Nonparametric estimation of volatility models with with serially dependent innovations

by

Christian M. Dahl
Purdue University

and

Michael Levine
Purdue University

Technical Report # 04-04

Department of Statistics
Purdue University

August 2004

Nonparametric estimation of volatility models with serially dependent innovations*

Christian M. Dahl† — Department of Economics, Purdue University

Michael Levine — Department of Statistics, Purdue University

July 23, 2004

Abstract

We propose a nonparametric estimator of the conditional volatility function in a time series model with serial correlated innovations. We establish the asymptotic properties of the nonparametric estimator, as well as the estimator of the parameterized innovation process. The main advantage of our approach is that it does not require any knowledge of the specific form of the conditional volatility function. As pointed out by Pagan and Hong (in Nonparametric and Semiparametric Methods in Economic Theory and Econometrics, Cambridge University Press, 1991), Pagan and Ullah (JAE, 1988) and Pagan and Schwert (JoE, 1990) most parametric models, including ARCH and GARCH models, do not adequately capture the functional relationship between volatility and underlying economic factors. By applying our more flexible approach/estimator these shortcomings may be avoided. Finally, some simulations are provided.

1 Introduction

In this paper we consider estimation of a zero mean stationary time series process with an unknown and possibly time varying conditional volatility function and serial correlated innovations. A novel nonparametric estimator of the conditional volatility function is proposed and its asymptotic properties are established. Secondly, we characterize the estimated parameters of the serially correlated innovation process as a solution to a weighted least squares (WLS) problem, where the weights are given by the infinite

*The scientific notation follows Abadir and Magnus (2002).

†Corresponding author. Address: 403 West State Street, Purdue University, West Lafayette, IN 47907-2056. E-mail: dahlc@mgmt.purdue.edu. Phone: 765-494-4503. Fax: 765-496-1778.

dimensional nonparametric estimator of the conditional volatility function. This (semi-) parametric estimator belongs to the class of so-called MINPIN estimators and by using the framework of Andrews (1994) the asymptotic properties of the estimated parameters in the innovation process are readily established.

The main advantage of our approach is that it does not require any knowledge of the specific form of the conditional volatility function. As pointed out by Pagan and Hong (1991), Pagan and Ullah (1988) and Pagan and Schwert (1990) most parametric models, including ARCH and GARCH models, do not adequately capture the functional relationship between volatility and underlying economic factors. By applying our more flexible approach/estimator these shortcomings may be avoided.

Nonparametric estimation of volatility models in economics and finance has up until recently attracted far less attention relative to parametric estimation of the well established (G)ARCH family of models. An important recent contribution has been made by Fan and Yao (1998), see also Ziegelmann (2002), who derive a fully adaptive local linear nonparametric estimator of the conditional volatility function. The approach allows for the inclusion of strong mixing random variables in the conditional volatility function (as well as in the conditional mean function) and consequently the model can encompass a variety of non-linear ARCH specifications. To our knowledge, however , this nonparametric approach has not been widely applied outside the original paper by Fan and Yao (1998), which seems somewhat surprising in the light of the above mentioned critique of the parametric approach.

A common feature shared by the (G)ARCH family of models as well as the very general non-parametric volatility model of Fan and Yao (199) is that the innovation process of the time series of interest is assumed to be i.i.d. In our view this is a very critical assumption when the volatility function is allowed to be time dependent since it will - as we will demonstrate by a simple example - imply that the "parameters" entering the conditional mean function will be time varying and proportional to the increase in the conditional volatility over the most recent time period. The implication is that if the conditional mean function is estimated assuming time invariant parameters it will be inconsistent and the effect of this misspecification will carry over into the volatility estimation. In addition, as pointed out by Halunga and Orme (2004), misspecification test in (G)ARCH type volatility models will be asymptotically sensitive to misspecification of the conditional mean. Based on the MINPIN estimator classical statistical inference regarding the presence of serial correlation in the innovation process - and a potential misspecification of the fixed parameter conditional mean function - is easily performed.

Instead of relying on the estimated mean function as in the above mentioned papers when computing the conditional volatility function, we introduce a nonparametric estimator of the conditional volatility function based on the squared differences of the time series of interest. The history of this approach goes back to Hall, Kay and Titterington

(1990) and Müller and Stadtmüller (1993) among others, but have mainly been restricted to the fixed design case with independent and identically distributed innovations.[1] We generalize this approach for nonparametric estimation of the conditional volatility function allowing for the possibility of serial correlated innovations.

The paper is organized as follows: In Section 2 the model is defined and described. Section 3 introduces the nonparametric estimator of the conditional volatility function and its asymptotic properties are established. In Section 4 the estimated parameters driving the innovation process are defined and the asymptotic properties are characterized. Section 5 contains simulation results and finally Section 6 concludes.

2 The Model

Consider the following process for the time series of interest denoted $y_t \in \mathbb{R}$, $t = 1, 2, ..., T$

$$y_t = \sqrt{f(x_t)}\epsilon_t, \quad (1)$$

$$\epsilon_t = \phi\epsilon_{t-1} + v_t, \quad (2)$$

where $v_t \sim i.i.d.\ \mathrm{N}(0,1)$, $\phi \in \Theta = (-1;1)$, $f(x_t) \in C^p[0,1]$ and $x_t \in [0,1]$. As for x_t, it is presumed ordered and equispaced, i.e., $x_1 \le x_2 \le \ldots \le x_T$, and $x_t = \frac{t}{T}, t = 1, \ldots, T$. This assumption is standard in nonparametric function analysis. We will refer to the function $f(x)$ as the volatility function although it does not fully describe the variance-covariance structure of the model (1)-(2). As it is common in nonparametric function estimation, we assume that there exist p continuous derivatives of $f(x)$. The assumption that the time series v_t is Gaussian is not restrictive and has been introduced mainly for the sake of technical convenience.

Nonparametric regression with correlated errors has been considered fairly extensively by S. Marron, see, e.g. Chu and Marron (1991). However, the main focus has been to analyse to what extent correlation between observations influence the performance of model-selection methods such as cross-validation. Conditional volatility function estimation in case of correlated data case was to our knowledge first rigorously approached by Fan and Yao (1998) assuming a random design; specifically, they consider the bivariate vector (y_t, x_t) to be generated by a two-dimensional strictly stationary process with $g(x) = \mathrm{E}(y_t|x_t = x)$ and $f(x) = \mathrm{var}(y_t|x_t = x)$. They proposed an estimation procedure that relies on first estimating the conditional mean function $g(x)$ and then constructing the estimator of the conditional variance function $f(x)$ based on the estimated squared residuals. Their estimator is asymptotically fully adaptive to the choice of the conditional

[1] Observations are assumed to have been ordered while the errors are independently generated from a distribution that satisfies some regularity conditions such as the existence of the fourth moment, see, e.g., Hall et al (1990).

mean. A slightly modified estimator was proposed in Ziegelmann (2002). A paper by Lu (1999) introduces a nonparametric regression model with martingale difference sequence errors but is concerned only with estimating the mean function.

Note that the model (1)-(2) can be re-written as

$$y_t = g(x_t, x_{t-1}, y_{t-1}; \phi) + \sqrt{f(x_t)} v_t, \tag{3}$$

where

$$g(x_t, x_{t-1}, y_{t-1}; \phi) = \sqrt{\frac{f(x_t)}{f(x_{t-1})}} \phi y_{t-1}. \tag{4}$$

Since the innovation term in (3) is now i.i.d. the model very closely resembles the model of Fan and Yao (1998). However, there are two important differences; Firstly, (3) potentially involves 4 variables namely $(y_t, y_{t-1}, x_t, x_{t-1})$, whereas the Fan and Yao (1998) model is bivariate. Secondly, the conditional mean function given by (4) is parametric. Only in the case where $\phi = 0$, the model given by (1)-(2) becomes a special case of the model in Fan and Yao (1998). It is also important to notice that if $\phi \neq 0$ one would be likely to obtain an inconsistent estimate of $\text{var}(y_t | x_t, x_{t-1}, y_{t-1})$ based on residuals from a least squares regression of y_t on y_{t-1} as one would assume that the parameter in this regression was constant when it actually is given as $\sqrt{\frac{f(x_t)}{f(x_{t-1})}}\phi$. Remarkably, this is exactly the standard procedure when estimating (G)ARCH models, as a result of the i.i.d. assumption on the innovation process. We recommend to test the hypothesis that $\phi = 0$ before undertaking such procedure and a test statistic will be provided in Section 4.

Our main interest is the estimation of the variance-covariance structure of the model (1)-(2) We approach the estimation problem by constructing a two stage procedure that first gives us the estimator of $f(x)$ - denoted $\hat{f}(x)$ - based on the differences of observations y_t and then construct the estimator of ϕ - denoted $\widehat{\phi}$ - that utilizes the estimated variance function $\hat{f}(x)$. It turns out that $\widehat{\phi}$ is a MINPIN estimator as defined by Andrews (1994) which is very convenient when characterizing its asymptotic properties as Andrews (1994) provides all the tools necessary.

3 The estimator of $f(x_t)$

We follow the so-called difference sequence-based approach by Hall et al.(1990). The underlying idea is as follows: first, estimate $f(x)$ at a point x_t by $\Delta^2_{t,r} = (\sum_{j=1}^{r} d_i y_{j+t})^2$ where $\{d_j\}$ is a sequence of real numbers such that

1. $\sum_{j=0}^{r} d_j = 0$
2. $\sum_{j=0}^{r} d_j^2 = 1$

The sequence d_i is usually called the difference sequence of order r.[2] Secondly, apply a local smoother (for example, the Nadaraya-Watson local average smoother) to all $\Delta_{t,r}$ and produce the estimator

$$\hat{f}(x) = \frac{\sum_{t=1}^{T-r} \Delta_{t,r}^2 K\left(\frac{x-x_t}{h}\right)}{\sum_{t=1}^{T-r} K\left(\frac{x-x_t}{h}\right)}, \tag{5}$$

where $K(\cdot)$ denotes the kernel function. Hall et al.(1990) show that, when the fixed variance $f(x) \equiv \sigma^2$ is estimated, both variance and mean squared error of the difference-based estimator (5) are of the order T^{-1}. In other words, such an estimator enjoys the parametric rate of convergence. These results were further extended by Levins (2003), showing that in the general case of the non-constant variance function $f(x)$ the following is true: if $f(x) \in C^p[0,1]$ and $g(x) \in C^{p-1}[0,1]$ for some integer $p > 0$, as $r \to \infty$ the variance slowly tapers off at the rate of $\frac{1}{r}$ and, asymptotically, the optimal rate of convergence $T^{-\frac{2p}{2p+1}}$ is achieved. Asymptotically, the estimator is fully adaptive w.r.t. the mean function.[3] Taking this approach the following nonparametric estimator of the conditional volatility function is proposed:

1. Define the pseudoresiduals η_t as

$$\eta_t = \frac{y_{t+2} - y_t}{\sqrt{2}}, t = 1, \ldots, T-2. \tag{6}$$

2. Based on (6), define the variance estimator $\hat{f}(x)$ as

$$\hat{f}(x) = \frac{\sum_{t=1}^{T-2} \eta_t^2 K\left(\frac{x-x_t}{h}\right)}{\sum_{t=1}^{T-2} K\left(\frac{x-x_t}{h}\right)}. \tag{7}$$

It may seem somewhat surprising that the observation differences considered use the second lag instead of the more "mundane" first lag as done, for example, in Levins (2003). The main reason is to ensure that the resulting estimator of the variance function $f(x_t)$ is consistent. Indeed, it is easy to check that if the pseudoresiduals are based on $\Delta_{t,1}$ instead of $\eta_t \equiv \Delta_{t,2}^2$ the resulting estimator of $f(x_t)$ will converge to the $\frac{f(x_t)}{1+\phi}$ asymptotically. An

[2] Conditions (1) and (2) are not the only possible constraints one may want to impose on the difference sequence $\{d_i\}$. For example, it may be sensible to consider difference sequences such that not only (1) is true, but, more generally, also $\sum_i d_i = 0$ $\sum_i i d_i = 0, \ldots, \sum_i i^{p-1} d_i = 0$ while $\sum_i i^p d_i \neq 0$ for some integer $p > 0$. If the mean function $g(x)$ has no more than p terms in its Taylor expansion (in other words, it is a polynomial of at least an order p), differences based on a sequence that satisfies these additional conditions ensure that the bias of the estimator $\hat{f}(x)$ does not depend on the mean.

[3] When estimating a constant variance $f(x) \equiv \sigma_2$, Dette, Munk and Wagner(1998) show that in small samples the MSE of the estimator (5) depends heavily on $\int [g'(x)]^2 \, dx$ and $\int [g''(x)]^2 \, dx$, in particular as the order of the sequence r increases. The choice of the proper order r therefore becomes a fairly delicate affair. In fact, r plays the role similar to the one of the smoothing parameter. For details, see Dette, Munk and Wagner(1998).

important property of the AR(1) time series is that the difference between its variance, $\gamma_0 = \text{var}(y_t)$, and covariance, $\gamma_2 = \text{cov}(y_t, y_{t-2})$, equals unity which becomes very handy and ensures the consistency of the estimator given in (7).[4] Notice that the estimator (7) looks very similar to the Nadaraya-Watson estimator; it is different, however, because the transformed data η_t that is used to construct this estimator is not independent which is usually the case with the standard Nadaraya-Watson estimator. For definitions, see for example, Fan and Gijbels (1995).

We next turn to describing the most important asymptotic properties of the estimator (7). The degree of precision of the estimator (7) is characterized by the integrated mean squared error (IMSE)

$$IMSE = \int_0^1 (\hat{f}(x) - f(x))^2 \, dx \tag{8}$$

We first establish consistency in mean squared and, as a consequence, consistency in probability. The asymptotic rate of convergence is obtained as the direct consequence of these results. As a second step, asymptotic normality of (7) is established next.

Theorem 1 Let data be generated according to the model (1)-(2). Assume that the conditional volatility function $f(x) > 0$ is an element of $C^2[0,1]$ and $K(u)$ is a second order non-negative kernel function:$K(u) \geq 0$ for any $u \in [-1,1]$, $\mu_1 = \int K(u)\,du = 0$ and $\sigma_K^2 \equiv \mu_2 = \int u^2 K(u)\,du \neq 0$ and $R_K = \int K(u)^2\,du$. Let the mean function $g(x) \in C[0,1]$. Then the estimator given by (7) is consistent and its mean squared convergence rate is $O(T^{-4/5})$ with asymptotic integrated mean squared error (AIMSE) at the optimal bandwidth value given as

$$\begin{aligned} AIMSE_o &= T^{-4/5} * \left[\frac{\sigma_K^4}{4^{19/5}} \left[C(\phi) \int_0^1 (f(t))^2 \, dt \right]^{4/5} \left[\int_0^1 \left[\mathrm{D}^2 f(t) - \frac{\gamma_2 [\mathrm{D}^2 f(t)]^2}{f(t)} \right]^2 dt \right]^{1/5} \right. \\ &\quad \left. + \frac{C(\phi) \int_0^1 (f(t))^2 \, dt \, R_K}{4} \right], \end{aligned}$$

[4]Clearly, any positive definite quadratic form in the observations y_t can be used to estimate the variance function. We already mentioned that (6) is needed to obtain the consistent estimator of $\hat{f}(x)$. Another purpose of using (6) and not, say, $\eta_t = y_t$ is that we hope to reduce the influence of the (potentially nonzero) unknown mean $g(x_t)$ on the bias of the variance function estimator $f(x_t)$; indeed, by using (6) the constant term in a Taylor series expansion of the function $g(x_t)$ cancels. Levins (2003) shows that in the case of i.i.d. innovations and $g(x_t) \neq 0$ the bias term of the estimator $\hat{f}(x_t)$ that is due to the mean $g(x_t)$ is proportional to $\int [g'(x)]^2 \, dx$ if pseudoresiduals defined by (6) are used. For more discussion on this topic, see Levins (2003).

where $C(\phi)$ is a constant that depends on ϕ only. The optimal bandwidth is of the order $T^{-1/5}$ and equals

$$h_o = T^{-1/5}\left[\frac{C(\phi)\int_0^1 (f(t))^2\,dt}{4\sigma_K^4 \int_0^1 \left[D^2 f(t) - \frac{\gamma_2 [D^2 f(t)]^2}{f(t)}\right]^2 dt}\right]^{1/5}.$$

Proof of Theorem 1 See the Mathematical Appendix. □

A few remarks are in order here. First, note that the quadratic functional

$$\int_0^1 \left[D^2 f(t) - \frac{\gamma_2 [D^2 f(t)]^2}{f(t)}\right]^2 dt \tag{9}$$

characterizes the degree of curvature of the function $f(x)$ corrected for the correlation present in the data. The larger (9), the smaller the bandwidth we have to choose according to (3).

Secondly, note that when the innovations are independently distributed we have $\gamma_2 = 0$, $C(0) = 12$ and the bias becomes $Bias\left(\hat{f}(x)\right) = \frac{h^2\sigma_K^2}{2} + o(h^2)$ exactly as in Levins (2003). The AIMSE in this case is also identical to Levins (2003). Levins' (2003) estimator is based on defining the pseudoresiduals as $(y_t - y_{t-1})^2$ but not surprisingly this now turns out not to matter asymptotically given the assumptions of Theorem 1, whenever $\phi = 0$.

Also, the rates of convergence are the same as those obtained for the kernel regression estimator of the mean function under identical smoothness requirements (see, for example, Simonoff (1996))

Finally, as an immediate consequence of this theorem , $\hat{f}(x)^{-k} \xrightarrow{p} f(x)^{-k}$ for any $k > 0$.

Since the estimator $\hat{f}(x)$ given by (6) and (7) converges in L_2-sense, it also converges in probability at the rate $O_p\left(\frac{1}{\sqrt{Th}}\right)$. In particular,

$$\sqrt{Th}\left(\hat{f}(x) - f(x) - Bias\left(\hat{f}(x)\right)\right) \xrightarrow{p} 0, \tag{10}$$

where

$$Bias\left(\hat{f}(x)\right) = \frac{h^2\sigma_K^2}{2}\left[D^2 f(x) - \frac{\gamma_2 [D^2 f(x)]^2}{f(x)}\right] + o(h^2). \tag{11}$$

In the following Theorem 2 we establish that $\hat{f}(x)$ is asymptotically normally distributed with mean

$$E\left(\hat{f}(x)\right) = f(x) + Bias\left(\hat{f}(x)\right). \tag{12}$$

and variance

$$\mathrm{var}\left(\hat{f}(x)\right) = \frac{C(\phi)\,(f(x))^2}{4Th} R_K. \tag{13}$$

Note that expressions (12) and (13) are derived and used in the proof of Theorem 1 in the Mathematical Appendix.

Theorem 2 Let the Assumptions of Theorem 1 hold. Then,

$$\hat{f}(x) \xrightarrow{d} \mathrm{N}\left(\mathrm{E}\left(\hat{f}(x)\right), \operatorname{var}\left(\hat{f}(x)\right)\right). \tag{14}$$

as $T \to \infty$, $h \to 0$ and $Th \to \infty$, where $\mathrm{E}\left(\hat{f}(x)\right)$ and $\operatorname{var}\left(\hat{f}(x)\right)$ are defined in (12) and (13) respectively.

Proof of Theorem 2 See the Mathematical Appendix. □

4 The estimator of ϕ

For notational simplicity we define and use $\sigma_t = \sqrt{f(x_t)}$ in this section. Following Andrews (1994) we use a GMM approach to estimate ϕ by defining the following loss function d_t

$$d_t(\sigma_t, \sigma_{t-1}, y_t, y_{t-1}; \phi) = (m_t(\sigma_t, \sigma_{t-1}, y_t, y_{t-1}; \phi))^2, \tag{15}$$

where m_t (denoting a moment condition) is given as

$$\begin{aligned} m_t(\sigma_t, \sigma_{t-1}, y_t, y_{t-1}; \phi) &= \left(\sigma_t^{-1} y_t - \sigma_{t-1}^{-1} \phi y_{t-1}\right)\left[\sigma_{t-1}^{-1} y_{t-1}\right] \\ &= \sigma_t^{-1}\sigma_{t-1}^{-1} y_t y_{t-1} - \sigma_{t-1}^{-2}\phi y_{t-1}^2 \\ &= v_t \epsilon_{t-1}. \end{aligned} \tag{16}$$

The so-called MINPIN estimator $\widehat{\phi}$, see Andrews (1994) for a definition, is then given as

$$\widehat{\phi}_T = \min_{\phi \in \Theta} \frac{1}{2T} \sum_{t=1}^{T} d_t(\widehat{\sigma}_t, \widehat{\sigma}_{t-1}, y_t, y_{t-1}; \phi).$$

or equivalently as a solution to

$$\frac{1}{T} \sum_{t=1}^{T} m_t\left(\widehat{\sigma}_t, \widehat{\sigma}_{t-1}, y_t, y_{t-1}; \widehat{\phi}_T\right) = 0. \tag{17}$$

Consequently, by solving (17), we can write

$$\widehat{\phi}_T = \left(\frac{1}{T} \sum_{t=1}^{T} \widehat{\sigma}_{t-1}^{-2} y_{t-1}^2\right)^{-1} \left(\frac{1}{T} \sum_{t=1}^{T} \widehat{\sigma}_t^{-1} \widehat{\sigma}_{t-1}^{-1} y_t y_{t-1}\right). \tag{18}$$

Immediately the following asymptotic results can be established.

Theorem 3 Let the Assumptions of Theorem 1 hold. Then, the MINPIN estimator given by (18) is consistent with respect to the true population parameter ϕ_0, i.e., $\hat{\phi}_T \xrightarrow{p} \phi_0$.

Proof of Theorem 3 See the Mathematical Appendix. □

Theorem 4 Let the Assumptions of Theorem 1 hold and assume in addition that $\mathrm{D}^k K(1) = \mathrm{D}^k K(-1) = 0$ and $\mathrm{D}^k \sigma_t^2 \in C^2[0,1]$ for $k = 1, 2$. Then,

$$\sqrt{T}\left(\widehat{\phi}_T - \phi_0\right) \xrightarrow{d} \mathrm{N}\left(0, 1 - \phi_0^2\right). \tag{19}$$

as $T \to \infty$, $h \to 0$ and $Th \to \infty$.

Proof of Theorem 4 See the Mathematical Appendix. □

Although the proof of Theorem 4 is based on the very general setup provided by Andrews (1994) it could also have been carried out using the approach by Robinson (1987), i.e., by establishing asymptotic equivalence of $\widehat{\phi}_T$ and

$$\widetilde{\phi}_T = \left(\frac{1}{T}\sum_{t=1}^{T} \sigma_{t-1}^{-2} y_{t-1}^2\right)^{-1} \left(\frac{1}{T}\sum_{t=1}^{T} \sigma_t^{-1}\sigma_{t-1}^{-1} y_t y_{t-1}\right) \tag{20}$$

which requires verifying the condition $\sqrt{T}\left(\widehat{\phi}_T - \widetilde{\phi}_T\right) \xrightarrow{p} 0$.

Now it is obvious that the estimator $\widehat{\phi}_T$ will not depend on the first stage estimator of the function $f(x_t)$ asymptotically and will be asymptotically equivalent to the maximum likelihood estimator of ϕ_0 given that ϵ_t was observable. Consequently the MINPIN estimator $\widehat{\phi}_T$ will be asymptotically efficient. In addition, since $\widehat{\phi}_T$ is an efficient estimator of ϕ_0, and ϕ_0 is the only unknown parameter in the model, the estimator will be adaptive, see, e.g., Andrews (1994) page 59. Finally, it is noteworthy that $\widehat{\phi}_T$, as many semiparametric estimators, will converge to ϕ_0 at the parametric $\sqrt{T}-$ rate.

5 Simulations

In this section the small sample properties of the estimators of $\widehat{f}(x)$ and $\widehat{\phi}_T$ are studied using simulations. We consider the observational data being generated by (1)−(2) for 6 alternative choices of volatility functions, assuming that the true population value of ϕ (denoted ϕ_0) equals 0.6. The volatility functions are specified is Table 1. The specifications of $f(x)$ applied in Model 1 - 3 are included as they are fundamental in econometrics/statistics and are typically included in graduate econometric textbook-chapters on heteroskedasticity in regression models, see, for example Ruud (2000) and Greene (2003). The specification of $f(x)$ in Model 4 is adapted from Example 1 in Fan and Yao (1998). They suggest this volatility function specification in modelling the yields of the US Treasury Bill from secondary markets. Model 5 is also inspired by Fan and Yao (1998), in particular, the choice of $f(x)$ is identical to the volatility function in their Example 2. Finally, the volatility function in Model 6 is taken from Haerdle and

Table 1: Alternative data generating processes

	Specifications
Model 1	$y_t = x_t \epsilon_t$
Model 2	$y_t = \sqrt{x_t^2} \epsilon_t$
Model 3	$y_t = \sqrt{exp(x_t)} \epsilon_t,$
Model 4	$y_t = \sqrt{0.02 x_t^{1.4}} \epsilon_t,$
Model 5	$y_t = \sqrt{0.4 \exp(-2x_t^2) + 0.2} \epsilon_t$
Model 6	$y_t = \sqrt{\varphi(x_t + 1.2) + 1.5\varphi(x_t - 1.2)} \epsilon_t$

Table 2: Simulated MSE of $\hat{f}(x_t)$ (as described by (21)) under alternative volatility function specifications and sample sizes. The number of Monte Carlo replications equals 1000.

	$T = 100$ MSE($\hat{f}(x_t)$)	$T = 1000$ MSE($\hat{f}(x_t)$)	$T = 2000$ MSE($\hat{f}(x_t)$)
Model 1	0.0736	0.0443	0.0407
Model 2	0.0443	0.0143	0.0098
Model 3	1.9330	1.8977	1.9062
Model 4	0.0002	0.0001	0.0001
Model 5	0.0126	0.0023	0.0014
Model 6	0.0315	0.0049	0.0029

Tsybakov (1997). We consider first the precision of the nonparametric estimator given by (7) based on the simulated mean squared error computed as

$$\text{MSE}\left(\hat{f}(x_t)\right) = \frac{1}{M}\sum_{s=1}^{M}\left(\frac{1}{T}\sum_{t=1}^{T}\left(\hat{f}_s(x_t) - f(x_t)\right)^2\right) \tag{21}$$

where M denotes the number of Monte Carlo replications, T equals the sample size. The results for the specifications of $f(x)$ given in Table 1 and $T = 100, 1000, 2000$ are summarized in Table 2. From the results in Table 2 we see that the precision of the nonparametric estimators improves substantially when the sample size increases from $T = 100$ to $T = 1000$ as expected. Overall the results are very encouraging. Only in terms of Model 3 the estimator seem to be performing less satisfying with very moderate improvements in precision as the sample size increases.

Next, lets turn to the properties of the MINPIN estimator given by (18). To first

Table 3: Simulated precision of $\hat{\phi}$ (as described by ()) under alternative volatility function specifications and sample sizes.

	$T = 100$		$T = 1000$		$T = 2000$	
	$\hat{\phi}$	(s.e.)	$\hat{\phi}$	(s.e.)	$\hat{\phi}$	(s.e.)
Model 1	0.5888	(0.0854)	0.5990	(0.0264)	0.5983	(0.0173)
Model 2	0.5891	(0.0948)	0.5990	(0.0300)	0.5986	(0.0200)
Model 3	0.5871	(0.0806)	0.5986	(0.0244)	0.5979	(0.0173)
Model 4	0.5715	(0.1235)	0.6025	(0.0390)	0.6033	(0.0258)
Model 5	0.5835	(0.0817)	0.5983	(0.0258)	0.5976	(0.0184)
Model 6	0.5846	(0.0807)	0.5982	(0.0249)	0.5978	(0.0183)

analyze the precision of the estimator in small samples we define

$$\hat{\phi}_{mc} = \frac{1}{M}\sum_{s=1}^{M}\hat{\phi}_s \tag{22}$$

$$\text{var}\left(\hat{\phi}_{mc}\right) = \frac{1}{M-1}\sum_{s=1}^{M}\left(\hat{\phi}_s - \hat{\phi}_{mc}\right)^2 \tag{23}$$

where

$$\hat{\phi}_s = \left(\sum_{t=2}^{T}\widehat{\epsilon}_{st-1}\right)^{-1}\left(\sum_{t=2}^{T}\widehat{\epsilon}_{st-1}\widehat{\epsilon}_{st}\right) \tag{24}$$

$$\widehat{\epsilon}_{st} = \frac{y_{st}}{\sqrt{\hat{f}_s(x_t)}}$$

and $\hat{f}_s(x_t)$ denotes the estimator of $f_s(x_t)$ for $s = 1, 2, ..., M$. Again data is generated according to the six models in Table 1 and for each replication $\hat{\phi}_s$ is computed according to (24). Based on each sequence $\left\{\hat{\phi}_s\right\}_{s=1}^{M}$ we compute the summary statistics given by (22) and (23). According to Theorem 3, we would expect to see $\hat{\phi}_{mc}$ getting closer to $\phi_0 = 0.6$ and var $\left(\hat{\phi}_{mc}\right)$ approaching zero as the sample size increases. The results are reported in Table 3. These results clearly indicate that the sample properties of the MINPIN estimator $\hat{\phi}_t$ are good across all the models considered and that the estimator works well even for small samples, i.e., for $T = 100$. Finally, we consider the sample density of $\widehat{d_T} = \sqrt{T}\left(\widehat{\phi}_T - \phi_0\right)/\sqrt{1-\phi_0^2}$ which according to Theorem 4 should converge to a standard normal density. In Figure 1 the density of $\widehat{d_T}$ for each of the six model of Table 1 based on $T = 100, 1000, 2000$ is depicted together with the standard normal density. From the figure we see clearly that the simulation results confirms the prediction of Theorem 4. No severe small sample biases seems to be present in any of the pictures and the small sample approximation to the standard normal in general seems to be very good.

The simulation results presented in this section all seem to indicate the small sample properties of the nonparametric estimator and the MINPIN estimator are very satisfactory.

6 Conclusion and directions for future research

In this paper we consider estimation of a zero mean stationary time series process with an unknown and possibly time varying conditional volatility function $f(x)$ and serial correlated innovations. A novel nonparametric estimator of the conditional volatility function is proposed and its asymptotic properties are established. The main advantage of this approach is that it does not require any knowledge of the specific form of the conditional volatility function. The form of the volatility function estimator $\hat{f}(x)$ ensures that, in case of the constant nonzero mean, the mean function $g(x)$ does not influence the mean squared error of the estimator $\hat{f}(x)$. The estimator possesses a rate of convergence equal to $T^{-2p/2p+1}$ if the function $f(x) \in C^p[0,1]$. This is the convergence rate that is commonly encountered in nonparametric function estimation. We conjecture that the estimator is asymptotic minimax among all possible estimators of $f(x)$. Research on verifying this conjecture is ongoing. It is also useful to recall that similar estimators ensuring the mean-related bias term of $\hat{f}(x)$ is zero can be constructed when the mean function $g(x)$ is a polynomial of an arbitrary order $p > 0$. Construction of such estimators in the case where the time series of interest possesses a nonzero trend is an interesting extension to be considered.

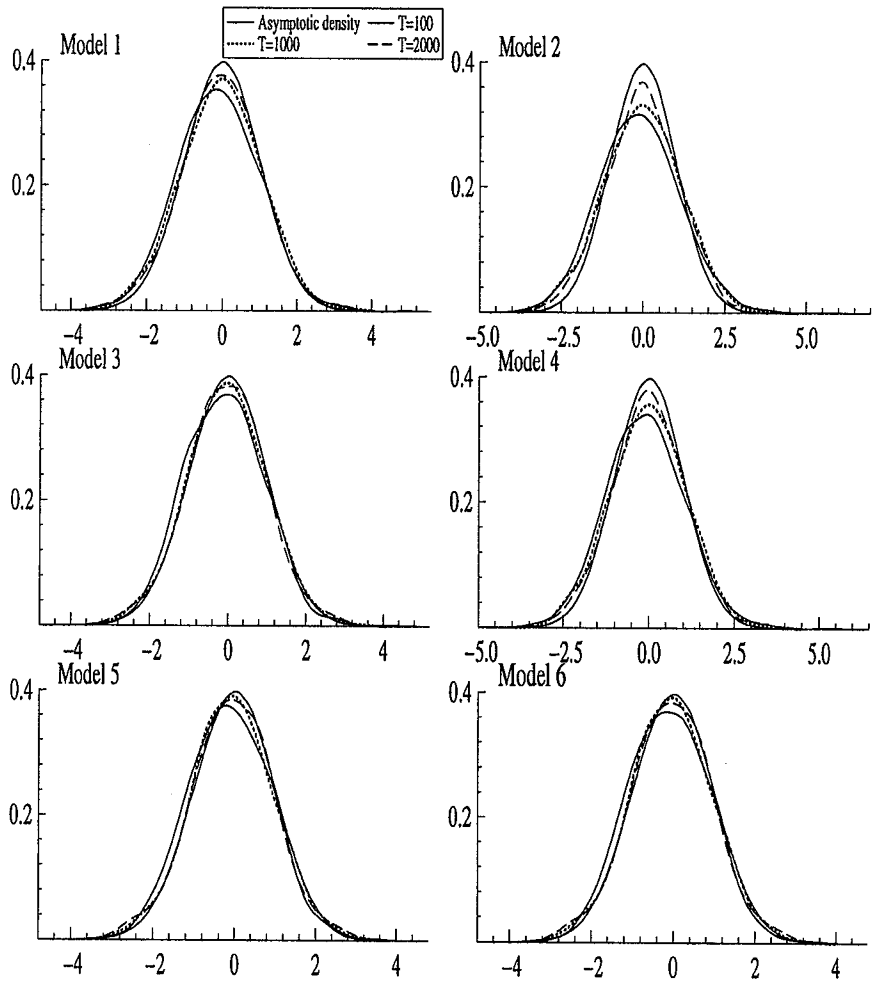
Secondly, we characterize the estimated parameter of the serially correlated innovation process as a solution to a weighted least squares (WLS) problem, where the weights are given by the infinite dimensional nonparametric estimator of the conditional volatility function. This semiparametric estimator belongs to the class of so-called MINPIN estimators and by using the framework of Andrews (1994) the asymptotic properties of the estimated parameter characterizing the innovation process are readily established. This estimator is asymptotically efficient, adaptive, and enjoys the parametric rate of convergence $T^{-1/2}$.

Based on simulation studies the finite sample properties of the proposed estimators are investigated and the findings are very encouraging.

It is important to stress that the present analysis has been limited to one particular dependence structure in the innovation process, namely the simple AR(1) structure. It is not entirely clear at this stage if the method used to construct the nonparametric estimator can be successfully generalized to any arbitrary AR(p) structure for $p > 0$.

Density

Figure 1: Small sample (simulated) densities and the asymptotic density of $\sqrt{T}\left(\widehat{\phi}_T - \phi_0\right)/\sqrt{1-\phi_0^2}$ under alternative volatility function specifications. The number of Monte Carlo replications equals 1000.



If, however, this is feasible we conjecture that a MINPIN estimator will exist with very similar properties to the one developed here.

An additional and natural extension to the nonparametric function estimation procedure would be to relax the assumption that the function domain is compact (e.g. $[0, 1]$). This requirement is rarely realistic in economics, where x_t cannot easily be bounded and in addition typically will be stochastic, i.e., contain lagged values of the dependent variable. It will be highly desirable to develop a method that would provide comparable asymptotic results when the domain of the function $f(x)$ is not compact and x is a stochastic vector, as in Fan and Yao (1998). Such a method would facilitate the estimation of nonparametric (G)ARCH models with serially dependent innovation processes. This is currently not feasible using our estimator but we believe that our model and the suggested approach provides an important first step in this direction.

References

Abadir, K. M. and J. R. Magnus (2002). Notation in econometrics: a proposal for a standard. *The Econometrics Journal 5*, 76–90.

Andrews, D. W. K. (1987). Consistency in nonlinear econometric models: A generic uniform law of large numbers. *Econometrica 55*, 1465–1471.

Andrews, D. W. K. (1994). Asymptotics for semiparametric econometric models via stochastic equicontinuity. *Econometrica 62*, 43–72.

Casella, G. and R.Berger (2001). *Statistical Inference.* Duxbury.

Chu, C. K. and J. S. Marron (1991). Choosing a kernel regression estimator. *Statistical Science 6*, 404–436.

Davidson, J. (1994). *Stochastic Limit Theory: An Introduction for Econometricians.* Oxford University Press.

Dow, M. (2003). Explicit inverses of toeplitz and associated matrices. *ANZIAM J. E 44*, 185215.

Fan, J. and I. Gijbels (1995). Data-driven bandwidthselection in local polynomial fitting: variable bandwidthand spatial adaptation. *J. Roy. Statist. Soc. Ser. B 57*, 371394.

Fan, J. and Q. Yao (1998). Efficient estimation of conditional variance functions in stochastic regression. *Biometrika 85*, 645–660.

Greene, W. H. (2003). *Econometric Analysis.* Prentice Hall.

Hall, P., J. W. Kay, and D. M. Titterington (1990). Asymptotically optimal difference based estimation of variance in nonparametric regression. *Biometrika 77*, 521–528.

Halunga, A. G. and C. Orme (2004). Testing for nonlinearities in garch models. *Unpublished manuscript, University of Manchester.*

Hardle, W. and A. Tsybakov (1997). Local polynomial estimators of the volatility function in nonparametric autoregression. *Journal of Econometrics 81*, 223–242.

Jacod, J. and P. Protter (2002). *Probability Essentials.* Springer-Verlag.

Levins, M. (2003). Variance estimation for nonparametric regression and its applications. *PhD Dissertation, University of Pennsylvania, 2003.*

Moser, B. K. (1996). *Linear Models: A Mean model Approach.* Academic Press.

Muller, H.-G. and U. Stadtmuller (1993). On variance function estimation with quadratic forms. *Journal of Statistical Planning and Inference 55*, 213–231.

Pagan, A. R. and Y. S. Hong (1991). Nonparametric estimation and the risk premium. In W. A. Barnett, J. Powell, and G. E. Tauchen (Eds.), *Nonparametric and*

Semiparametric Methods in Econometrics and Statistics, pp. 51–75. Cambridge, Cambridge University Press.

Pagan, A. R. and G. W. Schwert (1990). Alternative models for conditional stock volatility. *Journal of Econometrics 45*, 267–290.

Pagan, A. R. and A. Ullah (1988). The econometric analysis of models with risk terms. *Journal of Applied Econometrics 3*, 87–105.

Robinson, P. M. (1987). Asymptotically efficient estimation in the presence of heteroskedasticity of unknown form. *Econometrica 55*, 875–891.

Ruud, P. A. (2000). *An Introduction to Classical Econometric Theory.* Oxford.

Simonoff, J. S. (1996). *Smoothing Methods in Statistics.* Springer-Verlag.

White, H. (1984). *Asymptotic Inference for Econometricians.* Academic Press.

Ziegelmann, F. (2002). Nonparametric estimation of volatility functions: The local exponential estimator. *Econometric Theory 18*, 985–991.

Mathematical Appendix

Proof of Theorem 1 We begin by finding the expected value of η_t^2 given by (6). Since the function $f(x)$ is twice continuously differentiable on $[0,1]$, we can use the following Taylor series expansion

$$f(x_t) = f(x) - \mathrm{D}f(x)(x - x_t) + \frac{\mathrm{D}^2 f(x)(x-x_t)^2}{2} + o(x_t - x)^2,$$

It is easy to check that the existence of the two continuous derivatives of $f(x)$ guarantees that

$$f(x_t) = f(x) - \mathrm{D}f(x)(x - x_t) + \frac{\mathrm{D}^2 f(x)(x-x_t)^2}{2} + o(h)^2.$$

Note that we can write $\eta_t^2 = \frac{1}{2}\left(f(x_t)\epsilon_t^2 + f(x_{t-2})\epsilon_{t-2}^2 - \sqrt{f(x_t)f(x_{t-2})}\epsilon_t\epsilon_{t-1}\right)$. Using the Taylor expansion for $f(x_t)$ and $f(x_{t-2})$ we have

$$\begin{aligned}\sqrt{f(x_t)f(x_{t-2})} &= \Big((f(x))^2 + \mathrm{D}f(x)f(x)[(x-x_t)+(x-x_{t-2})] + [\mathrm{D}f(x)]^2(x-x_t)(x-x_{t-2}) \\ &\quad + \frac{f(x)\mathrm{D}^2 f(x)}{2}[(x-x_t)^2 + (x-x_{t-2})^2] + o(h^2)\Big)^{\frac{1}{2}}.\end{aligned}$$

As $\sqrt{1+x} = 1 + \frac{1}{2}x + o(x)$ for small x we obtain the following asymptotic expansion

$$\begin{aligned}\sqrt{f(x_t)f(x_{t-2})} &= f(x) + \frac{1}{2}\frac{\mathrm{D}f(x)}{f(x)}[(x-x_t)+(x_t-x_{t-2})] + \frac{1}{2}\frac{[\mathrm{D}f(x)]^2}{f(x)}(x-x_t)(x-x_{t-2}) \\ &\quad + \frac{1}{4}\mathrm{D}^2 f(x)[(x-x_t)^2 + (x-x_{t-2})^2] + o(h^2).\end{aligned}$$

Using that $\mathrm{E}(\epsilon_t) = 0$, $\mathrm{var}(\epsilon_t) \equiv \gamma_0 = \frac{1}{1-\phi^2}$ and $\mathrm{cov}(\epsilon_t, \epsilon_{t-l}) \equiv \gamma_l$ yields

$$\begin{aligned}\mathrm{E}(\eta_t) &= (\gamma_0 - \gamma_2)f(x) + \gamma_0 \mathrm{D}f(x)\left[(x-x_t)+(x-x_{t-2})\right] \qquad (25)\\ &\quad -\gamma_2\frac{\mathrm{D}f(x)}{f(x)}\left[(x-x_t)+(x-x_{t-2})\right] + \frac{1}{2}\gamma_0\mathrm{D}^2 f(x)\left[(x-x_t)^2 + (x-x_{t-2})^2\right] \\ &\quad -\gamma_2\frac{[\mathrm{D}f(x)]^2}{f(x)}\left[(x-x_t)(x-x_{t-2})\right] - \frac{1}{2}\gamma_2\mathrm{D}^2 f(x)\left[(x-x_t)^2 + (x-x_{t-2})^2\right].\end{aligned}$$

Since the expectation is linear

$$\mathrm{E}\left(\hat{f}(x)\right) = \frac{\sum_{t=1}^{T-2}\mathrm{E}\left(\eta_t^2\right)K\left(\frac{x-x_t}{h}\right)}{\sum_{t=1}^{T-2}K\left(\frac{x-x_t}{h}\right)}. \qquad (26)$$

Next, let us introduce the new variable $u_t = \frac{x-x_t}{h}$ and notice that $\sum_{i=1}^{T-2}K\left(\frac{x-x_t}{h}\right) = Th * \frac{1}{Th}\sum_{t=1}^{T-2}K\left(\frac{x-x_t}{h}\right) \approx Th\int K(u)\,du = Th$ asymptotically. As $(\gamma_0 - \gamma_2) = 1$, the first term in (25) is equal to $f(x)$ and consequently the bias can be expressed as

$$\begin{aligned} Bias\left(\hat{f}(x)\right) &= \frac{1}{2Th} * \left[2\gamma_0 \mathrm{D}f(x) \sum_{t=1}^{T-2} 2u_t K(u_t) - \gamma_2 \frac{\mathrm{D}f(x)}{f(x)} \sum_{t=1}^{T-2} 2u_t K(u_t)\right] + (27) \\ & \frac{1}{2Th} \left[\gamma_0 \mathrm{D}^2 f(x) h^2 \sum_{t=1}^{T-2} u_t^2 K(u_t) - \gamma_2 \frac{\mathrm{D}f(x)^2}{f(x)} \sum_{t=1}^{T-2} u_t^2 K(u_t) \right. \\ & \left. -\gamma_2 \mathrm{D}^2 f(x) \sum_{t=1}^{T-2} u_t^2 K(u_t) \right]. \end{aligned} \tag{28}$$

The first group in (27) consists of the first-order terms that are asymptotically equal to zero because our kernel $K(u)$ is the first-order kernel; indeed, the first one of these terms is equivalent to $4\gamma_0 f(x) \int uK(u)\,du = 0$, while the second one asymptotically equals $-2\gamma_2 \frac{\mathrm{D}f(x)}{f(x)} \int K(u)\,du = 0$. As a result, the bias only depend on the second order terms. After taking the limit as $T \to \infty$, $h \to 0$ and $Th \to \infty$ the Riemann sums on the right-hand side of (27) become integrals. In particular,

$$\begin{aligned} Bias\left(\hat{f}(x)\right) &= \frac{h^2\sigma_K^2}{2}\left[\gamma_0 \mathrm{D}^2 f(x) - \frac{\gamma_2[\mathrm{D}f(x)]^2}{f(x)} - \gamma_2 \mathrm{D}^2 f(x)\right] + o(h^2) \\ &= \frac{h^2\sigma_K^2}{2}\left[\mathrm{D}^2 f(x) - \frac{\gamma_2[\mathrm{D}^2 f(x)]^2}{f(x)}\right] + o(h^2), \end{aligned} \tag{29}$$

as $\gamma_0 - \gamma_2 = 1$.

Now let us proceed with computation of the asymptotic variance of $\hat{f}(x)$. First, recall that the denominator (7) is a constant and so we need only to compute the variance of the numerator. By definition of pseudoresiduals η_t^2, it is clear that they form a dependent data sequence, i.e., η_1^2 is correlated with η_3^2 and η_3^2 is correlated with η_5^2 etc., while η_2^2 is correlated with η_4^2 and η_4^2 is correlated with η_6^2 etc. Keeping this in mind we find that

$$\begin{aligned} \mathrm{var}\left(\sum_{t=1}^{T-2} \eta_t^2 K\left(\frac{x-x_t}{h}\right)\right) &= \sum_{t=1}^{T-2} \mathrm{var}(\eta_t^2)\left(K\left(\frac{x-x_t}{h}\right)\right)^2 \\ &+ \sum_{|t-u|=2} Cov\,(\eta_t^2, \eta_u^2) K\left(\frac{x-x_t}{h}\right) K\left(\frac{x-x_u}{h}\right). \end{aligned} \tag{30}$$

With respect to the first term in (30), note that $\mathrm{var}(\eta_t^2) \approx (f(x))^2 \mathrm{var}\left((\epsilon_t - \epsilon_{t-2})^2\right)$ asymptotically while it can be shown by straightforward calculations that

$$\mathrm{var}\left((\epsilon_t - \epsilon_{t-2})^2\right) = 2 + 6\phi^2 + 3(1+\phi^2)^2 + \frac{3(1-\phi^2)(1+2\phi^2)}{1+\phi^2} + \frac{(1-\phi^2)^2}{1+\phi^2} \equiv C_1(\phi), \tag{31}$$

where $C_1(\phi)$ depends only on ϕ. Therefore, up to the second order term

$$\mathrm{var}(\eta_t^2) = (f(x))^2\, C_1(\phi),$$

asymptotically and the first term divided by the denominator can be represented (recall that $\sum_{i=1}^{T-2} K\left(\frac{x-x_t}{h}\right) = Th$ asymptotically) as

$$\frac{C_1(\phi)\left(f(x)\right)^2}{4(Th)^2}\sum_{t=1}^{T-2}\left(K\left(\frac{x-x_t}{h}\right)\right)^2. \tag{32}$$

In the same way as before, introducing the new variable $u_t = \frac{x-x_t}{h}$ and treating (32) as a Riemann sum we obtain the asymptotic expression for the first term in (30) as

$$\frac{C_1(\phi)\left(f(x)\right)^2}{4(Th)^2}R_K, \tag{33}$$

where $R_K = \int \left(K(u)\right)^2 du$. Now, let us consider the second term in (30). In this case, again, up to the second order Taylor series term we have

$$\mathrm{cov}(\eta_t^2, \eta_{t-2}^2) = \left(f(x)\right)^2 \mathrm{cov}\left((\epsilon_t - \epsilon_{t-2})^2, (\epsilon_{t-2} - \epsilon_{t-4})^2\right).$$

Covariance calculations are fairly long and tedious but can be done in straightforward manner; the result is

$$\mathrm{cov}\left((\epsilon_t - \epsilon_{t-2})^2, (\epsilon_{t-2} - \epsilon_{t-4})^2\right) \equiv C_2(\phi) \tag{34}$$

$$= \frac{6\phi^4 - 2\phi^2 - 3}{(\phi^2 - 1)(1 - \phi^4)}. \tag{35}$$

Thus, $\mathrm{cov}(\eta_t^2, \eta_{t-2}^2) \approx C_2(\phi)\left(f(x)\right)^2$ and the second term after division by the denominator is

$$\frac{C_2(\phi)\left(f(x)\right)^2}{4(Th)^2}\sum_{t}^{T-2} K\left(\frac{x-x_t}{h}\right)K\left(\frac{x-x_{t-2}}{h}\right), \tag{36}$$

and, in the limit, it becomes

$$\frac{C_2(\phi)\left(f(x)\right)^2}{4(Th)^2}R_K. \tag{37}$$

Ultimately, the variance is

$$\mathrm{var}\left(\sum_{t=1}^{T-2}\eta_t^2 K\left(\frac{x-x_t}{h}\right)\right) = \frac{C(\phi)\left(f(x)\right)^2}{4(Th)^2}R_K, \tag{38}$$

where $C(\phi) = C_1(\phi) + C_2(\phi)$. Then, the asymptotic integrated mean squared error (AIMSE) becomes

$$AIMSE = \frac{h^4\sigma_K^4}{4}\int_0^1 \left[\mathrm{D}^2 f(t) - \frac{\gamma_2[\mathrm{D}^2 f(t)]^2}{f(t)}\right]^2 dt + \frac{C(\phi)\int_0^1 \left(f(t)\right)^2 dt}{4Th}R_K. \tag{39}$$

Differentiating this expression w.r.t. h and putting the result equal to zero we find the optimal (minimizing) bandwidth

$$h = T^{-1/5} \left[\frac{C(\phi) \int_0^1 (f(t))^2 \, dt}{4\sigma_K^4 \int_0^1 \left[\mathrm{D}^2 f(t) - \frac{\gamma_2 [\mathrm{D}^2 f(t)]^2}{f(t)} \right]^2 dt} \right]^{1/5} . \tag{40}$$

Thus, we confirm that $h = O(T^{-1/5})$. If we plug the above expression back into (39) we find that the optimal AIMSE is

$$\begin{aligned} AIMSE_o &= T^{-4/5} * \left[\frac{\sigma_K^4}{4^{19/5}} \left[C(\phi) \int_0^1 (f(t))^2 \, dt \right]^{4/5} \left[\int_0^1 \left[\mathrm{D}^2 f(t) - \frac{\gamma_2 [\mathrm{D}^2 f(t)]^2}{f(t)} \right]^2 dt \right]^{1/5} \right. \\ &\quad \left. + \frac{C(\phi) \int_0^1 (f(t))^2 \, dt \, R_K}{4} \right] . \end{aligned}$$

Hence the optimal AIMSE is of the order $O\left(T^{-4/5}\right)$. □

Proof of Theorem 2 As a first step, we note that the estimator in (7) can be represented as a (normalized) quadratic form, i.e.,

$$\hat{f}(x) = \frac{\boldsymbol{y}' \boldsymbol{D}(x) \boldsymbol{y}}{\mathrm{tr}\,(\boldsymbol{D}(x))}, \tag{41}$$

where $\boldsymbol{y} = (y_1, \ldots, y_T)'$ is a $(T, 1)$ vector of data generated by the model (1)-(2) while $\boldsymbol{D}(x)$ is the quadratic form matrix

$$\boldsymbol{D}(x) = \frac{1}{2} \begin{bmatrix} K(\frac{x-x_1}{h}) & 0 & -2K(\frac{x-x_1}{h}) & 0 & \cdots & \cdots & 0 \\ 0 & K(\frac{x-x_2}{h}) & 0 & -2K(\frac{x-x_2}{h}) & 0 & \cdots & 0 \\ -2K(\frac{x-x_1}{h}) & 0 & K(\frac{x-x_1}{h})+K(\frac{x-x_3}{h}) & 0 & -2K(\frac{x-x_3}{h}) \; 0 & \cdots & 0 \\ \vdots & \cdots & \cdots & \cdots & \cdots & \vdots & \\ 0 & \cdots & 0 & -2K(\frac{x-x_{T-2}}{h}) & 0 & & K(\frac{x-x_T}{h}) \end{bmatrix} . \tag{42}$$

Using the representation (41) and an elementary result about the quadratic form distribution (see Moser(1985)), we find that (41) is the linear combination of independent χ_1^2 variables. More precisely, let us denote $\boldsymbol{\Sigma}$ the variance-covariance matrix of $\boldsymbol{y}$ and $p = \mathrm{rk}\,(\boldsymbol{D}(x)\boldsymbol{\Sigma})$. Then we have

$$\boldsymbol{y}' \boldsymbol{D}(x) \boldsymbol{y} = \sum_{t=1}^{p} \lambda_t \chi_{1,t}^2, \tag{43}$$

with λ_t's being nonzero eigenvalues of the matrix $\boldsymbol{D}(x)\boldsymbol{\Sigma}$ and $\chi_{1,t}^2$ are independent (centered) χ_1^2 random variables. Applying a Taylor series expansion of the function $f(x)$

we find that up to the multiplicative factor $f(x)$ the variance-covariance matrix $\boldsymbol{\Sigma}$ is

$$\begin{bmatrix} 1 & \phi & \phi^2 & \cdots & \cdots & \phi^n \\ \phi & 1 & \phi & \phi^2 & \cdots & \phi^{n-1} \\ \vdots & \cdots & \cdots & \cdots & \cdots & \vdots \\ \phi^n & \phi^{n-1} & \phi^{n-2} & \cdots & \cdots & 1 \end{bmatrix},$$

which is a Töplitz matrix of a specific kind, namely the so-called Kac-Murdock-Szegö matrix. It is known that the determinant of this matrix is $(1-\phi^2)^{T-1}$ and therefore not equal to zero unless $\phi = 1$, see, e.g., Dow (2003). Thus, the matrix $\boldsymbol{\Sigma}$ is strictly positive-definite for any $\phi \in \Theta$ and as a consequence $\operatorname{rk}(\boldsymbol{D}(x)\boldsymbol{\Sigma}) = \operatorname{rk}(\boldsymbol{D}(x))$. Recall, that in order to derive asymptotic results we require $T \to \infty$, $h \to 0$ and $Th \to \infty$. The last requirement ensures that the number of points in the local neighborhood $T_h(x) = (x-h, x+h)$ about the point x remains infinite as the neighborhood shrinks, when $h \to \infty$. Assuming the bandwidth used is the optimal, i.e., $h = O(T^{-1/5})$, we find that each local neighborhood of x contains $O(T^{4/5})$ points. Since the design is equispaced, we have for $t = 1, \ldots, T$ that

$$\begin{aligned} K\left(\frac{x - x_t}{h}\right) &= K\left(O(T^{-3/5})\right), \\ &\to K(0), \end{aligned}$$

which is a constant term. This means that as $T \to \infty$, the rank of $\boldsymbol{D}(x)$ tends to the rank of

$$\boldsymbol{D} = \begin{bmatrix} 1 & 0 & -1 & 0 & \cdots & \cdots & 0 & \\ 0 & 1 & 0 & -1 & 0 & \cdots & 0 & \\ -1 & 0 & 1 & 0 & -1 & 0 & \cdots & 0 \\ \vdots & \cdots & \cdots & \cdots & \cdots & \vdots & & \\ 0 & \cdots & 0 & -1 & 0 & & 1 & \end{bmatrix}, \tag{44}$$

and consequently $\operatorname{rk}(\boldsymbol{D}(x)) \asymp T - 2$ where $\asymp$ stands for asymptotic equivalence. Thus,

$$\widehat{f}(x) \asymp \frac{1}{\operatorname{tr}(\boldsymbol{D}(x))} \sum_{t=1}^{T-2} \lambda_t \chi^2_{1,t}. \tag{45}$$

To handle (45) we use the CLT version for non-identically distributed random variables as described by Jacod and Protter (1998). To check that the conditions of the theorem we need to verify that i) $\sup \frac{\lambda_t^2}{(tr\, D(x))^2} < \infty$ and ii) $\lim_{T\to\infty} \sum_{t=1}^{T-2} \frac{\lambda_t^2}{(tr\, D(x))^2} = \infty$. Both of the conditions are satisfied immediately as we note that

$$\frac{\lambda_t^2}{\operatorname{tr}\,(\boldsymbol{D}(x))^2} \le \frac{2\operatorname{tr}(\boldsymbol{D}(x)\boldsymbol{\Sigma})^2}{\operatorname{tr}\,(\boldsymbol{D}(x))^2} \le \operatorname{var}\left(\frac{\boldsymbol{y}'\boldsymbol{D}(x)\boldsymbol{y}}{\operatorname{tr}\,(\boldsymbol{D}(x))}\right) = O\left(\frac{1}{Th}\right),$$

which completes the proof. □

Proposition A1 Let data be generated according to the model (1)-(2). Then,

$$\left(\frac{1}{T}\sum_{t=1}^{T}\widehat{\sigma}_{t-1}^{-2}y_{t-1}^{2}\right)-\left(\frac{1}{T}\sum_{t=1}^{T}\sigma_{t-1}^{-2}y_{t-1}^{2}\right)\xrightarrow{p}0. \tag{46}$$

Proof of Proposition A1 Rewrite the left hand side of (46) as

$$\frac{1}{T}\sum_{t=2}^{T}\left(\widehat{\sigma}_t^{-2}-\sigma_t^{-2}\right)y_{t-1}^2$$

and by the Holder inequality (see, e.g., B.5.14 in Davidson (1994)) we have that

$$\frac{1}{T}\left|\sum_{t=2}^{T}\left(\widehat{\sigma}_t^{-2}-\sigma_t^{-2}\right)y_{t-1}^2\right| \leq \left(\min_t\left(\widehat{\sigma}_t^2\right)\min_t\left(\sigma_t^2\right)\right)^{-1} \tag{47}$$

$$\times\sqrt{\frac{1}{T}\sum_{t=2}^{T}\left(\widehat{\sigma}_t^2-\sigma_t^2\right)^2}\sqrt{\frac{1}{T}\sum_{t=2}^{T}y_{t-1}^4} \tag{48}$$

$$= o_p(1)\sqrt{\frac{1}{T}\sum_{t=2}^{T}y_{t-1}^4}$$

Note that $\widehat{\sigma}_t^2-\sigma_t^2=o_p(1)$ and thus in order to complete the proof it suffices to show that $\frac{1}{T}\sum_{t=2}^{T}y_{t-1}^4=O_p(1)$. The following two steps are needed: (1) show that $\mathrm{E}(y_{t-1}^4)<\infty$ and (2) show that $\frac{1}{T}\sum_{t=2}^{T}y_{t-1}^4-\mathrm{E}(y_{t-1}^4)\xrightarrow{p}0$. First, notice that

$$\begin{aligned}
\mathrm{E}(y_{t-1}^4) &= \sigma_{t-1}^4\,\mathrm{E}\left(\sum_{i_1=0}^{\infty}\phi^{i_1}v_{t-i_1-1}\sum_{i_2=0}^{\infty}\phi^{i_2}v_{t-i_2-1}\sum_{i_3=0}^{\infty}\phi^{i_3}v_{t-i_3-1}\sum_{i_4=0}^{\infty}\phi^{i_4}v_{t-i_4-1}\right) \qquad (49)\\
&= \sigma_{t-1}^4\,\mathrm{E}\left(\sum_{i_1=0}^{\infty}\sum_{i_2=0}^{\infty}\sum_{i_3=0}^{\infty}\sum_{i_4=0}^{\infty}\phi^{i_1}\phi^{i_2}\phi^{i_3}\phi^{i_4}v_{t-i_1-1}v_{t-i_2-1}v_{t-i_3-1}v_{t-i_4-1}\right)\\
&= \sigma_{t-1}^4\sum_{i_1=0}^{\infty}\sum_{i_2=0}^{\infty}\sum_{i_3=0}^{\infty}\sum_{i_4=0}^{\infty}\phi^{i_1}\phi^{i_2}\phi^{i_3}\phi^{i_4}\,\mathrm{E}\left(v_{t-i_1-1}v_{t-i_2-1}v_{t-i_3-1}v_{t-i_4-1}\right)\\
&\leq \sigma_{t-1}^4\sum_{i_1=0}^{\infty}\sum_{i_2=0}^{\infty}\sum_{i_3=0}^{\infty}\sum_{i_4=0}^{\infty}\left|\phi^{i_1}\phi^{i_2}\phi^{i_3}\phi^{i_4}\right|\left|\mathrm{E}\left(v_{t-i_1-1}v_{t-i_2-1}v_{t-i_3-1}v_{t-i_4-1}\right)\right|
\end{aligned}$$

Since $\sum_{i_1=0}^{\infty}\sum_{i_2=0}^{\infty}\sum_{i_3=0}^{\infty}\sum_{i_4=0}^{\infty}\left|\phi^{i_1}\phi^{i_2}\phi^{i_3}\phi^{i_4}\right|=\sum_{i_1=0}^{\infty}\left|\phi^{i_1}\right|\sum_{i_2=0}^{\infty}\left|\phi^{i_2}\right|\sum_{i_3=0}^{\infty}\left|\phi^{i_3}\right|\sum_{i_4=0}^{\infty}\left|\phi^{i_4}\right|<\infty$ and $\left|\mathrm{E}\left(v_{t-i_1-1}v_{t-i_2-1}v_{t-i_3-1}v_{t-i_4-1}\right)\right|\leq\mathrm{E}\left(v_t^4\right)=\mu_4$ by strict stationarity of v_t, we have

$$\begin{aligned}
\mathrm{E}(y_{t-1}^4) &\leq \sigma_{t-1}^4\sum_{i_1=0}^{\infty}\sum_{i_2=0}^{\infty}\sum_{i_3=0}^{\infty}\sum_{i_4=0}^{\infty}\left|\phi^{i_1}\phi^{i_2}\phi^{i_3}\phi^{i_4}\right|\mu_4\\
&< \infty
\end{aligned}$$

as σ_{t-1}^4 is a bounded function. Secondly, define

$$
\begin{aligned}
Z_{t-1} &= y_{t-1}^4 - \mathrm{E}(y_{t-1}^4) \\
&= \sigma_{t-1}^4 \sum_{i_1=0}^{\infty} \sum_{i_2=0}^{\infty} \sum_{i_3=0}^{\infty} \sum_{i_4=0}^{\infty} \phi^{i_1}\phi^{i_2}\phi^{i_3}\phi^{i_4} \\
&\quad \times \left(v_{t-i_1-1}v_{t-i_2-1}v_{t-i_3-1}v_{t-i_4-1} - \mathrm{E}\left(v_{t-i_1-1}v_{t-i_2-1}v_{t-i_3-1}v_{t-i_4-1}\right)\right)
\end{aligned}
$$

Let $\Sigma_{t-m-1} = \{v_{t-m-1}, v_{t-m-2}, \ldots\}$ for $m > 1$. Consider a forecast of Z_{t-1} conditional on Σ_{t-m-1} :

$$
\begin{aligned}
\mathrm{E}\left(Z_{t-1}|\Sigma_{t-m-1}\right) &= \sigma_t^4 \sum_{i_1=m}^{\infty} \sum_{i_2=m}^{\infty} \sum_{i_3=m}^{\infty} \sum_{i_4=m}^{\infty} \phi^{i_1}\phi^{i_2}\phi^{i_3}\phi^{i_4} \\
&\quad \times \left(v_{t-i_1-1}v_{t-i_2-1}v_{t-i_3-1}v_{t-i_4-1} - \mathrm{E}\left(v_{t-i_1-1}v_{t-i_2-1}v_{t-i_3-1}v_{t-i_4-1}\right)\right).
\end{aligned}
$$

Then

$$
\begin{aligned}
\mathrm{E}\left|\mathrm{E}\left(Z_{t-1}|\Sigma_{t-m-1}\right)\right| &= \mathrm{E}\left|\sigma_{t-1}^4 \sum_{i_1=m}^{\infty} \sum_{i_2=m}^{\infty} \sum_{i_3=m}^{\infty} \sum_{i_4=m}^{\infty} \phi^{i_1}\phi^{i_2}\phi^{i_3}\phi^{i_4}\right. \\
&\quad \left. \times \left(v_{t-i_1-1}v_{t-i_2-1}v_{t-i_3-1}v_{t-i_4-1} - \mathrm{E}\left(v_{t-i_1-1}v_{t-i_2-1}v_{t-i_3-1}v_{t-i_4-1}\right)\right)\right| \\
&\leq \sigma_{t-1}^4 \sum_{i_1=m}^{\infty} \sum_{i_2=m}^{\infty} \sum_{i_3=m}^{\infty} \sum_{i_4=m}^{\infty} \left|\phi^{i_1}\phi^{i_2}\phi^{i_3}\phi^{i_4}\right| \\
&\quad \times \mathrm{E}\left(\left|v_{t-i_1-1}v_{t-i_2-1}v_{t-i_3-1}v_{t-i_4-1} - \mathrm{E}\left(v_{t-i_1-1}v_{t-i_2-1}v_{t-i_3-1}v_{t-i_4-1}\right)\right|\right) \\
&\leq \sigma_{t-1}^4 \sum_{i_1=m}^{\infty} \sum_{i_2=m}^{\infty} \sum_{i_3=m}^{\infty} \sum_{i_4=m}^{\infty} \left|\phi^{i_1}\phi^{i_2}\phi^{i_3}\phi^{i_4}\right| M \\
&= \xi_m c_{t-1}
\end{aligned}
$$

for some $M < \infty$, where

$$
\begin{aligned}
\xi_m &= \sum_{i_1=m}^{\infty} \sum_{i_2=m}^{\infty} \sum_{i_3=m}^{\infty} \sum_{i_4=m}^{\infty} \left|\phi^{i_1}\phi^{i_2}\phi^{i_3}\phi^{i_4}\right| \\
&= \sum_{i_1=m}^{\infty} \left|\phi^{i_1}\right| \sum_{i_2=m}^{\infty} \left|\phi^{i_2}\right| \sum_{i_3=m}^{\infty} \left|\phi^{i_3}\right| \sum_{i_4=m}^{\infty} \left|\phi^{i_4}\right|,
\end{aligned}
$$

and

$$
c_{t-1} = \sigma_{t-1}^4 M. \tag{50}
$$

Because $\left\{\phi^{i_1}\right\}_{i_1=0}^{\infty}$ is absolutely summable, $\lim_{m\to\infty} \sum_{i_1=m}^{\infty} \left|\phi^{i_1}\right| = 0$ implying that $\lim_{m\to\infty} \xi_m = 0$. Consequently, according to Andrews (1987), $\{Z_{t-1}\}$ is an L^1-mixingale. To apply Andrews (1987) LLN for L^1-mixingales we first need to show that y_{t-1}^4 is uniformly integrable, i.e., that $\mathrm{E}\left(\left|y_{t-1}^4\right|^2\right) < \infty$ (using $r = 2$). This can most easily be shown by noticing that since $y_{t-1}^4 = \left|y_{t-1}^4\right|$ the condition simplifies to showing that

$\mathrm{E}\left(y_t^8\right) < \infty$. Taking a similar approach to showing the existence of $\mathrm{E}\left(y_{t-1}^4\right)$ as in (49) the existence of $\mathrm{E}\left(y_{t-1}^8\right)$ will follow immediately due to the absolute summability of $\left\{\phi^{i_1}\right\}_{i_1=0}^{\infty}$ and the existence of $\mathrm{E}\left(v_{t-1}^8\right)$ (due to the assumption of normality). Finally, in order to apply the result of Andrews (1987) LLN we need to verify the condition

$$\lim_{T\to\infty}\frac{1}{T}\sum_{t=2}^{T}c_{t-1} = \lim_{T\to\infty}\frac{1}{T}\sum_{t=2}^{T}\sigma_{t-1}^4 M < \infty,$$

which will hold as σ_{t-1}^4 is bounded. We can therefore according to Andrews (1987) LLN conclude that

$$\frac{1}{T}\sum_{t=2}^{T}y_{t-1}^4 - \mathrm{E}(y_{t-1}^4) \xrightarrow{p} 0$$

where $\mathrm{E}(y_{t-1}^4) = O_p(1)$ and from (??) this implies that

$$\begin{aligned}\operatorname{plim}\frac{1}{T}\left|\sum_{t=2}^{T}\left(\widehat{\sigma}_t^{-2}-\sigma_t^{-2}\right)y_{t-1}^2\right| &= o_p(1)O_p(1)\\ &= o_p(1)\end{aligned}$$

as $T\to\infty$, $h\to 0$ and $Th\to\infty$, which completes the proof. □

Proposition A2 Let the Assumptions of Proposition A1 hold. Then

$$\left(\frac{1}{T}\sum_{t=1}^{T}\widehat{\sigma}_t^{-1}\widehat{\sigma}_{t-1}^{-1}y_t y_{t-1}\right) - \left(\frac{1}{T}\sum_{t=1}^{T}\sigma_t^{-1}\sigma_{t-1}^{-1}y_t y_{t-1}\right) \xrightarrow{p} 0. \tag{51}$$

as $T\to\infty$, $h\to 0$ and $Th\to\infty$.

Proof of Proposition A2 Rewrite the left hand side of (51) as

$$\frac{1}{T}\sum_{t=2}^{T}\left(\left(\widehat{\sigma}_{t-1}\widehat{\sigma}_t\right)^{-1}-\left(\sigma_{t-1}\sigma_t\right)^{-1}\right)y_{t-1}y_t,$$

and notice that (similar to the proof of Proposition A1) as $T\to\infty$, $h\to 0$ and $Th\to\infty$

$$\begin{aligned}\frac{1}{T}\left|\sum_{t=2}^{T}\left(\left(\widehat{\sigma}_{t-1}\widehat{\sigma}_t\right)^{-1}-\left(\sigma_{t-1}\sigma_t\right)^{-1}\right)y_{t-1}y_t\right| &\le \left(\min_t\left(\widehat{\sigma}_{t-1}\widehat{\sigma}_t\right)\min_t\left(\sigma_{t-1}\sigma_t\right)\right)^{-1}\\ &\quad\times\sqrt{\frac{1}{T}\sum_{t=2}^{T}\left(\widehat{\sigma}_{t-1}\widehat{\sigma}_t-\sigma_{t-1}\sigma_t\right)^2}\sqrt{\frac{1}{T}\sum_{t=2}^{T}y_{t-1}^2 y_t^2} \qquad (52)\\ &\le o_p(1)\sqrt{\frac{1}{T}\sum_{t=2}^{T}y_{t-1}^4}\sqrt{\frac{1}{T}\sum_{t=2}^{T}y_t^4} \qquad (53)\\ &= o_p(1), \qquad (54)\end{aligned}$$

since (from the proof of Proposition A1) $\frac{1}{T}\sum_{t=2}^{T} y_{t-1}^4 = O_p(1)$ and $\frac{1}{T}\sum_{t=2}^{T} y_t^4 = O_p(1)$ which completes the proof. □

Proof of Theorem 3 Write (18) as

$$\hat{\phi}_T = \left(\frac{1}{T}\sum_{t=1}^{T} \widehat{\sigma}_{t-1}^{-2} y_{t-1}^2\right)^{-1} \left(\frac{1}{T}\sum_{t=1}^{T} \widehat{\sigma}_t^{-1}\widehat{\sigma}_{t-1}^{-1} y_t y_{t-1}\right)$$

such that asymptotically

$$\operatorname{plim}\hat{\phi}_T = \left(\operatorname{plim}\frac{1}{T}\sum_{t=1}^{T} \widehat{\sigma}_{t-1}^{-2} y_{t-1}^2\right)^{-1} \left(\operatorname{plim}\frac{1}{T}\sum_{t=1}^{T} \widehat{\sigma}_t^{-1}\widehat{\sigma}_{t-1}^{-1} y_t y_{t-1}\right)$$

and by Proposition A1 and A2 we have

$$\begin{aligned}
\operatorname{plim}\hat{\phi}_T &= \left(\operatorname{plim}\frac{1}{T}\sum_{t=1}^{T} \sigma_{t-1}^{-2} y_{t-1}^2\right)^{-1} \left(\operatorname{plim}\frac{1}{T}\sum_{t=1}^{T} \sigma_t^{-1}\sigma_{t-1}^{-1} y_t y_{t-1}\right) \\
&= \phi + \left(\operatorname{plim}\frac{1}{T}\sum_{t=1}^{T} \epsilon_{t-1}^2\right)^{-1} \left(\operatorname{plim}\frac{1}{T}\sum_{t=1}^{T} \epsilon_{t-1} v_t\right) \\
&= \phi
\end{aligned}$$

as $T \to \infty$, $h \to 0$ and $Th \to \infty$. The results of the last equation follows from the fact that the random variable $\epsilon_{t-1}v_t$ is a martingale difference sequence with mean $\mathrm{E}(\epsilon_{t-1}v_t) = 0$, variance $\mathrm{E}\left((\epsilon_{t-1}v_t)^2\right) = \mathrm{E}\left(\epsilon_t^2\right)$ and with fourth moment $\mathrm{E}\left(\epsilon_t^4\right) < \infty$. Hence, from applying a LLN for martingale difference sequences, see, e.g., White (1984), it follows that $\operatorname{plim}\frac{1}{T}\sum_{t=2}^{T}\epsilon_{t-1}v_t = \mathrm{E}(\epsilon_{t-1}v_t)$ and $\operatorname{plim}\frac{1}{T}\sum_{t=2}^{T}(\epsilon_{t-1}v_t)^2 = \mathrm{E}((\epsilon_{t-1}v_t)^2)$. This completes the proof of consistency. □

Proof of Theorem 4 As $\hat{\phi}_T$ is a MINPIN estimator we will establish it asymptotic distribution by verifying, that given the assumptions of Theorem 1, all the conditions of Assumption N in Andrews (1994) is meet. According to Theorem 1 in Andrews (1994) this will be sufficient to provide the desired result. In what follows we will verify each of the conditions of Andrews (1994) Assumption N one by one:

Assumption N.a) Follows directly from Theorem 3.

Assumption N.b) In order to prove that $\lim_{T\to\infty} P\left(\widehat{f}(x) \in C[0,1]\right) = 1$ it suffices to show that i) $\widehat{f}(x) \xrightarrow{p} f(x)$ and ii) $\mathrm{D}^k f(x) \xrightarrow{p} \mathrm{D}^k \widehat{f}(x)$ for $k = 1, 2$. Condition i) has already been established. In order to prove ii), consider differentiating the estimator in

(7) w.r.t. x to obtain the following estimator

$$\mathrm{D}^k\widehat{f}(x) = h^{-k}\sum_{t=1}^{T-2}\eta_t^2\mathrm{D}^k W_t(x)$$

where

$$W_t(x) = \frac{K\left(\frac{x-x_t}{h}\right)}{\sum_{t=1}^{T-2}K\left(\frac{x-x_i}{h}\right)} \tag{55}$$

Taking expectations,

$$\mathrm{E}\left(\mathrm{D}^k\widehat{f}(x)\right) = h^{-k}\sum_{t=1}^{T-2}\mathrm{E}\left(\eta_t^2\right)W_i(x) \tag{56}$$

Define $u_t = \frac{x-x_t}{h}$ such that $x_t = x - hu_t$. Recall that $\sum_{t=1}^{T-2}K(u_t) \sim \int K(u)\,du = 1$ such that asymptotically, as $h \to 0$, $T \to \infty$ and $Th \to \infty$ we have

$$\begin{aligned}\mathrm{E}\left(\mathrm{D}^k\widehat{f}(x)\right) &= h^{-k}\int f(x-hu)\mathrm{D}^k K(u)\,du && (57)\\ &= \int \mathrm{D}^k f(x-hu)K(u)\,du && (58)\end{aligned}$$

Using the Taylor series expansion of $f(x)$, we immediately find that $\mathrm{E}\left(\mathrm{D}^k\widehat{f}(x)\right) = \mathrm{D}^k f(x) + o(1)$ and from Chebyshev's inequality we have

$$\begin{aligned}\lim_{T\to\infty}P\left(\left|\mathrm{D}^k\widehat{f}(x)-\mathrm{D}^k f(x)\right| \geq \varepsilon\right) &\leq \frac{\lim_{T\to\infty}E\left(\left|\mathrm{D}^k\widehat{f}(x)-\mathrm{D}^k f(x)\right|\right)}{\varepsilon}\\ &= 0\end{aligned}$$

for any $\varepsilon > 0$. This completes the verification of condition ii) and completes the verification of the assumption.

Assumption N.c) Let ϕ_0 denote the true value in the population of the parameter ϕ. Verifying this condition simplifies to showing that

$$\sqrt{T}\overline{m}_T^*\left(\widehat{\sigma}_t, \widehat{\sigma}_{t-1}, y_t, y_{t-1}, \phi_0\right) \xrightarrow{p} 0 \tag{59}$$

where

$$\begin{aligned}\sqrt{T}\overline{m}_T^*\left(\widehat{\sigma}_t, \widehat{\sigma}_{t-1}, y_t, y_{t-1}, \phi_0\right) &= \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\mathrm{E}\left(m_t\left(\widehat{\sigma}_t, \widehat{\sigma}_{t-1}, y_t, y_{t-1}, \phi_0\right)\right)\\ &= \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\mathrm{E}\left(\widehat{v}_t\widehat{\epsilon}_{t-1}\right)\end{aligned}$$

First notice that $\mathrm{E}\left(\widehat{v}_t\widehat{\epsilon}_{t-1}\right)$ will be a non-stochastic sequence (since the expectation is wrt the probability measure P_v), so only if $\mathrm{E}\left(\widehat{v}_t\widehat{\epsilon}_{t-1}\right) = 0$ or $\mathrm{E}\left(\widehat{v}_t\widehat{\epsilon}_{t-1}\right) = O(T^{-\delta})$ for

$\delta > \frac{1}{2}$ condition (59) will be satisfied. Consequently, we can also write condition (59) simply as

$$\lim_{T\to\infty} \sqrt{T}\overline{m}_T^* = 0$$

Next, define

$$\begin{aligned}
\widehat{v}_t &= \widehat{\epsilon}_t - \phi_0 \widehat{\epsilon}_{t-1} \\
\widehat{\epsilon}_t &= \widehat{\sigma}_t^{-1} y_t \\
\widehat{\epsilon}_{t-1} &= \widehat{\sigma}_{t-1}^{-1} y_{t-1}
\end{aligned}$$

Consider,

$$\begin{aligned}
W_T &= \frac{1}{\sqrt{T}} \sum_{t=1}^{T} (\widehat{v}_t \widehat{\epsilon}_{t-1}) - \frac{1}{\sqrt{T}} \sum_{t=1}^{T} v_t \epsilon_{t-1} \\
&= \frac{1}{\sqrt{T}} \sum_{t=1}^{T} (\widehat{v}_t \widehat{\epsilon}_{t-1} - v_t \epsilon_{t-1}) \\
&= I_{1T} - I_{2T}
\end{aligned}$$

where

$$\begin{aligned}
I_{1T} &= \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \left(\widehat{\sigma}_t^{-1} \widehat{\sigma}_{t-1}^{-1} - \sigma_t^{-1} \sigma_{t-1}^{-1}\right) y_t y_{t-1} \\
I_{2T} &= \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \left(\widehat{\sigma}_{t-1}^{-2} - \sigma_{t-1}^{-2}\right) y_{t-1}^2
\end{aligned}$$

Consequently,

$$\begin{aligned}
|I_{1T}| &\leq \left(\min_{t\in T} (\widehat{\sigma}_t \widehat{\sigma}_{t-1}) \min_{t\in T} (\sigma_t \sigma_{t-1})\right)^{-1} \left| \frac{1}{\sqrt{T}} \sum_{t=1}^{T} (\widehat{\sigma}_t \widehat{\sigma}_{t-1} - \sigma_t \sigma_{t-1}) y_t y_{t-1} \right| \\
&\leq \left(\min_{t\in T} (\widehat{\sigma}_t \widehat{\sigma}_{t-1}) \min_{t\in T} (\sigma_t \sigma_{t-1})\right)^{-1} \sqrt{\frac{1}{\sqrt{T}} \sum_{t=1}^{T} (\widehat{\sigma}_t \widehat{\sigma}_{t-1} - \sigma_t \sigma_{t-1})^2} \sqrt{\frac{1}{\sqrt{T}} \sum_{t=1}^{T} y_t^2 y_{t-1}^2} \\
&= \left(\min_{t\in T} (\widehat{\sigma}_t \widehat{\sigma}_{t-1}) \min_{t\in T} (\sigma_t \sigma_{t-1})\right)^{-1} \sqrt{O_p(T^{-\frac{3}{10}})} \sqrt{O_p(1)} \\
&= o_p(1)
\end{aligned}$$

since $(\widehat{\sigma}_t\widehat{\sigma}_{t-1} - \sigma_t\sigma_{t-1})^2 = O_p(T^{-\frac{4}{5}})$, $(y_t^2 y_{t-1}^2)$ satisfies a CLT and $\left(\min_{t\in T}(\widehat{\sigma}_t\widehat{\sigma}_{t-1})\min_{t\in T}(\sigma_t\sigma_{t-1})\right)^{-1}$ is bounded by the assumptions of Theorem 1. Similarly,

$$
\begin{aligned}
|I_{2T}| &\leq \left(\min_{t\in T}(\widehat{\sigma}_{t-1}^2)\min_{t\in T}(\sigma_{t-1}^2)\right)^{-1}\left|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}(\widehat{\sigma}_{t-1}^2 - \sigma_{t-1}^2)\,y_{t-1}^2\right| \\
&\leq \left(\min_{t\in T}(\widehat{\sigma}_{t-1}^2)\min_{t\in T}(\sigma_{t-1}^2)\right)^{-1}\sqrt{\frac{1}{\sqrt{T}}\sum_{t=1}^{T}(\widehat{\sigma}_{t-1}^2 - \sigma_{t-1}^2)^2}\sqrt{\frac{1}{\sqrt{T}}\sum_{t=1}^{T}y_{t-1}^4} \\
&= o_p(1)
\end{aligned}
$$

Consequently,

$$\operatorname*{plim}_{T\to\infty} W_T = 0$$

Secondly, notice that since $\mathrm{E}(v_t\epsilon_{t-1}) = 0$ we have that

$$\mathrm{E}(W_T) = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\mathrm{E}(\widehat{v}_t\widehat{\epsilon}_{t-1})$$

which is the expression we are interested in. Since, it is easy to verify that the random variable W_T is dominated (as required by Theorem A.DCT) we have according to Theorem A.DCT

$$
\begin{aligned}
\lim_{T\to\infty}\mathrm{E}(W_T) &= \mathrm{E}\left(\operatorname*{plim}_{T\to\infty} W_T\right) \\
&= \mathrm{E}(0) \\
&= 0
\end{aligned}
$$

which completes the verification of Assumption N.c.

Assumption N.d) Let m_t be given by (16) and define

$$
\begin{aligned}
v_T &= \sqrt{T}\left(\frac{1}{T}\sum_{t=1}^{T}m_t - \frac{1}{T}\sum_{t=1}^{T}\mathrm{E}(m_t)\right) \\
&= \frac{1}{\sqrt{T}}\sum_{t=1}^{T}v_t\epsilon_{t-1}.
\end{aligned}
$$

Notice that $v_t\epsilon_{t-1}$ is a martingale difference sequence. From straightforward application of CLT for martingale sequences, see, e.g., White (1984), we have that

$$v_T \xrightarrow{d} \mathrm{N}(0, S)$$

where $S = \frac{1}{1-\phi^2}$.

Assumption N.e) Define

$$\boldsymbol{W}_t = \begin{bmatrix} y_t y_{t-1} \\ \phi y_{t-1}^2 \end{bmatrix}$$
$$\tau = \begin{bmatrix} \sigma_t^{-1}\sigma_{t-1}^{-1} \\ \sigma_{t-1}^{-2} \end{bmatrix}.$$

and since $(\boldsymbol{W}_t - \mathrm{E}(\boldsymbol{W}_t))$ - as just defined - can easily be shown to (depending on the independent stochastic components $v_t\epsilon_{t-1}$ and ϵ_{t-1}^2 only) satisfy CLT's, Condition (e) is satisfied according to equation (2.4) page 46 in Andrews (1994).

Assumption N.f) Trivially satisfied.

Assumption N.g) Let m_t be given by (16). First we verify that m_t and $\partial m_t/\partial\phi$ satisfy the UWLLN over $\Theta \times C[0,1]$ using Andrews (1987). We begin by looking at m_t : Assumption A1 in Andrews (1987) is trivially satisfied. As

$$\begin{aligned} m_t &= v_t\epsilon_{t-1} \\ &= v_t \sum_{i=0}^{\infty} \phi^i v_{t-1-i} \end{aligned}$$

and $m_t \xrightarrow{p} 0$ uniformly on the interior of $\Theta \times C[0,1]$ (not only locally in a closed ball around ϕ) Assumption A2 in Andrews (1987) is satisfied. Next define $m_t^* = m_t\left(\sigma_t^*, \sigma_{t-1}^*, y_t, y_{t-1}; \phi^*\right)$ and consider

$$\begin{aligned} |m_t^* - m_t| &= \left| v_t \sum_{i=0}^{\infty} \phi^{*i} v_{t-1-i} - v_t \sum_{i=0}^{\infty} \phi^i v_{t-1-i} \right| \\ &= \left| \sum_{i=0}^{\infty} \left(\frac{\phi^{*i} - \phi^i}{\phi^i} \right) \phi^i v_{t-1-i} v_t \right| \\ &\leq \sqrt{\sum_{i=0}^{\infty} \phi^{2i} v_{t-1-i}^2 v_t^2} \sqrt{\sum_{i=0}^{\infty} \left(\frac{\phi^{*i} - \phi^i}{\phi^i} \right)^2} \end{aligned}$$

Defining

$$\begin{aligned} b_t(v_t, v_{t-1}, \phi) &= \sqrt{\sum_{i=0}^{\infty} \phi^{2i} v_{t-1-i}^2 v_t^2} \\ d(\phi^*, \phi) &= \sqrt{\sum_{i=0}^{\infty} \left(\frac{\phi^{*i} - \phi^i}{\phi^i} \right)^2} \end{aligned}$$

and noticing that

$$\begin{aligned}\sup_T \frac{1}{T}\sum_{t=2}^{T} \mathrm{E}\, b_t(v_t, v_{t-1}, \phi) &\leq \sup_T \frac{1}{T}\sum_{t=2}^{T} \sqrt{\mathrm{E}\left(\sum_{i=0}^{\infty} \phi^{2i} v_{t-1-i}^2 v_t^2\right)} \\ &= \sqrt{\frac{1}{1-\phi^2}}\end{aligned}$$

and $d(\phi^*, \phi) \downarrow 0$ as $\phi^* \to \phi$ we see that Assumption 4 in Andrews (1987) holds and according to Corollary 2 in Andrews (1987) we can conclude that m_t satisfy the UWLLN over $\Theta \times C[0,1]$. Next, we turn to $\partial m_t/\partial\phi$. Notice that

$$\frac{\partial m_t}{\partial \phi} = v_t \sum_{i=0}^{\infty} \phi^i v_{t-2-i}$$

and using similar steps as above it follows straightforwardly that also for $\partial m_t/\partial\phi$ Assumptions A1,A2 and A4 in Andrews (1987) applies hence it satisfies the UWLLN uniformly on $\Theta \times C[0,1]$. As m_t and $\partial m_t/\partial\phi$ does not depend on σ_t, Corollary 2 in Andrews (1987) also establishes uniform continuity of m_t given as

$$m = \lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} \mathrm{E}\, m_t(\phi, \sigma_t)$$

and M given by

$$\begin{aligned} M &= \lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} \mathrm{E}\left(\frac{\partial m_t}{\partial \phi}\right) \\ &= \frac{1}{1-\phi^2}. \end{aligned} \tag{60}$$

Finally notice that m_t is twice differentiable in ϕ uniformly on Θ which completes the verification of Assumption N.g).

Assumption N.h) Trivially satisfied on the interior of Θ.

Consequently, we have verified that all the conditions of Assumption N in Andrews (1994) holds which completes the proof.