NORMALISED LEAST-SQUARES ESTIMATION IN
TIME-VARYING ARCH MODELS

Piotr Fryzlewicz,
Department of Mathematics,
Imperial College,
London SW7 2AZ,
United Kingdom.
Email: p.fryzlewicz@imperial.ac.uk

Theofanis Sapatinas,
Department of Mathematics and Statistics,
University of Cyprus,
P.O. Box 20537,
CY 1678 Nicosia,
Cyprus.
Email: t.sapatinas@ucy.ac.cy

and

Suhasini Subba Rao,
Institut für Angewandte Mathematik,
Universität Heidelberg,
Im Neuenheimer Feld 294,
D-69120 Heidelberg,
Germany.
Email: suhasini@statlab.uni-heidelberg.de

August 13, 2004

Running Title: Estimation in Time-Varying ARCH Models
Abstract

We consider estimation in the recently proposed class of time-varying locally stationary ARCH\((p)\) models. We define a local normalised least-squares criterion which, unlike the previously proposed local quasi-maximum likelihood criterion, has the advantage of having a tractable, explicit solution. Under minimal moment conditions, we derive the asymptotic properties of the kernel-based normalised least-squares estimator. Despite its simplicity, tractability and ease of computation, it suffers from a number of drawbacks: we identify these and propose an adaptation of the local normalised least-squares criterion which yields an improved estimator of the time-varying ARCH\((p)\) parameters. We introduce this estimator as a two-stage scheme, which is computationally as simple to evaluate as the kernel-based normalised least-squares estimator. Under an additional mild moment condition, we show that the asymptotic properties of this two-stage kernel-based normalised least-squares estimator are similar to those of the kernel-based normalised least-squares estimator, as well as those of the kernel-based quasi-maximum likelihood estimator.

Since the stationary ARCH\((p)\) model belongs to the class of time-varying ARCH\((p)\) models, the estimators considered here can also be used to estimate stationary ARCH\((p)\) parameters. We summarise the normalised least-squares and two-stage normalised least-squares schemes for the estimation of stationary ARCH\((p)\) parameters, and state their asymptotic properties.

As an illustration, we use several exchange rate datasets to illustrate the forecasting ability of the time-varying ARCH\((1)\) model whose parameters are estimated using the two-stage kernel-NLS estimator, and make comparisons with the benchmark stationary GARCH\((1,1)\) model. Also, we fit the time-varying ARCH\((1)\) model to the USD/GBP exchange rate series and demonstrate its goodness-of-fit by examining the residuals.

Keywords: ARCH Models; GARCH Models; Quasi-Maximum Likelihood Estimation; Kernel Smoothing; Least Squares Estimation; Locally Stationary Models.
1 Introduction

Among models for log-returns \( X_t = \log(P_t) - \log(P_{t-1}), t = 1, 2, \ldots \), on speculative prices \( P_t \) (such as stock indices, foreign exchange rates, share prices, bond yields, etc.), the autoregressive conditionally heteroscedastic model of order \( p \) \((1 \leq p < \infty) \) (ARCH\((p)\), see Engle, 1982) and the generalised autoregressive conditionally heteroscedastic model of orders \( p \) and \( q \) \((1 \leq p < \infty, 0 \leq q < \infty) \) (GARCH\((p, q)\), see Bollerslev, 1986) have gained particular popularity and have become standard in the financial econometrics literature as they model well the volatility of financial markets over short periods of time. For a review of recent advances on ARCH\((p)\), GARCH\((p, q)\) and related models, which account for asymmetry, leverage effect, heavy tails and other stylised facts present in financial time series, we refer the reader to Fan & Yao (2003) and Giraitis et al. (2003).

However, underlying all these models is the assumption of stationarity. Given the changing pace of the world’s economy, modelling financial log-returns over long intervals using stationary time series models may be inappropriate. It is quite plausible that structural changes in financial time series may occur, causing the time series to deviate from stationarity over long intervals. It is therefore likely that by relaxing the assumption of stationarity in an adequate way, one may obtain a better fit. The modelling of financial data using nonstationary time series models has recently attracted considerable attention. Arguments for using such models were laid out, for example, in Mikosch & Stáricá (2000, 2003, 2004), Mercurio & Spokoiny (2004a, 2004b) and Stáricá & Granger (2003).

Recently, Dahlhaus & Subba Rao (2003) generalised the class of stationary ARCH\((p)\) models to models with time-varying parameters, i.e.,

\[ X_t = \sigma(t)Z_t, \quad \sigma^2(t) = a_0(t) + \sum_{j=1}^{p} a_j(t)X_{t-j}^2, \quad \text{for} \quad t = 1, 2, \ldots, \tag{1} \]

where \( \{Z_t : t = 1, 2, \ldots\} \) are independent and identically distributed random variables with \( \mathbb{E}(Z_t) = 0 \) and \( \mathbb{E}(Z_t^2) = 1 \). As in nonparametric regression and other work on nonparametric statistics, they used the rescaled time concept to develop an asymptotic theory for this class of models, that is the domain of the time-varying parameters was rescaled to the unit interval. The resulting model was called the time-varying ARCH\((p)\) (tvARCH\((p)\)) model. They assumed that the time-varying parameters were, in some sense, changing slowly through time, and used a local kernel quasi-maximum likelihood (kernel-QML) method to estimate them. The special case of tvARCH\((0)\) was also considered by Drees & Stáricá (2003), where the time-varying unconditional variance was estimated using a Nadaraya-Watson kernel estimator.

In this paper, we also consider the tvARCH\((p)\) process. In Section 2, we introduce our notation and assumptions, formally define the tvARCH\((p)\) process, and summarise the asymptotic properties of the kernel-QML estimator of the tvARCH\((p)\) parameters. As an extra motivation for applying the tvARCH\((p)\) model to financial log-returns, we show that despite the true autocovariance of the squares of the tvARCH\((p)\) process decaying geometrically fast to zero, the sample autocovariance, evaluated under the wrong premise of stationarity, does not decay to zero. Thus, the tvARCH\((p)\)
model, due to its nonstationarity, captures the long memory property present in many financial datasets: a feature also exhibited by a short memory GARCH (1,1) model with structural breaks (see Mikosch & Stårică, 2000, 2003, 2004).

The main focus of this paper is on the estimation of tvARCH($p$) parameters. If the dataset is large, evaluating the kernel-QML estimator at every time point can be computationally intensive. Moreover, the practical performance of this estimator depends on the numerical routine employed to minimise the criterion function, which is an obvious drawback of the estimation procedure. As an alternative to the local QML criterion, we define a local normalised least squares (NLS) criterion where the normalisation is introduced to avoid imposing excessively strong moment conditions on the process. The local NLS criterion has the advantage of having a tractable, explicit solution. In Section 3, we derive the asymptotic properties of a kernel-based NLS (kernel-NLS) estimator under minimal moment conditions. Despite its simplicity, tractability and ease of computation, the kernel-NLS estimator suffers from a number of drawbacks: we identify these and, motivated by this discussion and a small simulation study, propose an adaptation of the local NLS criterion which yields an improved estimator of the tvARCH($p$) parameters. In Section 4, we introduce this estimator as a two-stage scheme, which is computationally as simple to evaluate as the kernel-NLS estimator. Under an additional mild moment condition, we show that the asymptotic properties of this two-stage kernel-NLS estimator are similar to those of the kernel-NLS estimator, as well as those of the kernel-QML estimator.

We note that the stationary ARCH($p$) model belongs to the class of tvARCH($p$) models. Therefore, although the focus of this paper is on the estimation of tvARCH($p$) parameters, the estimators considered here can also be used to estimate stationary ARCH($p$) parameters. In Section 5, we summarise the NLS and two-stage NLS schemes for the estimation of stationary ARCH($p$) parameters, and state their asymptotic properties.

In Section 6, we use several exchange rate datasets to illustrate the forecasting ability of the tvARCH(1) model whose parameters are estimated using the two-stage kernel-NLS estimator, and make comparisons with the benchmark stationary GARCH(1,1) model. Also, we fit the tvARCH(1) model to the USD/GBP exchange rate series and demonstrate its goodness-of-fit by examining the residuals.

Concluding remarks and possible avenues for future research are provided in Section 7. The proofs are deferred to the Appendix.

2 The tvARCH($p$) model: preliminary results and motivation

In this section, we introduce our notation and assumptions, formally define the tvARCH($p$) process and state some of its properties, as well as summarising the asymptotic properties of the kernel-QML estimator of the tvARCH($p$) parameters. For more details on these and other probabilistic properties of the tvARCH($p$) process, and the kernel-QML estimator, we refer the reader to Dahlhaus & Subba Rao (2003).

As an extra motivation for applying this model to financial log-returns, we show that the
tvARCH($p$) model, due to its nonstationarity, captures the long memory property present in many financial datasets.

2.1 Notation and assumptions

Consider the nonstationary version of the ARCH($p$) process, where the parameters are time-dependent, as defined in (1). Suppose, for example, that the functions $a_j(t)$ can be written on some interval in a parametric form, for example $a_j(t) = b_j + c_j t$. Then, estimators for the parameters $b_j$ and $c_j$ can be obtained. However, classical asymptotic theory cannot be used to compare, for example, the efficiency of the parameter estimators. Furthermore, asymptotic considerations as $N \to \infty$ make little sense. For example, in the above case, $a_j(N) \to \infty$ as $N \to \infty$, which contradicts the condition $\sum_j a_j(t) < 1$ which ensures the stability of an ARCH($p$) process.

In order to obtain a framework for a meaningful asymptotic theory, which also does not affect estimation procedures, we rescale the domain of the time-varying parameters to the unit interval, as in nonparametric regression and other work on nonparametric statistics. That is, we assume

$$X_{t,N} = \sigma_{t,N} Z_t, \quad \sigma_{t,N}^2 = a_0 \left( \frac{t}{N} \right) + \sum_{j=1}^{p} a_j \left( \frac{t}{N} \right) X_{t-j,N}^2, \quad \text{for} \quad t = 1, 2, \ldots, N,$$

(2)

where $\{Z_t : t = 1, 2, \ldots, N\}$ are independent and identically distributed random variables with $\mathbb{E}(Z_t) = 0$ and $\mathbb{E}(Z_t^2) = 1$. The sequence of stochastic processes $\{X_{t,N} : t = 1, \ldots, N\}$ which satisfy (2) is called a tvARCH($p$) process.

Dahlhaus & Subba Rao (2003) show that the tvARCH($p$) model can locally be approximated by a stationary ARCH($p$) model and, therefore, this new class of models can be called locally stationary. The stationary approximation is useful since it can be used to transfer results on stationary ARCH($p$) processes to the tvARCH($p$) setup. Let $u \in (0, 1]$ and suppose that, for each fixed $u$, $\{\tilde{X}_t(u) : t = 1, 2, \ldots\}$ satisfies the model

$$\tilde{X}_t(u) = \tilde{\sigma}_t(u) Z_t, \quad \tilde{\sigma}_t^2(u) = a_0(u) + \sum_{j=1}^{p} a_j(u) \tilde{X}_{t-j}(u), \quad \text{for} \quad t = 1, 2, \ldots,$$

(3)

where $\{Z_t : t = 1, 2, \ldots\}$ are independent and identically distributed random variables with $\mathbb{E}(Z_t) = 0$ and $\mathbb{E}(Z_t^2) = 1$.

Below we show that $\{\tilde{X}_t(u) : t = 1, 2, \ldots\}$ can be regarded as a stationary approximation of the nonstationary process $\{X_{t,N}^2 : t = 1, 2, \ldots, N\}$ about $u \approx t/N$. Let $\bar{a}^T = (a_0, a_1, \ldots, a_p)$ and $\bar{a}^T(u) = (a_0(u), a_1(u), \ldots, a_p(u))$. Throughout the paper, $^T$ denotes the transpose of a vector.

**Assumption 2.1** Suppose $\{X_{t,N} : t = 1, 2, \ldots, N\}$ is a tvARCH($p$) process. We assume that the time-varying parameters $\{a_j(u) : j = 0, 1, \ldots, p\}$ and the innovation process $\{Z_t : t = 1, 2, \ldots\}$ satisfy the following conditions:

(ia) There exist $0 < \rho_1 \leq \rho_2 < \infty$ and $\delta > 0$ such that, for all $u \in (0, 1]$, $\rho_1 \leq a_0(u) \leq \rho_2$, $\rho_1 \leq a_p(u)$ and $\sup_u \sum_{j=1}^{p} a_j(u) \leq 1 - \delta$. 

5
(ib) There exist $0 < \rho_1 \leq \rho_2 < \infty$ and $\delta > 0$ such that, for all $u \in (0, 1]$, $\rho_1 \leq a_0(u) \leq \rho_2$, $p_1 \leq a_j(u)$ $(j = 1, 2, \ldots, p)$ and $\sup_u \sum_{j=1}^p a_j(u) \leq 1 - \delta$. Let $\Theta$ be the compact set defined as $\Theta = \{a : \rho_1 \leq a_0 \leq \rho_2, p_1 \leq a_j (j = 1, 2, \ldots, p), \sum_{j=1}^p a_j \leq 1 - \delta\}$. For each $u \in (0, 1]$, we assume that $a(u) \in \Theta$.

(ii) There exist $\beta \in (0, 1]$ and a constant $K > 0$ such that for $u, v \in (0, 1]$

$$|a_j(u) - a_j(v)| \leq K|u - v|^\beta$$  for each $j = 0, 1, \ldots, p$.

(iii) For some $\gamma > 0$, $\mathbb{E}\left(Z_t^{4(1+\gamma)}\right) < \infty$;

(iv) For some $0 < \eta \leq 1$ and $\delta > 0$, $\mathbb{E}\left(Z_t^{2(1+\eta)}\right)^{1/(1+\eta)} \sup_u \sum_{j=1}^p a_j(u) \leq 1 - \delta$.

Assumption 2.1(ia) (or Assumption 2.1(ib)) implies that $\sup_{t,N} \mathbb{E}(X_{t,N}^2) < \infty$ and $\sup_u \mathbb{E}(\tilde{X}_t^2(u)) < \infty$. Assumptions 2.1(ia, ii) (or Assumptions 2.1(ib, ii)) mean that the tvARCH($p$) process can locally be approximated by a stationary process. Under Assumptions 2.1(ia, ii), we are able to show consistency of the kernel-NLS estimator (defined in Section 2.1) and, additionally, under Assumption 2.1(iii), we can show its asymptotic normality. Additionally, we require Assumption 2.1(iv) to show consistency and asymptotic normality of the two-stage kernel-NLS estimator (defined in Section 4.1). Since these two results only require that for some $0 < \eta \leq 1$ and $\delta > 0$, $\mathbb{E}(Z_t^{2(1+\eta)})^{1/(1+\eta)} \sup_u \sum_{j=1}^p a_j(u) \leq 1 - \delta$, and we have that $\mathbb{E}(Z_t^2) \sup_u \sum_{j=1}^p a_j(u) \leq 1 - \delta$ by Assumption 2.1(ia), this additional assumption is only a very weak one. (Moreover, Assumption 2.1(iv) also implies that $\sup_{t,N} \mathbb{E}(X_{t,N}^{2(1+\eta)}) < \infty$ and $\sup_u \mathbb{E}(\tilde{X}_t^{2(1+\eta)}(u)) < \infty$.) Finally, Assumptions 2.1(ib, ii) and Assumptions 2.1(ib, ii, iii) are used in showing consistency and asymptotic normality, respectively, of the kernel-QML estimator (defined in Section 2.3).

The proof of the following lemma is a special case of Corollary 4.2 in Subba Rao (2004).

**Lemma 2.1** Suppose $\{X_{t,N} : t = 1, 2, \ldots, N\}$ is a tvARCH($p$) process which satisfies Assumptions 2.1(ia,ii), and let $\{\tilde{X}_t(u) : t = 1, 2, \ldots\}$ be defined as in (3). Then, we have that $\{\tilde{X}_t^2(u) : t = 1, 2, \ldots\}$ is a stationary, ergodic process such that

$$|X_{t,N}^2 - \tilde{X}_t^2(u)| \leq \frac{1}{N^{3/2}} V_{t,N} + \left|u - \frac{t}{N}\right|^\beta W_t, \quad \text{almost surely}, \quad (4)$$

and

$$|\tilde{X}_t^2(u) - \tilde{X}_t^2(v)| \leq |u - v|^\beta W_t, \quad \text{almost surely},$$

where $\{V_{t,N} : t = 1, 2, \ldots, N\}$ and $\{W_t : t = 1, 2, \ldots\}$ are well-defined positive processes.

In addition, if we assume that Assumption 2.1(iv) holds, then we have that

$$\sup_{t,N} \mathbb{E}|V_{t,N}|^{1+\eta} < \infty \quad \text{and} \quad \mathbb{E}|W_t|^{1+\eta} < \infty.$$
Remark 2.1 The explicit forms of the processes \( V_{t,N} \) and \( W_t \) in Lemma 2.1 are not important in the subsequent development, but we refer the reader to Corollary 4.2, (20) and (46) in Subba Rao (2004) for the explicit forms. It is worth noting, however, that \( \{ W_t : t = 1, 2, \ldots \} \) is a well-defined stationary process.

Lemma 2.1 is used in the following. To obtain asymptotic properties of the estimators considered in this paper, we use the idea of rescaled time (see, for example, Dahlhaus, 1997). To illustrate the idea, consider the local average of the process \( \{ X_{t,N}^2 : t = 1, 2, \ldots N \} \) about the time point \( t_0 = t_0(N) \). Since the time-varying parameters are only ‘slowly changing’ in some neighbourhood of the time point \( t_0 \), the observations \( \{ X_{t_0+k,N}^2 : k = -M, \ldots, M \text{ with } M << N \} \) are close to stationary. Therefore, the local average about the time point \( t_0 \), i.e.,

\[
\frac{1}{2M+1} \sum_{k=-M}^{M} X_{t_0+k,N}^2,
\]

could be considered as an estimator of \( \mathbb{E}(X_{t_0,N}^2) \). However, the limit as \( M, N \to \infty \), with \( M << N \), still makes no sense, since \( \mathbb{E}(X_{t_0,N}^2) \) may vary with \( t_0 \) and \( N \). To circumvent this problem, we appeal to Lemma 2.1. From (4) we have that the process \( \{ X_{t_0+k,N}^2 : k = -M, \ldots, M \text{ with } M << N \} \) can be approximated by the stationary process \( \{ \tilde{X}_{t_0+k}(t_0/N)^2 : k = -M, \ldots, M \text{ with } M << N \} \). Therefore, by replacing \( X_{t_0+k,N}^2 \) with \( \tilde{X}_{t_0+k}(t_0/N)^2 \), we have that

\[
\frac{1}{2M+1} \sum_{k=-M}^{M} X_{t_0+k,N}^2 = \frac{1}{2M+1} \sum_{k=-M}^{M} \tilde{X}_{t_0+k}\left(\frac{t_0}{N}\right) + O_p\left\{ \left(\frac{M}{N}\right)^\beta + \frac{1}{N^\beta} \right\},
\]

with \( M \to \infty \) and \( M/N \to 0 \) as \( N \to \infty \).

Suppose now that we impose the constraint \(|u_0 - t_0/N| < 1/N\), so that \( u_0 \approx t_0/N \), and let \( N \to \infty \). This means that for every \( N \) we have a new process (which in practice cannot be realised) and, as \( N \) increases, \( \{ X_{t_0+k,N}^2 : k = -M, \ldots, M \text{ with } M << N \} \) is closer to the stationary process \( \{ \tilde{X}_{t_0+k}^2(u_0) : k = -M, \ldots, M \text{ with } M << N \} \). Therefore, if \(|u_0 - t_0/N| < 1/N\), then, by using (5), we have that

\[
\frac{1}{2M+1} \sum_{k=-M}^{M} X_{t_0+k,N}^2 \overset{p}{\to} \mathbb{E}\left\{ \tilde{X}_{t_0}^2(u_0) \right\}, \text{ with } M \to \infty, \text{ } M/N \to 0 \text{ as } N \to \infty.
\]

In the asymptotic results derived in this paper, we will rely on the two ingredients described above: the time rescaling arguments as well as the approximation of the nonstationary process \( \{ X_{t,N}^2 : t = 1, 2, \ldots N \} \) by the stationary process \( \{ \tilde{X}_t^2(u) : t = 1, 2, \ldots \} \) about \( u \approx t/N \). (Throughout the paper, \( \overset{p}{\to} \) and \( \overset{D}{\to} \) denote convergence in probability and in distribution respectively.)

### 2.2 The long-memory effect

In this section, we show that the sample autocovariance of the squares of a tvARCH\((p)\) process, evaluated under the wrong premise of stationarity, does not typically decay to zero, which captures the long memory property present in many financial datasets.
The following proposition shows the behaviour of the true autocovariance function of the squares of a tvARCH($p$) process.

**Proposition 2.1** Suppose $\{X_{t,N} : t = 1, 2, \ldots, N\}$ is a tvARCH($p$) process which satisfies Assumptions 2.1(ia,ii), and assume that $\left\{ \mathbb{E}(Z_t^2) \right\}^{1/2} \sup_u \sum_{j=1}^{p} a_j(u) \leq 1 - \delta$, for some $\delta > 0$. Then, for some $\rho \in (1 - \delta, 1)$ and a fixed $h \geq 0$, we have that

$$\sup_{t,N} |\text{cov} \left\{ X_{t,N}^2, X_{t+h,N}^2 \right\}| \leq K \rho^h,$$

where the constant $K$ is independent of $h$.

If the fourth moment of the process $\{X_{t,N} : t = 1, \ldots, N\}$ exists, then Proposition 2.1 implies that the true autocovariance of the squares of the tvARCH($p$) process decays geometrically to zero. Therefore, $\{X_{t,N}^2 : t = 1, 2, \ldots, N\}$ is a short memory process. However, we now show that its sample autocovariance does not necessarily decay to zero. Typically, if we believed that the process $\{X_{t,N}^2 : t = 1, 2, \ldots, N\}$ was stationary, we would use $S_N(h)$ as an estimator of $\text{cov} \{X_{t,N}^2, X_{t+h,N}^2\}$, where

$$S_N(h) = \frac{1}{N-h} \sum_{t=1}^{N-h} X_{t,N}^2 X_{t+h,N}^2 - (\bar{X}_N)^2 \quad \text{and} \quad \bar{X}_N = \frac{1}{N-h} \sum_{t=1}^{N-h} X_{t,N}^2.$$

Denote $\mu(u) = \mathbb{E}(\tilde{X}_t^2(u))$ and $c(u, h) = \text{cov}(\tilde{X}_t^2(u), \tilde{X}_{t+h}(u))$ for each $u \in (0, 1]$ and $h \geq 0$. Note that, for each $u \in (0, 1]$, $c(u, h) \geq 0$ for all $h \geq 0$ by the associativity property of the process $\{\tilde{X}_t^2(u) : t = 1, 2, \ldots\}$ for each $u \in (0, 1]$ (see, for example, Giraitis et al., 2003).

The following proposition shows that the sample autocovariance, $S_N(h)$, evaluated under the wrong assumption of stationarity, does not decay to zero. Thus, the tvARCH($p$) model captures the long memory property: a feature also exhibited by short memory GARCH (1,1) models with structural breaks (see Mikosch & Stârică, 2000, 2003, 2004).

**Proposition 2.2** Suppose $\{X_{t,N} : t = 1, 2, \ldots, N\}$ is a tvARCH($p$) process which satisfies Assumptions 2.1(ia,ii), and assume that $\left\{ \mathbb{E}(Z_t^{2(2+\zeta)}) \right\}^{1/(2+\zeta)} \sup_u \sum_{j=1}^{p} a_j(u) \leq 1 - \delta$, for some $0 < \zeta \leq 2$ and $\delta > 0$. Then, for fixed $h > 0$, we have that

$$S_N(h) \quad \overset{p}{\rightarrow} \quad \int_0^1 c(u, h) du + \int \int_{\{0 \leq u < v \leq 1\}} \{\mu(u) - \mu(v)\}^2 du dv, \quad \text{as} \quad N \rightarrow \infty. \quad (6)$$

### 2.3 The kernel-QML estimator and its asymptotic properties

Typically, one would estimate the parameters of a stationary ARCH($p$) model using a QML estimation method. Analogously, Dahlhaus & Subba Rao (2003) used a kernel-QML estimation method to estimate the parameters of a tvARCH($p$) model. In this section, we define the kernel-QML estimator and summarise some of its asymptotic properties.

Throughout the paper, the kernel $W : [-1/2, 1/2] \rightarrow \mathbb{R}$ is a function of bounded variation which satisfies the standard conditions: $\int_{-1/2}^{1/2} W(x) dx = 1$, $\int_{-1/2}^{1/2} W(x) dx = 0$ and $\int_{-1/2}^{1/2} W^2(x) dx < \infty$. 

The kernel-QML is defined as follows. Following the notation in Dahlhaus & Subba Rao (2003), let

\[ L_{mle}^{t_0,N}(a(u)) = \sum_{k=p+1}^{N} \frac{1}{bN} W \left( \frac{t_0 - k}{bN} \right) \ell_{k,N}(a(u)), \tag{7} \]

where

\[ \ell_{k,N}(a(u)) = \frac{1}{2} \left( \log(w_{k,N}(a(u))) + \frac{X_{k,N}^2}{w_{k,N}(a(u))} \right) \]

and

\[ w_{k,N}(a(u)) = a_0 \left( \frac{k}{N} \right) + \sum_{j=1}^{p} a_j \left( \frac{k}{N} \right) X_{k-j,N}^2. \]

If \(|u_0 - t_0/N| < 1/N\), we use \( \hat{a}_{mle}^{t_0,N} \) as an estimator of \( a(u_0) \), where

\[ \hat{a}_{mle}^{t_0,N} = \arg\min_{a \in \Theta} L_{mle}^{t_0,N}(a), \tag{8} \]

where the parameter space \( \Theta \) is defined in Assumption 2.1.(ib). We call \( \hat{a}_{mle}^{t_0,N} \) the kernel-QML estimator.

In the derivation of the asymptotic properties of the kernel-QML estimator, we make use of the local approximation of the nonstationary process \( \{X_{t,N}^2 : t = 1, 2, \ldots N\} \) by the stationary process \( \{\tilde{X}_t^2(u) : t = 1, 2, \ldots\} \) about \( u \approx t/N \). Similarly to the above, let

\[ L_{mle}^{t_0}(u,a(u)) = \sum_{k=p+1}^{N} \frac{1}{bN} W \left( \frac{t_0 - k}{bN} \right) \tilde{\ell}_k(u,a(u)), \]

where

\[ \tilde{\ell}_k(u,a(u)) = \frac{1}{2} \left( \log(w_k(u,a(u))) + \frac{\tilde{X}_k^2(u)}{w_k(u,a(u))} \right) \]

and

\[ w_k(u,a(u)) = a_0(u) + \sum_{j=1}^{p} a_j(u) \tilde{X}_{k-j}^2(u). \]

We also define the following quantities

\[ B_{mle}^{t_0,N}(a(u)) = \sum_{k=p+1}^{N} \frac{1}{bN} W \left( \frac{t_0 - k}{bN} \right) \left[ \ell_{k,N}(a(u)) - \tilde{\ell}_k(u,a(u)) \right], \tag{9} \]

and

\[ \hat{\Sigma}^{mle}(u) = \frac{1}{N} \mathbb{E} \left\{ \frac{\tilde{X}_{k-1}(u) \tilde{X}_{k-1}^T(u)}{\sigma_k^4(u)} \right\}, \tag{10} \]

where \( \tilde{X}_t^T(u) = (1, \tilde{X}_t^2(u), \ldots, \tilde{X}_{t-p+1}^2(u)) \). Further, let \( (\nabla f)^T = (\frac{\partial f}{\partial x_1}, \ldots, \frac{\partial f}{\partial x_r})^T \) for any function \( f : \mathbb{R}^r \to \mathbb{R} \).
We now state the asymptotic properties of \( \hat{a}^{mle}_{t_0,N} \). The proof of the following proposition can be obtained by combining the arguments in Theorems 4.1, 4.2 and (39) in Dahlhaus & Subba Rao (2003). We note that the results given in Dahlhaus & Subba Rao (2003) are given under slightly stronger conditions on the time-varying parameters \( \{a_j(u) : j = 0,1,\ldots,p\} \). However, by using the methods laid out in Dahlhaus & Subba Rao (2003) and Lemma 2.1 above, it is straightforward to obtain the same results for the more general parameter space \( \Theta \), defined in Assumption 2.1.(ib).

**Proposition 2.3** Suppose \( \{X_{t,N} : t = 1,\ldots,N\} \) is a tvARCH(p) process which satisfies Assumptions 2.1.(ib,ii), and let \( \hat{a}^{mle}_{t_0,N} \), \( B^{mle}_{t_0,N}(a(u)) \) and \( \tilde{\Sigma}^{mle}(u) \) be defined as in (8), (9) and (10) respectively. If \( |u_0 - t_0/N| < 1/N \), then we have that

(i) \( \nabla B^{mle}_{t_0,N}(a(u_0)) = O_p(b^\beta) \), with \( b \to 0 \), \( bN \to \infty \) as \( N \to \infty \).

(ii) \( \hat{a}^{mle}_{t_0,N} \overset{D}{\to} a(u_0) \), with \( b \to 0 \), \( bN \to \infty \) as \( N \to \infty \).

(iii) In addition, if we assume that Assumption 2.1.(iii) holds, then we have that

\[
\sqrt{bN} \left( \hat{a}^{mle}_{t_0,N} - a(u_0) \right) + \sqrt{bN} \left\{ \tilde{\Sigma}^{mle}(u_0) \right\}^{-1} \nabla B^{mle}_{t_0,N}(a(u_0)) \overset{D}{\to} \mathcal{N} \left( 0, w_2 \frac{\mu_4}{2} \left\{ \tilde{\Sigma}^{mle}(u_0) \right\}^{-1} \right),
\]

with \( b \to 0 \), \( bN \to \infty \) as \( N \to \infty \), where \( w_2 = \int_{-1/2}^{1/2} W^2(x) dx \) and \( \mu_4 = \text{var} \left( Z_t^2 \right) \).

### 3 The kernel-NLS estimator and its asymptotic properties

The kernel-QML estimator defined in Section 2.3 is computationally intensive to evaluate, especially if it has to be computed at every time point to obtain estimates of \( \{a_j(t/N) : j = 0,1,\ldots,p\} \) at \( t = p+1,\ldots,N \). Also, no explicit form of (8) exists, meaning that (7) must be minimised numerically. For these reasons, we define a kernel-NLS estimator. Roughly speaking, this is a variation on the least-squares estimator, which has been adapted to the ARCH(p) environment where fewer moments are assumed to exist compared to typical AR(p) models.

If the usual least-squares estimator were used locally to obtain estimates of \( \{a_j(t/N) : j = 0,1,\ldots,p\} \) at \( t = p+1,\ldots,N \), consistency of the estimator would require the existence of the fourth moment of the process \( \{X_{t,N} : t = 1,2,\ldots,N\} \), whereas asymptotic normality would require the existence of its eighth moment. In other words, to show asymptotic normality of the local least-squares estimator we would require the condition \( \{\mathbb{E}(Z_t^8)\}^{1/4} \sup_{u} \sum_{j=1}^{p} a_j(u) \leq 1 - \delta \) for some \( \delta > 0 \), which is a very strong condition to place on the parameters. In contrast, the kernel-QML does not require such high-order moment conditions. Thus, we adapt the local least-squares method to include a normalisation term. The normalisation term allows us to obtain similar asymptotic results to the kernel-QML estimator, under the same moment conditions.

The main advantage of the local NLS criterion is that it has a tractable solution, unlike the solution to the local QML criterion which does not have an explicit form. We expand on this below.
3.1 The kernel-NLS estimator

In this section, we formally define the kernel-NLS estimator. Let $\| \cdot \|_1$ be the $\ell_1$-norm of a vector, and let $\mathcal{X}_{t,N}^T = (X_{t,N}^2, \ldots, X_{t-p+1,N}^2)$. The squares of the tvARCH($p$) process satisfy

$$X_{t,N}^2 = a_0 \left( \frac{t}{N} \right) + \sum_{j=1}^{p} a_j \left( \frac{t}{N} \right) X_{t-j,N}^2 + (Z_t^2 - 1) \sigma_{t,N}^2. \quad (12)$$

Normalising (12) with $(g + \|\mathcal{X}_{t-1,N}\|_1)$, we have the following representation

$$\frac{X_{t,N}^2}{(g + \|\mathcal{X}_{t-1,N}\|_1)} = \frac{a_0 \left( \frac{t}{N} \right) \left( g + \|\mathcal{X}_{t-1,N}\|_1 \right)}{(g + \|\mathcal{X}_{t-1,N}\|_1)} + \sum_{j=1}^{p} a_j \left( \frac{t}{N} \right) \frac{X_{t-j,N}^2}{(g + \|\mathcal{X}_{t-1,N}\|_1)} + \frac{(Z_t^2 - 1) \sigma_{t,N}^2}{(g + \|\mathcal{X}_{t-1,N}\|_1)}, \quad (13)$$

where $-1 < g < \infty$ is an arbitrary constant. Expression (13) motivates the following local NLS criterion

$$\mathcal{L}_{t_0,N}^{nls}(g(u)) = \sum_{k=p+1}^{N} \frac{1}{bN} W \left( \frac{t_0 - k}{bN} \right) h_{k,N}^{nls}(g(u)), \quad (14)$$

where

$$h_{k,N}^{nls}(g(u)) = \frac{1}{(g + \|\mathcal{X}_{k-1,N}\|_1)^2} \left\{ X_{k,N}^2 - a_0 \left( \frac{k}{N} \right) - \sum_{j=1}^{p} a_j \left( \frac{k}{N} \right) X_{k-j,N}^2 \right\}^2. \quad (15)$$

Typically, we would let $g = 0$ but, for reasons that will be clear later, we define the kernel-NLS criterion for general $g$. If $|u_0 - t_0/N| < 1/N$, we use $\hat{a}_{t_0,N}^{nls(g)}$ as an estimator of $a(u_0)$, where

$$\hat{a}_{t_0,N}^{nls(g)} = \text{argmin}_{a} \mathcal{L}_{t_0,N}^{nls}(g). \quad (16)$$

We call $\hat{a}_{t_0,N}^{nls(g)}$ the kernel-NLS($g$) estimator.

The significant advantage of $\hat{a}_{t_0,N}^{nls(g)}$ is that it can be explicitly written as

$$\hat{a}_{t_0,N}^{nls(g)} = \left\{ R_{t_0,N}^{nls(g)} \right\}^{-1} \sum_{t_0,N}^{nls(g)},$$

where

$$R_{t_0,N}^{nls(g)} = \sum_{k=p+1}^{N} \frac{1}{bN} W \left( \frac{t_0 - k}{bN} \right) \frac{X_{k-1,N}X_{k-1,N}^T}{(g + \|\mathcal{X}_{k-1,N}\|_1)^2},$$

$$\sum_{t_0,N}^{nls(g)} = \sum_{k=p+1}^{N} \frac{1}{bN} W \left( \frac{t_0 - k}{bN} \right) \frac{X_{k,N}^2X_{k-1,N}}{(g + \|\mathcal{X}_{k-1,N}\|_1)^2}.$$

Remark 3.1 We note that, unlike the kernel-QML estimator considered in Dahlhaus & Subba Rao (2003), for the kernel-NLS($g$) estimator we do not assume that there exists a constant $\rho > 0$ such that $\inf_{u_j} a_j(u) > \rho$ for $j = 1, 2, \ldots, p - 1$ (see Assumption 2.1(ia)). In other words, if $j \in \{1, 2, \ldots, p - 1\}$, then the case $a_j(u) = 0$ is allowed.
3.2 Asymptotic properties of the kernel-NLS estimator

We now consider the consistency and asymptotic normality of $\hat{\alpha}^{nls}_{t_0,N}$. We first show consistency of $\hat{\alpha}^{nls}_{t_0,N}$.

**Proposition 3.1** Suppose $\{X_{t,N} : t = 1, \ldots, N\}$ is a tvARCH($p$) process which satisfies Assumptions 2.1(i,a,ii), and let $\hat{\alpha}^{nls}_{t_0,N}$ be defined as in (16). If $|u_0 - t_0| < 1/N$, then we have that

$$\hat{\alpha}^{nls}_{t_0,N} \overset{p}{\rightarrow} \alpha(u_0), \text{ with } b \rightarrow 0, bN \rightarrow \infty \text{ as } N \rightarrow \infty.$$ 

Let us define the following quantities

$$A^{nls}(u) = \mathbb{E}\left\{ \tilde{X}_{k-1}(u)\tilde{X}_k^T(\cdot) \right\}, \quad \Sigma^{nls}(u) = \mathbb{E}\left\{ \tilde{\sigma}_k^4(u)\tilde{X}_{k-1}(u)\tilde{X}_k^T(\cdot) \right\},$$

and let

$$B^{nls}_{t_0,N}(\alpha(u)) = \sum_{k=p+1}^{N} \frac{1}{bN} W\left( \frac{t_0 - k}{bN} \right) \left[ h^{nls}_{k,N}(\alpha(u)) - h^{nls}_{k}(\alpha(u),\alpha(u)) \right],$$

where

$$h^{nls}_k(u,\alpha(u)) = \frac{1}{(g + ||\tilde{X}_{k-1}(u)||^2)} \left\{ \tilde{X}_k^2(u) - a_0(u) - \sum_{j=1}^{p} a_j(u)\tilde{X}_k(u) \right\}^2.$$ 

We now show asymptotic normality of $\hat{\alpha}^{nls}_{t_0,N}$.

**Proposition 3.2** Suppose $\{X_{t,N} : t = 1, \ldots, N\}$ is a tvARCH($p$) process which satisfies Assumptions 2.1(i,a,ii,iii), and let $\hat{\alpha}^{nls}_{t_0,N}$, $A^{nls}(u)$, $\Sigma^{nls}(u)$ and $B^{nls}_{t_0,N}(\alpha(u))$ be defined as in (16), (17), (18) and (19) respectively. If $|u_0 - t_0/N| < 1/N$, then we have that

(i)

$$\nabla B^{nls}_{t_0,N}(\alpha(u_0)) = O_p(b^3), \text{ with } b \rightarrow 0, bN \rightarrow \infty \text{ as } N \rightarrow \infty. \quad (21)$$

(ii)

$$\sqrt{bN} \left( \hat{\alpha}^{nls}_{t_0,N} - \alpha(u_0) \right) + \frac{1}{2}\sqrt{bN} \mathbb{E}\left\{ A^{nls}(u_0) \right\}^{-1} \nabla B^{nls}_{t_0,N}(\alpha(u_0)) \overset{p}{\rightarrow} \mathcal{N}(0, w_2 \mu_4 \mathbb{E}\left\{ A^{nls}(u_0) \right\}^{-1} \Sigma^{nls}(u_0) \mathbb{E}\left\{ A^{nls}(u_0) \right\}^{-1}), \quad (22)$$

with $b \rightarrow 0, bN \rightarrow \infty$ as $N \rightarrow \infty$, where $w_2 = \int_{-1/2}^{1/2} W^2(x)dx$ and $\mu_4 = \text{var} \left(Z_t^2\right)$.
Remark 3.2  (i) We note that $\nabla B_{t_0,N}^{nls(g)}(a(u))$ is the bias due to nonstationarity of the tvARCH($p$) process. Under stronger regularity assumptions on the time-varying parameters $\{a_j(u) : j = 0,1,\ldots,p\}$ (existence of derivatives of $a_j(u)$ for $j = 0,1,\ldots,p$) and higher moment conditions on the tvARCH($p$) process, we can obtain better rates of convergence for the bias. Since these rates are easily obtained by following the steps used in the proof of Proposition 4.2 in Dahlhaus & Subba Rao (2003), we do not elaborate on this.

(ii) Comparing (11) and (22), we immediately see that the kernel-QML and kernel-NLS($g$) estimators asymptotically have the same order of convergence. That is, if $|u_0 - t_0/N| < 1/N$, then we have that
\[
|\hat{a}_{u,N}^{mle} - a(u_0)| = O_p \left( b^2 + \frac{1}{(bN)^{1/2}} \right) \quad \text{and} \quad |\hat{a}_{u,N}^{nls(g)} - a(u_0)| = O_p \left( b^2 + \frac{1}{(bN)^{1/2}} \right),
\]
with $b \to 0$, $bN \to \infty$ as $N \to \infty$.

(iii) If the time-varying parameters $\{a_j(u) : j = 0,1,\ldots,p\}$ were constant over $u \in (0,1]$ for all $j = 0,1,\ldots,p$, then, obviously, $\nabla B_{t_0,N}^{nls(g)}(a(u)) = 0$. This yields the stationary ARCH($p$) process, which we consider in more detail in Section 5.

3.3 Drawbacks of the kernel-NLS estimator

As mentioned before, the kernel-NLS($g$) estimator possesses several attractive features, such as explicit form, ease of computation, and consistency under mild moment conditions. However, it also suffers from a number drawbacks, some of which are identified below. Two points seem to be of particular importance here:

Lack of scaling invariance. For simplicity, let us suppose that $\{X_t : t = 1,2,\ldots\}$ is a stationary ARCH($p$) process as defined in (35), with parameters $(a_0,a_1,\ldots,a_p)$. Let $Y_t = cX_t$, for $c > 0$. By Volterra expansion arguments (see, for example, Giraitis et al., 2003), $Y_t$ is a stationary ARCH($p$) process with parameters $(c^2a_0,a_1,\ldots,a_p)$. Thus, the rescaling of $X_t$ corresponds to a change in the $a_0$ parameter only, without affecting the other parameters. Therefore, it is desirable that a good estimate of $(a_1,\ldots,a_p)$ should be independent of the choice of $a_0$. This is, however, not true for the NLS($g$) estimator, as can easily be demonstrated by comparing the NLS($g$) estimators of $(a_1,\ldots,a_p)$ computed for $X_t$ and $cX_t$. Roughly speaking, this is due to the fact that the normalising factor $(g+1+\sum_{j=1}^{p}X_{t-j}^2)$ is not a linear functional of $\{X_t^2 : t = 1,2,\ldots\}$.

Behaviour for small $a_0$. Let us assume again that $\{X_t : t = 1,2,\ldots\}$ is a stationary ARCH($p$) process with parameters $(a_0,a_1,\ldots,a_p)$. Heuristically speaking, since the $a_0$ parameter corresponds to the “scaling” of $X_t$, one can expect that a small value of $a_0$ would result in $X_t^2$ being, on average, small. This in turn would cause the normalising factor $(g+1+\sum_{j=1}^{p}X_{t-j}^2)$ to be close to $g+1$. However, normalising the estimator by a factor which is “almost” independent of the data would reduce the NLS($g$) estimator to one which is close to the classical
least-squares estimator. This is undesirable: as was already pointed out, the classical least-squares estimator is an unappealing choice for ARCH($p$) processes due to the strong moment conditions required, as well as its poor practical performance.

To further substantiate these heuristic arguments, we now describe the outcome of a small simulation study, aimed at comparing the performance of the NLS(0) and QML estimators on stationary data. Each row of three plots in Figure 1 corresponds to a fixed triple of parameters $(a_0, a_1, a_2)$: their values are given in the titles. For each triple, we simulated 100 sample paths of length 1000 from a stationary ARCH(2) process with standard Gaussian innovations $Z_t$. In each plot, the left-hand boxplot shows the empirical distribution of the estimated values of the parameter specified in the title of the plot, computed using the NLS(0) estimator. The middle boxplot shows the analogous quantity for the QML estimator (the \texttt{garch} function from the S-Plus \texttt{garch} module was used). The dashed line indicates the true value of the given parameter.

It is interesting to observe that the NLS(0) estimator is positively biased for $a_0$ but negatively biased for $a_1$ and $a_2$ when $a_0$ is small. On the other hand, the NLS(0) estimator has a large variance when $a_0$ is large.

As a digression, observe the “unstable” behaviour of the QML estimator when $a_1$ is close to zero: note the heavy tails of the empirical distribution of the QML estimators of $a_0$ and $a_2$ in the top three rows. This is as expected, since in theory, the QML estimator requires that $a_j$ should be bounded away from zero for all $j = 0, 1, \ldots, p$ (see Assumption 2.1(ib)).

The above discussion of the properties of the NLS estimator seems to indicate the need for a data-driven choice of $g$ (or, alternatively, $1 + g$) in the NLS procedure. Indeed, in the following section, we propose a two-stage modification of the kernel-NLS estimation scheme for tvARCH($p$) processes: in the first step, $1 + g$ is estimated from the data; in the second step, the resulting estimate is used in the normalisation factor. In each boxplot in Figure 1, the right-hand column shows the empirical distribution of estimates obtained using this modified two-stage procedure, and clearly demonstrates its good performance. The large bias and/or variance, exhibited by the NLS(0) estimator, is not displayed by the modified two-step scheme. At the same time, the new procedure is stable wherever QML is not.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure1.png}
\caption{Figure 1 about here}
\end{figure}

4 The two-stage kernel-NLS estimator and its asymptotic properties

In this section, we introduce a two-stage local NLS estimation scheme, based on the arguments of Section 3.3, for estimating the tvARCH($p$) parameters. We also derive the consistency and asymptotic normality of the corresponding two-stage kernel-NLS($g$) estimator.
4.1 The two-stage kernel-NLS estimator

The discussion in Section 3.3 indicates the need for a data-driven choice of $1 + g$, possibly in such a way that $1 + g$ is “nearly proportional” to $a_0(u)$ (for example, by being a linear functional of $\{X^2_{t,N} : t = 1, 2, \ldots, N\}$). An estimate of $a_0(u)$ would be a natural candidate; however, to obtain a simple “raw” estimator of $a_0(u)$ from the observations $\{X_{t,N} : t = 1, 2, \ldots, N\}$ is not easy. Instead, we estimate $\mu(u) = \mathbb{E}\{X^2_t(u)\}$, which is easier to estimate and is of the same order as $a_0(u)$, since it satisfies the equation

$$\mu(u) = a_0(u) + \sum_{j=1}^{p} a_j(u)\mu(u).$$

Therefore, $\mu(u) = a_0(u)/\{1 - \sum_{j=1}^{p} a_j(u)\} > a_0(u)$. An estimate of $\mu(u)$ is easy to obtain, by evaluating a weighted average of $\{X^2_{t,N} : t = 1, 2, \ldots, N\}$ about the time point $[uN]$ (where $[x]$ denotes the largest integer less than or equal to $x$). That is, suppose $|u_0 - t_0/N| < 1/N$, then we use $\hat{\mu}_{t_0,N}$ as an estimator of $\mu(u_0)$, where

$$\hat{\mu}_{t_0,N} = \sum_{k=1}^{N} \frac{1}{bN} W(t_0 - k/bN) X^2_{k,N}. \quad (23)$$

Having obtained the pre-estimator $\hat{\mu}_{t_0,N}$, we minimise a modification of the local NLS criterion to obtain an estimator of the parameter vector $\mathcal{g}(u_0)$. We summarise the above estimation scheme in the following two-steps.

(i) Evaluate the pre-estimator $\hat{\mu}_{t_0,N}$, given in (23), of $\mu(u_0)$;

(ii) Let

$$L^{nls}(\hat{\mu}_{t_0,N}^{-1})(\mathcal{g}(u)) = \sum_{k=p+1}^{N} \frac{1}{bN} W(t_0 - k/bN) h^{nls}(\hat{\mu}_{t_0,N}^{-1})(\mathcal{g}(u)), \quad (24)$$

where

$$h^{nls}_{k,N}(\mathcal{g}(u)) = \frac{1}{(\hat{\mu}_{t_0,N}^{-1} + \|X_{k-1,N}\|)^2} \left\{ X^2_{k,N} - a_0\left(\frac{k}{N}\right) - \sum_{j=1}^{p} a_j\left(\frac{k}{N}\right) X^2_{k-j,N} \right\}^2. \quad (25)$$

If $|u_0 - t_0/N| < 1/N$, we use $\hat{\mu}_{t_0,N}^{-1}$ as an estimator of $\mathcal{g}(u_0)$, where

$$\hat{\mathcal{g}}^{nls}_{t_0,N} = \arg\min_{\mathcal{g}} L^{nls}_{t_0,N} \quad (26)$$

We call $\hat{\mathcal{g}}^{nls}_{t_0,N}$ the two-stage kernel-NLS$(\hat{\mu}_{t_0,N}^{-1})$ estimator.

Again, the significant advantage of $\hat{\mathcal{g}}^{nls}_{t_0,N}$ is that it can be explicitly written as

$$\hat{\mathcal{g}}^{nls}_{t_0,N} = \left\{ \mathcal{R}^{nls}_{t_0,N} \right\}^{-1} \mathcal{S}^{nls}_{t_0,N}.$$
Lemma 4.2

\[ R_{t_0,N}^{nls}(\mu_{t_0,N}^{-1}) = \sum_{k=p+1}^{N} \frac{1}{b_N} W \left( \frac{t_0 - k}{b_N} \right) \frac{X_{k-1,N} \gamma_{k-1,N}}{(\mu_{t_0,N} - 1 + \|X_{k-1,N}\|_1)^2}, \]  

(27)

In the following section, we show that the asymptotic properties of \( \hat{\mu}_{t_0,N}^{nls} \) are “close” to those of \( \hat{\mu}_{t_0,N}^{nls}(\mu(u))^{-1} \) (except that their biases are only asymptotically equivalent). In other words, the two-stage local NLS estimation scheme gives us an estimator that behaves almost as if \( \mu(u) \) was known \text{ a priori } and used in the estimate.

4.2 Asymptotic properties of the two-stage kernel-NLS estimator

In this section, we consider consistency and asymptotic normality of \( \hat{\mu}_{t_0,N}^{nls} \). To prove these asymptotic properties, we need the following lemmas.

Lemma 4.1 Suppose \( \{X_{t,N} : t = 1, \ldots, N\} \) is a tvARCH(p) process which satisfies Assumptions 2.1(ia,ii,iv), let \( \mu(u) = E\{\tilde{X}_i^2(u)\} \), and let \( \hat{\mu}_{t_0,N} \) be defined as in (23). If \( |u_0 - t_0/N| < 1/N \), then we have that

\[ |\hat{\mu}_{t_0,N} - \mu(u_0)| = O_P \left( b^3 + (bN)^{\gamma/\tau} \right), \text{ with } b \to 0, bN \to \infty \text{ as } N \to \infty. \]  

(29)

Lemma 4.2 Suppose \( \{X_{t,N} : t = 1, \ldots, N\} \) is a tvARCH(p) process which satisfies Assumptions 2.1(ia,ii,iv), let \( \mu(u) = E\{\tilde{X}_i^2(u)\} \), and let \( \hat{\mu}_{t_0,N} \) be defined as in (23). If \( |u_0 - t_0/N| < 1/N \), then, for \( 0 \leq i, j \leq p \), we have that

\[ \sum_{k=p+1}^{N} \frac{1}{b_N} W \left( \frac{t_0 - k}{b_N} \right) \frac{X_{k-i,N}^2 \gamma_{k-j,N}}{(\mu_{t_0,N} - 1 + \|X_{k-1,N}\|_1)^2} \overset{P}{\to} E \left( \frac{\tilde{X}_{k-i,N}(u_0) \tilde{X}_{k-j,N}(u_0)}{(\mu(u_0) - 1 + \|X_{k-1}(u_0)\|_1)^2} \right), \]  

(30)

with \( b \to 0, bN \to \infty \text{ as } N \to 0. \)

We first show consistency of \( \hat{\mu}_{t_0,N}^{nls} \).

Proposition 4.1 Suppose \( \{X_{t,N} : t = 1, \ldots, N\} \) is a tvARCH(p) process which satisfies Assumptions 2.1(ia,ii,iv), and let \( \hat{a}_{t_0,N}^{nls}(\hat{\mu}_{t_0,N}^{-1}) \) and \( \hat{\mu}_{t_0,N} \) be defined as in (26) and (23) respectively. If \( |u_0 - t_0/N| < 1/N \), then we have that

\[ \hat{a}_{t_0,N}^{nls}(\mu(u_0)) \overset{P}{\to} a(u_0), \text{ with } b \to 0, bN \to \infty \text{ as } N \to \infty. \]

Let us define the following quantity

\[ B_{t_0,N}^{nls}(\mu_{t_0,N}^{-1})(a(u)) = \sum_{k=p+1}^{N} \frac{1}{b_N} W \left( \frac{t_0 - k}{b_N} \right) \left[ h_{k,N}^{nls}(\mu_{t_0,N}^{-1})(a(u)) - h_{k,N}^{nls}(\mu_{t_0,N}^{-1})(u, a(u)) \right], \]  

(31)
where

\[ h_{k}^{nls(\hat{\mu}_{t0, N} - 1)}(u, \bar{a}(u)) = \frac{1}{(\hat{\mu}_{t0, N} - 1 + \|X_{k-1}(u)\|_1)^2} \left( \hat{X}_{k}(u) - a_0(u) - \sum_{j=1}^{p} a_j(u) \hat{X}_{k-j}(u) \right)^2. \] (32)

We now show asymptotic normality of \( \bar{a}_{t0, N}^{nls(\hat{\mu}_{t0, N} - 1)} \).

**Proposition 4.2** Suppose \( \{X_t, t = 1, \ldots, N\} \) is a tvARCH\((p)\) process which satisfies Assumptions 2.1(ia, ii, iii, iv), let \( \mu(u) = \mathbb{E}\{\hat{X}_{t}^{2}(u)\} \), and let \( \hat{\mu}_{t0, N}, \bar{a}_{t0, N}^{nls(\mu(u) - 1)}, A^{nls(\mu(u) - 1)}(u), \Sigma^{nls(\mu(u) - 1)}(u) \) and \( B_{t0, N}^{nls(\mu_{t0, N} - 1)}(\bar{a}(u)) \) be defined as in (23), (26), (17), (18) and (31) respectively. If \( |u_0 - t_0/N| < 1/N \), then we have that

(i)

\[ \nabla B_{t0, N}^{nls(\mu_{t0, N} - 1)}(\bar{a}(u_0)) = O_p(b^3), \text{ with } b \to 0, bN \to \infty \text{ as } N \to \infty. \] (33)

(ii)

\[ \sqrt{bN} \{ \bar{a}_{t0, N}^{nls(\mu_{t0, N} - 1)} - \bar{a}(u_0) \} + \frac{1}{\sqrt{bN}} \{ A^{nls(\mu(u_0) - 1)}(u_0) \}^{-1} \nabla B_{t0, N}^{nls(\mu_{t0, N} - 1)}(\bar{a}(u_0)) \]

\[ \overset{D}{\to} N(0, \frac{2}{\mu_{t0, N}} \{ A^{nls(\mu(u_0) - 1)}(u_0) \}^{-1} \Sigma^{nls(\mu(u_0) - 1)}(u_0) \{ A^{nls(\mu(u_0) - 1)}(u_0) \}^{-1}), \] (34)

with \( b \to 0, bN \to \infty \) as \( N \to \infty \), where \( w_2 = \int_{-1/2}^{1/2} W^2(z) dz \) and \( \mu_4 = \text{var} \{ Z^2 \} \).

**Remark 4.1** (i) We observe that the two-stage kernel-NLS(\( \hat{\mu}_{t0, N} - 1 \)) estimator is “close” to the kernel-NLS(\( \mu(u) - 1 \)) estimator, where \( \mu(u) \) is known. The difference lies in that their biases are only asymptotically equivalent, in the sense that they asymptotically have the same order of convergence.

(ii) Comparing (22) and (34), and taking into account the discussion above, we immediately see that the kernel-NLS(\( \mu(u) - 1 \)) and two-stage kernel-NLS(\( \hat{\mu}_{t0, N} - 1 \)) estimators asymptotically have the same order of convergence. That is, if \( |u_0 - t_0/N| < 1/N \), then we have that

\[ |\bar{a}_{t0, N}^{nls(\mu(u) - 1)} - \bar{a}(u_0)| = O_p \left( b^\beta + \frac{1}{(bN)^{1/2}} \right) \] and

\[ |\bar{a}_{t0, N}^{nls(\hat{\mu}_{t0, N} - 1)} - \bar{a}(u_0)| = O_p \left( b^\beta + \frac{1}{(bN)^{1/2}} \right), \]

with \( b \to 0, bN \to \infty \) as \( N \to \infty \).

## 5 Estimation for the stationary ARCH(p) process

In this section, we consider the stationary ARCH(p) process, which is a member of the class of tvARCH\((p)\) processes discussed earlier, where the time-varying parameters \( \{a_j(u) : j = 0, 1, \ldots, p\} \) are now constant over \( u \in (0, 1] \) for all \( j = 0, 1, \ldots, p \). We show that both the kernel-NLS(\( g \)) and the two-stage kernel-NLS(\( \hat{\mu}_{t0, N} - 1 \)) estimators can be adapted to estimate the parameters of a stationary ARCH(p) process. However, because the process is stationary, we can use the whole set
Assumption 5.1

Consider the following NLS criterion

\[ X_t = \sigma(t)Z_t, \quad \sigma^2(t) = a_0 + \sum_{j=1}^{p} a_j X_{t-j}^2, \quad \text{for} \quad t = 1, 2, \ldots, \]

where \( \sum_{j=1}^{p} a_j < 1 \) and \( \{Z_t : t = 1, 2, \ldots\} \) are independent and identically distributed random variables with \( \mathbb{E}(Z_t) = 0 \) and \( \mathbb{E}(Z_t^2) = 1 \). Let \( X_t^T = (1, X_t^2, \ldots, X_t^{2p+1}) \), and let \( \mu = \mathbb{E}(X_t^2) \).

Assumption 5.1 Suppose \( \{X_t : t = 1, 2, \ldots\} \) is a stationary ARCH\((p)\) process. We assume that the parameters \( \{a_j : j = 0, 1, \ldots, p\} \) and the innovation process \( \{Z_t : t = 1, 2, \ldots\} \) satisfy the following conditions:

(i) There exist \( 0 < \rho_1 \leq \rho_2 < \infty \) and \( \delta > 0 \) such that \( \rho_1 \leq a_0 \leq \rho_2, \rho_1 \leq a_p \) and \( \sum_{j=1}^{p} a_j \leq 1 - \delta \).

(ii) For some \( \gamma > 0 \), \( \mathbb{E}\left( Z_t^{k(1+\gamma)} \right) < \infty \);

(iii) For some \( 0 < \eta \leq 1 \) and \( \delta > 0 \), \( \mathbb{E}\left( Z_t^{2(1+\eta)} \right)^{1/(1+\eta)} \sum_{j=1}^{p} a_j \leq 1 - \delta \).

5.1 The NLS estimator and its asymptotic properties

Consider the following NLS criterion

\[ \mathcal{L}_N^{nls(g)}(\theta) = \sum_{k=p+1}^{N} \left( \frac{1}{(g + \|X_{k-1}\|_1)^2} \left( X_k^2 - \theta_0 - \sum_{j=1}^{p} \theta_j X_{k-j}^2 \right) \right)^2, \]

where \(-1 < g < \infty\) is an arbitrary constant. We use \( \hat{\theta}_N^{nls(g)} \) as an estimator of \( \theta \), where

\[ \hat{\theta}_N^{nls(g)} = \arg \min_{\theta} \mathcal{L}_N^{nls(g)}(\theta), \]

and call \( \hat{\theta}_N^{nls(g)} \) the NLS\((g)\) estimator. We note that the \( \hat{\theta}_N^{nls(0)} \) estimator was also recently discussed by Horváth & Liese (2004).

The estimator \( \hat{\theta}_N^{nls(g)} \) can be explicitly written as

\[ \hat{\theta}_N^{nls(g)} = \left\{ \mathcal{R}_N^{nls(g)} \right\}^{-1} \mathcal{L}_N^{nls(g)}, \]

where

\[ \mathcal{R}_N^{nls(g)} = \frac{1}{N-p-1} \sum_{k=p}^{N-1} \frac{X_k X_k^T}{(g + \|X_{k-1}\|_1)^2}, \]

\[ \mathcal{L}_N^{nls(g)} = \frac{1}{N-p-1} \sum_{k=p+1}^{N} \frac{X_k^2 X_{k-1}}{(g + \|X_{k-1}\|_1)^2}. \]
Let us define the following quantities

\[ A^{nls}(g) = \mathbb{E} \left( \frac{X_{k-1}X_{k-1}^T}{(g + \|X_{k-1}\|_1)^2} \right), \]  \hspace{1cm} (37)

\[ \Sigma^{nls}(g) = \mathbb{E} \left( \frac{\sigma^4_k X_{k-1}X_{k-1}^T}{(g + \|X_{k-1}\|_1)^4} \right). \]  \hspace{1cm} (38)

We now discuss consistency and asymptotic normality of \( \hat{a}^{nls}(g) \). Since these asymptotic properties follow easily from the results of Section 3.2, we state them without proof.

**Proposition 5.1** Suppose \( \{X_t : t = 1, 2, \ldots \} \) is a stationary ARCH(p) process which satisfies Assumption 5.1(i), and let \( \hat{a}^{nls}(g) \), \( A^{nls}(g) \) and \( \Sigma^{nls}(g) \) be defined as in (36), (37) and (38) respectively. Then, we have that

(i) \( \hat{a}^{nls}(g) \xrightarrow{P} a \), with \( N \to \infty \).

(ii) In addition, if we assume that Assumption 5.1(ii) holds, then we have that

\[ \sqrt{N} \{ \hat{a}^{nls}(g) - a \} \xrightarrow{D} \mathcal{N} \left( 0, \mu_4 \{ A^{nls}(g) \}^{-1} \Sigma^{nls}(g) \{ A^{nls}(g) \}^{-1} \right), \]  \hspace{1cm} with \( N \to \infty \),

where \( \mu_4 = \text{var} \left( Z_t^2 \right) \).

### 5.2 The two-stage NLS estimator and its asymptotic properties

We summarise this estimator in the following two steps.

(i) Evaluate the pre-estimator \( \hat{\mu}_N \) of \( \mu \), where

\[ \hat{\mu}_N = \frac{1}{N} \sum_{k=1}^{N} X_k^2; \]  \hspace{1cm} (39)

(ii) Let

\[ L^{nls}(\hat{\mu}_N^{-1})(a) = \sum_{k=p+1}^{N} \frac{1}{(\hat{\mu}_N - 1 + \|X_{k-1}\|_1)^2} \left\{ X_k^2 - a_0 - \sum_{j=1}^{p} a_j X_{k-j}^2 \right\}^2. \]

We use \( \hat{a}^{nls}(\hat{\mu}_N^{-1}) \) as an estimator of \( a \), where

\[ \hat{a}^{nls}(\hat{\mu}_N^{-1}) = \arg\min_a L^{nls}(\hat{\mu}_N^{-1})(a), \]  \hspace{1cm} (40)

and call \( \hat{a}^{nls}(\hat{\mu}_N^{-1}) \) the two-stage NLS(\( \hat{\mu}_N - 1 \)) estimator.

Again, the estimator \( \hat{a}^{nls}(\hat{\mu}_N^{-1}) \) can be explicitly written as

\[ \hat{a}^{nls}(\hat{\mu}_N^{-1}) = \left( \mathcal{R}^{nls}(\hat{\mu}_N^{-1}) \right)^{-1} nls(\hat{\mu}_N^{-1}) \]
where

\[
R^{nls}(\hat{\mu}_N^{-1}) = \frac{1}{N - p - 1} \sum_{k=p}^{N-1} \frac{X_k A_k^T}{(\hat{\mu}_N - 1 + \|A_k\|_2)^2},
\]

\[
\Sigma^{nls}(\hat{\mu}_N^{-1}) = \frac{1}{N - p - 1} \sum_{k=p+1}^{N} \frac{X_k^2 A_k^{-1}}{(\hat{\mu}_N - 1 + \|A_k^{-1}\|_2)^2}.
\]

We now discuss consistency and asymptotic normality of \(\hat{\mu}_N^{nls}(\hat{\mu}_N^{-1})\). Since these asymptotic properties follow easily from the results of Section 4.2, we state them without proof.

**Proposition 5.2** Suppose \(\{X_t : t = 1, 2, \ldots\}\) is a stationary ARCH\((p)\) process which satisfies Assumption 5.1(i,iii), let \(\mu = \mathbb{E}(X_t^2)\), and let \(\hat{\mu}_N, \hat{\mu}_N^{nls}(\hat{\mu}_N^{-1}), A^{nls}(\mu^{-1})\) and \(\Sigma^{nls}(\mu^{-1})\) be defined as in (39), (40), (37) and (38) respectively. Then, we have that

(i) \(\hat{\mu}_N^{nls}(\hat{\mu}_N^{-1}) \xrightarrow{P} \mu\) with \(N \to \infty\).

(ii) In addition, if we assume that Assumption 5.1(ii) holds, then we have that

\[
\sqrt{N}(\hat{\mu}_N^{nls}(\hat{\mu}_N^{-1}) - \mu) \xrightarrow{D} \mathcal{N}(0, \mu_4 \{A^{nls}(\mu^{-1})\}^{-1} \Sigma^{nls}(\mu^{-1}) \{A^{nls}(\mu^{-1})\}^{-1}),\]

with \(N \to \infty\), where \(\mu_4 = \text{var}(Z_t^2)\).

### 6 Volatility estimation and forecasting for exchange rate datasets

In this section, we describe a numerical study whereby the forecasting ability of the \(\text{tvARCH}(p)\) model is compared to that of the benchmark stationary GARCH(1,1) process with standard Gaussian innovations. Suppose that we observe \(X_{1,N}, X_{2,N}, \ldots, X_{t,N}\) from a \(\text{tvARCH}(p)\) process, and want to forecast \(\sigma_{t+1,N}^2, \ldots, \sigma_{t+h,N}^2\) where \(t + h \leq N\). Denote the Mean Square Error-optimal forecasts by \(\sigma_{t+h,N}^{2,\text{tvARCH}(p)}\). If the process \(\{X_{t,N} : t = 1, 2, \ldots, N\}\) was stationary, then by classical theory (see, for example, Bera & Higgins, 1993), the forecasts would be obtained as

\[
\sigma_{t+h,N}^{2,\text{ARCH}(p)} = \sigma^2 + \left(\sum_{j=1}^{p} a_j \right)^h (\sigma_{t,N}^2 - \sigma^2),
\]

where \(\sigma^2 = a_0/(1 - \sum_{j=1}^{p} a_j)\) was the unconditional variance. Suppose now that the process \(\{X_{t,N} : t = 1, 2, \ldots, N\}\) is non-stationary. In general, in order to obtain the forecasts, estimates of the time-varying parameters \(\{a_j(u) : j = 0, 1, \ldots, p\}\) would have to be extrapolated on the interval \([t/N, 1]\). However, to simplify matters, we assume that the true time-varying parameters \(\{a_j(u) : j = 0, 1, \ldots, p\}\) are constant for \(u \in I = [t/N - \delta, 1]\) for \(\delta > 0\). In this simplified setup, it is easy to see that formula (41) generalises to

\[
\sigma_{t+h,N}^{2,\text{tvARCH}(p)} = \sigma^2(t/N) + \left(\sum_{j=1}^{p} a_j(t/N) \right)^h (\sigma_{t,N}^2 - \sigma^2(t/N)),
\]

In this simplified setup, it is easy to see that formula (41) generalises to
with \( \sigma^2(t/N) = a_0(t/N)/(1 - \sum_{j=1}^{p} a_j(t/N)) \).

In practice, we estimate the quantities \( \{a_j(t/N) : j = 0, 1, \ldots, p\} \) in formula (42) by \( \hat{a}_{nls}(\hat{\mu}_{tb,N} - 1) \), where \( t_b = t - \lfloor bN/2 \rfloor \). For \( bN \) and \( N \) sufficiently large, this is valid due to the \( a_j(u) \)'s being constant in \( I \) for all \( j = 0, 1, \ldots, p \). Note that \( t_b \) is the "last" time point for which the two-stage kernel-NLS(\( \hat{\mu}_{tb,N} - 1 \)) estimator can be computed without the support of the kernel extending beyond the domain of the observed dataset.

Below, we use several exchange rate datasets (details are given below) to test the long-term forecasting ability of the tvARCH(\( p \)) model against the benchmark stationary GARCH(1,1) model with standard Gaussian innovations. For the tvARCH(\( p \)) model, we take \( p = 1 \) for simplicity, and use the forecasting procedure described above with a rectangular kernel and bandwidth \( bN = 250 \).

For the stationary GARCH(1,1) prediction, we use the standard S-Plus \texttt{garch} and \texttt{predict} routines to obtain forecasts which we denote here by \( \hat{\sigma}^2_{t|t+h} \), \( \hat{\sigma}^2_{t|t+250} \), \( \hat{\sigma}^2_{t|t+250} \)

and compare them to the "realised" volatility

\[
\bar{\sigma}^2_{t|t+250,N} = \sum_{h=1}^{250} \sigma^2_{t|t+h,N} \]

\[
\bar{\sigma}^2_{t|t+250} = \sum_{h=1}^{250} \sigma^2_{t|t+h} \]

using the scaled Aggregated Mean Square Error (AMSE)

\[
\text{AMSE}^{\text{tvARCH}(1)}_{250,1000,N} = \sum_{t=1000}^{N-250} \left( \sigma^2_{t|t+250,N} - \bar{\sigma}^2_{t|t+250,N} \right)^2 \]

\[
\text{AMSE}^{\text{GARCH}(1,1)}_{250,1000,N} = \sum_{t=1000}^{N-250} \left( \sigma^2_{t|t+250} - \bar{\sigma}^2_{t|t+250} \right)^2 \]

where the scaling is by the factor of \( 1/(N - 1000) \). The stationary GARCH(1,1) parameters are re-estimated for each \( t \). Our datasets are logged and differenced daily exchange rates between USD and a number of other currencies running from 01/01/1990 to 31/12/1999: the data are available from the US Federal Reserve website

\url{http://www.federalreserve.gov/releases/h10/Hist/default1999.htm}

In the sequel, we use the following acronyms: CHF (Switzerland Franc), GBP (United Kingdom Pound), HKD (Hong Kong Dollar), JPY (Japan Yen), NOK (Norway Kroner), NZD (New Zealand Dollar), SEK (Sweden Kronor), TWD (Taiwan New Dollar). Table 1 lists the AMSEs attained by the proposed tvARCH(1) and stationary GARCH(1,1) models: the better results are boxed.
For most examples considered, the proposed tvARCH(1) model significantly outperformed the stationary GARCH(1,1) model. Moreover, the stationary GARCH(1,1) model failed to fit at several points of the USD/HKD and USD/TWD exchange rate series (this is marked by the “bullets” in the table), producing forecasts which were extremely inaccurate. The inaccuracy was due to the fact that the estimated parameters fell outside the stationarity region.

<table>
<thead>
<tr>
<th>Currency</th>
<th>AMSE$_{\text{GARCH}(1,1)}^{250,1000,N}$</th>
<th>AMSE$_{\text{tvARCH}(1)}^{250,1000,N}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>CHF</td>
<td>2426 $\times 10^{-8}$</td>
<td>4369 $\times 10^{-8}$</td>
</tr>
<tr>
<td>GBP</td>
<td>20282 $\times 10^{-9}$</td>
<td>8119 $\times 10^{-9}$</td>
</tr>
<tr>
<td>HKD</td>
<td></td>
<td>3155 $\times 10^{-12}$</td>
</tr>
<tr>
<td>JPY</td>
<td>8687 $\times 10^{-8}$</td>
<td>11211 $\times 10^{-8}$</td>
</tr>
<tr>
<td>NOK</td>
<td>3016 $\times 10^{-8}$</td>
<td>2013 $\times 10^{-8}$</td>
</tr>
<tr>
<td>NZD</td>
<td>11890 $\times 10^{-8}$</td>
<td>8358 $\times 10^{-8}$</td>
</tr>
<tr>
<td>SEK</td>
<td>37720 $\times 10^{-9}$</td>
<td>6783 $\times 10^{-9}$</td>
</tr>
<tr>
<td>TWD</td>
<td></td>
<td>3262 $\times 10^{-7}$</td>
</tr>
</tbody>
</table>

Table 1: AMSE for long-term forecasts using stationary GARCH(1,1) and tvARCH(1) models.

We conclude this section by exhibiting the performance of the two-stage Kernel-NLS($\hat{\mu}_{b,N} - 1$) estimator on the USD/GBP exchange rate series. The parameters are as before: we assume $p = 1$ and use a rectangular kernel with $bN = 250$. The top left plot in Figure 2 shows the estimated curve $\hat{a}_0(u)$, against the squared data. The top right plot shows the positive part of the estimated curve $\hat{a}_1(u)$, which displays interesting “periodic” behaviour. The negative values arise since our estimator is not guaranteed to be nonnegative. The middle left plot shows the empirical residuals from the fit, and the middle right plot shows the Q-Q plot of the residuals against the quantiles of the standard normal (the variance of the empirical residuals is 0.97). Finally, the bottom plots show the sample autocorrelations of the residuals, and of their squares, respectively. The above plots indicate a very good fit. In particular, the dependence in the squares, observed in the original data, is (almost) completely absent from the residuals. Note also the heavy-tailedness of the empirical residuals.

7 Concluding Remarks

In this paper, we proposed a new methodology for estimating the parameters of a tvARCH($p$) process. Our technique was based on minimising a local NLS criterion, where the normalisation was introduced to avoid imposing excessively strong moment conditions on the process.
Unlike the previously proposed local kernel-QML estimator, we showed that the resulting kernel-NLS estimator had an explicit form and was rapidly computable. We also demonstrated the consistency and asymptotic normality of the kernel-NLS estimator under mild moment conditions.

Despite these attractive features, the kernel-NLS estimator suffers from a number of drawbacks: we identified them, and used simulation to verify our heuristic observations. It turned out that the kernel-NLS estimator exhibited large bias or large variance for certain parameter values. Roughly speaking, the reason for that behaviour was that the normalising factor was not chosen in a data-dependent way.

Motivated by this discussion, we proposed a modification of the local NLS criterion in which the normalising factor was selected in a data-driven way. We introduced the resulting estimator as a two-stage scheme, and showed its consistency and asymptotic normality, under an additional mild moment condition. Using simulation, we showed that the numerical performance of the two-stage kernel-NLS estimator was very good, and often superior to the traditional kernel-QML estimator. Moreover, we demonstrated that the two-stage kernel-NLS estimator inherited the explicit form and ease of computation from the non-data-driven kernel-NLS estimator.

Since stationary ARCH($p$) processes form a subclass of the tvARCH($p$) class, a separate section was devoted to the behaviour of the NLS and two-stage NLS estimators for stationary processes. Consistency and asymptotic results were stated for both estimators.

As an application of the proposed methodology, we modelled several exchange rate datasets as tvARCH(1) processes, and used the two-stage kernel-NLS scheme to produce long term volatility forecasts. A numerical study showed that our method outperformed the benchmark stationary GARCH(1,1) approach in most of the cases. Also, we applied the two-stage kernel-NLS estimator to the estimation of the (possibly) time-varying parameters of the USD/GBP exchange rate series. The residuals indicated a very good fit.

Future research. In future work, we intend to investigate the problem of choosing the order $p$, and the bandwidth $b$, in a data-driven way. Also, the fact that the time-varying parameters $\{a_j(u) : j = 1, 2, \ldots, p - 1\}$ do not need to be bounded away from zero in order for the local NLS schemes to work (unlike the local QML criterion), provides opportunity for testing the nullity of $a_j(u)$ for all $j = 1, 2, \ldots, p - 1$. Moreover, a possible recursive “online” scheme (based on the modified local NLS criterion) for the estimation of the time-varying parameters $\{a_j(u) : j = 0, 1, \ldots, p\}$ merits further investigation, as does the possibility of applying wavelets, or other modern nonlinear techniques, in the estimation procedure. Finally, we are fascinated by the challenge of developing analogous theory and estimation techniques for time-varying GARCH($p$, $q$) processes.
A Appendix

The aim of this Appendix is to prove the theoretical results stated in the previous sections: mainly the long-memory effect of tvARCH\((p)\) processes (Section 2.2), and the consistency and asymptotic normality of kernel-NLS\((g)\) (Section 3.2) and two-stage kernel-NLS\(\hat{\mu}_{t_0,N} - 1\) (Section 4.2) estimators of the tvARCH\((p)\) parameters.

Before proving these results, we first obtain some results related to weighted sums of tvARCH\((p)\) processes that we use below.

In what follows, we use \(K\) to denote a generic finite positive constant, not necessarily the same each time it is used, even within a single equation.

A.1 Properties of tvARCH\((p)\) processes

Let \(f : \mathbb{R}^n \to \mathbb{R}\). If \(|f(x_1, \ldots, x_n) - f(y_1, \ldots, y_n)| \leq K \sum_{i=1}^{n} |x_i - y_i|\), then we say that \(f\) is Lipschitz continuous of order 1 (\(f \in \text{Lip}(1)\)).

**Lemma A.1** Suppose \(\{X_{t,N} : t = 1, \ldots, N\}\) is a tvARCH\((p)\) process which satisfies Assumptions 2.1(ia,ii), let \(\tilde{X}_t(u) : t = 1, 2, \ldots\) be defined as in (3), and let \(f : \mathbb{R}^n \to \mathbb{R}\) be such that \(f \in \text{Lip}(1)\). If \(|u_0 - t_0/N| < 1/N\), then, for fixed \(i_1, \ldots, i_n\), we have that

\[
\sum_{k=p+1}^{N} \frac{1}{bN} W \left( \frac{t_0 - k}{bN} \right) f \left( X_{k,N}^2, X_{k+i_1,N}^2, \ldots, X_{k+i_n,N}^2 \right) \xrightarrow{P} \mathbb{E} \left\{ f \left( \tilde{X}_k^2(u_0), \tilde{X}_{k+i_1}^2(u_0), \ldots, \tilde{X}_{k+i_n}^2(u_0) \right) \right\},
\]

with \(b \to 0\), \(bN \to \infty\) as \(N \to \infty\).

**PROOF.** It follows easily by using Lemma 2.1 and the same methods as those given in the proof of Lemma A.6 in Dahlhaus & Subba Rao (2003). We omit the details. \(\square\)

Let us now define the following quantity

\[
\mathcal{L}^{nls(g)}(u) = \mathbb{E} \left\{ \frac{\tilde{X}_k^2(u) \tilde{X}_{k-1}(u)}{(g + \|\tilde{X}_{k-1}(u)\|_1)^2} \right\}.
\]

**Lemma A.2** Suppose \(\{X_{t,N} : t = 1, \ldots, N\}\) is a tvARCH\((p)\) process which satisfies Assumptions 2.1(ia,ii), let \(\mu(u) = \mathbb{E}\{\tilde{X}_t^2(u)\}\), and let \(\mathcal{A}^{nls(g)}(u)\), \(\Sigma^{nls(g)}(u)\) and \(\mathcal{L}^{nls(g)}(u)\) be defined as in (17), (18) and (44) respectively. If \(|u_0 - t_0/N| < 1/N\), then we have that

(i)

\[
\sum_{k=1}^{N} \frac{1}{bN} W \left( \frac{t_0 - k}{bN} \right) X_{k,N}^2 \xrightarrow{P} \mu(u_0);
\]

(45)
satisfies the representation

\[ \mathcal{R}_{t_0,N}^{nls(g)} = \sum_{k=p+1}^{N} \frac{1}{bN} W \left( \frac{t_0 - k}{bN} \right) \frac{\mathcal{X}_{k-1,N}^{T} \mathcal{X}_{k-1,N}}{(g + \| \mathcal{X}_{k-1,N} \|_1)^2} \mathcal{P} \mathcal{A}_{t_0,N}^{nls(g)}(u_0); \]  

(46)

\[(iii)\]

\[ \mathcal{I}_{t_0,N}^{nls(g)} = \sum_{k=p+1}^{N} \frac{1}{bN} W \left( \frac{t_0 - k}{bN} \right) \frac{\mathcal{X}_{k,N}^{2} \mathcal{X}_{k-1,N}}{(g + \| \mathcal{X}_{k-1,N} \|_1)^2} \mathcal{P} \mathcal{I}_{t_0,N}^{nls(g)}(u_0); \]  

(47)

\[(iv)\]

\[ \sum_{k=p+1}^{N} \frac{1}{bN} W^2 \left( \frac{t_0 - k}{bN} \right) \frac{\sigma_{k,N}^{4} \mathcal{X}_{k-1,N}^{T} \mathcal{X}_{k-1,N}}{(g + \| \mathcal{X}_{k-1,N} \|_1)^4} \mathcal{P} \mathcal{W}_{nls}^{nls(g)}(u_0), \]  

(48)

with \( b \to 0, bN \to \infty \) as \( N \to \infty \), where \( w_2 = \int_{-1/2}^{1/2} W^2(x)dx \).

PROOF. The proof of (i), (ii) and (iii) is a straightforward application of Lemma A.1, while the proof of (iv) uses a minor modification of Lemmas A.1 and A.2 in Dahlhaus & Subba Rao (2003), with \( W(\cdot) \) replaced by \( W^2(\cdot) \). We omit the details. \(\square\)

Remark A.1 By using similar arguments, it can be easily seen that the asymptotic results of Lemma A.2 also hold with \( X_{k,N}^2, X_{k-1,N} \) and \( \sigma_{k,N}^{4} \) replaced by \( \tilde{X}_{k}(u_0), \tilde{X}_{k-1}(u_0) \) and \( \tilde{\sigma}_{k}^{4}(u_0) \) respectively.

We now give some mixingale properties of the stationary approximation \( \{ \tilde{X}_t(u) : t = 1, 2, \ldots \} \) of a tvARCH\((p)\) process. Let \( 0 < q < \infty \), and define \( \| X \|_q = (\mathbb{E}(\| X \|_q^q))^{1/q} \) to be the \( L_q \)-norm of the random vector \( X^T = (X_1, X_2, \ldots, X_s) \), where \( \| X \|_q = (\sum_{i=1}^{s} |X_i|^q)^{1/q} \). Furthermore, for each \( t \in \mathbb{Z} \), let \( \sigma(X_t, X_{t-1}, \ldots) \) be the \( \sigma \)-field generated by the sequence of random variables \( \{X_k\}_{k=-\infty}^{1} \) defined on the same probability space \( (\Omega, \mathcal{F}, \mathbb{P}) \).

Lemma A.3 Suppose \( \{ \phi_k : k = 1, 2, \ldots \} \) is a stochastic process which satisfies \( \mathbb{E}(\phi_k) = 0 \) and \( \mathbb{E}(\phi_k^q) < \infty \) for some \( 1 < q \leq 2 \). Further, let \( \mathcal{F}_t = \sigma(\phi_t, \phi_{t-1}, \ldots) \), and suppose that there exists a \( \rho \in (0, 1) \) such that \( \| \mathbb{E}(\phi_k|\mathcal{F}_{k-j}) \|_q \leq K \rho^j \). Then we have that

\[ \left\| \sum_{k=1}^{s} a_k \phi_k \right\|_q^E \leq \frac{K}{1 - \rho} \left( \sum_{k=1}^{s} |a_k|^q \right)^{1/q}. \]  

(49)

PROOF. Under the stated assumptions, it is not difficult to see that \( \{ (\phi_k, \mathcal{F}_k) : k = 1, 2, \ldots \} \) is a mixingale (see, for example, Davidson, 1994, Chapter 16). Therefore \( \{ (\phi_k, \mathcal{F}_k) : k = 1, 2, \ldots \} \) satisfies the representation

\[ \phi_k = \sum_{j=0}^{\infty} \left[ \mathbb{E}_{k-j}(\phi_k) - \mathbb{E}_{k-j-1}(\phi_k) \right], \text{ almost surely}, \]  

(50)
where $\mathbb{E}_{k-j}(\phi_k) = \mathbb{E}(\phi_k|\mathcal{F}_{k-j})$. By substituting (50) into the sum $\sum_{k=1}^s a_k \phi_k$, we obtain

$$
\sum_{k=1}^s a_k \phi_k = \sum_{k=1}^s a_k \sum_{j=0}^\infty \mathbb{E}_{k-j}(\phi_k) - \mathbb{E}_{k-j-1}(\phi_k)
= \sum_{j=0}^\infty \left( \sum_{k=1}^s a_k \mathbb{E}_{k-j}(\phi_k) - \mathbb{E}_{k-j-1}(\phi_k) \right), \quad \text{almost surely.}
$$

Keeping $j$ constant, we see that $\{(\mathbb{E}_{k-j}(\phi_k) - \mathbb{E}_{k-j-1}(\phi_k), \mathcal{F}_{k-j}) : k = 1, 2, \ldots \}$ is a martingale difference (see, for example, Davidson, 1994, p. 250). Therefore, we can apply inequality (15.52) in Davidson (1994, Theorem 15.17) to (51), and get

$$
\left\| \sum_{k=1}^s a_k \phi_k \right\|_q^E \leq \sum_{j=0}^\infty \left( 2 \sum_{k=1}^s |a_k|^q \left( \|\mathbb{E}_{k-j}(\phi_k) - \mathbb{E}_{k-j-1}(\phi_k)\|_q \right)^K \right)^{1/q}.
$$

Under the stated assumption, $\|\mathbb{E}_{k-j}(\phi_k) - \mathbb{E}_{k-j-1}(\phi_k)\|_q \leq 2K \rho^j$. Substituting this inequality into the above gives

$$
\left\| \sum_{k=1}^s a_k \phi_k \right\|_q^E \leq \sum_{j=0}^\infty \left( 2 \sum_{k=1}^s |a_k|^q (2K \rho^j)^q \right)^{1/q} \leq 2^{1+q} K \sum_{j=0}^\infty \rho^j \left( \sum_{k=1}^s |a_k|^q \right)^{1/q}.
$$

This completes the proof of Lemma A.3.
and set $\mu_1(u, 0, h) = \mu_1(u, h)$ and $c(u, 0, h) = c(u, h)$.

Moreover, we apply Lemma A.3 in the proof of the second part of Lemma A.4 below (with $a_k = 1$, $\phi_k = \{\tilde{X}_k^2(u)\tilde{X}_{k+h}^2(u) - \mu_1(u, h, d)\}$ and $q = 1+\zeta/2$). This requires the following proposition, which is a variant of Proposition A.1 above.

**Proposition A.2** Suppose $\{X_{t,N} : t = 1, 2, \ldots, N\}$ is a tvARCH($p$) process which satisfies Assumptions 2.1(ia,ii), and let $\{\tilde{X}_t(u) : t = 1, 2, \ldots\}$ be defined as in (3). Let $\mu_1(u, d, h)$ be defined as in (53), and suppose that $\mathbb{E}\left(\mathcal{Z}_t^{(2+\zeta)}\right)^{1/(2+\zeta)} \sum_{j=1}^p a_j(u) \leq 1 - \delta$ for some $0 < \zeta \leq 2$ and $\delta > 0$. Let also $F_t = \sigma(\tilde{X}_1^2(u), \tilde{X}_{t-1}^2(u), \ldots)$. Then, for all $0 < u + d \leq 1$, there exists a $\rho \in (1 - \delta, 1)$ such that

$$
\left\| \mathbb{E}\left[\tilde{X}_t^2(u)\tilde{X}_{t+h}^2(u + d) \mid F_{t-k}\right] - \mu_1(u, d, h)\right\|_{1+\zeta/2}^E \leq K\rho^k \left(1 + \left\|\tilde{X}_{t-k}(u)\right\|_{1+\zeta/2}^E + \left\|\tilde{X}_{t-k}(u + d)\right\|_{1+\zeta/2}^E + \left\|\tilde{X}_{t-k}(u)\tilde{X}_{t-k}^T(u + d)\right\|_{1+\zeta/2}^E\right),
$$

(55)

where the constant $K$ is independent of $u, d, k$ and $t$.

**PROOF.** It follows easily by using the same steps as in the proof of Proposition A.1. We omit the details. \qed

Define also the following quantities

$$
\mathcal{S}_{k,bN}(u) = \frac{1}{bN} \sum_{s=kbN}^{(k+1)bN-1} \tilde{X}_s^2(u),
$$

(56)

$$
\mathcal{S}_{k,bN}(u, h, d) = \frac{1}{bN} \sum_{s=kbN}^{(k+1)bN-1} \tilde{X}_s^2(u)\tilde{X}_{s+h}^2(u + d).
$$

(57)

**Lemma A.4** Suppose $\{X_{t,N} : t = 1, 2, \ldots, N\}$ is a tvARCH($p$) process which satisfies Assumptions 2.1(ia,ii,iv), and let $\{\tilde{X}_t(u) : t = 1, 2, \ldots\}$ be defined as in (3). Let $\mu(u) = \mathbb{E}\{\tilde{X}_1^2(u)\}$, and let $\mu_1(u, d, h)$, $\mathcal{S}_{k,bN}(u)$ and $\mathcal{S}_{k,bN}(u, h, d)$ be defined as in (53), (56) and (57) respectively. Then, we have that

$$
\left\| \sum_{k=p+1}^N \frac{1}{bN} W\left(\frac{t-k}{bN}\right) \{\tilde{X}_k^2(u) - \mu(u)\}\right\|_{1+\eta}^E \leq K(bN)^{-\frac{\eta}{1+\eta}}
$$

(58)

and

$$
\left\|\mathcal{S}_{k,bN}(kb) - \mu(kb)\right\|_{1+\eta}^E \leq K(bN)^{-\frac{\eta}{1+\eta}},
$$

(59)

where the constant $K$ is independent of $u$. 

27
Further, if $\mathbb{E} \left( Z_t^{2(2+\zeta)} \right)^{1/(2+\zeta)} \sup_a \sum_{j=1}^p a_j(u) \leq 1 - \delta$ for some $0 < \zeta \leq 2$ and $\delta > 0$, then we have that

$$\| S_{k,bN}(u,h,d) - \mu_1(u,d,h) \|_{1+\zeta/2}^E \leq K(bN)^{-\frac{\zeta}{2+\zeta}},$$

where the constant $K$ is independent of $u$ and $d$.

PROOF. We will first prove (58). We use Lemma A.3, with $a_k = W \left( \frac{t-k}{bN} \right)$, $\phi_k = \{ \tilde{X}_k^2(u) - \mu(u) \}$ and $q = 1 + \eta$, and take into consideration inequality (52). Note that $\mathbb{E}(\phi_k) = 0$ and, by using (52), we have that

$$\left\| \mathbb{E} \left( \tilde{X}_k^2(u) | F_{k-j} \right) - \mu(u) \right\|_{1+\eta}^E \leq K \rho^j \left( 1 + \| \tilde{X}_{k-j}(u) \|_{1+\eta}^E \right),$$

where $F_t = \sigma(\tilde{X}_t^2(u), \tilde{X}_{t-1}^2(u), \ldots)$. Since $\{ \tilde{X}_t^2(u) : t = 1, 2, \ldots \}$ is a stationary process, then by using (49) and that the support of $W \left( \frac{t-k}{bN} \right)$ is proportional to $bN$, we have that

$$\left\| \sum_{k=p+1}^{N} \frac{1}{bN} W \left( \frac{t-k}{bN} \right) \{ \tilde{X}_k^2(u) - \mu(u) \} \right\|_{1+\eta}^E \leq \frac{1}{bN} K \left( \sum_{k=p+1}^{N} \left| W \left( \frac{t-k}{bN} \right) \right|^{1+\eta} \right)^{1/(1+\eta)} \leq K(bN)^{-\frac{\eta}{1+\eta}}.$$

Thus, we have proved (58). The proof of (59) is identical to the proof of (45), hence we omit the details.

The proof of (60) uses Lemma A.3, with $a_k = 1$, $\phi_k = \{ \tilde{X}_k^2(u)\tilde{X}_{k+h}^2(u) - \mu_1(u,h,d) \}$ and $q = 1 + \zeta/2$, takes into account (55), and is similar to the proof of (45), hence we omit the details. This completes the proof of Lemma A.4.

A.2 The long-memory effect of tvARCH($p$) processes

Proof of Proposition 2.1. It follows easily by making a time-varying Volterra series expansion of the tvARCH($p$) process (see Section 5 in Dahlhaus & Subba Rao, 2003) and using Lemma 2.1 in Giraitis et al. (2000). We omit the details.

The following lemma is used to prove Proposition 2.2.

Lemma A.5 Suppose $\{X_{t,N} : t = 1, 2, \ldots, N\}$ is a tvARCH($p$) which satisfies Assumptions 2.1(ia,ii,iv), and let $\{\tilde{X}_t(u) : t = 1, 2, \ldots\}$ be defined as in (3). Let $h := h(N)$ be such that $h/N \rightarrow d \in [0, 1)$ as $N \rightarrow \infty$. Then we have that

$$\frac{1}{N-h} \sum_{s=1}^{N-h} \tilde{X}_{s,N}^2 \xrightarrow{p} \int_0^{1-d} \mathbb{E}\{\tilde{X}_t^2(u)\} du.$$

Further, if $\mathbb{E} \left( Z_t^{2(2+\zeta)} \right)^{1/(2+\zeta)} \sup_a \sum_{j=1}^p a_j(u) \leq 1 - \delta$ for some $0 < \zeta \leq 2$ and $\delta > 0$, then we have that

$$\frac{1}{N-h} \sum_{s=1}^{N-h} \tilde{X}_{s,N}^2 \tilde{X}_{s+h,N}^2 \xrightarrow{p} \int_0^{1-d} \mathbb{E}\{\tilde{X}_t^2(u)\tilde{X}_{t+h}^2(u+d)\} du.$$
PROOF. We first prove (62). Let \( b := b(N) \) be such that \( \frac{1}{b} \) is an integer, \( b \to 0 \) and \( b(N-h) \to \infty \) as \( N \to \infty \). We partition the left hand side of (62) into \( \frac{1}{b} \) blocks, i.e.,

\[
\frac{1}{N-h} \sum_{s=1}^{N-h} X_{s,N}^2 X_{s+h,N}^2 = b \sum_{k=0}^{1/b-1} \frac{1}{b(N-h)} \sum_{r=0}^{b(N-h)-1} X_{k,b(N-h)+r,N}^2 X_{k,b(N-h)+r+h,N}^2.
\]

(63)

Let \( k_b = kb[1-d] \), and replace the terms \( X_{k,b(N-h)+r,N}^2 \) and \( X_{k,b(N-h)+r+h,N}^2 \) with \( \tilde{X}_{k,b(N-h)+r}^2(k_b) \) and \( \tilde{X}_{k,b(N-h)+r+h}^2(k_b+d) \) respectively. Then, we obtain

\[
\frac{1}{N-h} \sum_{s=1}^{N-h} X_{s,N}^2 X_{s+h,N}^2 = b \sum_{k=0}^{1/b-1} \frac{1}{b(N-h)} \sum_{r=0}^{b(N-h)-1} \tilde{X}_{k,b(N-h)+r}^2(k_b) \tilde{X}_{k,b(N-h)+r+h}^2(k_b+d) + R_N,
\]

(64)

where, by using Lemma 2.1, we have that

\[
\mathbb{E}[R_N] \leq K \left\{ \left( b + kb \left| d - \frac{h}{N} \right| \right)^\beta + \frac{1}{N^\beta} \right\}.
\]

Therefore, \( R_N \nrightarrow 0 \) as \( N \to \infty \). Recall the notation \( S_{k,b,N}(u, h, d) \) and \( \mu_1(u, d, h) \) given in (53) and (57) respectively. By using (63) and (64), we have that

\[
\left\| \frac{1}{N-h} \sum_{s=1}^{N-h} X_{s,N}^2 X_{s+h,N}^2 - \int_0^{1-d} \mu_1(u, d, h) du \right\|_{1+\zeta/2}^E \leq b \sum_{k=0}^{1/b-1} \|S_{k,b(N-h)}(k_b, h, d) - \mu_1(k_b, d, h)\|_{1+\zeta/2}^E + \left\| \sum_{k=0}^{1/b-1} \mu_1(k_b, d, h) - \int_0^{1-d} \mu_1(k_b, d, h) du \right\|_{1+\zeta/2}^E + O \left\{ \left( b + kb \left| d - \frac{h}{N} \right| \right)^\beta + \frac{1}{N^\beta} \right\}.
\]

(65)

We note that, by using (60), we have that

\[
b \sum_{k=0}^{1/b-1} \|S_{k,b(N-h)}(k_b, h, d) - \mu_1(k_b, d, h)\|_{1+\zeta/2}^E \leq K(bN)^{-\frac{1}{1+\zeta}}.
\]

(66)

Now, by substituting (66) and

\[
b \sum_{k=0}^{1/b-1} \mu_1(k_b, d, h) = \int_0^{1-d} \mathbb{E} \left\{ \tilde{X}_t^2(u) \tilde{X}_{t+h}^2(u + d) \right\} du + O(b)
\]

into (65), we have that

\[
\left\| \frac{1}{N-h} \sum_{s=1}^{N-h} X_{s,N}^2 X_{s+h,N}^2 - \int_0^{1-d} \mathbb{E} \left\{ \tilde{X}_t^2(u) \tilde{X}_{t+h}^2(u + d) \right\} du \right\|_{1+\zeta/2}^E \to 0,
\]

29
which give us (62). The proof of (61) is similar and we omit the details. This completes the proof of Lemma A.5.

**Proof of Proposition 2.2.** We first consider the more general case where \( h := h(N) \) is such that \( h/N \to d \in [0, 1) \) as \( N \to \infty \). Then, for fixed \( h > 0 \), we obtain (6) as special case with \( d = 0 \).

Let \( S_N(h) = A_N - B_N \), where

\[
A_N = \frac{1}{N - h} \sum_{t=1}^{N-h} X_{t,N}^2 X_{t+h,N}^2 \quad \text{and} \quad B_N = (\bar{X}_N)^2.
\]

We consider the asymptotic behaviour of the terms \( A_N \) and \( B_N \) separately. By using (61) and (62), we have that

\[
A_N \xrightarrow{P} \int_0^{1-d} \mu_1(u, h, d) \, du \quad \text{and} \quad B_N \xrightarrow{P} \int_0^{1-d} \int_0^{1-d} \mu(u) \mu(v) \, du \, dv.
\]

Recall that \( \mu(u) = \mathbb{E} \{ \bar{X}_1^2(u) \} \), and that \( \mu_1(u, d, h) \) and \( c(u, d, h) \) are defined in (53) and (54) respectively. By using the formula \( \mu_1(u, d, h) = c(u, d, h) + \mu(u) \mu(u + d) \), we obtain

\[
S_N(h) \xrightarrow{P} \int_0^{1-d} \{ c(u, d, h) + \mu(u) \mu(u + d) \} \, du - \left[ \int_0^{1-d} \mu(u) \, du \right]^2.
\]

Let us now consider the special case of (67) where \( d = 0 \). Then, for fixed \( h > 0 \), we have that

\[
S_N(h) \xrightarrow{P} \int_0^1 c(u, h) \, du + \int_0^1 \int_0^1 \mu^2(u) \, du \, dv - \int_0^1 \int_0^1 \mu(u) \mu(v) \, du \, dv
\]

\[
= \int_0^1 c(u, h) \, du + \int_0^1 \left\{ \mu^2(u) - \int_0^1 \int_0^1 \mu(u) \mu(v) \, du \, dv \right\} \, du
\]

\[
= \int_0^1 c(u, h) \, du + \int \int_{\{0 \leq u < v \leq 1\}} \left\{ \mu^2(u) - \int_0^1 \int_0^1 \mu(u) \mu(v) \, du \, dv \right\} \, du \, dv
\]

\[
+ \int \int_{\{0 \leq v < u \leq 1\}} \left\{ \mu^2(u) - \int_0^1 \int_0^1 \mu(u) \mu(v) \, du \, dv \right\} \, du \, dv
\]

\[
= \int_0^1 c(u, h) \, du + \int \int_{\{0 \leq u < v \leq 1\}} \{ \mu(v) - \mu(u) \}^2 \, du \, dv, \text{ as } N \to \infty.
\]

This proves (6) and, hence, the proof of Proposition 2.2 is completed.

**A.3 Asymptotic properties of the kernel-NLS estimator**

In this section, we prove consistency and asymptotic normality of the kernel-NLS\((g)\) estimator.

**Proof of Proposition 3.1** By using (46), (47) and Slutsky’s theorem, we have that

\[
\hat{\alpha}_{nls,N}^{nls(g)} = \left\{ R_{nls,(g)}^{nls(g)} \right\}^{-1} \xrightarrow{P} \left\{ A_{nls(g)}(u_0) \right\}^{-1} \xrightarrow{P} \overline{\alpha}_{nls(g)}(u_0).
\]

Finally, to show that \( \hat{\alpha}_{nls,N}^{nls(g)} \xrightarrow{P} \overline{\alpha}(u_0) \), we show that \( \alpha(u_0) = \left\{ A_{nls(g)}(u_0) \right\}^{-1} \overline{\alpha}_{nls(g)}(u_0) \). By using (3), \( \{ \bar{X}_k^2(u_0) : k = 1, 2, \ldots \} \) satisfies the representation

\[
\bar{X}_k^2(u_0) = \overline{\alpha}^T(u_0) \hat{\alpha}_{k-1}(u_0) + (Z_k^2 - 1) \overline{\alpha}_k^2(u_0).
\]
where

\[
\frac{\tilde{X}_k^2(u_0)}{(g + \|\tilde{X}_{k-1}(u_0)\|_1)^2} = \frac{\tilde{g}^T(u_0)\tilde{X}_{k-1}(u_0)}{(g + \|\tilde{X}_{k-1}(u_0)\|_1)^2} + (Z_k^2 - 1) \frac{\tilde{\sigma}_k^2(u_0)}{(g + \|\tilde{X}_{k-1}(u_0)\|_1)^2}.
\]  

(69)

Finally, multiplying (69) by \(\tilde{X}_{k-1}(u_0)\) and taking expectations, we have that

\[
E \left\{ \tilde{X}_k(u_0)\tilde{X}_{k-1}(u_0) \right\} = E \left\{ \tilde{X}_{k-1}(u_0)\tilde{X}_k^T(u_0) \right\} \tilde{a}(u_0).
\]

Therefore, \(\tilde{a}(u_0) = \left\{ A^{nls}(u_0) \right\}^{-1} \Sigma^{nls}(u_0)\), completing the proof of Proposition 3.1.

To show asymptotic normality, we first show asymptotic normality of the gradient of the stationary likelihood \(L_{t_0}^{nls}(u, \tilde{a}(u))\), where

\[
L_{t_0}^{nls}(u, \tilde{a}(u)) = \sum_{k=p+1}^{N} \frac{1}{bN} W \left( \frac{t_0 - k}{bN} \right) h_k^{nls}(u, \tilde{a}(u)),
\]

(70)

where \(h_k^{nls}(u, \tilde{a}(u))\) is defined in (20).

The following lemma is used to prove Proposition 3.2.

**Lemma A.6** Suppose \(\{X_{t,N} : t = 1, \ldots, N\}\) is a tvARCH(p) process which satisfies Assumptions 2.1(ia,ii,iii). Let \(L_{t_0}^{nls}(u, \tilde{a}(u))\) and \(\Sigma^{nls}(u)\) be defined as in (70) and (18) respectively. If \(|u_0 - t_0/N| < 1/N\), then we have that

\[
\sqrt{bN} \nabla L_{t_0}^{nls}(u_0, \tilde{a}(u_0)) \xrightarrow{D} N \left( 0, 4w_2\mu_4\Sigma^{nls}(u_0) \right),
\]

(71)

with \(b \to 0, bN \to \infty\) as \(N \to \infty\), where \(w_2 = \int_{-1/2}^{1/2} W^2(x)dx\) and \(\mu_4 = \text{var}(Z_t^2)\).

**PROOF.** We easily see that

\[
\sqrt{bN} \nabla L_{t_0}^{nls}(u_0, \tilde{a}(u_0)) = -2 \sum_{k=p+1}^{N} \frac{1}{\sqrt{bN}} W \left( \frac{t_0 - k}{bN} \right) \frac{(Z_k^2 - 1)\tilde{\sigma}_k^2(u_0)}{(g + \|\tilde{X}_{k-1}(u_0)\|_1)^2} \tilde{X}_{k-1}(u_0).
\]

(72)

Since \(\sqrt{bN} \nabla L_{t_0}^{nls}(u_0, \tilde{a}(u_0))\) is a weighted sum of martingale differences, we use the martingale central limit theorem (see, for example, Hall & Heyde, 1980, Corollary 3.1) to prove asymptotic normality. First, it is seen that the conditional Lindeberg condition is satisfied. By using (48) and Remark A.1, we have that the conditional variance satisfies

\[
\sum_{k=p+1}^{N} \frac{4}{bN} W^2 \left( \frac{t_0 - k}{bN} \right) E \left\{ \left( Z_k^2 - 1 \right) \tilde{\sigma}_k^2(u_0) \tilde{X}_{k-1}(u_0) \tilde{X}_k^T(u_0) \mid \mathcal{F}_{k-1} \right\} \overset{P}{=} 4w_2\mu_4\Sigma^{nls}(u_0),
\]

where \(\mathcal{F}_{k-1} = \sigma \left( \tilde{X}_{k-1}^2(u_0), \tilde{X}_{k-2}^2(u_0) \ldots \right)\). By using the Cramér-Wold device (see, for example, Billingsley, 1995, Theorem 29.4), we now get (71). This completes the proof of Lemma A.6. \(\square\)
PROOF of Proposition 3.2 We first prove part (ii). By using the decomposition
\[ \nabla L_{t_0,N}^{nls(g)}(\tilde{a}(u_0)) = \nabla L_{t_0}^{nls(g)}(u_0, \tilde{a}(u_0)) + \nabla B_{t_0,N}^{nls(g)}(\tilde{a}(u_0)), \tag{73} \]
we have that
\[ \nabla L_{t_0,N}^{nls(g)}(\tilde{a}_{t_0,N}) = \nabla L_{t_0}^{nls(g)}(\tilde{a}(u_0)) + \nabla^2 L_{t_0,N}^{nls(g)}(\tilde{a}_{t_0,N} - \tilde{a}(u_0)) \]
\[ = \left\{ \nabla L_{t_0}^{nls(g)}(u_0, \tilde{a}(u_0)) + \nabla B_{t_0,N}^{nls(g)}(\tilde{a}(u_0)) \right\} \]
\[ + \nabla^2 L_{t_0,N}^{nls(g)}(\tilde{a}_{t_0,N} - \tilde{a}(u_0)). \]
By using (46), we easily see that \( \nabla^2 L_{t_0,N}^{nls(g)}(\tilde{a}(u_0)) \xrightarrow{P} 2 \mathcal{A}^{nls(g)}(u_0) \), and since \( \nabla L_{t_0,N}^{nls(g)}(\tilde{a}_{t_0,N}) = 0 \), we have that
\[ \{ \tilde{a}_{t_0,N} - \tilde{a}(u_0) \} = \left\{ -\nabla L_{t_0}^{nls(g)}(u_0, \tilde{a}(u_0)) - \nabla B_{t_0,N}^{nls(g)}(\tilde{a}(u_0)) \right\} \times \]
\[ \times \left\{ \frac{1}{2} \left( A^{nls(g)}(u_0) \right)^{-1} + o_p(1) \right\}, \]
which leads to
\[ \sqrt{bN} \{ \tilde{a}_{t_0,N} - \tilde{a}(u_0) \} + \frac{1}{2} \sqrt{bN} \left( A^{nls(g)}(u_0) \right)^{-1} \nabla B_{t_0,N}^{nls(g)}(\tilde{a}(u_0)) \]
\[ = -\frac{1}{2} \sqrt{bN} \left( A^{nls(g)}(u_0) \right)^{-1} \nabla L_{t_0}^{nls(g)}(u_0, \tilde{a}(u_0)) + o_p(1). \tag{74} \]
By combining (71) and (74), we get (22).

We now prove part (i). By using decomposition (73), applying Lemma 2.1, and following the arguments of Lemma A.6 in Dahlhaus & Subba Rao (2003), it can be shown that \( \nabla B_{t_0,N}^{nls(g)}(\tilde{a}(u_0)) = O_p(b^3) \), therefore we get (21). We omit the details. This completes the proof of Proposition 3.2. \( \square \)

A.4 Asymptotic properties of the two-stage kernel-NLS estimator

In this section, we prove consistency and asymptotic normality of the two-stage kernel-NLS(\( \hat{\mu}_{t_0,N} - 1 \)) estimator.

PROOF of Lemma 4.1 Since the kernel \( W(\cdot) \) is of bounded variation satisfying \( \int_{-1/2}^{1/2} W(x)dx = 1 \), it is easy to see that
\[ \mu(u_0) = \mu(u_0) \sum_{k=p+1}^{N} \frac{1}{bN} W \left( \frac{t_0 - k}{bN} \right) + O \left( \frac{1}{bN} \right). \tag{75} \]
By replacing \( \{ X_{k,N}^2 : k = 1, 2, \ldots, N \} \) with its stationary approximation \( \{ \tilde{X}_k^2(u) : k = 1, 2, \ldots \} \)
about \( u \approx t/N \), and using (75), we have that

\[
\hat{\mu}_{t_0,N} - \mu(u_0) = \sum_{k=p+1}^{N} \frac{1}{bN} W \left( \frac{t_0 - k}{bN} \right) \left\{ X^2_{k,N} - \hat{X}^2_k(u_0) \right\} + \sum_{k=p+1}^{N} \frac{1}{bN} W \left( \frac{t_0 - k}{bN} \right) \left\{ \hat{X}^2_k(u_0) - \mu(u_0) \right\} + O \left( \frac{1}{bN} \right) \\
:= B_{t_0,N}(u_0) + H_{t_0}(u_0) + O \left( \frac{1}{bN} \right). \tag{76}
\]

We consider the first term in (76). Under the stated assumptions, and by using Lemma 2.1, we have that

\[
\| B_{t_0,N}(u_0) \|^E_{1+\eta} = O \left( b^\beta \right). \tag{77}
\]

We now consider the second term in (76). By using (58), we have that

\[
\| H_{t_0}(u_0) \|^E_{1+\eta} = O \left( (bN)^{-\frac{\eta}{1+\eta}} \right). \tag{78}
\]

Therefore, by using (77) and (78), we have that

\[
\| \hat{\mu}_{t_0,N} - \mu(u_0) \|^E_{1+\eta} = O \left( b^\beta + (bN)^{-\frac{\eta}{1+\eta}} \right),
\]

which gives us (29). This completes the proof of Lemma 4.1. \( \square \)

Let us define the following quantities

\[
S_{k-1,N} = \sum_{j=1}^{p} X^2_{k-j,N} \quad \text{and} \quad S_{k-1}(u) = \sum_{j=1}^{p} \hat{X}^2_{k-j}(u). \tag{79}
\]

**PROOF of Lemma 4.2** Let us consider the following quantity

\[
Y_{k,N} = \frac{X^2_{k-i,N}X^2_{k-j,N}}{(\hat{\mu}_{t_0,N} + S_{k-1,N})^2}. \]

Suppose that \( |\hat{\mu}_{t_0,N} - \mu(u_0)| < \delta \), where for some \( \nu > 0 \), \( \delta < (1 - \nu)\mu(u_0) \). Then, \( \delta \) is sufficiently small, such that we can make the following geometric expansion

\[
Y_{k,N} = \frac{X^2_{k-i,N}X^2_{k-j,N}}{(\mu(u_0) + S_{k-1,N})^2} \times \frac{1}{\left[ 1 + \frac{\hat{\mu}_{t_0,N} - \mu(u_0)}{\mu(u_0) + S_{k-1,N}} \right]^2} = \frac{X^2_{k-i,N}X^2_{k-j,N}}{(\mu(u_0) + S_{k-1,N})^2} \left\{ 1 + \sum_{m=1}^{\infty} (-1)^m(m+1)(m+2) \left( \frac{\hat{\mu}_{t_0,N} - \mu(u_0)}{\mu(u_0) + S_{k-1,N}} \right)^m \right\}. \tag{80}
\]

We note, if \( \delta < (1 - \nu)\mu(u_0) \), then

\[
\left| \sum_{m=1}^{\infty} (-1)^m(m+1)(m+2) \left( \frac{\hat{\mu}_{t_0,N} - \mu(u_0)}{\mu(u_0) + S_{k-1,N}} \right)^m \right| \leq K \left| \hat{\mu}_{t_0,N} - \mu(u_0) \right| \frac{1}{\nu^\beta}. \tag{81}
\]

33
Let us now define the following quantities
\[ \Gamma(u_0) = \mathbb{E}\left( \frac{X_{k-1}^2(u_0)X_{k-1}^2(u_0)}{(\mu(u_0) + S_{k-1}(u_0))^2} \right), \]
\[ A_{t_0,N} = \sum_{k=p+1}^{N} \frac{1}{bN} W\left( \frac{t_0 - k}{bN} \right) \frac{X_{k-i,N}^2X_{k-j,N}^2}{(\mu(u_0) + S_{k-1,N})^2}, \]
\[ C_{t_0,N}(u_0) = \left| \sum_{k=p+1}^{N} \frac{1}{bN} W\left( \frac{t_0 - k}{bN} \right) \frac{X_{k-i,N}^2X_{k-j,N}^2}{(\mu(u_0) + S_{k-1,N})^2} - \Gamma(u_0) \right|. \]

Suppose, without loss of generality, that \( W(\cdot) \) is a positive function. Since
\[ \frac{\hat{\mu}(u_0) - \mu(u_0)}{\mu(u_0) + S_{k-1,N}} < 1 - \nu, \]
then, by using the expansion in (80) and (81), we have that
\[ \left| \sum_{k=p+1}^{N} \frac{1}{bN} W\left( \frac{t_0 - k}{bN} \right) Y_{k,N} - \Gamma(u_0) \right| \leq \left| \sum_{k=p+1}^{N} \frac{1}{bN} W\left( \frac{t_0 - k}{bN} \right) \frac{X_{k-i,N}^2X_{k-j,N}^2}{(\mu(u_0) + S_{k-1,N})^2} - \Gamma(u_0) \right| 
+ \frac{1}{\nu^3} \left| \hat{\mu}(u_0) - \mu(u_0) \right| \sum_{k=p+1}^{N} \frac{1}{bN} W\left( \frac{t_0 - k}{bN} \right) \times \frac{X_{k-i,N}^2X_{k-j,N}^2}{(\mu(u_0) + S_{k-1,N})^2}. \]

By adding and subtracting \( \Gamma(u_0) \) into (82), we have that
\[ |A_{t_0,N} - \Gamma(u_0)| \leq C_{t_0,N}(u_0) + \frac{1}{\nu^3} \left| \hat{\mu}(u_0) - \mu(u_0) \right| C_{t_0,N}(u_0) + \frac{\left| \hat{\mu}(u_0) - \mu(u_0) \right|}{\nu^3 \mu(u_0)} \Gamma(u_0). \]

Suppose \( \varepsilon > 0 \). We will show that \( \mathbb{P}(|A_{t_0,N} - \Gamma(u_0)| > \varepsilon) \xrightarrow{P} 0 \). Using that \( \mathbb{P}(A) \leq \mathbb{P}(A|B)\mathbb{P}(B) + \mathbb{P}(B^c) \), and conditioning on the events \( |\hat{\mu}(u_0) - \mu(u_0)| < \delta \) and \( |\hat{\mu}(u_0) - \mu(u_0)| > \delta \) separately, we have that
\[ \mathbb{P}(|A_{t_0,N} - \Gamma(u_0)| > \varepsilon) \leq \mathbb{P}\left( \big| A_{t_0,N} - \Gamma(u_0) \big| > \varepsilon \big| \hat{\mu}(u_0) - \mu(u_0) \big| < \delta \right) \mathbb{P}(|\hat{\mu}(u_0) - \mu(u_0)| \leq \delta) 
+ \mathbb{P}(|\hat{\mu}(u_0) - \mu(u_0)| \geq \delta). \]

Let \( \delta < (1 - \nu)\mu(u_0) \). Then, by using (83), we have that
\[ \mathbb{P}\left( \big| A_{t_0,N} - \Gamma(u_0) \big| > \varepsilon \big| \hat{\mu}(u_0) - \mu(u_0) \big| \leq \delta \right) \leq \mathbb{P}\left( \big| C_{t_0,N}(u_0) + \frac{1}{\nu^3} \left| \hat{\mu}(u_0) - \mu(u_0) \right| C_{t_0,N}(u_0) + \frac{\left| \hat{\mu}(u_0) - \mu(u_0) \right|}{\nu^3 \mu(u_0)} \Gamma(u_0) \big| > \varepsilon \big| \hat{\mu}(u_0) - \mu(u_0) \big| \leq \delta \right) 
+ \mathbb{P}\left( \big| \frac{\left| \hat{\mu}(u_0) - \mu(u_0) \right|}{\nu^3 \mu(u_0)} \Gamma(u_0) \big| > \frac{\varepsilon}{3} \big| \hat{\mu}(u_0) - \mu(u_0) \big| \leq \delta \right). \]
Since $P(A \mid B) \leq P(A)/P(B)$, by using the above, we have that
\[
P \left( \left| A_{t_0,N} - \Gamma(u_0) \right| > \varepsilon \left| \hat{\mu}_{t_0,N} - \mu(u_0) \right| \leq \delta \right) P \left( \left| \hat{\mu}_{t_0,N} - \mu(u_0) \right| \leq \delta \right)
\leq P \left( C_{t_0,N}(u_0) > \frac{\varepsilon}{3} \right) + P \left( \left| \hat{\mu}_{t_0,N} - \mu(u_0) \right| C_{t_0,N}(u_0) > \frac{\varepsilon}{3} \right)
+ P \left( \left| \hat{\mu}_{t_0,N} - \mu(u_0) \right| \Gamma(u_0) > \frac{\varepsilon}{3} \right).
\] (85)

Substituting (85) into (84), we have that
\[
P \left( \left| A_{t_0,N} - \Gamma(u_0) \right| > \varepsilon \right) \leq P \left( C_{t_0,N}(u_0) > \frac{\varepsilon}{3} \right) + P \left( \left| \hat{\mu}_{t_0,N} - \mu(u_0) \right| C_{t_0,N}(u_0) > \frac{\varepsilon}{3} \right)
+ P \left( \left| \hat{\mu}_{t_0,N} - \mu(u_0) \right| \Gamma(u_0) > \frac{\varepsilon}{3} \right) + P \left( \left| \hat{\mu}_{t_0,N} - \mu(u_0) \right| \geq \delta \right).
\] (86)

By using (43) and (45), if $t_0/N - u_0 < 1/N$, then, for every $\varepsilon_1 > 0$, we have that
\[
P \left( C_{t_0,N}(u_0) > \varepsilon_1 \right) \to 0 \quad \text{and} \quad P \left( \left| \hat{\mu}_{t_0,N} - \mu(u_0) \right| > \varepsilon_1 \right) \to 0.
\] (87)

Therefore, by using (86) and (87), we have that
\[
P \left( \left| A_{t_0,N} - \Gamma(u_0) \right| > \varepsilon \right) \to 0,
\]
which gives (30). This completes the proof of Lemma 4.2. \( \square \)

**Corollary A.1** Let $\mathcal{R}^{nls(\hat{\mu}_{t_0,N}^{-1})}_{t_0,N}$, $\mathcal{L}^{nls(\mu(u)^{-1})}_{t_0,N}$, $\mathcal{A}^{nls(\mu(u)^{-1})}_{t_0,N}$ and $\mathcal{L}^{nls(\mu(u)^{-1})}_{t_0,N}$ be defined as in (27), (28), (17) and (44) respectively. If $|u_0 - t_0|/N < 1/N$, then we have that
\[
\mathcal{R}^{nls(\hat{\mu}_{t_0,N}^{-1})}_{t_0,N} \xrightarrow{P} \mathcal{A}^{nls(\mu(u)^{-1})}_{t_0,N}, \text{ with } b \to 0, bN \to \infty \text{ as } N \to \infty,
\] (88)
\[
\mathcal{L}^{nls(\hat{\mu}_{t_0,N}^{-1})}_{t_0,N} \xrightarrow{P} \mathcal{L}^{nls(\mu(u)^{-1})}_{t_0,N}, \text{ with } b \to 0, bN \to \infty \text{ as } N \to \infty.
\] (89)

**PROOF.** It follows easily from Lemma 4.2. We omit the details. \( \square \)

**PROOF of Proposition 4.1** It is straightforward to show consistency using (27), (28), (88) and (89). We omit the details. \( \square \)

Let us now define the following quantities
\[
\mathcal{L}^{nls(\mu(u)^{-1})}_{t_0} \left( u, a(u) \right) = \sum_{k=p+1}^{N} \frac{1}{bN} W \left( \frac{t_0 - k}{bN} \right) \left\{ \frac{1}{\mu(u) + S_{k-1}(u)} \right\}^2 \times \left\{ \hat{X}_k(u)^2 - a_0(u) - \sum_{j=1}^{p} a_j(u) \hat{X}_{k-j}(u)^2 \right\}^2,
\] (90)
\[
\mathcal{L}^{nls(\hat{\mu}_{t_0,N}^{-1})}_{t_0} \left( u, a(u) \right) = \sum_{k=p+1}^{N} \frac{1}{bN} W \left( \frac{t_0 - k}{bN} \right) \left\{ \frac{1}{\hat{\mu}_{t_0,N} + S_{k-1}(u)} \right\}^2 \times \left\{ \hat{X}_k(u)^2 - a_0(u) - \sum_{j=1}^{p} a_j(u) \hat{X}_{k-j}(u)^2 \right\}^2.
\] (91)
We have shown in Lemma A.6 that the asymptotic normality of \( \sqrt{bN} \nabla L_{t_0}^{nls} (\mu(u_0) - 1) (u_0, g(u_0)) \) can easily be established by verifying the conditions of the martingale central limit theorem. However, the same theorem cannot be used to show the asymptotic normality of \( \sqrt{bN} \nabla L_{t_0}^{nls} (\tilde{\mu}_{t_0,N} - 1) (u_0, g(u_0)) \), since \( L_{t_0}^{nls}(\tilde{\mu}_{t_0,N} - 1) (u_0, g(u_0)) \) is not a sum of martingale differences. In Lemma A.7 below, we overcome this problem by showing that \( L_{t_0}^{nls}(\tilde{\mu}_{t_0,N} - 1) (u_0, g(u_0)) \) is ‘close’ enough to \( L_{t_0}^{nls} (\mu(u_0) - 1) (u_0, g(u_0)) \) for us to replace \( L_{t_0}^{nls}(\tilde{\mu}_{t_0,N} - 1) (u_0, g(u_0)) \) with \( L_{t_0}^{nls}(\mu(u_0) - 1) (u, g(u)) \), and then use it to prove Proposition 4.2.

**Lemma A.7** Suppose \( \{X_{t,N} : t = 1, \ldots, N\} \) is a tvARCH\((p)\) process which satisfies Assumptions 2.1(iii,iii,iv). Let \( L_{t_0}^{nls}(\tilde{\mu}_{t_0,N} - 1)(u_0, g(u_0)) \) and \( L_{t_0}^{nls}(\mu(u_0) - 1)(u_0, g(u_0)) \) be defined as in (90) and (91) respectively. If \( |u_0 - t_0| < 1/N \), then we have that

\[
\sqrt{bN} \left[ \nabla L_{t_0}^{nls}(\tilde{\mu}_{t_0,N} - 1)(u_0, g(u_0)) - \nabla L_{t_0}^{nls}(\mu(u_0) - 1)(u_0, g(u_0)) \right] = o_p(1),
\]

with \( b \to 0, bN \to \infty \) as \( N \to \infty \).

**PROOF.** Let \( M_k(u_0) = (Z_k^2 - 1) \hat{\sigma}_k^2 (u_0) \tilde{X}_{k-1}(u_0) \). Then it is easily seen that

\[
\sqrt{bN} \left[ \nabla L_{t_0}^{nls}(\tilde{\mu}_{t_0,N} - 1)(u_0, g(u_0)) - \nabla L_{t_0}^{nls}(\mu(u_0) - 1)(u_0, g(u_0)) \right]
= -2 \sum_{k=p+1}^{N} \frac{1}{bN} W \left( \frac{t_0 - k}{bN} \right) \left\{ \frac{1}{[\tilde{\mu}_{t_0,N} + S_{k-1}(u_0)]^2} - \frac{1}{[\mu(u_0) + S_{k-1}(u_0)]^2} \right\} M_k(u_0)
= 2 \sum_{k=p+1}^{N} \frac{1}{bN} W \left( \frac{t_0 - k}{bN} \right) \left\{ \frac{[\mu(u_0) + S_{k-1}(u_0)]^2}{[\tilde{\mu}_{t_0,N} + S_{k-1}(u_0)]^2} \right\} M_k(u_0)
= 2[\tilde{\mu}_{t_0,N} - \mu(u_0)] \sum_{k=p+1}^{N} \frac{1}{bN} W \left( \frac{t_0 - k}{bN} \right) \left\{ \frac{\{\tilde{\mu}_{t_0,N} + \mu(u_0)\} + 2S_{k-1}(u_0)}{[\mu(u_0) + S_{k-1}(u_0)]^4} \right\} M_k(u_0)
-2[\tilde{\mu}_{t_0,N} - \mu(u_0)]^2 \sum_{k=p+1}^{N} \frac{1}{bN} W \left( \frac{t_0 - k}{bN} \right) \left\{ \frac{\{\tilde{\mu}_{t_0,N} + \mu(u_0)\} + 2S_{k-1}(u_0)}{[\mu(u_0) + S_{k-1}(u_0)]^4} \right\} M_k(u_0)
:= 2 \{[\tilde{\mu}_{t_0,N} - \mu(u_0)] I_{t_0,N} - J_{t_0,N} \}.
\]

We first consider \( I_{t_0,N} \). It can be rewritten as

\[
I_{t_0,N} = [\tilde{\mu}_{t_0,N} + \mu(u_0)] \sum_{k=p+1}^{N} \frac{1}{bN} W \left( \frac{t_0 - k}{bN} \right) \frac{M_k(u_0)}{[\mu(u_0) + S_{k-1}(u_0)]^4}
+ 2 \sum_{k=p+1}^{N} \frac{1}{bN} W \left( \frac{t_0 - k}{bN} \right) \frac{M_k(u_0)S_{k-1}(u_0)}{[\mu(u_0) + S_{k-1}(u_0)]^4}.
\]

We note that if \( \{(Z_k, \mathcal{F}_k) : k = 1, 2, \ldots\} \), where \( \mathcal{F}_k = \sigma(Z_k, Z_{k-1}, \ldots) \), is a sequence of martingale differences with \( \mathbb{E} \left( Z_k^2 \right) < \infty \), then

\[
\mathbb{E} \left\{ \sum_{k=p+1}^{N} \frac{1}{bN} W \left( \frac{t_0 - k}{bN} \right) Z_k \right\}^2 = \sum_{k=p+1}^{N} \frac{1}{bN} W \left( \frac{t_0 - k}{bN} \right)^2 \mathbb{E} \left( Z_k^2 \right) = O(1).
\]

36
By taking \( Z_k = M_k(u_0)/[\mu(u_0) + S_{k-1}(u_0)]^4 \) or \( Z_k = M_k(u_0)S_{k-1}(u_0)/[\mu(u_0) + S_{k-1}(u_0)]^4 \), and applying the bound given in (95) to both parts of (94), we have that
\[
\sum_{k=p+1}^{N} \frac{1}{\sqrt{bN}} W \left( \frac{t_0 - k}{bN} \right) \frac{M_k(u_0)}{[\mu(u_0) + S_{k-1}(u_0)]^4} = O_p(1),
\]
\[
\sum_{k=p+1}^{N} \frac{1}{\sqrt{bN}} W \left( \frac{t_0 - k}{bN} \right) \frac{M_k(u_0)S_{k-1}(u_0)}{[\mu(u_0) + S_{k-1}(u_0)]^4} = O_p(1).
\]

Since \( \tilde{\mu}_{t_0,N} \overset{P}{\to} \mu(u_0) \), we have that \( I_{t_0,N} = O_p(1) \) and
\[
\| (\tilde{\mu}_{t_0,N} - \mu(u_0)) I_{t_0,N} \|_1 \leq |\tilde{\mu}_{t_0,N} - \mu(u_0)| \| I_{t_0,N} \|_1 \overset{P}{\to} 0.
\]

(96)

We now consider the second term \( J_{t_0,N} \). The problem here is the appearance of \( \tilde{\mu}_{t_0,N} \) in the denominator of \( J_{t_0,N} \). Let
\[
K(u_0) = \mathbb{E} \left( \left\{ \frac{[2\mu(u_0) + 2S_{k-1}(u_0)]^2}{[\mu(u_0) + S_{k-1}(u_0)]^6} \right\} (Z_k^2 + 1)\tilde{\sigma}_k^2(u_0) \| \tilde{X}_{k-1}(u_0) \|_1 \right).
\]

By using arguments identical to those given in the proof of Lemma 4.2, we can show that
\[
K_{t_0,N} := \sum_{k=p+1}^{N} \frac{1}{bN} W \left( \frac{t_0 - k}{bN} \right) \left\{ \frac{[\tilde{\mu}_{t_0,N} + \mu(u_0)] + 2S_{k-1}(u_0)]^2}{[\tilde{\mu}_{t_0,N} + S_{k-1}(u_0)]^2[\mu(u_0) + S_{k-1}(u_0)]^4} \right\} (Z_k^2 + 1)\tilde{\sigma}_k^2(u_0) \| \tilde{X}_{k-1}(u_0) \|_1 \overset{P}{\to} K(u_0).
\]

(97)

From the definition of \( K_{t_0,N} \), and using (29), we have that
\[
\| J_{t_0,N} \|_1 \leq |\tilde{\mu}_{t_0,N} - \mu(u_0)|^2 \sqrt{bN} K_{t_0,N} \leq K \sqrt{bN} \left\{ b^\beta + (bN)^{\frac{\beta}{1+\gamma}} \right\}^2 K_{t_0,N}
\]

If \( b \) is such that
\[
\sqrt{bN} \left\{ b^\beta + (bN)^{\frac{\beta}{1+\gamma}} \right\}^2 \to 0, \quad \text{as } N \to \infty,
\]
and taking into account (97), then \( \| J_{t_0,N} \|_1 \overset{P}{\to} 0 \). However, if (98) is not satisfied, we need to go through the same procedure in (93) of replacing the denominator in \( [\tilde{\mu}_{t_0,N} + S_{k-1}(u_0)]^2 \) in \( J_{t_0,N} \) with \( [\mu(u_0) + S_{k-1}(u_0)]^2 \) and taking differences as was done in (93). We must iterate this \( n \) times until
\[
\sqrt{bN} \left\{ b^\beta + (bN)^{\frac{\beta}{1+\gamma}} \right\}^n \to 0.
\]

(99)

At this point all the terms which contain \( \mu(u_0) \) in the denominator will converge to zero (using the martingale argument given above). Moreover, it is straightforward to show that the term which contains \( \tilde{\mu}_{t_0,N} \) in the denominator will be of order \( \sqrt{bN} \left\{ b^\beta + (bN)^{\frac{\beta}{1+\gamma}} \right\}^n \). By (99) this term goes to zero, hence we have that \( \| J_{t_0,N} \|_1 \overset{P}{\to} 0 \). Since the details for the \( n \)th iteration are similar to those of the first iteration we omit the details.
In summary, for every $b$ and $\eta$, there will always be an $n$ which satisfies (99), it then follows that
\[
\|Jt_{0,N}\|_1 \overset{p}{\rightarrow} 0.
\] (100)

Now by using (96) and (100) we have that
\[
\sqrt{bN}\|\nabla L_{t_{0}}^{nls(\hat{\mu}_{0,N}^{-1})}(u_0, \tilde{a}(u_0)) - \nabla L_{t_{0}}^{nls(\mu(u_0))}(u_0, \tilde{a}(u_0))\|_1 \leq 2(\|\hat{\mu}_{0,N} - \mu(u_0)\|I_{t_{0},N}\|_1 + \|J_{t_{0},N}\|_1) \overset{p}{\rightarrow} 0,
\]
which yields (92). This completes the proof of Lemma A.7. \(\square\)

**PROOF of Proposition 4.2** We first prove part (ii). It is easy to see that the following decomposition holds
\[
\nabla L_{t_{0},N}^{nls(\hat{\mu}_{0,N}^{-1})}(\tilde{a}(u_0)) = \nabla L_{t_{0}}^{nls(\mu(u_0))}(u_0, \tilde{a}(u_0)) + \nabla B_{t_{0},N}^{nls(\hat{\mu}_{0,N}^{-1})}(\tilde{a}(u_0)) + \{\nabla L_{t_{0}}^{nls(\hat{\mu}_{0,N}^{-1})}(u_0, \tilde{a}(u_0)) - \nabla L_{t_{0}}^{nls(\mu(u_0))}(u_0, \tilde{a}(u_0))\}.
\]

Since $\nabla L_{t_{0},N}^{nls(\hat{\mu}_{0,N}^{-1})}(\tilde{a}(u_0)) = 0$, we have that
\[
-\nabla L_{t_{0}}^{nls(\mu(u_0))}(u_0, \tilde{a}(u_0)) + \{\nabla L_{t_{0}}^{nls(\hat{\mu}_{0,N}^{-1})}(u_0, \tilde{a}(u_0)) - \nabla L_{t_{0}}^{nls(\mu(u_0))}(u_0, \tilde{a}(u_0))\} - \nabla B_{t_{0},N}^{nls(\hat{\mu}_{0,N}^{-1})}(\tilde{a}(u_0)) = \nabla^{2} L_{t_{0},N}^{nls(\hat{\mu}_{0,N}^{-1})}(\tilde{a}(u_0)) - \nabla L_{t_{0}}^{nls(\mu(u_0))}(u_0, \tilde{a}(u_0)).
\]

By using (88), we easily see that $\nabla^{2} L_{t_{0},N}^{nls(\hat{\mu}_{0,N}^{-1})}(\tilde{a}(u_0)) \overset{p}{\rightarrow} 2\ A_{nls(\mu(u_0))}(u_0)$, and, using (92), we have that
\[
\left\{\tilde{a}(u_0) - \tilde{a}(u_0)\right\} = \left\{-\nabla L_{t_{0}}^{nls(\mu(u_0))}(u_0, \tilde{a}(u_0)) - \nabla B_{t_{0},N}^{nls(\hat{\mu}_{0,N}^{-1})}(\tilde{a}(u_0))\right\} \times \left\{\frac{1}{2}\left\{A_{nls(\mu(u_0))}(u_0)\right\}^{-1} + o_p(1)\right\},
\]
which leads to
\[
\sqrt{bN}\left\{\tilde{a}(u_0) - \tilde{a}(u_0)\right\} + \frac{1}{2}\sqrt{bN}\left\{A_{nls(\mu(u_0))}(u_0)\right\}^{-1} \nabla B_{t_{0},N}^{nls(\hat{\mu}_{0,N}^{-1})} = \frac{1}{2}\sqrt{bN}\left\{A_{nls(\mu(u_0))}(u_0)\right\}^{-1} \nabla L_{t_{0}}^{nls(\mu(u_0))}(u_0, \tilde{a}(u_0)) + o_p(1).
\] (101)

By combining (71) and (101), we easily get (34).

We now prove part (i). By using (25) and (32), we see that
\[
-\frac{1}{2}\left\{\nabla h_{k,N}^{nls(\hat{\mu}_{0,N}^{-1})}(\tilde{a}(u_0)) - \nabla h_{k,N}^{nls(\hat{\mu}_{0,N}^{-1})}(u_0, \tilde{a}(u_0))\right\}
= \left(\frac{X_{k,N}^{2} - a^T(u_0)X_{k-1,N}}{\hat{\mu}_{0,N} + S_{k-1,N}} - \frac{\hat{X}_{k}^{2}(u_0) - a^T(u_0)\hat{X}_{k-1}(u_0)}{\hat{\mu}_{0,N} + S_{k-1}(u_0)}\right) \frac{X_{k-1,N}}{\hat{\mu}_{0,N} + S_{k-1,N}} + \left(\frac{X_{k-1,N}}{\hat{\mu}_{0,N} + S_{k-1,N}} - \frac{\hat{X}_{k-1}(u_0)}{\hat{\mu}_{0,N} + S_{k-1}(u_0)}\right) \left(\frac{\hat{X}_{k}^{2}(u_0) - a^T(u_0)\hat{X}_{k-1}(u_0)}{\hat{\mu}_{0,N} + S_{k-1}(u_0)}\right),
\] (102)
where $S_{k-1,N}$ and $S_{k-1}(u_0)$ are defined in (79). We now consider

$$\frac{X_{k,N}^2 - a^T(u_0)X_{k-1,N}}{\mu_{t_0,N} + S_{k-1,N}} - \frac{\tilde{X}_{k}^2(u_0) - a^T(u_0)\tilde{X}_{k-1}(u_0)}{\hat{\mu}_{t_0,N} + \hat{S}_{k-1}(u_0)}$$

$$= \left\{ X_{k,N}^2 - \tilde{X}_{k}^2(u_0) \right\} - a^T(u_0)\left\{ X_{k-1,N} - \tilde{X}_{k-1}(u_0) \right\}$$

By using Lemma 2.1, we have that

$$\left| \frac{X_{k,N}^2 - a^T(u_0)X_{k-1,N}}{\mu_{t_0,N} + S_{k-1,N}} - \frac{\tilde{X}_{k}^2(u_0) - a^T(u_0)\tilde{X}_{k-1}(u_0)}{\mu_{t_0,N} + S_{k-1}(u_0)} \right|$$

$$\leq K \left( \left| \frac{k}{N} - u_0 \right|^\beta + \left( \frac{p + 1}{N} \right)^\beta \right) \left\{ V_{k,N} + W_k \right\} + \left( 1 + \frac{1 + Z_k^2}{\mu_{t_0,N}} \right) \sum_{j=1}^p (V_{k,j,N} + W_{k-j}) \right\}$$

Similarly, we can show that

$$\left| \frac{\tilde{X}_{k-1,N}}{\mu_{t_0,N} + S_{k-1,N}} - \frac{\tilde{X}_{k-1}(u_0)}{\hat{\mu}_{t_0,N} + S_{k-1}(u_0)} \right|_1$$

$$\leq K \left( \left| \frac{k}{N} - u_0 \right|^\beta + \left( \frac{p + 1}{N} \right)^\beta \right) \left( 1 + \frac{1}{\mu_{t_0,N}} \right) \sum_{j=1}^p (V_{k-j,N} + W_{k-j}) \right\}$$

Since

$$\left| \frac{\tilde{X}_{k}^2(u_0) - a^T(u_0)\tilde{X}_{k-1}(u_0)}{\mu_{t_0,N} + S_{k-1}(u_0)} \right| \leq K \left( 1 + \frac{1 + Z_k^2}{\mu_{t_0,N}} \right)$$

by using substituting (103) and (104) into (102), we have that

$$\left\| \nabla h_k^{nls(\hat{\mu}_{t_0,N}^{-1})}(a(u_0)) - \nabla h_k^{nls(\mu_{t_0,N}^{-1})}(u_0, a(u_0)) \right\|_1$$

$$\leq 2K \left( \left| \frac{k}{N} - u_0 \right|^\beta + \left( \frac{p + 1}{N} \right)^\beta \right) \left\{ V_{k,N} + W_k \right\} + \left( 1 + \frac{1 + Z_k^2}{\mu_{t_0,N}} \right) \sum_{j=1}^p (V_{k,j,N} + W_{k-j})$$

Finally, since

$$\left\| \nabla B_{0,N}^{nls(\hat{\mu}_{t_0,N}^{-1})}(a(u_0)) \right\|_1 \leq \sum_{k=p+1}^N \frac{1}{bN} \left| \frac{t_0 - k}{bN} \right| \times$$

$$\times \left\| \nabla h_k^{nls(\hat{\mu}_{t_0,N}^{-1})}(a(u_0)) - \nabla h_k^{nls(\mu_{t_0,N}^{-1})}(u_0, a(u_0)) \right\|_1$$

by substituting (105) into (106), it can be shown that $\nabla B_{0,N}^{nls(\hat{\mu}_{t_0,N}^{-1})}(a(u_0)) = O_p(b^\beta)$, which proves (33). This completes the proof of Proposition 4.2. \qed
Acknowledgements

Piotr Fryzlewicz gratefully acknowledges the financial support of the Department of Mathematics, Imperial College, and David Hand. The work of Theofanis Sapatinas was partly supported by a ‘University of Cyprus – 2004 Research Grant’. The work of Suhasini Subba Rao was supported by DFG (DA 187/12-1) grant. Piotr Fryzlewicz would like to thank Theofanis Sapatinas for financial support and warm hospitality whilst visiting Nicosia to carry out part of this work. The authors would like to thank Rainer Dahlhaus and Efstathios Paparoditis for many useful and interesting comments.

References


41
Figure 1: Boxplots of the estimated parameter values for a tvARCH(2) process using the NLS(0) [NLS], QML [QML] and two-stage NLS [NLS-2] estimators for the various parameter values shown in the titles. The dashed lines indicate the true values of the given parameters.
Figure 2: Performance of the two-stage Kernel-NLS estimator on the USD/GBP exchange rate series. Top plots: estimated curve $\delta_0(u)$ against the squared data (left), positive part of the estimated curve $\delta_1(u)$ (right). Middle plots: empirical residuals from the fit (left), and Q-Q plot of the residuals against the quantiles of the standard normal (right). Bottom plots: sample autocorrelations of the residuals (left), and of their squares (right).