Autoregression With Heteroscedastic Errors: A Study in Concentration

Gabriel Chandler
Department of Mathematics
Pomona College
Claremont, CA USA
E-mail: gabriel.chandler@pomona.edu

ABSTRACT: Many time series exhibit non-stationary behavior, i.e. either their mean or covariances are functions of time. The so-called modulated autoregressive process, an AR process where the variance is allowed to be a function of time, is a model that allows certain forms of non-stationarity to be modeled in a convenient way. For instance, this model has been used for seismic events such as earthquakes and mining explosions, as well as animal footfall data recorded via a geo-phone buried in the ground ground. What turns out to be important regarding estimation of parameters and order selection can be described as the concentration of the variance function. Highly concentrated variance functions affect the ability to estimate parameters via standard techniques such as the Yule-Walker equations, as well as making selection of the correct order for the model difficult. Though Pötscher (1989) has proven that, regardless of the shape of the variance function, order selection can be made consistently, these procedures do not account for the non-stationary behavior. As a result, the order selection procedure under performs. We show that the variance function can be estimated consistently up to a proportionality constant without specifying an order. Dividing through the original series by this estimate essentially makes the data stationary, improving order selection procedures as well as the standard error of the estimates.

1. Introduction

There exists a lot of work in the literature dealing with time series that exhibit conditional heteroscedasticity, i.e. the current volatility of a time series is dependent on past values of itself or its underlying properties. Such models take the form $Y_t = \epsilon_t$ where ϵ_t is a function of the past values of the process and its underlying properties. The celebrated ARCH model of Engle (1982), its extension to the GARCH model of Bollerslev (1986) and subsequent extensions have made these models very popular in finance problems. Models assuming unconditional heteroscedasticity have not garnered as much attention, as their application in finance is thought to be limited. Such models are applicable, for instance, to time series generated by seismic events such as earthquakes or explosions. For example, Fujita and Shibara (1978) and Dargahi-Noubary, Laycock, and Subba Rao (1978) suggested modeling these time series by $X(t) = c(t)Y(t)$ where $Y(t)$ is a stationary processes and $c(t)$ is a deterministic function. A similar model to this is the p^{th} order autoregressive model

(AR(p)) with unconditional heteroscedastic errors, i.e.

$$X_t = \sum_{j=1}^{p} \phi_j X_{t-j} + \sigma(t)\epsilon_t, \quad t = l, ..., T, \tag{1}$$

where the ϵ_t are assumed to be either identically distributed independent random variables with mean 0 and variance 1, or similar constraints under some dependency to allow the variance function to be identifiable. $\sigma^2(t)$ varies in time, independent of the past. This model has been considered earlier by, among others, Wichern, Miller and Hsu (1976), Tyssedal and Tjøsthein (1982), Paulsen and Tjøstheim (1985), and Pötscher (1989). More recently, Drees and Stărică (2002) and Fryzlewicz et. al (2006) used model (1) for stock returns, Dahlhaus and Polonik (2007) looked at estimation of this model under shape constraints, and Chandler and Polonik (2006) used this as a foundation for their discrimination procedure, with applications to seismic events. Under appropriate smoothness conditions on the variance function $\sigma^2(\cdot)$, this model is a special case of the locally stationary model defined by Dahlhaus (1997). Phillips and Xu (2006) derived the asymptotic behavior of the AR parameter estimates and considered testing strategies for the values of the AR parameters. One might hope that such locally stationary AR models would provide a good approximation for many time series exhibiting non-stationary behavior. In addition to the financial application, model (1) can be used to model non-stationary time series such as seismic events, animal footfalls recorded using a geo-phone (as found in Wood, O'Connell-Rodwell and Klemperer, 2005) or EEG recordings of seizures.

What turns out to be important regarding estimation is not the magnitude of the variance function, but rather the concentration of the variance function. Concentration is important in a variety of concepts. For example, the ARCH(m) model describes the concentration of the volatility clusters through the order m. A large value for m denotes longer range dependence and thus the series will tend to exhibit volatility clusters that are dispersed versus smaller values of m where the clusters will tend to be more concentrated. In establishing the effect of concentration on estimation of the asymptotic distribution of our estimates, we provide some insight into how concentration might be measured. Of particular interest in this paper is order selection, i.e. a data selected choice for p. In the stationary case, order selection is often based on a variant of the AIC (Akaike 1974), which seeks a balance between explanatory power and model size. The AIC method selects the order which minimizes

$$\log \hat{\sigma}^2(k) + \frac{kC(T)}{T}, \tag{2}$$

where k is the order of the model and $\hat{\sigma}^2$ is an estimate of the global variance. Selecting $C(T) = 2$ yields Akaike's AIC criterion, while selecting $C(T) = \log(T)$ gives Schwarz's (1978) BIC criterion. Many variants have been proposed with different values for $C(T)$ (see for example Hannan and Quinn, 1979), in particular to make the order selection consistent. Paulsen and Tjøstheim (1985) considered estimation and order selection of model (1). Pötscher (1989) went on to show that, under mild assumptions, a generalized AIC procedure can be constructed which is consistent for the correct order of a class of non-stationary autoregressions, with model (1) as a special case. However, in the case of a heteroscedastic model, the appearance of $\hat{\sigma}^2(\cdot)$ in the criterion loses its standard interpretation, since the true variance is a function of time. One might search for a modification of AIC that takes into account the non-stationarity of the time series. Recently in the literature, work dealing with the time-varying autoregressive process has offered some ideas for order selection for the more general case (see Dahlhaus, 1997 and Van Bellegem and Dahlhaus, 2005), in which the autoregressive parameters are also allowed to vary in time. These techniques tend to be computationally intensive. By way of a simple, but still useful, model, this paper proposes an order selection criterion which takes into account the non-stationary structure of the data while also being computationally inexpensive. First, the problem of heteroscedasticity and its effects on estimation are discussed in section 2. Section 3 deals with the order selection procedure for our model (1). Section 4 contains simulation studies. All proofs are deferred to section 5.

2. Estimation of Autoregressive Processes with Heteroscedastic Errors

In this section, we consider the model (1). This model constitutes a non-stationary class of processes which, under some smoothness conditions on σ^2, is a special case of the locally stationary model of Dahlhaus (1997). This definition was extended to only require bounded variation of σ^2 (Dahlhaus and Polonik, 2007). Similarly, the smoothness assumptions are relaxed in what follows. Following Dahlhaus' methodology for establishing a meaningful asymptotic theory, we rescale the time series to the unit interval, i.e. we assume we observe the process

$$X_{t,T} = \sum_{j=1}^{p} \phi_j X_{t-j,T} + \epsilon_t \sigma(\frac{t}{T}). \tag{3}$$

Standard estimation of parameters in a stationary model is usually done via the Yule-Walker estimators, which are based on the empirical autocovariances $\hat{\gamma}(h) = \frac{1}{T}\sum_{t=1}^{T-h} X_t X_{t+h}$ at values $h = 0, \ldots, p$ of the process. Because σ^2 is a function of time, the interpretation of $\hat{\gamma}(h)$ as a global covariance estimate is lost. In order to make sense of what the effect of heteroscedasticity is, we need some idea of what stationary processes

we should compare this non-stationary process to. It can be shown that the empirical autocovariance function

$$\hat{\gamma}(h) \to \int_0^1 \sigma^2(u)du \sum_{i=0}^{\infty} \psi_i \psi_{i+h}$$

in probability for all $h < \infty$ fixed, where the ψ_i are the coefficients in the MA(∞) representation under causality, i.e. $X_{t,T} = \sum_{i=0}^{\infty} \psi_i \epsilon_{t-i}$. Phillips and Xu (2006) established a similar result under fairly weak conditions (finitely many points of discontinuity of $\sigma(u)$ and slightly greater than 4^{th} moments). In other words, the first order properties of the ACF are the same for all processes with identical autoregressive coefficients and integrated variance functions. In terms of estimation of the autoregressive parameters, the magnitude of the error variance does not affect the estimation. It will be seen that what is of importance is what we might call the concentration of the variance function. The following asymptotic result quantifies the effect of concentration on the accuracy of the estimated parameters for the model (3).

$$\sqrt{T}(\hat{\phi} - \phi) \to N(0, \Gamma_0^{-1} \int_0^1 (\bar{\sigma}^2(u))^2 du)$$

in distribution, where $\phi := (\phi_1, \ldots, \phi_p)', \hat{\phi} := (\hat{\phi}_1, \ldots, \hat{\phi}_p)'$ is the vector of conditional least squares estimators, $\Gamma_0 = [\sum_{k=1}^{\infty} \psi^k \psi^{k+|i-j|}]_{i,j=1}^p$ and $\bar{\sigma}^2(u) = \frac{\sigma^2(u)}{\int_0^1 \sigma^2(u)du}$, i.e. the normalized (to integrate to unity) variance function.

Paulsen and Tjøstheim (1985) established the asymptotic normality of the unweighted least squares estimate under similar conditions. Phillips and Xu (2006) derived the asymptotic variance, which is of interest here. What we see is that the effect of concentration on the asymptotic variance of the estimate is equivalent to the L_2-norm of the normalized variance function. Since only the normalized variance function comes into play, only the concentration of $\sigma^2(\cdot)$ matters. This is proportional to one of the measures of concentration suggested by Chandler and Polonik (2006). Should we be able to make the volatility less concentrated, for instance by dividing through by a model-free estimate of the variance function, we might expect the standard error of the estimate to be reduced. This idea is made clear in the following section, and simulations results are shown below.

3. Order Selection for the Modulated AR Model

In the homoscedastic case, the problem of selecting the order of an autoregressive process has been well studied. For example, Akaike (1969, 1970) considered minimization of the final prediction error

$$FPE(k) = \frac{T+k}{T-k}\hat{\sigma}^2(k) \tag{4}$$

where T is the length of the observed time series, k is the order of the model, and $\hat{\sigma}^2(k)$ is an estimate of the (global) variance of the errors. Akaike (1974) also proposed the AIC, and subsequent authors have gone on to offer variations of this criterion which take the form of (2).

For properly chosen $C(T)$, these criteria are strongly consistent for the correct order in model (1). However, they are based on a global measure of variation ($\hat{\sigma}^2(k)$) and thus do not account for heteroscedasticity. As a consequence, for smaller sample sizes, they will under perform. In fact, similar to the problems seen in estimation of the AR parameters, the concentration of the variance function plays a role in our ability to correctly specify the order of our model. For example, consider the two variance functions presented in figure 1, both with identical values of $\int_0^1 \sigma^2(u)du$. In the homoscedastic case, order selection is based on

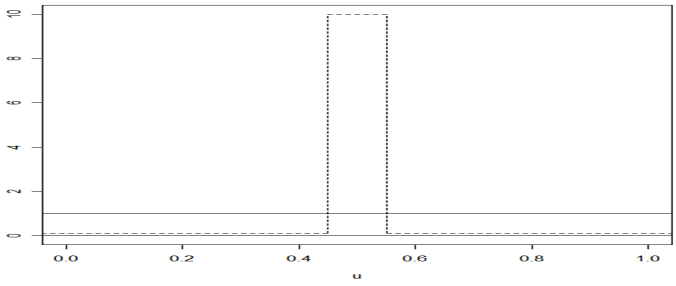

Figure 1: Two variance functions with equal integral, the solid line a homoscedastic variance function, the broken line a concentrated heteroscedastic variance function

all $T = 100$ observations equally. In the heteroscedastic case, the lack of fit term $\hat{\sigma}^2(k)$ will be dominated by the residuals where the variance is large. Thus, the selected order is virtually determined by the 10 observations lying in this interval, drastically reducing the effective sample size. As a result, a standard AIC model selection approach seems unreasonable. Order selection for locally stationary processes, specifically the time-varying autoregressive (tvAR) model,

$$V_t = \sum_{j=1}^{p} \phi_j(t) V_{t-j} + \sigma(t)\epsilon_t \tag{5}$$

has been discussed in the literature (Dahlhaus, 1997, Van Bellegem and Dahlhaus, 2005). tvAR models constitute a more general class of models than of interest here, and pose additional difficulties in estimation due to the fact that the autoregressive functions are not directly "observed", in the sense of the function plus error. As a result, estimation and order selection is done in the spectral domain, essentially based on a penalized (Whittle) likelihood, which is a computationally intensive procedure. In contrast to the autoregressive parameters, the variance function is directly observed in this sense. Additionally, because model (1) has a wide array of applications even without the generalization of model (4), a simpler, more computationally friendly order selection procedure which takes into account the non-stationarity of the process would be useful. The procedure described below is all done in the time domain, thus not requiring computation of time dependent spectrum, perhaps offering some intuition for the user. Furthermore, it accounts for the non-stationarity of the process being observed, and thus would be expected to outperform AIC directly on the non-stationary data.

To illustrate what is meant by the variance function being "directly observed", the following theorem provides some insight.

THEOREM **1.** *Define* $\hat{\Sigma}^0(\alpha) = \frac{1}{T}\sum_{t=1}^{[T\alpha]} X_{t,T}^2$ *and define* $\Sigma^0(\alpha) = c^2 \int_0^\alpha \sigma^2(u)du$, *where* $c^2 := \sum_{j=0}^\infty \psi_j^2$. *Under the assumptions of Phillips and Xu (2006),*

$$\sup_{\alpha\in[0,1]} |\hat{\Sigma}^0(\alpha) - \Sigma^0(\alpha)| = o_p(1)$$

The above theorem is of interest, for instance, in the case of Chandler and Polonik (2006), who considered functionals of the partial sum process of the squared residuals after normalization for measuring concentration of a function. As only the normalized estimate of the variance function appeared in their method, this result shows that correct identification of the order is not needed. In general however, correctly identifying the order of an autoregression is necessary. It seems clear from this theorem that without specifying the order, we can define an estimator $\hat{\eta}^2(\frac{t}{T})$, which estimates a function that is proportional to the true variance function. We essentially study the process

$$X_{t,T}^* := \frac{X_{t,T}}{\hat{\eta}(\frac{t}{T})}, \tag{6}$$

which behaves similarly to a stationary AR-process with the same autoregressive parameters as $X_{t,T}$ for large T under some conditions on the estimator. The exact sense in which we mean it behaves similarly is motivated by (2). In order to perform order selection, we only need the estimator of the global variance

to behave as in the stationary case. This criterion seems reasonable to use for the $X^*_{t,T}$ since the data is essentially homoscedastic. As will be shown, our order selection procedure is consistent for the correct order p, where p is the true order of the underlying process, as the following theorem states.

ASSUMPTION **1.** *(i) $L > p$, where L is the largest order under consideration and p is the correct order. (ii) $\hat{\eta}^2(u)$ is a uniformly consistent estimator of $c^2\sigma^2(u)$ on any compact subinterval $I \in (0,1)$ not containing the points of discontinuity.*

THEOREM **2.** *Let*

$$s_T^2(k) = \frac{1}{T}\sum_{t=k+1}^{T} \frac{1}{\hat{\eta}^2(\frac{t}{T})}(X_{t,T} - \sum_{j=1}^{k}\hat{\phi}_j X_{t-j,T})^2. \tag{7}$$

Define

$$p = \arg\min_{0\leq k\leq L}(\log(s_T^2(k)) + \frac{kC(T)}{T})$$

with $C(T)$ such that order selection in the stationary case is consistent (see for example Schwarz (1978), Hannan and Quinn (1979)). Under assumptions 1 and the assumptions of Phillips and Xu (2006), $\hat{p} \to p$ as $T \to \infty$ in probability.

While the definition of $s_T^2(k)$ differs slightly from using AIC on the $X^*_{t,T}$ as defined earlier, the idea is the same. We divide through by a local estimate of the variance in an attempt to make the data stationary. The difference between using (7) or the sample variance based on (6) will be asymptotically negligible.

REMARK **1.** *In light of theorem 2, the assumption regarding a uniformly consistent estimator is readily satisfied. For instance, consider a kernel estimator of the X_t^2 with a box kernel. In this case, the estimator can be written as*

$$\hat{\eta}_T^2(u) = \frac{\hat{\Sigma}^0(u+b) - \hat{\Sigma}^0(u-b)}{2b}$$

Thus,

$$\begin{aligned}\sup_{u\in I}|\hat{\eta}_T^2(u) - c^2\sigma^2(u)| &\leq \sup_{u\in I}|\frac{\hat{\Sigma}^0(u+b) - \hat{\Sigma}^0(u-b)}{2b} - \frac{\Sigma^0(u+b) - \Sigma^0(u-b)}{2b}| \\ &+ \sup_{u\in I}|\frac{\Sigma^0(u+b) - \Sigma^0(u-b)}{2b} - c^2\sigma^2(u)|\end{aligned}$$

which, for an appropriately chosen sequence b, converges to zero due to Theorem 1 and an application of the mean value theorem combined with the uniform continuity of $\sigma^2(u)$ on I.

The consistency of the estimated order is perhaps not surprising, since the order based on $X_{t,T}$ is also consistent. However, $X^*_{t,T}$ itself is not an autoregressive process though it does behave similarly to an autoregressive process if the model is correct. Also, $s^2_T(k)$ behaves similarly to the estimated variance of an autoregressive procedure. More importantly, they behave as if the errors are homoscedastic, and thus should be more likely to select the correct order than the heteroscedastic model.

4. Simulations

In this section, we present some simulation studies to show the affect of concentration in the variance function on our ability to select the order of the autoregression. We compare order selection on the corrected-for time series against the case where the heteroscedastic nature of the process is completely ignored as well as accounting for it by estimating the AR parameters using weighted conditional least squares. We show that incorporating the non-constant variance improves our ability to correctly select the order, especially when the variance function is highly concentrated. The variance function was estimated using a gaussian kernel with data selected bandwidth.

Consider the following modulated AR(1) model,

$$X_t = .9X_{t-1} + \epsilon_t\sigma(t), \qquad \text{where} \qquad \sigma^2(t) = \exp(\frac{2at}{T}) \tag{8}$$

for different values of a. $T = 1024$ observations were generated from this model, and $M = 1000$ simulations were run using the criterion (2) with $C(T) = 4$. The proportion of the trials for which each of the 3 methods selected the correct order are given in table 1.

Method	a=0(d=1.0)	a=1(d=1.1)	a=2(d=1.3)	a=3(d=1.6)	a=4(d=2.1)	a=5(d=2.5)
$X_{t,T}$	0.99	0.94	0.76	0.53	0.34	0.22
$X_{t,T}$ w/WLS	0.99	0.95	0.80	0.56	0.40	0.27
$X^*_{t,T}$	0.99	0.99	0.98	0.98	0.98	0.97

Table 1: Simulation results (proportion correct) with an AR(1) and variance function (7). WLS is weighted least squares, and d is the concentration of the variance function, i.e. the L_2-norm of the normalized variance function.

The second order selection simulation presented mimics a real data application. In figure 2 (top), a footstep from an elephant recorded with a geo-phone is shown. The model (1) is assumed, and the procedure described above selects an AR(1) model for the data, with $\phi = .94$. The estimated variance function is presented in figure 2 (bottom), and has a concentration of 2.35, as measured by the L_2 norm of the normalized function.

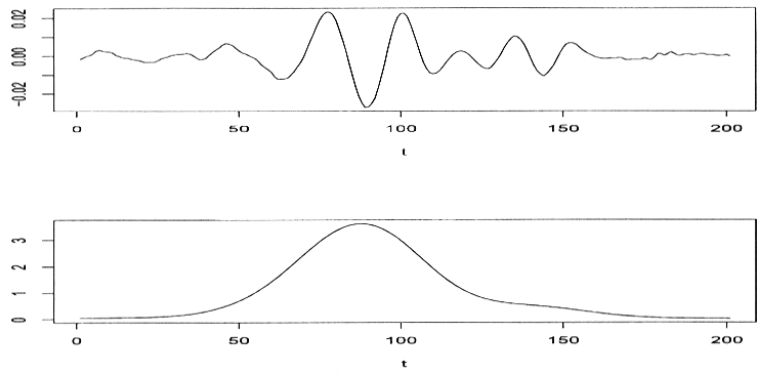

Figure 2: A geo-phone recording of an elephant's footfall (above) and the estimated (normalized) time-dependent variance function (below).

10,000 simulations from an AR(1) with $T = 201$ modulated by this variance function were performed, each resulting in an order selected for the 3 methods. The results of the simulation are given in table 2.

Method	Correct
$X_{t,T}$	0.78
$X_{t,T}$ w/WLS	0.79
$X^*_{t,T}$	0.91

Table 2: Simulation results with an AR(1) and variance function pictured in figure 2 (bottom). WLS is weighted least squares.

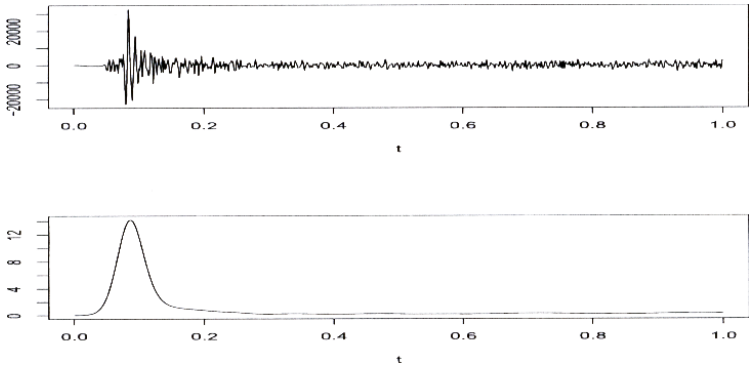

Figure 3: A seismic recording of a mining explosion (above) and the estimated time-dependent variance function (below).

The third order selection simulation presented also mimics a real data application. In figure 3 (top), a seismic recording of a mining explosion is shown. The model (1) is again assumed and, following Chandler and Polonik (2006), an AR(2) model is used with $\phi_1 = 1.65$ and $\phi_2 = -.94$. The estimated variance function is presented in figure 3 (bottom), and has a concentration of 7.65. 5,000 simulations from an AR(2) with $T = 1024$ modulated by this variance function were done, each resulting in an order selected for the 3

methods. The results of the simulation are given in table 3.

Method	Correct
$X_{t,T}$	0.52
$X_{t,T}$ w/WLS	0.58
$X^*_{t,T}$	0.90

Table 3: Simulation results with an AR(2) and variance function pictured in figure 3 (bottom). WLS is weighted least squares.

From the simulations, the effect of heteroscedasticity on the selected order can be quite strong. Weighted least squares seems to improve the chance of selecting the correct order, though the non-constant variance still seems problematic. The proposed method outperforms the others in all instances studied.

With regards to estimation, consider the model (8) with $a = 5$ and $X^*_{t,T}$ as defined above. Using a bandwidth of .2 on the squared observations to estimate the variance function, we compare the regular series, the series divided by the estimate, and the series divided by the true variance over 1000 simulations. The results are shown in table 4. Having the true variance function on hand allows for a further reduction in the standard deviation of the estimated AR parameter, though using the estimated function provides a great reduction from the unadulterated data. Consistency of the estimate will follow for the same reasons as in Theorem 2, primarily the smoothness of the variance function.

Method	Mean	Standard Deviation
X_t	.885	.049
X^*_t	.886	.019
$X_t/\sigma(t)$	.890	.014

Table 4: Simulation results for estimation of an AR(1) model with heteroscedastic errors.

5. Proofs

Proof of Theorem 1

PROOF. We first establish that $\hat{\Sigma}^0(\alpha)$ is pointwise consistent for $\Sigma^0(\alpha) = c^2 \int_0^\alpha \sigma^2(u)du$. This is simply achieved by using the technique of Phillips and Xu (2006) and dealing with partial sums to show:

$$\begin{aligned} E(\frac{1}{T}\sum_{t=0}^{[T\alpha]} X_t^2) &= \frac{1}{T}\sum_{t=0}^{[T\alpha]} E(\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\psi_i\psi_j\epsilon_{t-i}\epsilon_{t-j}) = \frac{1}{T}\sum_{t=0}^{[T\alpha]}\sum_{i=0}^{\infty}\psi_i^2\sigma^2(\frac{t-i}{T}) \\ &= \frac{1}{T}\sum_{t=0}^{[T\alpha]}\sigma^2(\frac{t}{T})\sum_{i=0}^{\infty}\psi_i^2 + o(1) \to \sum_{i=0}^{\infty}\psi_i^2\int_0^\alpha \sigma^2(u)du \end{aligned}$$

Next, uniform consistency in α can be shown using a similar technique to that used in the proof of the Glivenko-Cantelli theorem. The key idea is that both the partially integrated variance function and its empirical version are monotonically increasing functions. Details are omitted.

Proof of Theorem 2

PROOF. Let $Y_t, t = 1, \ldots, T$, be a AR(p) time series with the same AR coefficients as in the series of interest $X_{t,T}$ and constant variance $\sigma_Y^2 = \frac{1}{c^2}$, with $c^2 = \sum_{j=0}^{\infty} \psi_j^2$.

The conditions on the error process (for example in Hannan and Quinn (1979)) are readily satisfied for the process Y_t. Notice that the only randomness that appears in the criterion (2) is through the sum of squared errors, or equivalently, the estimate of the "residual standard error" $s^2(k)$. If the proposed weighted standard error (7) has an equivalent limit to that of the Y_t, then order selection based on the $X_{t,T}$ in this sense is consistent as well, since the Y_t is a stationary AR process.

To do this, we will show first that error due to estimation of the AR coefficients ϕ_{jk} is asymptotically negligible, where ϕ_{jk} is the j^{th} AR parameter when an AR(k) model is fit to the data. Next, we will show that error from estimation of the variance function is also asymptotically negligible.

Assume first that the variance function is continuous. Thus I can be chosen of the form $[\delta, 1-\delta]$ for any $\delta > 0$, in particular so the contribution of points outside I is small.

Define

$$V_k(t) = (\frac{X_t}{\hat{\eta}(\frac{t+1}{T})}, \frac{X_{t-1}}{\hat{\eta}(\frac{t+1}{T})}, \ldots, \frac{X_{t-k+1}}{\hat{\eta}(\frac{t+1}{T})})', \quad Y_k = (\frac{X_{k+1}}{\hat{\eta}(\frac{k+1}{T})}, \frac{X_{k+2}}{\hat{\eta}(\frac{k+2}{T})}, \ldots, \frac{X_T}{\hat{\eta}(\frac{T}{T})})',$$

$\hat{\Phi}_k = (\hat{\phi}_1, \hat{\phi}_2, \ldots, \hat{\phi}_k)'$, $\Phi_k = (\phi_1, \phi_2, \ldots, \phi_k)'$ and $\mathbf{V}_k = (V_k(k)', V_k(k+1)', \ldots, V_k(T-1)')$.

Consider the difference

$$\frac{1}{T}(Y - \mathbf{V}_k\hat{\Phi}_k)'(Y - \mathbf{V}_k\hat{\Phi}_k) - \frac{1}{T}(Y - \mathbf{V}_k\Phi_k)'(Y - \mathbf{V}_k\Phi_k) = -\frac{2}{T}(\hat{\Phi}_k - \Phi_k)'\mathbf{V}_k'(Y - \mathbf{V}_k\Phi_k) + o_p(1) = o_p(1) \quad (9)$$

by the consistency of the $\hat{\phi}_k$ (as they are functions of the autocovariances via the Yule-Walker equations) and consistency of $\hat{\eta}(u)$, which ensures all the $V(t)$ are finite. Next, we show that the error from the estimation of the variance function is asymptotically negligible.

Define $\epsilon_{t,k} = X_{t,T} - \sum_{j=1}^{k} \phi_{jk} X_{t-j,T}$ and $f_t(x) = \frac{\epsilon_{t,k}^2}{x}$. Thus, $f_t'(x) = -\frac{\epsilon_{t,k}^2}{x^2}$. Consider, on I,

$$\begin{aligned} \frac{1}{T}\sum_{t/T\in I}\left|f_t\left(\hat{\eta}^2(\frac{t}{T})\right) - f_t\left(c^2\sigma^2(\frac{t}{T})\right)\right| &= \frac{1}{T}\sum_{t/T\in I}\left|f_t'(\xi_t)\left(\hat{\eta}^2(\frac{t}{T}) - c^2\sigma^2(\frac{t}{T})\right)\right| \\ &\leq \sup_{u\in I}\left|\left(\hat{\eta}^2(u) - c^2\sigma^2(u)\right)\right| \frac{1}{T}\sum_{t/T\in I}|f_t'(\xi_t)| \end{aligned} \quad (10)$$

by the mean value theorem, where ξ_t is between $\hat{\eta}^2(\frac{t}{T})$ and $c^2\sigma^2(\frac{t}{T})$. (10) is $o_p(1)$ due to the uniform consistency of the estimator of the variance function, the moment condition on ϵ_t and the bounds on the variance function. Lastly, we consider

$$\frac{1}{T}\sum_{t=1+k}^{T}\frac{1}{c^2\sigma^2(\frac{t}{T})}\left(X_{t,T}-\sum_{j=1}^{k}\phi_{jk}X_{t-j,T}\right)^2,$$

which is an average of a finite sum of weighted empirical covariance terms. Letting $\phi_{0k}=-1$, we have

$$\begin{aligned}
E\quad(\quad & \frac{1}{c^2T}\sum_{t=1+k}^{T}\frac{1}{\sigma^2(\frac{t}{T})}\sum_{j=0}^{k}\sum_{l=0}^{k}\phi_{jk}\phi_{lk}X_{t-j,T}X_{t-l,T}\quad) \\
= \quad & \frac{1}{c^2T}\sum_{t=1+k}^{T}\frac{1}{\sigma^2(\frac{t}{T})}\sum_{j=0}^{k}\sum_{l=0}^{k}\phi_{jk}\phi_{lk}\sum_{m}\psi_m\psi_{j+m-l}\sigma^2(\frac{t-j-m}{T}) \\
= \quad & \frac{1}{c^2T}\sum_{t=1+k}^{T}\sum_{j=0}^{k}\sum_{l=0}^{k}\phi_{jk}\phi_{lk}\sum_{m}\psi_m\psi_{j+m-l}+o(1) \qquad (11)
\end{aligned}$$

using the technique of Phillips and Xu (2006). The first term in (11) is identical to the expectation of $s_T^2(k)$ based on the $Y_{t,T}$ series. Furthermore, the variance of these quantities is $o(1)$ due to the moment condition. Thus, these two quantities are asymptotically equivalent. In the case there are finitely many points of discontinuity, I can be chosen so that the contribution of these points is arbitrarily small asymptotically. Finally, since we are only considering finitely many k, it follows that the estimated order based on the normalized data is asymptotically equivalent to the order based on autoregressive data of the same order, and hence the selected order based on the normalized data is consistent for the true order.

6. References

AKAIKE, H. (1969) "Fitting autoregressive models for prediction", Ann. Inst. Statist. Math., **21**, 243-247.

AKAIKE, H. (1970) "Statistical predictor identification", Ann. Inst. Statist. Math., **22**, 203-217.

AKAIKE, H. (1974) "A new look at the statistical model identification, " I.E.E.E. Transactions on Automatic Control, **19** (6), 716-723.

BOLLERSLEV, T. (1986), "General autoregressive conditional heteroscedasticity," J. Econ., **31**, 307-327.

CHANDLER, G. and POLONIK, W. (2006), "Discrimination of locally stationary time series based on the excess mass functional," Journal of the American Statistical Association, **101**, 240- 253.

DAHLHAUS, R. (1997) "Fitting time series models to nonstationary processes," Ann. Statist. **25**, 1-37.

DAHLHAUS, R. and POLONIK, W. (2007), "Nonparametric quasi maximum likelihood estimation for

Gaussian locally stationary processes", Annals of Statistics, **34**, 2790-2824.

DARGAHI-NOUBARY, G.R., LAYCOCK, P.J. and SUBBA RAO, T. (1978), "Non-linear stocahstic models for seismic events with applications in event identification", Geophys. J. Roy. Astron. Soc., **55**, 655-668.

DREES, H. and STARICA, C. (2002), "A simple non-stationary model for stock returns," preprint.

ENGLE, R.F. (1982), "Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation," Econometrica, **50**, 987-1007.

FRYZLEWICZ, P., SAPATINAS, T., and SUBBA RAO, S. (2006). A Haar-Fisz technique for locally stationary volatility estimation. *Biometrika* **93**, 687-704.

FUJITA, T. and SHIBARA, H. (1978), "On a model of earthquake ground motions for response analysis and some examples of analysis through experiment," Conf. on Engineering Design for Earthquake Environments **12**, 139-148.

HANNAN, E. J. and QUINN, E.G. (1979) "The determination of the order of an autoregression," J. R. Statist. Soc. **541**, 190-195.

PAULSEN, J. and TJØSTHEIM D. (1985) "Least squares estimates and order determination procedures for autoregressive processes with a time dependent variance," Journal of Time Series Analysis **6** 117-133.

PHILLIPS, P.C.B. and XU, K.-L. (2006) "Inference in autoregression under heteroskedasticity," Journal of Time Series Analysis **27**(2), 289-308.

PÖTSCHER, B. M. (1989) "Model selection under nonstationarity: autoregressive models and stochastic linear regression models," The Annals of Statistics **17**, 1257-1274.

SCHWARZ, G. (1978) "Estimating the dimension of a model," Annals of Statistics, **6**, 461-464.

TYSSEDAL, J.S. and TJØSTHEIM D. (1982) "Autoregressive processes with a time dependent variance," Journal of Time Series Analysis **3**, 209-217.

VAN BELLEGEM, S. and DAHLHAUS, R., "Semiparametric estimation by model selection for locally stationary processes," Journal Of The Royal Statistical Society Series B, **68**(5), 721-746.

WICHERN, D.W., MILLER, R.B. and HSU, D.A. (1976) "Changes of variance in first-order autoregressive time series models - with an application," Applied Statistics **25**, 248-256.

WOOD, J.D., O'CONNELL-RODWELL, C.E., and KLEMPERER, S.L., (2005)"Using seismic sensors to detect elephants and other large mammals: a potential census technique," *Journal of Applied Ecology* (42), pp. 587-594.