (1984, pp. 156-165), the association between the one-step-ahead forecasts and the corresponding outcome is measured using concordant and discordant pairs of observations. For example, a forecast-outcome combination (1,1) forms a concordant pair with any (0,0), (0,-1), (-1,0), or (-1,-1) combination. The discordant pairs are the ones where one of the two elements is higher and the other lower than the corresponding element in the other observation forming the pair. The remaining paired observations are ties. Two measures of association that are based on the differences between concordant and discordant pairs, Somers' d and Kendall's tau-b equal 0.65 and 0.66, respectively. This indicates a positive association between the sign of the one-step-ahead forecasts with the true outcome. An ordinal test of independence based on concordant and discordant pairs can be found in Agresti (1984, pp. 180-181). If we denote the number of concordant and discordant pairs by C and D the test statistic is

$$z = \frac{C - D}{\sigma(C - D)} \tag{5.4}$$

where $\sigma(C - D)$ is the standard error of $C - D$. Under the null hypothesis of no association the statistic (5.4) is approximately normally distributed in large samples with zero mean and unit variance. For the OMX index the value of this statistic equals $\widehat{z} = 22.9$, which is strong evidence in favour of positive ordinal association between the forecasts and outcomes.

5.2. JPY/USD exchange rate

Even here, the analysis was started by estimating an AR model including weekday dummies and a constant term for the series. No remaining serial dependence was found. Various characteristics of the (leptokurtic) errors for the linear models can be found in Table C.1 as well as the result that the BDS test heavily rejects the null hypothesis of IID errors.

5.2.1. Estimation and evaluation

As before, testing linearity was the first step of the specification procedure. In these tests the maximum lag of the linear autoregressive part was 7 and the alternative an LSTAR(3) model. The results from these linearity tests for delay parameter $0 < d \leq 5$ appear in Table C.1, and show that linearity is rejected.

As before, we first estimated a STAR-GARCH model without asymmetry in the conditional variance. As in the linear model, the weekday effects in the conditional mean are represented by daily dummies and a dummy for holiday effects. The parameter estimates (standard deviations in parentheses) of the model are reported in Table C.2 and shows that linearity is rejected. A few characteristic features of the standardized residuals of the estimated model can be found in Table C.3. The leptokurtosis has increased compared to that in the original series. This is because with our univariate model we cannot predict the actual shock, but we can model the average response to it. The actual shock may thus become more conspicuous in the standardized residuals than in the original data, which leads to increased leptokurtosis in the residuals. The condition for the error process having a finite fourth moment, given in Bollerslev (1986) or, more generally, He and Teräsvirta (1999a), is valid for the estimated parameter combination.

The results of the misspecification tests in Section 4 for the conditional mean can be found in Table C.4. There is no remaining autocorrelation, the parameters seem constant, and there is no evidence of remaining additional nonlinearity either. The results of the misspecification tests for the conditional variance appear in Table C.6. The test against remaining multiplicative ARCH structure indicates that there is some structure left that the STAR-GARCH model does not capture. On the other hand, the test of the functional form does not indicate remaining serial dependence of GARCH type and the results of the parameter constancy test are satisfactory. The BDS test in Table C.3 does not reject the IID hypothesis of the errors. Finally, the linearity test against smooth transition GARCH in Table C.6 does not reject the null of symmetry. Thus we tentatively accept the STAR-GARCH model and do not consider the STAR-STGARCH one.

5.2.2. Interpretation

A feature of the results is that the coefficient estimates of all weekday dummies are negative and significant as a 5 percent level which suggests a positive "Friday effect". The conditional mean of the model is asymmetric. Most of the time there exists some linear structure in the process. However, a sufficient large positive shock causes a nonlinear response in the series. This is clearly seen from the sliced spectrum in Figure C.8 which demonstrates the emergence of a local cycle with the period of about 8 days. On the other hand, such a large shock is a relatively rare event. This is best seen from the graph of the transition function in Figure C.7 where every circle represents a single observation. As discussed above there is no asymmetry to be modelled in the conditional variance.

5.2.3. Forecasting one business day ahead

We computed 1460 one-step-ahead predictions of the returns of the JPY/USD exchange rate both with a linear and a nonlinear model. We did not find any linear dependence in the conditional mean and assume therefore in the linear case the process to be a random walk. The RMSE for the conditional mean part of the STAR-GARCH model and the deviation of the actual observations from zero are both 0.27. We computed the Granger and Newbold (1986) RMSE test using statistic (5.3). It is seen from Table C.7 that we cannot reject the null hypothesis $\mathrm{H}_0 : r = 0$ against $\mathrm{H}_1 : r > 0$. As nonlinearity is only required to characterize the response of the process to large shocks, there is no general improvement in the predictive performance compared to the random walk model.

We use the same ordinal observations as to compare one-step-ahead forecasts with the true outcomes. The boundaries for "zero" are now -0.0005 and 0.0005. These ordinal data are cross-tabulated in Table C.9. The Somers' d and the Kendall's tau-b equal 0.02 and 0.01 respectively, and do not suggest any association between the direction of the one-step-ahead forecasts and that of the true outcome. For the JPY/USD exchange rate the value of the ordinal test statistic (5.4) is $\widehat{z} = 0.57$, which strengthens this conclusion.

6. Conclusions

Our STAR-STGARCH model is intended to help us characterize the behaviour of high-frequency economic time series. Many modellers of such series tend to ignore the first moment, but in this paper the first and the second moment are modelled jointly. A coherent modelling strategy is a key to doing that in a systematic way, and such a strategy is designed and applied to data here. An advantage of the proposed strategy is that the specification and misspecification tests we use only require standard asymptotic theory and are easy to perform. The tests for the conditional variance are discussed in detail in the companion paper Lundbergh and Teräsvirta (1998).

The empirical examples indicate that there is nonlinear structure in the conditional mean to be modelled. In the case of the OMX index this leads to improved forecasts. For the JPY/USD exchange rate return series the nonlinear parameters only characterize some extreme events in the series. Because such events by definition are rare, the forecast accuracy, when measured from a large number of forecasts, is not improved by extra parameters. The aftermath of a large positive shock is the only occasion in which the estimated STAR-STGARCH model may generate better forecasts than a linear autoregressive model for this series.

References

Agresti, A. (1984). *Analysis of Ordinal Categorial Data*, Wiley, New York.

Akaike, H. (1974). A new look at the statistical model identification, *IEEE Transactions of Automatic Control* **19**: 716–723.

Bacon, D. W. and Watts, D. G. (1971). Estimating the transition between two intersecting straight lines, *Biometrika* **58**: 525–534.

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

Bollerslev, T., Engle, R. F. and Nelson, D. B. (1994). ARCH models, *in* R. F. Engle and D. L. McFadden (eds), *Handbook of Econometrics*, Vol. 4, Elsevier, Amsterdam, pp. 2959–3038.

Box, G. and Jenkins, G. M. (1970). *Time Series Analysis, Forecasting and Control*, Holden Day, San Francisco.

Breusch, T. S. and Pagan, A. R. (1980). The Lagrange mulitiplier test and its applications to model specification in econometrics, *Review of Economic Studies* **47**: 239–253.

Brock, W. A., Dechert, W., Scheinkman, J. and LeBaron, B. (1996). A test for independence based on the correlation dimension, *Econometric Reviews* **15**: 197–235.

Chan, K. S. and Tong, H. (1986). On estimating thresholds in autoregressive models, *Journal of Time Series Analysis* **7**: 178–190.

Chappell, D. and Peel, D. A. (1998). A note on some properties of the ESTAR model, *Economics Letters* **60**: 311–315.

Eitrheim, Ø. and Teräsvirta, T. (1996). Testing the adequacy of smooth transition autoregressive models, *Journal of Econometrics* **74**: 59–75.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation, *Econometrica* **50**: 987–1007.

Escribano, A. and Jordá, O. (1996). Improved testing and specification of smooth transition regression models. Universidad Carlos III de Madrid, Working Paper.

Ghysels, E., Harvey, A. C. and Renault, E. (1996). Stochastic volatility, *in* G. S. Maddala and C. R. Rao (eds), *Handbook of Statistics*, Vol. 14, Elsevier, Amsterdam, pp. 119–191.

Glosten, L., Jagannathan, R. and Runkle, D. (1993). On the relation between expected value and the volatility of the nominal excess return on stocks, *Journal of Finance* **48**: 1779–1801.

González-Rivera, G. (1998). Smooth transition GARCH models, *Studies in Nonlinear Dynamics and Econometrics* (forthcoming) .

Granger, C. W. J. and Newbold, P. (1986). *Forecasting Economic Time Series*, Academic Press, London and San Diego.

Granger, C. W. J. and Teräsvirta, T. (1993). *Modelling Nonlinear Economic Relationships*, Oxford University Press, Oxford.

Hagerud, G. E. (1997). *A New Non-Linear GARCH Model*, EFI, Stockholm School of Economics.

Hansen, B. E. (1996). Inference when a nuisance parameter is not identified under the null hypothesis, *Econometrica* **64**: 413–430.

He, C. and Teräsvirta, T. (1999a). Fourth moment structure of the GARCH(p,q) process, *Econometric Theory* (forthcoming) .

He, C. and Teräsvirta, T. (1999b). Properties of the autocorrelation function of squared observations for second order GARCH processes under two sets of parameter constraints, *Journal of Time Series Analysis* (forthcoming) .

LeBaron, B. (1997). A fast algorithm for the BDS statistic, *Studies in Nonlinear Dynamics and Econometrics* **2**: 53–59.

Lee, Y. N. and Li, W. K. (1998). On smooth transition double threshold models. Department of Statistics, University of Hong Kong, Research Report No. 198.

Li, C. W. and Li, W. K. (1996). On a double threshold autoregressive heteroscedastic time series model, *Journal of Applied Econometrics* **11**: 253–274.

Li, W. K. and Lam, K. (1995). Modelling asymmetry in stock returns by a threshold autoregressive conditional heteroscedastic model, *The Statistician* **44**: 333–341.

Li, W. K. and Mak, T. K. (1994). On the squared residual autocorrelations in non-linear time series with conditional heteroskedasticity, *Journal of Time Series Analysis* **15**: 627–636.

Lin, C.-F. and Teräsvirta, T. (1994). Testing the constancy of regression parameters against continuous structural change, *Journal of Econometrics* **62**: 211–228.

Lundbergh, S. and Teräsvirta, T. (1998). Evaluating GARCH models. Stockholm School of Economics, Working Paper Series in Economics and Finance, No. 292.

McLeod, A. I. and Li, W. K. (1983). Diagnostic checking ARMA time series models using squared residual autocorrelations, *Journal of Time Series Analysis* **4**: 269–273.

Mittnik, S. and Rachev, S. T. (1993). Modeling asset returns with alternative stable distributions, *Econometric Reviews* **12**: 261–330.

Nelson, D. B. and Cao, C. Q. (1992). Inequality constraints in the univariate GARCH model, *Journal of Business and Economic Statistics* **10**: 229–235.

Pagan, A. (1996). The econometrics of financial markets, *Journal of Empirical Finance* **3**: 15–102.

Palm, F. (1996). GARCH models of volatility, *in* G. Maddala and C. Rao (eds), *Handbook of Statistics*, Vol. 14, Elsevier, Amsterdam, pp. 209–240.

Sentana, E. (1995). Quadratic ARCH models, *Review of Economic Studies* **62**: 639–661.

Skalin, J. and Teräsvirta, T. (1999). Another look at Swedish business cyles, 1861-1988, *Journal of Applied Econometrics* (forthcoming) .

Taylor, S. (1986). *Modelling Financial Time Series*, Wiley, Chichester.

Teräsvirta, T. (1994). Specification, estimation, and evaluation of smooth transition autoregressive models, *Journal of the American Statistical Association* **89**: 208–218.

Tong, H. (1990). *Non-Linear Time Series. A Dynamical System Approach*, Oxford University Press, Oxford.

Tsay, R. S. (1989). Testing and modeling threshold autoregressive processes, *Journal of the American Statistical Association* **84**: 231–240.

Tsay, R. S. (1998). Testing and modeling multivariate threshold models, *Journal of the Americal Statistical Association* **93**: 1188–1202.

Wong, C. S. and Li, W. K. (1997). Testing for threshold autoregression with conditional heteroscedasticity, *Biometrika* **84**: 407–418.

Zakoïan, J.-M. (1994). Threshold heteroskedastic models, *Journal of Economic Dynamics and Control* **18**: 931–955.

A. The linearity test in presence of GARCH

The linearity test in Section 3 is derived under the assumption of constant variance. The aim with this simple simulation study is to examine how the test performs when the conditional variance follows a GARCH-type process.

We let the conditional mean follow an AR(4) model, with parameter values that correspond to one of the estimated regimes found in the Swedish OMX index. This yields the conditional mean model $y_t = 0.0055 - 0.038y_{t-4} + u_t$ where $u_t = \sqrt{h_t}\varepsilon_t$. The IID error term, ε_t, is assumed $N(0,1)$. In the simulations we let the conditional variance follow the asymmetric GJR-GARCH specification suggested by Glosten et al. (1993). This model may be written as

$$h_t = \alpha_0 + \alpha_1 \left[|u_{t-1}| + \omega u_{t-1}\right]^2 + \beta_1 h_{t-1} \tag{A.1}$$

where ω is the asymmetry parameter. If $\omega = 0$ the model reduces to the standard GARCH(1,1) model, and with $\alpha_0 > 0$, $\alpha_1 = \beta_1 = \omega = 0$ the model has a constant conditional variance. The values of the parameters match the estimated ones for the Swedish OMX-index. The parameter values for the different conditional variance DGPs can be found in Table A.1, where a missing value denotes that the corresponding parameter value in (A.1) equals zero.

DGP	Parameter				
	α_0	α_1	β_1	$\alpha_1 + \beta_1$	ω
DGP0	2×10^{-4}	.	.	.	.
DGP1	7×10^{-6}	0.15	0.84	0.99	.
DGP2	7×10^{-6}	0.06	0.84	0.90	.
DGP3	7×10^{-6}	0.15	0.84	0.90 or 0.99	-0.10

Table A.1: Simulation design for the GARCH model (A.1).

The number of replications in the simulation study is set to 5000. The length of the generated time series is 1000 observations after removing the first 500 observations from the beginning of the series to eliminate the effects of the initial values. For each replicate we compute two versions of the linearity test (3.1), one described in Section 3 and another one mentioned in Granger and Teräsvirta (1993, p. 69) which is robust against unspecified heteroscedasticity.

The empirical size of the standard linearity test can be found in Table A.2. When the conditional variance is generated by DGP2 it is found that the size is only marginally affected. As to DGP1 and DGP3, the standard linearity test quite often erroneously detects nonlinear structure in the conditional mean, whereas the corresponding robust version is slightly undersized.

	Standard version			Robust version		
T=1000	1 %	5 %	10 %	1 %	5 %	10 %
DGP0	0.008	0.043	0.085	0.004	0.032	0.069
DGP1	0.166	0.276	0.356	0.003	0.023	0.062
DGP2	0.011	0.054	0.100	0.004	0.035	0.079
DGP3	0.189	0.312	0.391	0.003	0.024	0.063

Table A.2: Empirical size of the linearity test. Each cell represents the proportion of rejections at the given nominal significance level. The alternative to linearity is a nonlinearity of LSTAR(3)-type. The transition variable used in the linearity test is y_{t-1}.

To see how the two versions of the linearity test behave when a nonlinearity is present in the conditional mean we generated data from an LSTAR model whose parameters values match the estimated ones for the Swedish OMX index, see Table C.2. The model is

$$y_t = 0.072 y_{t-4} + (0.0055 - 0.11 y_{t-4})(1 + \exp(-6.1(y_{t-1} - 0.0039))/\widehat{\sigma}(y))^{-1} + \sqrt{h_t}\varepsilon_t$$

where $\widehat{\sigma}(y) = 0.013$. The conditional variance is generated by the DGPs in Table A.1. In this situation we find that the robustification against heteroskedasticity partly absorbs the nonlinearity in the conditional mean, see Table A.3. In the case when the conditional variance follows DGP1 or DGP3 the robust version of the linearity test has very little power against the nonlinearity in the conditional mean. Because of this disadvantage we shall not apply the robust linearity test in this paper.

	Standard version			Robust version		
T=1000	1 %	5 %	10 %	1 %	5 %	10 %
DGP0	0.10	0.26	0.38	0.065	0.22	0.36
DGP1	0.81	0.90	0.93	0.014	0.073	0.15
DGP2	0.40	0.63	0.73	0.13	0.33	0.49
DGP3	0.82	0.91	0.94	0.012	0.069	0.15

Table A.3: Empirical power of the linearity test. Two versions are computed, the standard linearity test for the conditional mean and a version robust against unspecified heteroskedasticity. The alternative to linearity is a nonlinearity of LSTAR(3)-type. The transition variable used in the linearity test is y_{t-1}.

B. Gradient and Hessian of the log-likelihood function

The analytical gradient and the analytical Hessian of the STAR-GARCH model are reported here. The specification tests in Section 4.1 consists of, at most, two additional linear terms and are thus straight forward to implement.

Consider the STAR-STGARCH model defined in (2.1) and (2.2).

$$\begin{aligned} y_t &= \boldsymbol{\varphi}'\mathbf{w}_t + f(\mathbf{w}_t;\boldsymbol{\theta}) + u_t \\ u_t &= \varepsilon_t\sqrt{h(\mathbf{w}_t;\boldsymbol{\varphi},\boldsymbol{\theta},\boldsymbol{\eta},\boldsymbol{\zeta})} \end{aligned}$$

where $h(\mathbf{w}_t,\boldsymbol{\varphi},\boldsymbol{\theta},\boldsymbol{\eta},\boldsymbol{\zeta}) = h_t = \boldsymbol{\eta}'\mathbf{z}_t + g(\mathbf{z}_t;\boldsymbol{\zeta})$. The nonlinear functions $f(\mathbf{w}_t;\boldsymbol{\theta})$ and $g(\mathbf{z}_t;\boldsymbol{\zeta})$ are defined in (2.4) respectively (2.5). Assuming that the sequence $\{\varepsilon_t\}$ is identically normal distributed, the log-likelihood function at time t is

$$l_t = -\frac{1}{2}\ln 2\pi - \frac{1}{2}\ln h_t - \frac{1}{2}\frac{u_t^2}{h_t}$$

where $u_t = y_t - \boldsymbol{\varphi}'\mathbf{w}_t - f(\mathbf{w}_t;\boldsymbol{\theta})$ as neither $\boldsymbol{\varphi}$ nor $\boldsymbol{\theta}$ is assumed to depend on either $\boldsymbol{\eta}$ or $\boldsymbol{\zeta}$. The derivatives of the log-likelihood function are reported in B.1 and the derivatives of the nonlinear functions $f(\mathbf{w}_t;\boldsymbol{\theta})$ and $g(\mathbf{z}_t;\boldsymbol{\zeta})$ are reported in B.2 and B.3.

B.1. Partial derivatives of the log-likelihood function

The first and second order partial derivatives is to be found in this section. The second order partial derivatives are also given in expectation which is useful when computing the specification tests. Also recall that $E[u_t] = 0$ and $E\left[u_t^2\right] = E[g_t] =$ unconditional variance.

B.1.1. First order partial derivative of l_t

The gradient of the log-likelihood function at time t is given by the derivative with respect to the parameters of; the nonlinear function, the linear part and the variance model.

$$G_t = \begin{pmatrix} \frac{\partial l_t}{\partial \boldsymbol{\theta}'} & \frac{\partial l_t}{\partial \boldsymbol{\varphi}'} & \frac{\partial l_t}{\partial \boldsymbol{\eta}'} & \frac{\partial l_t}{\partial \boldsymbol{\zeta}'} \end{pmatrix}$$

where

$$\frac{\partial l_t}{\partial \boldsymbol{\theta}'} = \frac{u_t}{h_t}\frac{\partial f(\mathbf{w}_t;\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'} + \frac{1}{2h_t}\left(\frac{u_t^2}{h_t} - 1\right)\frac{\partial h_t}{\partial \boldsymbol{\theta}'}$$

$$
\begin{aligned}
\frac{\partial l_t}{\partial \boldsymbol{\varphi}'} &= \frac{u_t}{h_t}\mathbf{w}_t' + \frac{1}{2h_t}\left(\frac{u_t^2}{h_t}-1\right)\frac{\partial h_t}{\partial \boldsymbol{\varphi}'} \\
\frac{\partial l_t}{\partial \boldsymbol{\eta}'} &= \frac{1}{2h_t}\left(\frac{u_t^2}{h_t}-1\right)\frac{\partial h_t}{\partial \boldsymbol{\eta}'} \\
\frac{\partial l_t}{\partial \boldsymbol{\zeta}'} &= \frac{1}{2h_t}\left(\frac{u_t^2}{h_t}-1\right)\frac{\partial h_t}{\partial \boldsymbol{\zeta}'}
\end{aligned}
$$

B.1.2. Second order partial derivative of l_t

The Hessian of the log-likelihood function at time t is given by:

$$
H_t = \left(\begin{array}{cc|cc}
\frac{\partial^2 l_t}{\partial \boldsymbol{\varphi}\partial \boldsymbol{\varphi}'} & \frac{\partial^2 l_t}{\partial \boldsymbol{\varphi}\partial \boldsymbol{\theta}'} & \frac{\partial^2 l_t}{\partial \boldsymbol{\varphi}\partial \boldsymbol{\eta}'} & \frac{\partial^2 l_t}{\partial \boldsymbol{\varphi}\partial \boldsymbol{\zeta}'} \\
\frac{\partial^2 l_t}{\partial \boldsymbol{\theta}\partial \boldsymbol{\varphi}'} & \frac{\partial^2 l_t}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}'} & \frac{\partial^2 l_t}{\partial \boldsymbol{\theta}\partial \boldsymbol{\eta}'} & \frac{\partial^2 l_t}{\partial \boldsymbol{\theta}\partial \boldsymbol{\zeta}'} \\
\hline
\frac{\partial^2 l_t}{\partial \boldsymbol{\eta}\partial \boldsymbol{\varphi}'} & \frac{\partial^2 l_t}{\partial \boldsymbol{\eta}\partial \boldsymbol{\theta}'} & \frac{\partial^2 l_t}{\partial \boldsymbol{\eta}\partial \boldsymbol{\eta}'} & \frac{\partial^2 l_t}{\partial \boldsymbol{\eta}\partial \boldsymbol{\zeta}'} \\
\frac{\partial^2 l_t}{\partial \boldsymbol{\zeta}\partial \boldsymbol{\varphi}'} & \frac{\partial^2 l_t}{\partial \boldsymbol{\zeta}\partial \boldsymbol{\theta}'} & \frac{\partial^2 l_t}{\partial \boldsymbol{\zeta}\partial \boldsymbol{\eta}'} & \frac{\partial^2 l_t}{\partial \boldsymbol{\zeta}\partial \boldsymbol{\zeta}'}
\end{array}\right)
$$

We first report the derivatives with respect only to the parameters in the conditional mean block, followed by the derivatives with respect only to the parameters of the conditional variance block. Then we report the cross-term derivatives.

Derivative with respect to the parameters of the linear part in the conditional mean

$$
\begin{aligned}
\frac{\partial^2 l_t}{\partial \boldsymbol{\varphi}\partial \boldsymbol{\varphi}'} &= -\frac{1}{h_t}\mathbf{w}_t\mathbf{w}_t' - \frac{u_t^2}{2h_t^3}\frac{\partial h_t}{\partial \boldsymbol{\varphi}}\frac{\partial h_t}{\partial \boldsymbol{\varphi}'} - \frac{u_t}{h_t^2}\left(\frac{\partial h_t}{\partial \boldsymbol{\varphi}}\mathbf{w}_t' + \mathbf{w}_t\frac{\partial h_t}{\partial \boldsymbol{\varphi}'}\right) \\
&\quad + \frac{1}{2h_t}\left(\frac{u_t^2}{h_t}-1\right)\left(\frac{\partial^2 h_t}{\partial \boldsymbol{\varphi}\partial \boldsymbol{\varphi}'} - \frac{1}{g_t}\frac{\partial h_t}{\partial \boldsymbol{\varphi}}\frac{\partial h_t}{\partial \boldsymbol{\varphi}'}\right)
\end{aligned}
$$

and the expectation of it,

$$
E\left[\frac{\partial^2 l_t}{\partial \boldsymbol{\varphi}\partial \boldsymbol{\varphi}'}\right] = -E\left[\frac{1}{h_t}\mathbf{w}_t\mathbf{w}_t' + \frac{1}{2h_t^2}\frac{\partial h_t}{\partial \boldsymbol{\varphi}}\frac{\partial h_t}{\partial \boldsymbol{\varphi}'}\right]
$$

Derivative with respect to the parameters of the nonlinear function in the conditional mean

$$
\begin{aligned}
\frac{\partial^2 l_t}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}'} &= -\frac{1}{h_t}\frac{\partial f(\mathbf{w}_t;\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\frac{\partial f(\mathbf{w}_t;\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'} + \frac{u_t}{h_t}\left(\frac{\partial^2 f(\mathbf{w}_t;\boldsymbol{\theta})}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}'} - \frac{u_t}{2h_t^2}\frac{\partial h_t}{\partial \boldsymbol{\theta}}\frac{\partial h_t}{\partial \boldsymbol{\theta}'}\right) \\
&\quad - \frac{u_t}{h_t^2}\left(\frac{\partial h_t}{\partial \boldsymbol{\theta}}\frac{\partial f(\mathbf{w}_t;\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'} + \frac{\partial f(\mathbf{w}_t;\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\frac{\partial h_t}{\partial \boldsymbol{\theta}'}\right) \\
&\quad + \frac{1}{2h_t}\left(\frac{u_t^2}{h_t}-1\right)\left(\frac{\partial^2 h_t}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}'} - \frac{1}{g_t}\frac{\partial h_t}{\partial \boldsymbol{\theta}}\frac{\partial h_t}{\partial \boldsymbol{\theta}'}\right)
\end{aligned}
$$

and the expectation of it,

$$E\left[\frac{\partial^2 l_t}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\right]=-E\left[\frac{1}{h_t}\frac{\partial f\left(\mathbf{w}_t;\boldsymbol{\theta}\right)}{\partial\boldsymbol{\theta}}\frac{\partial f\left(\mathbf{w}_t;\boldsymbol{\theta}\right)}{\partial\boldsymbol{\theta}'}+\frac{1}{2h_t^2}\frac{\partial h_t}{\partial\boldsymbol{\theta}}\frac{\partial h_t}{\partial\boldsymbol{\theta}'}\right]$$

Cross-term derivative within the conditional mean

$$\begin{aligned}\frac{\partial^2 l_t}{\partial\boldsymbol{\varphi}\partial\boldsymbol{\theta}'} &= -\frac{1}{h_t}\mathbf{w}_t\frac{\partial f\left(\mathbf{w}_t;\boldsymbol{\theta}\right)}{\partial\boldsymbol{\theta}'}-\frac{u_t^2}{2h_t^3}\frac{\partial h_t}{\partial\boldsymbol{\varphi}}\frac{\partial h_t}{\partial\boldsymbol{\varphi}'}-\frac{u_t}{h_t^2}\left(\frac{\partial h_t}{\partial\boldsymbol{\varphi}}\frac{\partial f\left(\mathbf{w}_t;\boldsymbol{\theta}\right)}{\partial\boldsymbol{\theta}'}+\mathbf{w}_t\frac{\partial h_t}{\partial\boldsymbol{\theta}'}\right)\\ &\quad+\frac{1}{2h_t}\left(\frac{u_t^2}{h_t}-1\right)\left(\frac{\partial^2 h_t}{\partial\boldsymbol{\varphi}\partial\boldsymbol{\theta}'}-\frac{1}{h_t}\frac{\partial h_t}{\partial\boldsymbol{\varphi}}\frac{\partial h_t}{\partial\boldsymbol{\theta}'}\right)\end{aligned}$$

and the expectation of it,

$$E\left[\frac{\partial^2 l_t}{\partial\boldsymbol{\varphi}\partial\boldsymbol{\theta}'}\right]=-E\left[\frac{1}{h_t}\mathbf{w}_t\frac{\partial f\left(\mathbf{w}_t;\boldsymbol{\theta}\right)}{\partial\boldsymbol{\theta}'}+\frac{1}{2h_t^2}\frac{\partial h_t}{\partial\boldsymbol{\varphi}}\frac{\partial h_t}{\partial\boldsymbol{\theta}'}\right]$$

Derivative with respect to the parameters of the linear part of the conditional variance

$$\frac{\partial^2 l_t}{\partial\boldsymbol{\eta}\partial\boldsymbol{\eta}'}=-\frac{u_t^2}{2h_t^3}\frac{\partial h_t}{\partial\boldsymbol{\eta}}\frac{\partial h_t}{\partial\boldsymbol{\eta}'}+\frac{1}{2h_t}\left(\frac{u_t^2}{h_t}-1\right)\left(\frac{\partial^2 h_t}{\partial\boldsymbol{\eta}\partial\boldsymbol{\eta}'}-\frac{1}{h_t}\frac{\partial h_t}{\partial\boldsymbol{\eta}}\frac{\partial h_t}{\partial\boldsymbol{\eta}'}\right)$$

and the expectation of it,

$$E\left[\frac{\partial^2 l_t}{\partial\boldsymbol{\eta}\partial\boldsymbol{\eta}'}\right]=-E\left[\frac{1}{2h_t^2}\frac{\partial h_t}{\partial\boldsymbol{\eta}}\frac{\partial h_t}{\partial\boldsymbol{\eta}'}\right]$$

Derivative with respect to the parameters of the nonlinear part of the conditional variance

$$\frac{\partial^2 l_t}{\partial\boldsymbol{\zeta}\partial\boldsymbol{\zeta}'}=-\frac{u_t^2}{2h_t^3}\frac{\partial h_t}{\partial\boldsymbol{\zeta}}\frac{\partial h_t}{\partial\boldsymbol{\zeta}'}+\frac{1}{2h_t}\left(\frac{u_t^2}{h_t}-1\right)\left(\frac{\partial^2 h_t}{\partial\boldsymbol{\zeta}\partial\boldsymbol{\zeta}'}-\frac{1}{h_t}\frac{\partial h_t}{\partial\boldsymbol{\zeta}}\frac{\partial h_t}{\partial\boldsymbol{\zeta}'}\right)$$

and the expectation of it,

$$E\left[\frac{\partial^2 l_t}{\partial\boldsymbol{\zeta}\partial\boldsymbol{\zeta}'}\right]=-E\left[\frac{1}{2h_t^2}\frac{\partial h_t}{\partial\boldsymbol{\zeta}}\frac{\partial h_t}{\partial\boldsymbol{\zeta}'}\right]$$

Cross-term derivative within the conditional variance

$$\frac{\partial^2 l_t}{\partial\boldsymbol{\eta}\partial\boldsymbol{\zeta}'}=-\frac{u_t^2}{2h_t^3}\frac{\partial h_t}{\partial\boldsymbol{\eta}}\frac{\partial h_t}{\partial\boldsymbol{\zeta}'}+\frac{1}{2h_t}\left(\frac{u_t^2}{h_t}-1\right)\left(\frac{\partial^2 h_t}{\partial\boldsymbol{\eta}\partial\boldsymbol{\zeta}'}-\frac{1}{h_t}\frac{\partial h_t}{\partial\boldsymbol{\eta}}\frac{\partial h_t}{\partial\boldsymbol{\zeta}'}\right)$$

and the expectation of it,

$$E\left[\frac{\partial^2 l_t}{\partial\boldsymbol{\eta}\partial\boldsymbol{\zeta}'}\right]=-E\left[\frac{1}{2h_t^2}\frac{\partial h_t}{\partial\boldsymbol{\eta}}\frac{\partial h_t}{\partial\boldsymbol{\zeta}'}\right]$$

Cross-terms derivatives between the conditional mean and the conditional variance.

$$
\begin{aligned}
\frac{\partial^2 l_t}{\partial \boldsymbol{\varphi} \partial \boldsymbol{\eta}'} &= -\frac{u_t^2}{2h_t^3}\frac{\partial h_t}{\partial \boldsymbol{\varphi}}\frac{\partial h_t}{\partial \boldsymbol{\eta}'} - \frac{u_t}{h_t^2}\mathbf{w}_t\frac{\partial h_t}{\partial \boldsymbol{\eta}'} + \frac{1}{2h_t}\left(\frac{u_t^2}{h_t}-1\right)\left(\frac{\partial^2 h_t}{\partial \boldsymbol{\varphi}\partial \boldsymbol{\eta}'} - \frac{1}{h_t}\frac{\partial h_t}{\partial \boldsymbol{\varphi}}\frac{\partial h_t}{\partial \boldsymbol{\eta}'}\right) \\
\frac{\partial^2 l_t}{\partial \boldsymbol{\varphi} \partial \boldsymbol{\zeta}'} &= -\frac{u_t^2}{2h_t^3}\frac{\partial h_t}{\partial \boldsymbol{\varphi}}\frac{\partial h_t}{\partial \boldsymbol{\zeta}'} - \frac{u_t}{h_t^2}\mathbf{w}_t\frac{\partial h_t}{\partial \boldsymbol{\zeta}'} + \frac{1}{2h_t}\left(\frac{u_t^2}{h_t}-1\right)\left(\frac{\partial^2 h_t}{\partial \boldsymbol{\varphi}\partial \boldsymbol{\zeta}'} - \frac{1}{h_t}\frac{\partial h_t}{\partial \boldsymbol{\varphi}}\frac{\partial h_t}{\partial \boldsymbol{\zeta}'}\right) \\
\frac{\partial^2 l_t}{\partial \boldsymbol{\theta} \partial \boldsymbol{\eta}'} &= -\frac{u_t^2}{2h_t^3}\frac{\partial h_t}{\partial \boldsymbol{\theta}}\frac{\partial h_t}{\partial \boldsymbol{\eta}'} - \frac{u_t}{h_t^2}\frac{\partial f(\mathbf{w}_t;\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\frac{\partial h_t}{\partial \boldsymbol{\eta}'} + \frac{1}{2h_t}\left(\frac{u_t^2}{h_t}-1\right)\left(\frac{\partial^2 h_t}{\partial \boldsymbol{\theta}\partial \boldsymbol{\eta}'} - \frac{1}{g_t}\frac{\partial h_t}{\partial \boldsymbol{\theta}}\frac{\partial h_t}{\partial \boldsymbol{\eta}'}\right) \\
\frac{\partial^2 l_t}{\partial \boldsymbol{\theta} \partial \boldsymbol{\zeta}'} &= -\frac{u_t^2}{2h_t^3}\frac{\partial h_t}{\partial \boldsymbol{\theta}}\frac{\partial h_t}{\partial \boldsymbol{\zeta}'} - \frac{u_t}{h_t^2}\frac{\partial f(\mathbf{w}_t;\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\frac{\partial h_t}{\partial \boldsymbol{\zeta}'} + \frac{1}{2h_t}\left(\frac{u_t^2}{h_t}-1\right)\left(\frac{\partial^2 h_t}{\partial \boldsymbol{\theta}\partial \boldsymbol{\zeta}'} - \frac{1}{g_t}\frac{\partial h_t}{\partial \boldsymbol{\theta}}\frac{\partial h_t}{\partial \boldsymbol{\zeta}'}\right)
\end{aligned}
$$

and the corresponding expectations of them,

$$
\begin{aligned}
E\left[\frac{\partial^2 l_t}{\partial \boldsymbol{\varphi} \partial \boldsymbol{\eta}'}\right] &= -E\left[\frac{1}{2g_t^2}\frac{\partial h_t}{\partial \boldsymbol{\varphi}}\frac{\partial h_t}{\partial \boldsymbol{\eta}'}\right] \\
E\left[\frac{\partial^2 l_t}{\partial \boldsymbol{\varphi} \partial \boldsymbol{\zeta}'}\right] &= -E\left[\frac{1}{2g_t^2}\frac{\partial h_t}{\partial \boldsymbol{\varphi}}\frac{\partial h_t}{\partial \boldsymbol{\zeta}'}\right] \\
E\left[\frac{\partial^2 l_t}{\partial \boldsymbol{\theta} \partial \boldsymbol{\eta}'}\right] &= -E\left[\frac{1}{2h_t^2}\frac{\partial h_t}{\partial \boldsymbol{\theta}}\frac{\partial h_t}{\partial \boldsymbol{\eta}'}\right] \\
E\left[\frac{\partial^2 l_t}{\partial \boldsymbol{\theta} \partial \boldsymbol{\zeta}'}\right] &= -E\left[\frac{1}{2h_t^2}\frac{\partial h_t}{\partial \boldsymbol{\theta}}\frac{\partial h_t}{\partial \boldsymbol{\zeta}'}\right]
\end{aligned}
$$

If the conditional variance model is symmetric and satisfies certain regularity conditions then the expectation of these matrices will be zero; see Theorem 4 in Engle (1982).

B.2. Partial derivatives of the nonlinear function $f(\mathbf{w}_t;\boldsymbol{\theta})$.

As neither $\boldsymbol{\varphi}$ nor $\boldsymbol{\theta}$ is assumed to depend on either $\boldsymbol{\eta}$ or $\boldsymbol{\zeta}$, the partial derivatives of the conditional mean does not depend of the parameterization of the conditional variance. Assume that the nonlinear function $f(\mathbf{w}_t;\boldsymbol{\theta})$ in the conditional mean is parameterized as in (2.4):

$$
f(\mathbf{w}_t;\boldsymbol{\theta}) = \boldsymbol{\phi}'\mathbf{w}_t\left(1+\exp(-\gamma\prod_{l=1}^{n}(y_{t-d}-c_l))\right)^{-1}
$$

To simplify the calculation of the derivatives rewrite function $f(\mathbf{w}_t;\boldsymbol{\theta})$ as:

$$
f(\mathbf{w}_t;\boldsymbol{\theta}) = \boldsymbol{\phi}'\mathbf{w}_t\frac{1}{2\exp(-\gamma\prod_{l=1}^{n}(y_{t-d}-c_l))\cosh(\gamma\prod_{l=1}^{n}(y_{t-d}-c_l))}
$$

Let $\xi_t(y_{t-d};\gamma,c_1,...,c_n) = \frac{\gamma}{2}\prod_{l=1}^{n}(y_{t-d}-c_l)$, drop the arguments and write the function as

$$f(\mathbf{w}_t;\boldsymbol{\theta}) = \boldsymbol{\phi}'\mathbf{w}_t \frac{e^{\xi_t}}{2\cosh\xi_t}$$

B.2.1. First order partial derivative of $f(\mathbf{w}_t;\boldsymbol{\theta})$

Derivative with respect to the parameters of the autoregressive part, the slope and the location vector.

$$\frac{\partial f(\mathbf{w}_t;\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \begin{pmatrix} \frac{\partial f(\mathbf{w}_t;\boldsymbol{\theta})}{\partial \boldsymbol{\phi}'} & \frac{\partial f(\mathbf{w}_t;\boldsymbol{\theta})}{\partial \gamma} & \frac{\partial f(\mathbf{w}_t;\boldsymbol{\theta})}{\partial \mathbf{c}'} \end{pmatrix}$$

where

$$\begin{aligned}
\frac{\partial f(\mathbf{w}_t;\boldsymbol{\theta})}{\partial \boldsymbol{\phi}'} &= \mathbf{w}_t' \frac{e^{\xi_t}}{2\cosh\xi_t} \\
\frac{\partial f(\mathbf{w}_t;\boldsymbol{\theta})}{\partial \gamma} &= \frac{\boldsymbol{\phi}'\mathbf{w}_t}{2\cosh^2\xi_t}\frac{\partial \xi_t}{\partial \gamma} \\
\frac{\partial f(\mathbf{w}_t;\boldsymbol{\theta})}{\partial \mathbf{c}'} &= \frac{\boldsymbol{\phi}'\mathbf{w}_t}{2\cosh^2\xi_t}\frac{\partial \xi_t}{\partial \mathbf{c}'}
\end{aligned}$$

by letting c_k denote the k^{th} element in the location vector $\mathbf{c}$, we can write the first order derivative of ξ_t as

$$\tfrac{\partial \xi_t}{\partial \gamma} = \tfrac{1}{2}\prod_{l=1}^{k}(y_{t-d}-c_l), \quad \tfrac{\partial \xi_t}{\partial c_i} = -\tfrac{\gamma}{2}\prod_{l=1, l\neq i}^{k}(y_{t-d}-c_l)$$

B.2.2. Second order partial derivative of $f(\mathbf{w}_t;\boldsymbol{\theta})$

These derivatives are only interesting when computing the full Hessian, they are not needed for the expectation.

$$\frac{\partial^2 f(\mathbf{w}_t;\boldsymbol{\theta})}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}'} = \begin{pmatrix}
\frac{\partial^2 f(\mathbf{w}_t;\boldsymbol{\theta})}{\partial \boldsymbol{\phi}\partial \boldsymbol{\phi}'} & \frac{\partial^2 f(\mathbf{w}_t;\boldsymbol{\theta})}{\partial \boldsymbol{\phi}\partial \gamma'} & \frac{\partial^2 f(\mathbf{w}_t;\boldsymbol{\theta})}{\partial \boldsymbol{\phi}\partial \mathbf{c}'} \\
\frac{\partial^2 f(\mathbf{w}_t;\boldsymbol{\theta})}{\partial \gamma\partial \boldsymbol{\phi}'} & \frac{\partial^2 f(\mathbf{w}_t;\boldsymbol{\theta})}{\partial \gamma\partial \gamma'} & \frac{\partial^2 f(\mathbf{w}_t;\boldsymbol{\theta})}{\partial \gamma\partial \mathbf{c}'} \\
\frac{\partial^2 f(\mathbf{w}_t;\boldsymbol{\theta})}{\partial \mathbf{c}\partial \boldsymbol{\phi}'} & \frac{\partial^2 f(\mathbf{w}_t;\boldsymbol{\theta})}{\partial \mathbf{c}\partial \gamma'} & \frac{\partial^2 f(\mathbf{w}_t;\boldsymbol{\theta})}{\partial \mathbf{c}\partial \mathbf{c}'}
\end{pmatrix} \tag{B.1}$$

where the elements in (B.1) are given by:

$$\begin{aligned}
\frac{\partial^2 f(\mathbf{w}_t;\boldsymbol{\theta})}{\partial \boldsymbol{\phi}\partial \boldsymbol{\phi}'} &= 0 \\
\frac{\partial^2 f(\mathbf{w}_t;\boldsymbol{\theta})}{\partial \gamma\partial \boldsymbol{\phi}'} &= \frac{\mathbf{w}_t'}{2\cosh^2\xi_t}\frac{\partial \xi_t}{\partial \gamma} \\
\frac{\partial^2 f(\mathbf{w}_t;\boldsymbol{\theta})}{\partial \mathbf{c}\partial \boldsymbol{\phi}'} &= \frac{\mathbf{w}_t'}{2\cosh^2\xi_t}\frac{\partial \xi_t}{\partial \mathbf{c}} \\
\frac{\partial^2 f(\mathbf{w}_t;\boldsymbol{\theta})}{\partial \gamma\partial \gamma} &= \frac{\boldsymbol{\phi}'\mathbf{w}_t}{\cosh^2\xi_t}\left(\frac{1}{2}\frac{\partial^2 \xi_t}{\partial \gamma\partial \gamma} - \frac{\partial \xi_t}{\partial \gamma}\frac{\partial \xi_t}{\partial \gamma}\tanh\xi_t\right)
\end{aligned}$$

$$
\begin{aligned}
\frac{\partial^2 f(\mathbf{w}_t;\boldsymbol{\theta})}{\partial \mathbf{c}\partial\gamma} &= \frac{\boldsymbol{\phi}'\mathbf{w}_t}{\cosh^2\xi_t}\left(\frac{1}{2}\frac{\partial^2\xi_t}{\partial\mathbf{c}\partial\gamma} - \frac{\partial\xi_t}{\partial\mathbf{c}}\frac{\partial\xi_t}{\partial\gamma}\tanh\xi_t\right)\\
\frac{\partial^2 f(\mathbf{w}_t;\boldsymbol{\theta})}{\partial \mathbf{c}\partial\mathbf{c}'} &= \frac{\boldsymbol{\phi}'\mathbf{w}_t}{\cosh^2\xi_t}\left(\frac{1}{2}\frac{\partial^2\xi_t}{\partial\mathbf{c}\partial\mathbf{c}'} - \frac{\partial\xi_t}{\partial\mathbf{c}}\frac{\partial\xi_t}{\partial\mathbf{c}'}\tanh\xi_t\right).
\end{aligned}
$$

Let c_k again denote the k^{th} element of the location vector $\mathbf{c}$. The second order derivative of ξ_t is then

$$
\frac{\partial^2\xi_t}{\partial\gamma\partial\gamma} = 0, \quad \frac{\partial^2\xi_t}{\partial\gamma\partial c_i} = -\frac{1}{2}\prod_{l=1,l\neq i}^{k}(y_{t-d}-c_l), \quad \frac{\partial^2\xi_t}{\partial c_j\partial c_i} = \frac{\gamma}{2}\prod_{l=1,l\neq i,l\neq j}^{k}(y_{t-d}-c_l)
$$

B.3. Partial derivative of the conditional variance h_t.

Assume that the conditional variance, $h(w_t;\boldsymbol{\varphi},\boldsymbol{\theta},\boldsymbol{\eta},\boldsymbol{\zeta}) = h_t$, is parameterized as in (2.2) and (2.5):

$$
\begin{aligned}
h_t &= \boldsymbol{\eta}'\mathbf{z}_t + g(\mathbf{z}_t;\boldsymbol{\zeta})\\
&= \alpha_{01} + \sum_{j=1}^{q}\alpha_{0j}H_{n^*}(u_{t-j}) + \sum_{j=1}^{q}(\alpha_{1j}+\alpha_{2j}H_{n^*}(u_{t-j}))u_{t-j}^2 + \sum_{j=1}^{p}\beta_j h_{t-j}
\end{aligned}
$$

To initialize the iterative computation pre-sample values of h_t are estimated by the sample (unconditional) variance. This is done for all $t \leq 0$ by setting $h_t = u_t^2 = \frac{1}{T}\sum_{s=1}^{T}u_s^2$ where $u_s = y_s - \boldsymbol{\varphi}'\mathbf{w}_s - f(\mathbf{w}_s;\boldsymbol{\theta})$.

B.3.1. First order partial derivative of h_t

The first order derivatives may be computed iteratively by using the following expressions

$$
\begin{aligned}
\frac{\partial h_t}{\partial\boldsymbol{\theta}'} &= \sum_{j=1}^{q}(\alpha_{0j}+\alpha_{2j}u_{t-i}^2)\frac{\partial H_{n^*}(u_{t-j})}{\partial\boldsymbol{\theta}'}\\
&\quad -2\sum_{j=1}^{q}(\alpha_{1j}+\alpha_{2j}H_{n^*}(u_{t-j}))u_{t-j}\frac{\partial f(\mathbf{w}_{t-j};\boldsymbol{\theta})}{\partial\boldsymbol{\theta}'} + \sum_{j=1}^{p}\beta_j\frac{\partial h_{t-j}}{\partial\boldsymbol{\theta}'}\\
\frac{\partial h_t}{\partial\boldsymbol{\varphi}'} &= \sum_{j=1}^{q}(\alpha_{0j}+\alpha_{2j}u_{t-i}^2)\frac{\partial H_{n^*}(u_{t-j})}{\partial\boldsymbol{\varphi}'}\\
&\quad -2\sum_{j=1}^{q}(\alpha_{1j}+\alpha_{2j}H_{n^*}(u_{t-j}))u_{t-j}\mathbf{w}_{t-j}' + \sum_{j=1}^{p}\beta_j\frac{\partial h_{t-j}}{\partial\boldsymbol{\varphi}'}\\
\frac{\partial h_t}{\partial\boldsymbol{\eta}'} &= \mathbf{z}_t' + \sum_{j=1}^{q}(\alpha_{0j}+\alpha_{2j}u_{t-j}^2)\frac{\partial H_{n^*}(u_{t-j})}{\partial\boldsymbol{\eta}'}
\end{aligned}
$$

$$+\sum_{j=1}^{p}\beta_j\frac{\partial h_{t-j}}{\partial \boldsymbol{\zeta}'}$$

$$\frac{\partial h_t}{\partial \boldsymbol{\zeta}'} = \sum_{j=1}^{q}\frac{\partial(\alpha_{0j}+\alpha_{2j}u_{t-j}^2)}{\partial \boldsymbol{\zeta}'}+\sum_{j=1}^{q}(\alpha_{0j}+\alpha_{2j}u_{t-j}^2)\frac{\partial H_{n^*}(u_{t-j})}{\partial \boldsymbol{\zeta}'} +\sum_{j=1}^{p}\beta_j\frac{\partial h_{t-j}}{\partial \boldsymbol{\zeta}'}$$

where the pre-sample values are given by $-2u_t\frac{\partial f(\mathbf{w}_t;\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'} = \frac{\partial h_t}{\partial \boldsymbol{\theta}'} = -\frac{2}{T}\sum_{s=1}^{T} u_s\frac{\partial f(\mathbf{w}_s;\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'}$, $-2u_t\mathbf{w}_t' = \frac{\partial h_t}{\partial \boldsymbol{\varphi}'} = -\frac{2}{T}\sum_{s=1}^{T} u_s\mathbf{w}_t'$, $\frac{\partial h_t}{\partial \boldsymbol{\eta}'} = 0$

B.3.2. Second order partial derivative of $h_t = h(w_t;\boldsymbol{\varphi},\boldsymbol{\theta},\boldsymbol{\eta},\boldsymbol{\zeta})$

$$\frac{\partial^2 g_t}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'} = 2\sum_{i=1}^{q}\alpha_i\left(\frac{\partial f(\mathbf{w}_{t-i};\theta)}{\partial\boldsymbol{\theta}}\frac{\partial f(\mathbf{w}_{t-i};\boldsymbol{\theta})}{\partial\boldsymbol{\theta}'} - u_{t-i}\frac{\partial^2 f(\mathbf{w}_{t-i};\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\right)+\sum_{j=1}^{p}\beta_j\frac{\partial^2 g_{t-j}}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}$$

where the pre-sample values are given by $2\left(\frac{\partial f(\mathbf{w}_t;\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\frac{\partial f(\mathbf{w}_t;\boldsymbol{\theta})}{\partial\boldsymbol{\theta}'} - u_{t-i}\frac{\partial^2 f(\mathbf{w}_t;\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\right) = \frac{\partial^2 g_t}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'} = \frac{2}{T}\sum_{s=1}^{T}\left(\frac{\partial f(\mathbf{w}_s;\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\frac{\partial f(\mathbf{w}_s;\boldsymbol{\theta})}{\partial\boldsymbol{\theta}'} - u_s\frac{\partial^2 f(\mathbf{w}_s;\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\right)$

$$\frac{\partial^2 g_t}{\partial\boldsymbol{\varphi}\partial\boldsymbol{\varphi}'} = 2\sum_{i=1}^{q}\alpha_i\mathbf{w}_{t-i}\mathbf{w}_{t-i}'+\sum_{j=1}^{p}\beta_j\frac{\partial^2 g_{t-j}}{\partial\boldsymbol{\varphi}\partial\boldsymbol{\varphi}'}$$

where the pre-sample values are given by $2\mathbf{w}_t\mathbf{w}_t' = \frac{\partial^2 g_t}{\partial\boldsymbol{\varphi}\partial\boldsymbol{\varphi}'} = \frac{2}{T}\sum_{s=1}^{T}\mathbf{w}_s\mathbf{w}_s'$

$$\frac{\partial^2 g_t}{\partial\boldsymbol{\eta}\partial\boldsymbol{\eta}'} = \frac{\partial\mathbf{z}_t'}{\partial\boldsymbol{\eta}}+\sum_{j=1}^{p}\left(\frac{\partial\beta_j}{\partial\boldsymbol{\eta}}\frac{\partial g_{t-j}}{\partial\boldsymbol{\eta}'}+\beta_j\frac{\partial^2 g_{t-j}}{\partial\boldsymbol{\eta}\partial\boldsymbol{\eta}'}\right)$$

where the pre-sample values are given by $\frac{\partial^2 g_{t-j}}{\partial\boldsymbol{\eta}\partial\boldsymbol{\eta}'} = 0$. The partial derivative of z_t is: $\frac{\partial z_t'}{\partial\eta} = \left(0,0,...,0,\frac{\partial g_{t-1}}{\partial\eta},...,\frac{\partial g_{t-p}}{\partial\eta}\right)$ and the derivative of β_i is: $\frac{\partial\beta_i}{\partial\eta} = (0,0,...,1^*,...,0)'$ where 1^* correspond to element $1+q+i$.

$$\frac{\partial^2 g_t}{\partial\boldsymbol{\varphi}\partial\boldsymbol{\theta}'} = 2\sum_{i=1}^{q}\alpha_i\mathbf{w}_{t-i}\frac{\partial f(\mathbf{w}_{t-i};\boldsymbol{\theta})}{\partial\boldsymbol{\theta}'}+\sum_{j=1}^{p}\beta_j\frac{\partial^2 g_{t-j}}{\partial\boldsymbol{\varphi}\partial\boldsymbol{\theta}'}$$

where the pre-sample values are given by $2\mathbf{w}_t\frac{\partial f(\mathbf{w}_t;\boldsymbol{\theta})}{\partial\boldsymbol{\theta}'} = \frac{\partial^2 g_{t-j}}{\partial\boldsymbol{\varphi}\partial\boldsymbol{\theta}'} = \frac{2}{T}\sum_{s=1}^{T} w_s\frac{\partial f(\mathbf{w}_s;\boldsymbol{\theta})}{\partial\boldsymbol{\theta}'}$

$$\frac{\partial^2 g_t}{\partial\boldsymbol{\varphi}\partial\boldsymbol{\eta}'} = \frac{\partial\mathbf{z}_t'}{\partial\boldsymbol{\varphi}}+\sum_{j=1}^{p}\beta_j\frac{\partial^2 g_{t-j}}{\partial\boldsymbol{\varphi}\partial\boldsymbol{\eta}'}$$

where the pre-sample values are given by $\frac{\partial^2 g_{t-j}}{\partial \boldsymbol{\varphi} \partial \boldsymbol{\eta}'} = 0$. The partial derivative of $\mathbf{z}_t$ is: $\frac{\partial \mathbf{z}_t'}{\partial \boldsymbol{\varphi}} = \left(0, -2u_{t-1}\mathbf{w}_{t-1}, ..., -2u_{t-q}\mathbf{w}_{t-q}, \frac{\partial g_{t-1}}{\partial \boldsymbol{\varphi}}, ..., \frac{\partial g_{t-p}}{\partial \boldsymbol{\varphi}}\right)$

$$\frac{\partial^2 g_t}{\partial \boldsymbol{\theta} \partial \boldsymbol{\eta}'} = \frac{\partial \mathbf{z}_t'}{\partial \boldsymbol{\theta}} + \sum_{j=1}^{p} \beta_j \frac{\partial^2 g_{t-j}}{\partial \boldsymbol{\theta} \partial \boldsymbol{\eta}'}$$

where the pre-sample values are given by $\frac{\partial^2 g_{t-j}}{\partial \boldsymbol{\theta} \partial \boldsymbol{\eta}'} = 0$. The partial derivative of $\mathbf{z}_t$ is: $\frac{\partial \mathbf{z}_t'}{\partial \boldsymbol{\theta}} = \left(0, -2u_{t-1}\frac{\partial f(\mathbf{w}_{t-1};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'}, ..., -2u_{t-q}\frac{\partial f(\mathbf{w}_{t-q};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'}, \frac{\partial g_{t-1}}{\partial \boldsymbol{\varphi}}, ..., \frac{\partial g_{t-p}}{\partial \boldsymbol{\varphi}}\right)$

B.4. Partial derivatives of the logistic function $H_n(s_t; \gamma, \mathbf{c})$

The logistic function (2.3) is the key term in the parameterization of the nonlinearities in both the conditional mean and the conditional variance. On the one hand, the nonlinear function in the conditional mean is constructed as $f(\mathbf{w}_t; \boldsymbol{\theta}) = \boldsymbol{\phi}'\mathbf{w}_t H_n(s_t; \gamma, \mathbf{c})$, with $s_t = y_{t-d}$. On the other hand, the logistic function $H_{n^*}(s_t; \delta, \mathbf{k})$ that imposes the nonlinearity in the conditional variance is parameterized with $s_t = u_{t-d}$ which depends on $\boldsymbol{\varphi}$ and $\boldsymbol{\theta}$. The logistic function is defined as:

$$H_n(s_t; \gamma, \mathbf{c}) = \left(1 + \exp(-\gamma \prod_{l=1}^{n} (s_t - c_l))\right)^{-1}, \gamma > 0, c_1 \leq ... \leq c_n.$$

To simplify the calculation of the derivatives the logistic function is rewritten as:

$$H_n(s_t; \gamma, \mathbf{c}) = \frac{1}{2\exp(-\gamma \prod_{l=1}^{n} (s_t - c_l)) \cosh(\gamma \prod_{l=1}^{n} (s_t - c_l))}$$

Let $\xi_t(s_t; \gamma, c_1, ..., c_n) = \frac{\gamma}{2} \prod_{l=1}^{n} (s_t - c_l)$, drop the arguments and write the function as:

$$H_n(s_t; \gamma, \mathbf{c}) = \frac{e^{\xi_t}}{2\cosh \xi_t}$$

B.4.1. First order partial derivative of $H_n(s_t; \gamma, \mathbf{c})$

All of the first order derivatives have the same structure, as an example we consider the derivative with respect to $\boldsymbol{\varphi}$.

$$\frac{\partial H_n(s_t)}{\partial \boldsymbol{\varphi}'} = \frac{1}{2\cosh^2 \xi_t} \frac{\partial \xi_t}{\partial \boldsymbol{\varphi}}$$

Let c_i denote the i^{th} element of the location vector $\mathbf{c}$, then the derivatives of ξ_t with respect to the parameters of the model are

$$
\begin{aligned}
\frac{\partial \xi_t}{\partial \boldsymbol{\varphi}} &= \frac{\delta}{2} \frac{\partial s_t}{\partial \boldsymbol{\varphi}} \times Sum \\
\frac{\partial \xi_t}{\partial \boldsymbol{\theta}} &= \frac{\delta}{2} \frac{\partial s_t}{\partial \boldsymbol{\theta}} \times Sum \\
\frac{\partial \xi_t}{\partial \delta} &= \frac{1}{2} \prod_{l=1}^{n} (s_t - c_l) \\
\frac{\partial \xi_t}{\partial c_i} &= -\frac{\delta}{2} \prod_{l=1, l \neq i}^{n} (s_t - c_l)
\end{aligned}
$$

where $Sum = \sum_{j=1}^{n} \left(\prod_{l=1, l \neq j}^{n} (s_t - c_l) \right)$ if $n > 1$, otherwise $Sum = 1$. Note that $\frac{\partial \xi_t}{\partial \boldsymbol{\varphi}} = \mathbf{0}$ and $\frac{\partial \xi_t}{\partial \boldsymbol{\theta}} = \mathbf{0}$ in the conditional mean case due to the fact that $s_t = y_{t-d}$ which depends neither on $\boldsymbol{\varphi}$ nor $\boldsymbol{\theta}$.

B.4.2. Second order partial derivative of $H_n(s_t; \gamma, \mathbf{c})$

These derivatives are only interesting when computing the full Hessian:

$$
\frac{\partial^2 f(\mathbf{w}_t; \boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} = \begin{pmatrix} \frac{\partial^2 f(\mathbf{w}_t; \boldsymbol{\theta})}{\partial \phi \partial \phi'} & \frac{\partial^2 f(\mathbf{w}_t; \boldsymbol{\theta})}{\partial \phi \partial \gamma'} & \frac{\partial^2 f(\mathbf{w}_t; \boldsymbol{\theta})}{\partial \phi \partial \mathbf{c}'} \\ \frac{\partial^2 f(\mathbf{w}_t; \boldsymbol{\theta})}{\partial \gamma \partial \phi'} & \frac{\partial^2 f(\mathbf{w}_t; \boldsymbol{\theta})}{\partial \gamma \partial \gamma'} & \frac{\partial^2 f(\mathbf{w}_t; \boldsymbol{\theta})}{\partial \gamma \partial \mathbf{c}'} \\ \frac{\partial^2 f(\mathbf{w}_t; \boldsymbol{\theta})}{\partial \mathbf{c} \partial \phi'} & \frac{\partial^2 f(\mathbf{w}_t; \boldsymbol{\theta})}{\partial \mathbf{c} \partial \gamma'} & \frac{\partial^2 f(\mathbf{w}_t; \boldsymbol{\theta})}{\partial \mathbf{c} \partial \mathbf{c}'} \end{pmatrix}
$$

C. Tables and Figures

	OMX index	JPY/USD
Characteristics:		
Min	-0.083	-0.049
Max	0.091	0.039
Mean	0	0
Variance	1.6×10^{-4}	4.7×10^{-5}
Skewness	0.056	-0.38
Kurtosis	8.7	6.4
Non-parametric test of IID (BDS)		
statistic	12.4	7.8
LM-test against nonlinearity of STAR type:		
	p-value	p-value
d =1	7.9×10^{-24}	0.0014
d =2	9.5×10^{-14}	0.00017
d =3	1.8×10^{-20}	0.00044
d =4	3.6×10^{-9}	7.6×10^{-6}
d =5	7.6×10^{-16}	0.17

Table C.1: Certain characteristics, the BDS test and linearity tests for the OMX-index and the JPY/USD exchange rate. The characteristics and the BDS test are computed from the residuals of a linear model (a constant, daily dummy variables and an AR polynomial). The BDS statistic is asymptotically normally distributed with zero expectation and unit variance. The LM test against nonlinearity of LSTAR(3) type uses a lag length 7 for the autoregressive part. The test is computed, assuming constant conditional variance, against the alternative for, $1 \leq d \leq 5$. The null hypotesis is given in the table.

Figure C.1: The daily Swedish OMX index, December 30, 1983 to September 30, 1998. The dashed vertical line corresponds to October 5, 1994. Observations preceding this date are used for estimation and the remaining ones for one-step-ahead forecasting.

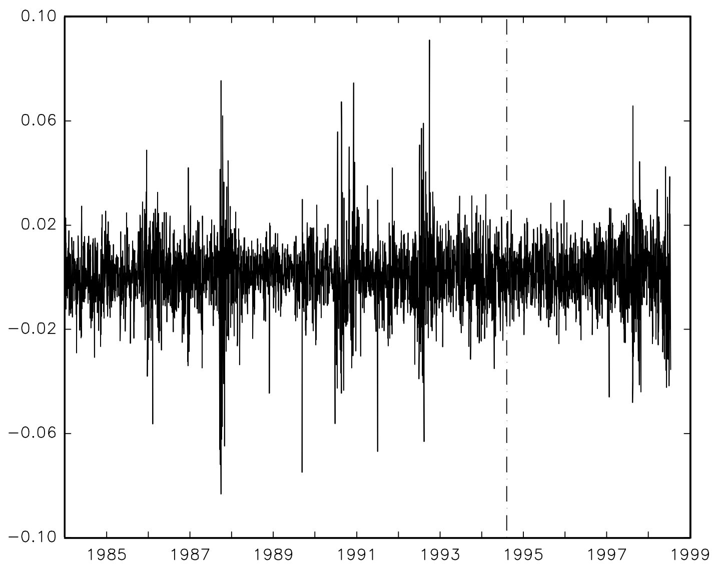

Figure C.2: Returns of the daily Swedish OMX index (first differences), December 31, 1983 to September 30, 1998. The dashed line corresponds to October 5, 1994. Observations preceding this date are used for estimation and the remaining ones for one-step-ahead forecasting.

Parameter	OMX index	JPY/USD
Conditional mean model		
D_{Mo}	-0.0033 (0.00052)	-0.00092 (0.00034)
D_{Tu}	-0.0023 (0.00050)	-0.00070 (0.00035)
D_{We}	-0.00048 (0.00050)	-0.0010 (0.00035)
D_{Th}	-0.00083 (0.00052)	-0.0013 (0.00035)
D_{Hol}	-0.00073 (0.0028)	-0.0015 (0.0013)
φ_0	.	0.00082 (0.00025)
φ_2	.	0.029 (0.019)
φ_3	.	0.042 (0.019)
φ_4	0.072 (0.028)	.
ϕ_0	0.0055 (0.00063)	-0.011 (0.0043)
ϕ_2	.	-0.55 (0.32)
ϕ_3	.	-0.43 (0.30)
ϕ_4	-0.11 (0.049)	.
d	1	4
γ	6.1 (2.1)	4.2 (3.7)
c	0.0039 (0.00087)	0.019 (0.0024)
Conditional variance model		
α_0	7.1×10^{-6} (1.2×10^{-6})	3.7×10^{-6} (6.4×10^{-7})
α_{11}	0.16 (0.021)	0.094 (0.013)
β_1	0.84 (0.018)	0.83 (0.022)
α_{01}	.	.
α_{21}	-0.10 (0.021)	.
δ	7.7 (5.5)	.
k	.	.

Table C.2: Parameter estimates of the STAR-STGARCH models (standard deviations in parentheses) for the OMX-index and the JPY/USD exchange rate.

	OMX index	JPY/USD
Characteristics:		
Min	-9.2	-9.1
Max	5.6	4.2
Mean	-0.018	-0.0089
Variance	1.0	1.0
Skewness	-0.44	-0.63
Kurtosis	7.7	7.9
Nonparametric test of IID (BDS)		
BDS statistic	0.28	1.1
p-value (asymptotic)	0.78	0.29
p-value (bootstrap)	0.78	0.30

Table C.3: Characteristics of the standardized residuals of the STAR-STGARCH model and the BDS test of independence for the standardized errors. For the BDS test a bootstrapped probability value based on 1000 resampled series is reported as well.

	OMX index	JPY/USD
Remaining autocorrelation (p-values)		
l =1	0.27	0.11
l =2	0.54	0.037
l =3	0.69	0.086
l =4	0.81	0.15
l =5	0.90	0.13
Parameter constancy (p-values)		
All	0.076	0.13
Dummies	0.0080	0.80
Linear	0.73	0.033
Non-linear	0.82	0.57
Remaining nonlinearity (p-values)		
d =1	0.11	0.49
d =2	0.44	0.36
d =3	0.38	0.019
d =4	0.17	0.36
d =5	0.35	0.72

Table C.4: p-values of specification tests for the conditional mean for the estimated STAR-STGARCH model. LM tests for the conditional mean: The test of no remaining autocorrelation is computed against the alternative of remaining autocorrelation up to the given lag, l. The test of parameter constancy is computed against the alternative of time-dependence given by an LSTAR(3) parametrization with time as the transition variable. The test against nonlinearity of LSTAR(3) type uses a lag length 7 for the autoregressive part. The test is computed separately against the alternatives $1 \leq d \leq 5$.

	OMX index
Remaining nonlinearity of STGARCH type (p-values)	
All parameters	1.8×10^{-8}
Constant intercept	9.9×10^{-7}
Remaining nonlinearity with a fixed delay (p-values)	
$d=1$	1.8×10^{-8}
$d=2$	2.3×10^{-5}
$d=3$	0.017

Table C.5: p-values of the test against nonlinearity in the conditional variance for a STAR-GARCH model estimated for the OMX index. The tests against remaining nonlinearity make use of a third order Taylor approximation of the transition function. One test is against a STGARCH structure and the other one against a nonlinearity with a fixed delay. The latter one is computed separately against $1 \leq d \leq 3$.

	OMX index	JPY/USD
Remaining ARCH (p-values)		
$l=1$	0.15	0.17
$l=2$	0.32	0.18
$l=3$	0.51	0.0013
$l=10$	0.45	0.022
Test of the functional form (p-values)		
$l=1$	0.60	0.49
$l=2$	0.44	0.52
$l=3$	0.45	0.15
$l=10$	0.11	0.33
Parameter constancy (p-values)		
All	0.012	0.053
Intercept	0.060	0.090
Alfa	0.13	0.74
Beta	0.055	0.38
Smooth	0.25	.
Remaining nonlinearity of STGARCH type (p-values)		
All parameters	0.0021	0.080
Constant intercept	0.0013	0.69
Remaining nonlinearity with a fixed delay (p-values)		
$d=1$	0.0021	0.080
$d=2$	0.60	0.53
$d=3$	0.13	0.21

Table C.6: p-values of specification tests for the conditional variance of the estimated STAR-STGARCH model. LM tests for the conditional variance: The tests of no remaining serial dependence in the squared and standardized residuals are computed against the alternative of remaining dependence up to the given lag, l. The test of parameter constancy is computed against the alternative of time-dependence given by an LSTAR(2) parametrization with time as the transition variable. The tests against remaining nonlinearity make use of a third order Taylor approximation of the transition function. One test is against a STGARCH structure and the other one against a nonlinearity with a fixed delay. The latter one is computed separately against the alternatives $1 \leq d \leq 3$.

Asset	RMSE		Test
	Our model	Competitor	p-value
OMX index	0.32	0.39	0
JPY/USD	0.27	0.27	0.99

Table C.7: Root mean square errors (RMSE) of the conditional mean part of the STAR-STGARCH model and its linear competitor. Also the p-value of the test of the nullhypothesis $\mathrm{H}_0 : r = 0$ against $\mathrm{H}_1 : r > 0$ is given. For the Swedish OMX index we use a linear AR(7) model as the competitor. For the JPY/USD exchange rate we use the deviation of the actual observations from zero as the competitor.

		The model			
		-1	0	1	Total
The observed data	-1	108	281	3	392
	0	27	117	3	147
	1	1	126	334	461
	Total	136	524	340	1000

Table C.8: Cross-tabulation of the ordered actual observations and one-step-ahead forecasts of the Swedish OMX index. A value between -0.002 and 0.002 is denoted 0, otherwise positive and negative values are represented by 1 and -1.

		The model			
		-1	0	1	Total
The observed data	-1	89	453	104	646
	0	18	89	18	125
	1	95	471	123	689
	Total	202	1013	245	1460

Table C.9: Cross-tabulation of the ordered actual observations and one-step-ahead forecasts of the JPY/USD exchange rate. A value between -0.0005 and 0.0005 is denoted 0, otherwise positive and negative values are represented by 1 and -1.

Figure C.3: The JPY/USD daily exchange rate, December 28, 1978 to September 30, 1997. The dashed vertical line corresponds to January 1, 1992. Observations preceding this date are used for estimation and the remaining ones for one-step-ahead forecasting.

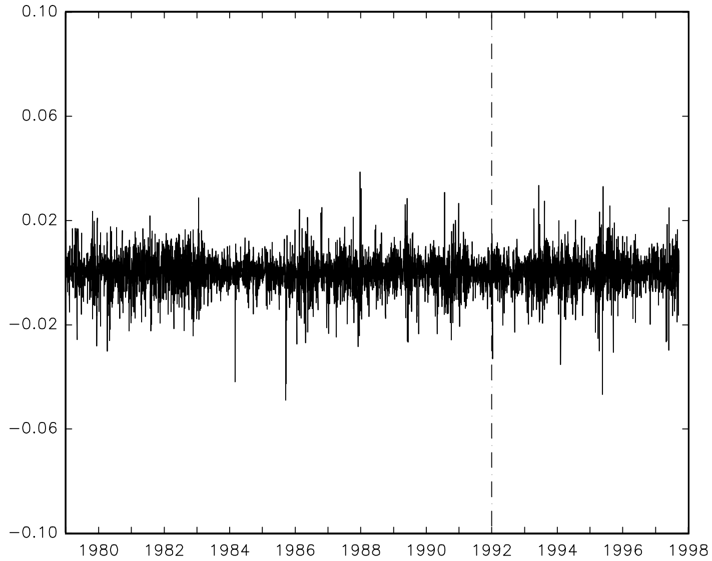

Figure C.4: Returns of the JPY/USD daily exchange rate (first differences), December 29, 1978 to September 30 1997. The dashed line corresponds to January 1, 1992. Observations preceding this date are used for estimation and the remaining ones for one-step-ahead forecasting.

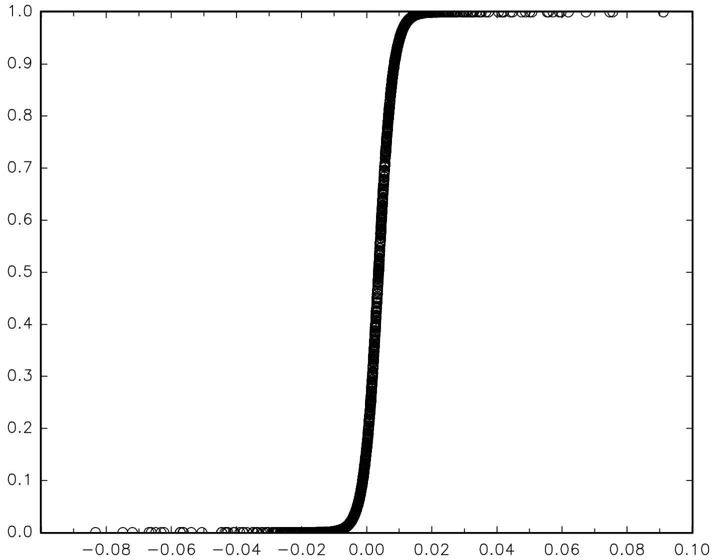

Figure C.5: Values of the transition function for the conditional mean part of model (2.4) for the Swedish OMX index return series. Each circle indicates an observation.

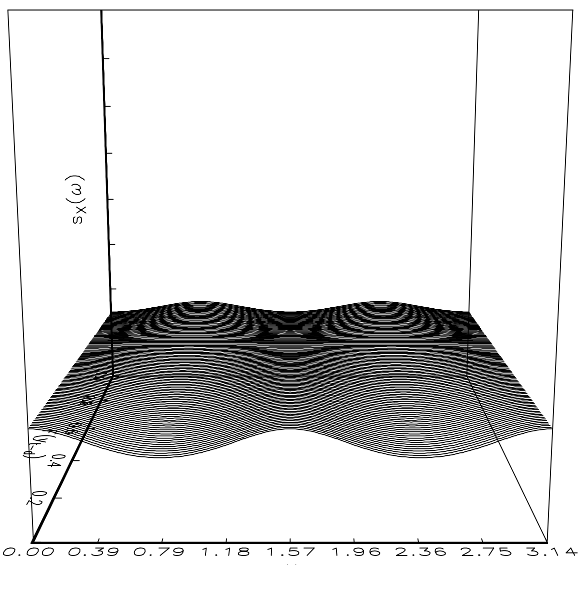

Figure C.6: The estimated 'sliced' spectrum of model (2.4) for the Swedish OMX index return series. The x-axis gives the frequency and the y-axis gives the value of the transition function. A slice (solid curve) represents at least one observed value of the transition function.

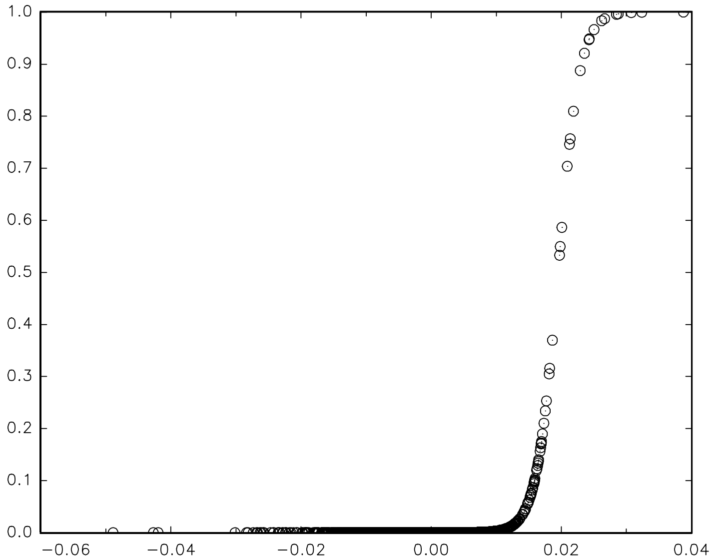

Figure C.7: Values of the transition function for the conditional mean part of model (2.4) for the JPY/USD exchange rate return series. Each circle indicates an observation.

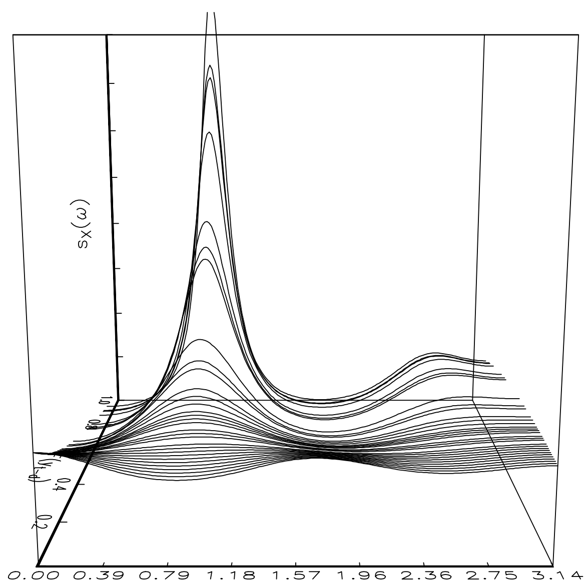

Figure C.8: The estimated 'sliced' spectrum of model (2.4) for the JPY/USD exchange rate return series. The x-axis gives the frequency and the y-axis gives the value of the transition function. A slice (solid curve) represents at least one observed value of the transition function.