Bootstrapping autoregression under non-stationary volatility

KE-LI XU

Department of Finance and Management Science, School of Business, University of Alberta
Edmonton, Alberta, T6G 2R6 Canada
E-mail: keli.xu@ualberta.ca

First version received: January 2006; final version accepted: October 2007

Summary  This paper studies robust inference in autoregression around a polynomial trend with stable autoregressive roots under non-stationary volatility. The formulation of the volatility process is quite general including many existing deterministic and stochastic non-stationary volatility specifications. The aim of the paper is two-fold. First, it develops a limit theory for least squares estimators and shows how non-stationary volatility affects the consistency, convergence rates and asymptotic distributions of the slope and trend coefficients estimators in different ways. This complements the results recently obtained by Chung and Park (2007, Journal of Econometrics 137, 230–59). Second, it studies the recursive wild bootstrap procedure of Gonçalves and Kilian (2004, Journal of Econometrics 123, 89–120) in the presence of non-stationary volatility, and shows its validity when the estimates are asymptotically mixed Gaussian. Simulations are performed to compare favourably the recursive wild bootstrap with other inference procedures under non-stationary volatility.

Key words: Autoregression, Bootstrap, Deterministic trend, Mixed Gaussian, Non-stationary volatility, Robust inference, Spurious regression, Stochastic volatility, Wild bootstrap.

1. INTRODUCTION

Recently increasing attention has been paid to modelling non-stationarity in the variances or conditional variances of a time series. One of the main reasons for that is, although the widely used ARCH-type volatility models are successful in capturing many important features in macroeconomic or financial time series such as volatility clustering and persistent autocorrelation, they are generally unable to account for the potential time-varying feature of the unconditional variances. In this paper I provide a unified non-parametric framework to analyze the effects of non-stationary volatility on the inference of the finite-order stable autoregressive model (AR(p)) around a polynomial trend. The specification of the volatility process is quite general nesting several non-stationary volatility models existing in the literature. I also establish the conditions under which the conventional Eicker–White robust procedure is valid and suggest a consistent residual-based recursive wild bootstrap method that can be used in the presence of general non-stationary volatility.

Several approaches to modelling non-stationary variances coexist in the literature. The simplest (but effective) one is to assume the unconditional variances to be a known or unknown, continuous or discontinuous function of time. For instance, see Hsu et al. (1974), Officer (1976),
Wichern et al. (1976), Merton (1980), French et al. (1987), and recently Drees and Starica (2002) and Stărică (2003). Phillips and Xu (2006) studied the robust inference in the autoregressive model under weak assumptions on this kind of unknown deterministic non-stationary volatility allowing for abrupt structural changes, trending behaviour, seasonal effects or other smooth transitions in the volatility process. In their simulation experiments, Phillips and Xu (2006) found that although the standard Eicker–White type robust procedures are asymptotically correct, they may yield poor distributional approximations in finite samples, e.g. when there are step changes in the volatility process.

Alternatively, non-stationary variances may also be modelled by assuming the volatility (viz. conditional standard deviation) process to be a transformation of a latent integrated stochastic process. Hansen (1995) formulated the conditional variances of the martingale difference (m.d.) innovations in a regression model as a non-parametric continuous function of a scaled nearly integrated autoregressive process. The non-stationary version of the popular log-normal model of stochastic volatility (SV) falls in this category. To account for stochastic trending behaviour explicitly in the variables that generate volatilities, more recently Park (2002) proposed the non-linear and non-stationary heteroskedasticity (NNH) models which treated the volatility as a function of an unscaled integrated time series. He showed that such volatility models are able to capture non-stationarity and many other stylized features in economic and financial time series such as volatility clustering and leptokurtosis. Chung and Park (2007) studied the estimation and inference of the linear regression coefficients when the volatility process follows Park’s (2002) NNH model, but they ruled out the trend and lagged dependent variables as regressors.

The general non-stationary volatility specification I use here is similar to that of Boswijk (2005), who studied the robust testing of the unit root process under non-stationary volatility. The key assumption (c.f. Section 2) requires that the scaled volatility process, when re-indexed on the unit interval, converges weakly as the sample size diverges. It allows for many previously proposed deterministic or stochastic non-stationary volatility formulations such as ones mentioned above. The asymptotic results presented there extend and complement the earlier ones obtained in Hansen (1995), Phillips and Xu (2006) and Chung and Park (2007). It makes clear that the slope and trend estimators are affected by non-stationary volatility in different ways. In particular, non-stationary volatility generally changes the convergence rates of the trend estimates to the true values and may lead to inconsistent estimation, but the convergence rates of the autoregressive slope estimates are not penalized. The limiting distributions are generally non-Gaussian, involving stochastic integral representations.

The recommended recursive wild bootstrap method presented here builds on the work of Gonçalves and Kilian (2004), who established the validity conditions for the recursive wild bootstrap in zero mean stable autoregression under conditional heteroskedasticity allowing for a variety of GARCH-type models and stationary SV models. They demonstrated via simulations the superior performance of the wild bootstrap over the standard Eicker–White robust procedures. There are several important differences between this paper and theirs. First, they assumed the unconditional variance of the time series to be constant over time thus ruled out the non-stationary behaviour in volatilities. The current paper extends their approach by allowing for unconditional heteroskedasticity in addition to conditional heteroskedasticity of the innovations. Second, I incorporate the trend regressors in our model which offers more flexibility to account for the trending behaviour in economic time series. I also demonstrate how non-stationary volatility is involved in estimating the trend coefficients and autoregressive slopes differently. Third, I give a different set of conditions under which the Eicker–White robust procedure and the recursive wild bootstrap are valid under non-stationary volatility.
Parallel to the line of this research, other authors have considered the effects of the non-stationary volatility on inference of regression models where the regressors are integrated, e.g. unit root tests (Hamori and Tokihisa 1997; Boswijk 2001, 2005; Kim et al. 2002; Cavaliere 2004b) and stationarity tests (Busetti and Taylor 2003; Cavaliere 2004a), and some robust versions have been proposed (see Cavaliere and Taylor 2007a; Beare 2005, for robust unit root tests, and Cavaliere and Taylor 2005; for robust stationarity tests). In a closely related paper, Cavaliere and Taylor (2008) considered the wild bootstrap in unit root testing in the presence of a class of deterministic permanent variance changes.

The remainder of the paper is organized as follows. Section 2 introduces the general volatility specification in the stable AR($p$) model with a polynomial trend and relates it to some existing non-stationary volatility models in the literature. A limit theory is developed in Section 3.1 and the effects of non-stationary volatility on the trend and slope estimators are discussed in details. Section 3.2 studies the wild bootstrap procedure and establishes its first-order validity condition. Some simulation results are reported in Section 4, and they reveal that more accurate coverage rates of confidence intervals are achieved by the wild bootstrap than those based on the asymptotic approximation or conventional naive bootstrap under the non-stationary volatility models considered. Section 5 concludes and discusses some possible extensions. The proofs of the main results are collected in the Appendix.

2. THE MODEL AND ASSUMPTIONS

This paper is concerned with the following AR($p$) process around a deterministic $m$th order polynomial trend (with intercept)

$$Y_t = \sum_{i=0}^{m} \delta_i t^i + \sum_{j=1}^{p} \theta_j Y_{t-j} + \varepsilon_t,$$

(2.1)

where $p$ is known and finite. We assume the autoregressive (slope) coefficients $\theta = (\theta_1, \ldots, \theta_p)^t$ are such that all roots of $\theta(L) = 1 - (\sum_{j=1}^{p} \theta_j L^j)$ are outside the unit circle. Considering a polynomial time trend, not merely a linear trend, gives more flexibility in modelling deterministically trending behaviour of economic time series since any smooth non-linear function of time admits a polynomial approximation. The error term $\varepsilon_t$ in (2.1) satisfies

$$\varepsilon_t = \sigma_t \eta_t,$$

(2.2)

where $\eta_t$ is i.i.d. with zero mean and unit variance, and $\sigma_t$ is a strictly positive stochastic volatility process to be specified. Let $\mathcal{F}_{t-1}$ be the sigma-field generated by $\eta_{t-j}$, $\sigma_{t+1-j}$, $j \geq 1$ and other latent random processes occurred up to time $t-1$, and we also assume $\eta_t$ is independent of $\mathcal{F}_{t-1}$. Then $\sigma_t^2 = \text{Var}(Y_t|\mathcal{F}_{t-1})$, viz. conditional variance of the time series $Y_t$ is fully characterized by the stochastic process $\sigma_t^2$. Since $\sigma_t^2$ may depend on the past events, the formulation (2.2) also features conditional heteroskedasticity.

To focus on the non-stationarity of the volatility process $\sigma_t$, we make the following assumptions. Throughout the paper, integrals are always taken over the limits 0 to 1 unless specified otherwise. The symbol $[\ldots]$ means integer part, and $\Rightarrow$ means weak convergence with respect to the uniform metric on $[0, 1]$. We denote a standard Brownian motion by $W(r)$ or $W^j(r)$ for a positive integer $j$. For brevity stochastic integrals like $\int X(r)dY(r)$ and integrals with respect to Lebesgue measure like $\int X(r)dr$ will be written as $\int XdY$ and $\int X$, respectively.

© Royal Economic Society 2008
ASSUMPTION 2.1. (i). There exists a non-negative process \( \sigma(\cdot) \) with piecewise continuous sample paths, such that \( \mathbb{E}(\int \sigma^2) < \infty \) and the sequence \( \sigma_t \) and \( \eta_t \) satisfy

\[
\left( n^{-1/2} \sum_{\gamma_n^{-1} \sigma_t}^{t=\lfloor ns \rfloor} f_{\tau_j} \eta_\tau \right) \overset{d}{=} \left( n^{-1/2} \sum_{\gamma_n^{-1} \sigma_{\lfloor ns \rfloor}}^{t=\lfloor ns \rfloor} \eta_\tau f_{\tau_j} \right) \Rightarrow \left( \frac{W^j(s)}{\sigma(s)} \right), \tag{2.3}
\]

as \( n \to \infty \), for any non-negative integer \( j \) and some non-random sequence \( \gamma_n \), where

\[
f_{\tau_j} = \begin{cases} 1, & \text{if } j = 0, \\ \eta_{\tau-j}, & \text{if } j \geq 1. \end{cases}
\]

(ii). For some \( a > 0 \), \( \max_{1 \leq t \leq n} \mathbb{E} \eta_t^{4+a} < \infty \) and \( \max_{1 \leq t \leq n} \mathbb{E} [\gamma_n^{-1} \sigma_t]^{4+a} < \infty \) uniformly in \( n \).

Assumption 2.1 (i) is similar in spirit to Boswijk (2005). Note that joint convergence to \( W^j(s) \) and \( \sigma(s) \) is assumed in (2.3). Under the independence between sequences \( \{\sigma_t\} \) and \( \{\eta_t\} \), (2.3) reduces to marginal convergence

\[
\gamma_n^{-1} \sigma_t := \gamma_n^{-1} \sigma_{\lfloor ns \rfloor} \Rightarrow \sigma(s), \quad \tag{2.4}
\]

\( n \to \infty \), since \( n^{-1/2} \sum_{\gamma_n^{-1} \sigma_t}^{\lfloor ns \rfloor} f_{\tau} \eta_\tau \Rightarrow W^j(s) \) holds trivially by the invariance principle for m.d. sequences. Here the non-random sequence \( \gamma_n \) reflects the explosive behaviour of the volatility process, and is referred to as the asymptotic order of non-stationary volatility in what follows. The key assumption is (2.4), viz. the volatility sequence \( \gamma_n \), redefined as a stochastic process indexed on the unit interval \([0, 1]\) after proper standardization, converges weakly when the sample size approaches infinity. In fact, it is convenient to assume (2.4) since \( \sigma_t \) can be generated by discretely sampling any continuous-time (potentially non-stationary) stochastic volatility process (Nelson 1990a; Shephard 2005). While we focus on non-stationarity here, of course stationary volatility is also allowed; e.g. (2.4) holds trivially when \( \sigma_t \) is strictly stationary such as the stationary log-normal SV model and stationary GARCH models (Nelson 1990b; Meitz 2006).

The weak convergence (2.4) is satisfied by a variety of non-stationary volatility models that have been studied in the literature, e.g. deterministic non-stationary volatility such as abrupt changes or smooth transitions (Phillips and Xu, 2006), (nearly-)integrated GARCH models, Hansen’s (1995) (nearly-)integrated autoregressive SV models (in which cases \( \gamma_n = 1 \)) and Park’s (2002) NNH volatility model (in which case \( \gamma_n \) may diverge) - see Cavaliere and Taylor (2007b) for detailed discussions. It should be noted, however, that Assumption 2.1 does not include all non-stationary volatility models existing in the literature. For instance, it rules out the GARCH models with time-varying coefficients and regime switching GARCH models.

For the validity of the robust Eicker–White procedures and the bootstrap results that will be introduced in Section 3, we need the following stronger assumption:

ASSUMPTION 2.2. \( \sigma_t \) is independent of \( \eta_t \) for any \( t \) and \( t' \).

Assumption 2.2 ensures the least squares estimates of (2.1) to be asymptotically mixed Gaussian, and the recursive wild bootstrap (c.f. Section 3.2) recovers the correct limit distribution with some further restrictions on the growth rate of \( \gamma_n \). Deterministic non-stationary volatility satisfies Assumption 2.2 trivially and for particular non-stationary stochastic volatility models, weaker versions of Assumption 2.2 ensuring mixed Gaussianity may be explored (e.g. see Hansen 1995; Assumption 4).
3. MAIN RESULTS

3.1. Limit theory

This subsection develops an asymptotic theory of the model of (2.1) under the non-stationary volatility satisfying Assumption 2.1. Before proceeding, I introduce a useful transformation of the regressors in (2.1). Hamilton (1994, section 16.3) used this transformation to derive the limit distribution for an AR\(^p\) model around a linear trend, and with some extra effort we can build by induction a more general representation that gives us the flexibility to analyze the AR\(^p\) process around a polynomial trend as in (2.1). Rewrite \(Y_t\) as

\[
Y_t = \sum_{i=0}^{m} \mu_i t^i + \sum_{j=1}^{p} \theta_j \tilde{Y}_{t-j} + \varepsilon_t, \quad (3.1)
\]

where \(\mu_m = \delta_m \frac{1}{1 - \sum_{j=1}^{p} \theta_j} \), \(\mu_{m-i} = \frac{\delta_{m-i} + \sum_{j=1}^{i} (-1)^j \binom{m-i}{j} \binom{p}{j} \theta_j \mu_{m-i-j}}{1 - \sum_{j=1}^{p} \theta_j}, 1 \leq i \leq m. \quad (3.2)\)

Then the regressors in the transformed regression equation (3.1) involve time and lags of a zero mean stable AR\(^p\) process \(\tilde{Y}_t\) which follows

\[
\theta(L) \tilde{Y}_t = \varepsilon_t. \quad (3.3)
\]

In matrix form, the original and transformed regression equations (2.1) and (3.1) can be written as

\[
Y_t = X_t' \beta + \varepsilon_t = \tilde{X}_t' \alpha + \varepsilon_t = X_t' G' (G')^{-1} \beta + \varepsilon_t, \quad (3.4)
\]

where

\[
X_t = (Y_{t-1}, \ldots, Y_{t-p}, 1, t, \ldots, t^m)', \quad \beta = (\theta_1, \ldots, \theta_p, \delta_0, \ldots, \delta_m)', \quad \tilde{X}_t = G X_t = (\tilde{Y}_{t-1}, \ldots, \tilde{Y}_{t-p}, 1, t, \ldots, t^m)', \quad \alpha = (G')^{-1} \beta = (\theta_1, \ldots, \theta_p, \mu_0, \ldots, \mu_m)',
\]

and the \((m + p + 1) \times (m + p + 1)\) matrix

\[
G' = \begin{pmatrix} I_p & 0 \\ H'_{(m+1) \times p} & I_{m+1} \end{pmatrix},
\]

with \(H = (H_{jk}) = -\sum_{h=0}^{m-k+1} \mu_{k+h-1} \binom{k+h-1}{h} (-j)^h. \)

Let \(\hat{\beta}\) and \(\hat{\alpha}\) denote the OLS estimators of regressions of \(Y_t\) on \(X_t\) and \(\tilde{X}_t(t = 1, \ldots, n)\), respectively.\(^1\) It is easy to show \(\hat{\alpha} = (G')^{-1} \hat{\beta}\) and thus we have

\[
\hat{\beta} - \beta = G'(\hat{\alpha} - \alpha). \quad (3.5)
\]

\(^1\)In fact, \(\hat{\alpha}\) is not directly computable since \(\{\tilde{X}_t\}\) involve unknown parameters and thus are unobservable. Here we need \(\hat{\alpha}\) only for theoretical purposes which will be clear in what follows.
Under the assumptions in Section 2, we can write \( \theta(L)^{-1} = \sum_{j=0}^{\infty} \psi_j L^j \) with \( \psi_0 = 1 \), where the coefficients \( \psi_j \) decline exponentially satisfying \( \sum_{j=0}^{\infty} j|\psi_j| < \infty \). Following the notation in Gonçalves and Kilian (2004), we write \( b_j = (\psi_{j-1}, \ldots, \psi_{j-p}) \) for \( j \geq 1 \), with \( \psi_j = 0 \) for \( j < 0 \), then \( (\tilde{Y}_{t-1}, \ldots, \tilde{Y}_{t-p})' = \sum_{j=1}^{\infty} b_j \varepsilon_{t-j} \). Denote \( \Omega = \sum_{j=1}^{\infty} b_j b_j' \). The following lemma gives the asymptotic distribution of the OLS estimator \( \hat{\alpha} \) of the transformed regression (3.1).

**Lemma 3.1.** (i) Define the \((p + m + 1) \times (p + m + 1)\) diagonal matrix

\[
\Upsilon = \text{diag}(\sqrt{n}, \ldots, \sqrt{n} \gamma_n^{-1}, n^{3/2} \gamma_n^{-1}, \ldots, n^{(2m+1)/2} \gamma_n^{-1})
\]

and the \((m + 1) \times (m + 1)\) (Hilbert) matrix \( R = (R_{ij}) \), where \( R_{ij} = (i + j - 1)^{-1} \) for \( i, j = 1, \ldots, m + 1 \). Then under Assumption 2.1, as \( n \to \infty \)

\[
\Upsilon(\hat{\alpha} - \alpha) \Rightarrow Q^{-1} \int V(r) dB_{p+m+1}(r),
\]

where \( B_{p+m+1}(r) \) is a vector Brownian motion with variance matrix \( \Lambda = \begin{pmatrix} \Omega & 0_{p \times (m+1)} \\ 0_{(m+1) \times p} & I_{(m+1)}' \end{pmatrix} \) with \( t = (1, \ldots, 1)' \),

\[
Q = \begin{pmatrix} \Omega \int \sigma^2 \\ 0_{(m+1) \times p} \end{pmatrix} R \text{ and } V(r) = \text{diag}(\sigma^2(r)I_p, \sigma(r), r\sigma(r), \ldots, r^m \sigma(r)).
\]

(ii) Under Assumptions 2.1 and 2.2, as \( n \to \infty \)

\[
\Upsilon(\hat{\alpha} - \alpha) \Rightarrow \mathcal{MN}_{p+m+1}(0, Q^{-1} \left( \int V \Lambda V \right) Q^{-1}),
\]

where \( \mathcal{MN}_{p+m+1}(0, Q^{-1} \left( \int V \Lambda V \right) Q^{-1}) \) denotes a \((p + m + 1)\)-dimensional mixed Gaussian distribution with mixing variate \( Q^{-1} \left( \int V \Lambda V \right) Q^{-1} \).

Lemma 3.1 shows that under Assumption 2.1 the OLS estimator of the slope parameter vector \( \theta, \hat{\theta} \), either in the untransformed regression or transformed regression, is \( \sqrt{n} \)-consistent, unaffected by the non-stationary volatility. This can be intuitively understood by noting that the non-stationary volatility affects the dependent variable and the regressors (as lagged dependent variables) simultaneously so that it leaves the convergence rates of the least square estimates unchanged. In contrast, the convergence rates of the trend coefficient estimators \( \hat{\mu}_0, \ldots, \hat{\mu}_m \) in the transformed regression involve \( \gamma_n' \), being affected by multiplying the usual convergence rates under homoskedasticity by the factor \( \gamma_n^{-1} \).

Our purpose is to estimate the parameters in the original regression (2.1). The limit behaviour of the OLS estimator \( \hat{\beta} = (\hat{\theta}_1, \ldots, \hat{\theta}_p, \hat{\delta}_0, \ldots, \hat{\delta}_m)' \) of the original regression (2.1) can be derived as a consequence of Lemma 3.1. By (3.5), each \( \hat{\delta}_i \) is a linear combination of elements in \( \hat{\alpha} \); so its convergence rate is determined by the slowest one among elements in \( \hat{\alpha} \); in other words, for each \( i = 0, \ldots, m \), we need to compare the growing rates of \( n^{(2i+1)/2} \gamma_n^{-1} \) and \( \sqrt{n} \), or to see whether \( n^{-1} \gamma_n \) diminishes or not as \( n \) goes to infinity. The asymptotic distribution of each element in \( \hat{\beta} \) is summarized in the following theorem.
Theorem 3.1. (i) Under Assumption 2.1, as \( n \to \infty \)
\[
\sqrt{n}(\hat{\theta} - \theta) \Rightarrow \frac{\Omega^{-1/2}}{f \sigma^2} \int \sigma^2 dW_p;
\]
\[
\sqrt{n}(\hat{\delta}_i - \delta_i) \Rightarrow \bar{g}_{p+i+1} Q^{-1} \int V dB_{p+m+1}, \text{ if } n^{-i} \gamma_n \to c, \ |c| < \infty,
\]
\[
n^{(2i+1)/2} \gamma_n^{-1}(\hat{\delta}_i - \delta_i) \Rightarrow \bar{g}_{p+i+1} Q^{-1} \int V dB_{p+m+1}, \text{ if } n^i \gamma_n^{-1} \to 0,
\]
for \( i = 0, \ldots, m \), where \( W_p \) is a standard \( p \)-dimensional standard Brownian motion, \( \bar{g}_{p+i+1} \) is the same as the \( (p+i+1) \)-th row of \( G' \) except with \( (p+i+1) \)-th element \( c \) rather than \( 1 \) (when \( c = 1 \), \( \bar{g}_{p+i+1} \) is the \( (p+i+1) \)-th row of \( G' \)), and \( \bar{g}_{p+i+1} \) is the same as \( (p+i+1) \)-th row of \( G' \) except with first \( p \) elements \( 0 \)'s rather than elements in \( (i+1) \)-th row of \( H' \).

(ii) Under Assumptions 2.1 and 2.2, as \( n \to \infty \)
\[
\sqrt{n}(\hat{\delta}_i - \delta_i) \Rightarrow \mathcal{MN}_p \left(0, \frac{f \sigma^4}{(f \sigma^2)^2} \Omega^{-1} \right)
\]
\[
n^{(2i+1)/2} \gamma_n^{-1}(\hat{\delta}_i - \delta_i) \Rightarrow \mathcal{MN}_{p+m+1}(0, \bar{g}_{p+i+1} Q^{-1} (\int V^\Lambda V) Q^{-1} \bar{g}_{p+i+1}), \text{ if } n^{-i} \gamma_n \to c, \ |c| < \infty,
\]
\[n^{(2i+1)/2} \gamma_n^{-1}(\hat{\delta}_i - \delta_i) \Rightarrow \mathcal{MN}_{p+m+1}(0, \bar{g}_{p+i+1} Q^{-1} (\int V^\Lambda V) Q^{-1} \bar{g}_{p+i+1}), \text{ if } n^i \gamma_n^{-1} \to 0,
\]
for \( i = 0, \ldots, m \), where \( \bar{g}_{p+i+1} \) and \( \bar{g}_{p+i+1} \) are defined in part (i).

Remark 3.1. Theorem 3.1 shows that under Assumption 2.1 the convergence rate of the estimate of the trend coefficient, \( \hat{\delta}_i \) (\( 0 \leq i \leq m \)) depends on the asymptotic order of non-stationary volatility \( \gamma_n \). Specifically, \( \hat{\delta}_i \) is \( \sqrt{n} \)-consistent if
\[
\gamma_n \propto n^k, \ k \in (-\infty, i]
\]
and its limit distribution is that in (3.7); \( \hat{\delta}_i \) is \( n^{i+1/2} \gamma_n^{-1} \)-consistent (slower than \( \sqrt{n} \) but still consistent) if
\[
\gamma_n \propto n^k, \ k \in \left(i, i + \frac{1}{2}\right)
\]
and its limit distribution is that in (3.8). In other words, for a particular \( i \), we need to control the explosive behaviour of the volatility process to ensure the consistency of \( \hat{\delta}_i, \hat{\delta}_i \) is consistent if
\[
\gamma_n \propto n^k, \ k \in \left(-\infty, i + \frac{1}{2}\right).
\]
Conversely, if the non-stationary volatility diverges so fast that \( \gamma_n \) increases at least with the same rate as \( n^{i+1/2} \) does, then \( \hat{\delta}_i \) is inconsistent. This is called spurious regression by Granger and Newbold (1974) and Phillips (1986) in that we do not have meaningful estimators for model
Table 1. Convergence rates of $\hat{\delta}_0$ and $\hat{\delta}_1$ under non-stationary volatility when a linear trend is assumed in model (2.1). NA means inconsistent estimation.

<table>
<thead>
<tr>
<th>$\gamma_n \propto n^k$</th>
<th>$k \leq 0$</th>
<th>$0 &lt; k &lt; \frac{1}{2}$</th>
<th>$\frac{1}{2} \leq k \leq 1$</th>
<th>$1 &lt; k &lt; \frac{3}{2}$</th>
<th>$k \geq \frac{3}{2}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{\delta}_0$</td>
<td>$\sqrt{n}$</td>
<td>$n^{2-k}$</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>$\hat{\delta}_1$</td>
<td>$\sqrt{n}$</td>
<td>$\sqrt{n}$</td>
<td>$\sqrt{n}$</td>
<td>$n^{2-k}$</td>
<td>NA</td>
</tr>
</tbody>
</table>

parameters. From (3.12) we can see the estimate of the trend coefficient of higher order has wider consistency range in terms of the growing rate of $\gamma_n$. Thus all $\hat{\delta}_i (0 \leq i \leq m)$'s are consistent if the $\hat{\delta}_0$ is consistent, viz.

$$\gamma_n \propto n^k, k \in \left( -\infty, \frac{1}{2} \right).$$

(3.13)

These results complement the findings in Chung and Park (2007), who studied asymptotics of linear regressions with strictly exogenous regressors under non-stationary volatility. The results in their paper only apply to the case of $\hat{\delta}_0$ in our context.

REMARK 3.2. Consider a simple case of model (2.1) with the linear trend ($m = 1$). Convergence rates of the trend estimators $\hat{\delta}_0$ and $\hat{\delta}_1$ under non-stationary volatility of difference asymptotic orders $\gamma_n$ are listed in Table 1. For example, suppose that $\sigma_t$ has the form of non-stationary volatility satisfying Assumption 2.1 with $\gamma_n = 1$ and all coefficient estimates $\hat{\delta}_i$ and $\hat{\theta}$ are $\sqrt{n}$--consistent, just as in the case of homogeneous innovations (Hamilton 1994, section 16.3). Non-stationary volatility only plays roles in their limit distributions. Now suppose $\sigma_t$ follow the NNH model of Park (2002) with $\sigma_t = g(S_0^t)$, where $S_0^t$ is an integrated process and $g(\cdot) = |\ldots|^K$. Then when $0 < K < 1$, $\hat{\delta}_1$ is $\sqrt{n}$-consistent and $\hat{\delta}_0$ is also consistent but with a slower rate $n^{1/2-K/2}$. When $K = 1$, $\hat{\delta}_0$ becomes inconsistent but $\hat{\delta}_1$ is still $\sqrt{n}$-consistent.

REMARK 3.3. Theorem 3.1 extends the earlier contributions on regression under non-stationary volatility. Hansen (1995) considered a special case of non-stationary volatility allowed by Assumption 2.1, and (3.6) applies to his model with $\sigma(\cdot)$ being a function of an Ornstein–Uhlenbeck diffusion process. Phillips and Xu (2006) considered deterministic structural changes in volatility and the limit theory developed in their paper can be obtained by (3.9) with $\sigma(\cdot)$ being a continuous or discontinuous non-random function.

Non-Gaussianity of the limiting distribution of OLS estimates under non-stationary volatility has been noted earlier by Hansen (1995) and Chung and Park (2007). They also discussed the conditions like Assumption 2.2 here to ensure mixed Gaussianity of the OLS estimates.

REMARK 3.4. Theorem 3.1 also extends the results in estimating the model (2.1) under homogeneous errors. Hamilton (1994) shows that, assuming i.i.d. errors, all the trend and the slope estimators have the same convergence rate $\sqrt{n}$. This result continues to be true when the innovations exhibit mild non-stationary volatility, as shown in (3.6) and (3.7). However, excessive non-stationary volatility would lead to different convergence rates of $\hat{\delta}_i$ over $i$.

Under Assumptions 2.1 and 2.2, the mixed Gaussianity of OLS estimates implies that we can construct tests of hypotheses on regression coefficients provided that $\hat{\beta}$ is consistent. For instance, $t$-tests are asymptotically standard normal under the null hypothesis provided that (3.13) holds.
COROLLARY 3.1. Define \( \hat{\varepsilon}_t = Y_t - X'_t \hat{\beta} = Y_t - \hat{X}'_t \hat{\alpha} \) and \( \gamma \) as in Lemma 3.1. Then under Assumptions 2.1, 2.2 and (3.13), as \( n \to \infty \),

\[
\begin{align*}
t_{\theta_j} &= \frac{\sqrt{n}(\hat{\theta}_j - \theta_j)}{\sqrt{\hat{C}_{jj}}} \Rightarrow N(0, 1), \text{ for } j = 1, \ldots, p; \\
t_{\delta_0} &= \begin{cases} 
\sqrt{n}(\hat{\delta}_0 - \delta_0)/\sqrt{\hat{C}_{p+1,p+1}} \Rightarrow N(0, 1), \text{ if } \gamma_n \to c, |c| < \infty; \\
\sqrt{n}\gamma_n^{-1}(\hat{\delta}_0 - \delta_0)/\sqrt{\hat{C}_{p+1,p+1}} \Rightarrow N(0, 1), \text{ if } \gamma_n \to \infty \text{ and } \frac{\gamma_n}{\sqrt{n}} \to 0; 
\end{cases}
\end{align*}
\]

where \( \hat{C} \) is the \((j, j)-\) element of \( \hat{C} \) and

\[
\hat{C} = \left( \gamma^{-1} \sum_{i=1}^{n} X_i X'_i \right)^{-1} \left( \sum_{i=1}^{n} X_i \hat{\varepsilon}'_i \right) \left( \sum_{i=1}^{n} X_i X'_i \gamma^{-1} \right)^{-1}
\]

is the robust estimator of the asymptotic covariance matrix of \( \hat{\beta} \).

Corollary 3.1 shows that \( t \)-tests can be constructed using coefficient estimates and estimated residuals from the regression (2.1), rather than the transformed regression (3.1); no knowledge of the matrix \( G \) is required. Note that \( \gamma_n \) is involved in the variance estimator \( \hat{C} \). So the construction of the test statistics depends on which non-stationary volatility model is used.

### 3.2. Residual-based bootstrap

In Section 3.1, we have shown that under non-stationary volatility when the estimates are asymptotically mixed Gaussian, valid inference can be achieved using Eicker–White-type standard errors. But this method based on asymptotic approximation may lead to inaccurate inference in small samples, as demonstrated in our simulation experiments in Section 4. To improve it, in this subsection we analyze the residual-based bootstrap methods of Gonçalves and Kilian (2004) for the autoregressive model (2.1) in the presence of non-stationary volatility.

Let \( \hat{\varepsilon}_t = Y_t - \sum_{i=0}^{m} \hat{\delta}_i t^i - \sum_{j=1}^{p} \hat{\theta}_j Y_{t-j} \) denote the OLS regression residuals of model (2.1). The recursive design bootstrap data generating process (DGP) follows

\[
Y^*_t = \sum_{i=0}^{m} \hat{\delta}_i t^i + \sum_{j=1}^{p} \hat{\theta}_j Y^*_{t-j} + \hat{\varepsilon}^*_t = X_t' \hat{\beta} + \hat{\varepsilon}^*_t,
\]

where \( X'_t = (Y^*_{t-1}, \ldots, Y^*_{t-p}, 1, t, \ldots, t^m)' \) and each time \( \hat{\varepsilon}^*_t (t = 1, \ldots, n) \) is drawn randomly from some distribution utilizing the information in \( \hat{\varepsilon}_t \). The sequence \( \{Y^*_t\} \) is initialized with \( \{Y^*_0, \ldots, Y^*_0\} = \{0, \ldots, 0\} \). The bootstrap estimator of \( \hat{\beta} \), \( \hat{\beta}^* = (\hat{\theta}_1^*, \ldots, \hat{\theta}_p^*, \hat{\delta}_0^*, \ldots, \hat{\delta}_m^*)' \) is obtained by applying OLS to the bootstrap data \( \{Y^*_t : t = 1, \ldots, n\} \). Bootstrap versions of the statistics of interest, such as normalized sampling errors or \( t \)-statistics, are constructed by using the bootstrap estimator \( \hat{\beta}^* \) and seeing \( \hat{\beta} \) as the true value of the bootstrap DGP. The empirical distribution of these bootstrap statistics is used to approximate the distribution of the original statistics of interest.

© Royal Economic Society 2008
For the conventional residual-based bootstrap (i.i.d. bootstrap) (see, e.g. Bose 1988; 1990; Kreiss and Franke 1992), for each $t = 1, \ldots, n$, $\hat{\varepsilon}^*_{it}$ is drawn independently from the same distribution—the empirical distribution of $\{\hat{\varepsilon}_t: t = 1, \ldots, n\}$. Following the arguments in Gonçalves and Kilian (2004), it can be shown that under the bootstrap probability measure, the conventional residual-based bootstrap estimator $\hat{\beta}^*$ converges weakly to the same Gaussian limit distribution as in the case of i.i.d. innovations (Hamilton 1994, section 16.3), neglecting potential non-stationary volatility. It is not surprising since the conventional bootstrap treats $\varepsilon_t$ as independent and identically distributed, and thus fails to capture the potential heteroskedasticity in the errors. This is analogous to Gonçalves and Kilian (2004), in which the conventional bootstrap is shown to fail to recover the correct asymptotic covariance matrix in the presence of conditional heteroskedasticity.

The wild bootstrap proposed by Wu (1986) and Liu (1988) is designed to allow heteroskedasticity in linear regression models, and seems to be the better choice when the volatility is shown to fail to recover the correct asymptotic covariance matrix in the presence of conditional heteroskedasticity. Unlike the standard bootstrap, the wild bootstrap draws each $\hat{\varepsilon}^*_i \sim \mathcal{N}(0, \sigma^2_i)$ from a different distribution with mean zero and variance $\hat{\varepsilon}^2_i$. The following theorem evaluates the limit behaviour of wild bootstrap analogs of the sampling errors in model (2.1) $\hat{\beta} - \beta$ and $\hat{\delta}_i - \delta_i$, $0 \leq i \leq m$.

**Theorem 3.2.** Under Assumption 2.1 and (3.13), as $n \to \infty$,

$$\sqrt{n}(\hat{\theta}^* - \theta) \Rightarrow \mathcal{M}_p \left(0, \frac{f}{f \sigma^4} \Omega^{-1} \right), \text{ in probability;}$$

$$\sqrt{n}(\hat{\delta}^*_0 - \delta_0) \Rightarrow \mathcal{M}_{p+m+1} \left(0, \bar{g}_{p+1} Q^{-1} \left(\int \Lambda V \Lambda V \right) Q^{-1} \bar{g}_{p+1} \right), \text{ if } \gamma_n \to c, |c| < \infty;$$

$$\sqrt{n} \gamma_n^{-1} (\hat{\delta}^*_0 - \delta_0) \Rightarrow \mathcal{M}_{p+m+1} \left(0, \bar{g}_{p+1} Q^{-1} \left(\int \Lambda V \Lambda V \right) Q^{-1} \bar{g}_{p+1} \right), \text{ if } \gamma_n \to \infty \text{ and } \frac{\gamma_n}{\sqrt{n}} \to 0;$$

$$\sqrt{n}(\hat{\delta}_i^* - \delta_i) \Rightarrow \mathcal{M}_{p+m+1} \left(0, \bar{g}_{p+i+1} Q^{-1} \left(\int \Lambda V \Lambda V \right) Q^{-1} \bar{g}_{p+i+1} \right), \text{ in probability},$$

where $\Rightarrow$ in probability means weak convergence under the bootstrap probability measure occurring in a set with probability converging to one, and $\bar{g}'_{p+i+1}$ and $\bar{g}_{p+i+1}$ are defined in Theorem 3.1.

Note that Theorem 3.2 holds without imposing Assumption 2.2. It shows that the centred and standardized residual-based recursive wild bootstrap estimators converge to a mixed Gaussian distribution, implicitly neglecting the potential dependence between $\{\sigma_t\}$ and $\{\eta_t\}$. In other words, the residual-based recursive wild bootstrap procedure described above recovers the correct limiting distributions of sampling errors if the innovations in model (2.1) with non-stationary volatility satisfy both Assumptions 2.1 and 2.2.

Theorem 3.2 still holds if we use $Y_{t-j}$ in place of $Y_{t-j}^*$ in the bootstrap DGP in (3.16). We do not pursue this (fixed-design) bootstrap scheme further here, since we observe in the simulations
that its finite-sample performance in constructing percentile-\(t\) confidence interval is dominated by the recursive-design bootstrap scheme as in (3.16).\(^2\)

The wild bootstrap analogs of the \(t\)-test statistics are constructed similarly. Let \(\hat{C}^*_j\) be the \((j, j)\)-element of \(\hat{C}^*\), where

\[
\hat{C}^* = \left( \gamma^{-1} \sum_{t=1}^{n} X_t^* X_t^{*\prime} \right)^{-1} \left( \sum_{t=1}^{n} X_t^* X_t^{*\prime} \hat{\epsilon}^*_t \right) \left( \sum_{t=1}^{n} X_t^* X_t^{*\prime} \gamma^{-1} \right)^{-1}
\]

with bootstrap residuals \(\hat{\epsilon}^*_t = Y_t^* - X_t^{*\prime} \hat{\beta}^*\). The following corollary shows that when both Assumptions 2.1 and 2.2 hold, the empirical distribution of \(t^*_{\hat{\theta}_j}\) and \(t^*_{\hat{\delta}_i}\) can be used to approximate the distribution of \(t_{\hat{\theta}_j}\) and \(t_{\hat{\delta}_i}\) in Corollary 3.1.

**COROLLARY 3.2.** Under Assumption 2.1 and (3.13), as \(n \to \infty\),

\[
t^*_{\hat{\theta}_j} = \sqrt{n}(\hat{\theta}^*_j - \hat{\theta}_j) / \sqrt{\hat{C}^*_j} \overset{d}{\Rightarrow} N(0, 1), \text{ for } j = 1, \ldots, p;
\]

\[
t^*_{\hat{\delta}_i} = \begin{cases} 
\sqrt{n}(\hat{\delta}^*_0 - \hat{\delta}_0) / \sqrt{\hat{C}^*_{p+1,p+1}} \overset{d}{\Rightarrow} N(0, 1), & \text{if } \gamma_n \to c, |c| < \infty; \\
\sqrt{n}\gamma_n^{-1}(\hat{\delta}^*_0 - \hat{\delta}_0) / \sqrt{\hat{C}^*_{p+1,p+1}} \overset{d}{\Rightarrow} N(0, 1), & \text{if } \gamma_n \to \infty \text{ and } \frac{\sqrt{n}}{\gamma_n} \to 0; \\
\sqrt{n}(\hat{\delta}^*_{i+1} - \hat{\delta}_{i+1}) / \sqrt{\hat{C}^*_{p+1+p+1}} \overset{d}{\Rightarrow} N(0, 1), & \text{for } i = 1, \ldots, m.
\end{cases}
\]

Assumption 2.2, which is needed in validating the recursive wild bootstrap procedure in our context, is closely related to the symmetry assumption of Gonçalves and Kilian (2004, p. 96, assumption A (iv)). Assumption 2.2 excludes the so-called leverage effect in the stochastic volatility literature (Black 1976; Shephard 2008). Under non-stationary volatility, the symmetry-type condition Assumption 2.2 plays two roles: it not only ensures the mixed Gaussianity of the OLS estimators which always holds in the bootstrap world no matter whether Assumption 2.2 is imposed, but also ensures the wild bootstrap to recover the correct asymptotic variance. Note that assumption 2.2 in our paper implies assumption A (iv) in Gonçalves and Kilian (2004). Although Assumption 2.1 here is much weaker than the maintained assumptions, assumption A of Gonçalves and Kilian (2004), we pay a price by imposing a stronger additional condition than theirs to ensure the validity of the wild bootstrap. The intuition provided in their paper regarding why the recursive wild bootstrap requires the symmetry condition is also useful in explaining why Assumption 2.2 is needed here.\(^3\)

### 4. SIMULATIONS

In this section, we perform simulation experiments to assess the finite-sample accuracy of the wild bootstrap approximation, compared with the methods based on i.i.d. bootstrap and asymptotic approximation. We focus on the zero mean AR(1) model. The DGP is

\[^2\]Gonçalves and Kilian (2004) also considered the pairwise bootstrap under conditional heteroskedasticity. Under our non-stationary volatility framework, the results of Theorem 3.2 type can be also established for this bootstrap scheme.  
\[^3\]I thank a referee for pointing out this to me.
The first one involves a deterministic abrupt shift in variances. The design used here is due to Cavaliere (2004b). We assume the non-stationarity of the volatility process is due to a step variance change of \( \varepsilon_t = \sigma_t \eta_t, \theta \in \{0.5, 0.9\} \). We consider two kinds of non-stationary volatility.

The second kind of non-stationary volatility we consider is Park’s (2002) NNH stochastic volatility model. Suppose \( \sigma_t = g(s) \) with \( g(s) = |s|^K \). The integrated process \( S_t^0 \) is such that \( (1 - L)S_t^0 = \varepsilon_t^0 \), where \( \varepsilon_t^0 = 0.5z_t \) and \( (z_t^0) \sim \text{i.i.d. } \mathcal{N}(0, 2) \). Then (2.4) is satisfied with \( \gamma_n = n^{K/2} \) and \( \sigma(s) \) being a function of a standard Brownian motion (see Park and Phillips 1999, 2001). We set \( K = \{0.5, 1\} \) corresponding to the case \( \gamma_n = n^{0.25} \) and \( \sqrt{n} \). We take \( \sigma_{25} \in \{0, 0.1, 0.2, 0.3\} \) to check the robustness of methods for inference under consideration to Assumption 2.2.

The second kind of non-stationary volatility we consider is Park’s (2002) NNH stochastic volatility model. Suppose \( \sigma_t = g(S_t^0) \) with \( g(s) = |s|^K \). The integrated process \( S_t^0 \) is such that \( (1 - L)S_t^0 = \varepsilon_t^0 \), where \( \varepsilon_t^0 = 0.5z_t \) and \( (z_t^0) \sim \text{i.i.d. } \mathcal{N}(0, 2) \). Then (2.4) is satisfied with \( \gamma_n = n^{K/2} \) and \( \sigma(s) \) being a function of a standard Brownian motion (see Park and Phillips 1999, 2001). We set \( K = \{0.5, 1\} \) corresponding to the case \( \gamma_n = n^{0.25} \) and \( \sqrt{n} \). We take \( \sigma_{25} \in \{0, 0.1, 0.2, 0.3\} \) to check the robustness of methods for inference under consideration to Assumption 2.2.

The sample size is set to be 50, 100 and 250. For the bootstrap method, the bootstrap replication is 999 and we choose \( v_i \sim \text{i.i.d. } \mathcal{N}(0, 1) \) for the wild bootstrap.

We are interested in the coverage accuracy of the nominal 90% symmetric percentile—\( t \) confidence interval for \( \theta \). Three methods of constructing confidence intervals are compared and all of them are based on least squares estimation of (4.1) with an intercept. We report the actual type I errors of confidence intervals using the asymptotic distribution (ASY) (Eicker–White–type standard errors are used), recursive design conventional bootstrap (CB) and wild bootstrap (WB) procedures, based on 4000 replications. For each method we report the percentage of samples in which the true value of \( \theta \) lay to the left (right) of estimated confidence interval in the row labelled \( L \) (\( R \)) (So ideally the entries should be close to 0.05). All results are contained in Table 2 for a jump in volatility, and Tables 3 and 4 for Park’s (2002) NNH model with \( K = 0.5 \) and 1, respectively.

Several observations are in order. For the deterministic non-stationary volatility under consideration, we can see the performance of the asymptotic interval is poor for both cases \( \theta = 0.5 \) and 0.9. When \( \theta = 0.5 \), the left side type I error is much larger than 0.05, and when \( \theta = 0.9 \) both sides errors are quite different from 0.05. The small sample distributions appear to be seriously skewed to the left and the actual coverage rates of the resulting confidence intervals are much lower than the nominal level 90%. As sample sizes increase, the performance of the asymptotic interval improves when \( \theta = 0.5 \), but the small sample distributions continue to be skewed when \( \theta = 0.9 \).

The wild bootstrap intervals have much less coverage distortion on both sides than asymptotic intervals for both cases of \( \theta = 0.5 \) and 0.9. As the sample size increases, the type I errors of the wild bootstrap intervals are fairly close to 0.05 in all cases considered, and appear to be robust to the position and the magnitude of the shift. Surprisingly we also observe that the conventional bootstrap intervals perform fairly well and lead to improvements over asymptotic intervals, although significant type I errors are observed when \( \theta = 0.9 \) and a positive variance shift occurs at the end of the sample (\( \tau = 0.8, \delta = 5 \)).
Table 2. Actual type I errors of confidence intervals based on asymptotic distribution (ASY), recursive design conventional bootstrap (CB), and wild bootstrap (WB), under deterministic non-stationary volatility in (4.2), for $\theta \in \{0.5, 0.9\}$, $\tau \in \{0.2, 0.5, 0.8\}$ $\delta \in \{0.2, 1, 5\}$ and the sample size $n \in \{50, 100, 250\}$.

<table>
<thead>
<tr>
<th>$\tau$</th>
<th>$\delta$</th>
<th>$\theta = 0.5$</th>
<th>$\theta = 0.9$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>ASY</td>
<td>CB</td>
</tr>
<tr>
<td>.2</td>
<td>.2</td>
<td>118</td>
<td>035</td>
</tr>
<tr>
<td>1</td>
<td>.120</td>
<td>0.071</td>
<td>0.072</td>
</tr>
<tr>
<td>5</td>
<td>.105</td>
<td>0.058</td>
<td>0.067</td>
</tr>
<tr>
<td>.5</td>
<td>.5</td>
<td>0.057</td>
<td>0.072</td>
</tr>
<tr>
<td>.8</td>
<td>.2</td>
<td>0.053</td>
<td>0.065</td>
</tr>
<tr>
<td>5</td>
<td>.138</td>
<td>0.087</td>
<td>0.108</td>
</tr>
<tr>
<td></td>
<td></td>
<td>.2</td>
<td>.2</td>
</tr>
<tr>
<td>1</td>
<td>.097</td>
<td>0.066</td>
<td>0.053</td>
</tr>
<tr>
<td>5</td>
<td>.085</td>
<td>0.058</td>
<td>0.049</td>
</tr>
<tr>
<td>.5</td>
<td>.2</td>
<td>0.054</td>
<td>0.047</td>
</tr>
<tr>
<td>5</td>
<td>.086</td>
<td>0.045</td>
<td>0.047</td>
</tr>
<tr>
<td>.8</td>
<td>.2</td>
<td>0.043</td>
<td>0.055</td>
</tr>
<tr>
<td>5</td>
<td>.094</td>
<td>0.065</td>
<td>0.084</td>
</tr>
<tr>
<td></td>
<td></td>
<td>.2</td>
<td>.2</td>
</tr>
<tr>
<td>1</td>
<td>.082</td>
<td>0.053</td>
<td>0.048</td>
</tr>
<tr>
<td>5</td>
<td>.081</td>
<td>0.064</td>
<td>0.050</td>
</tr>
<tr>
<td>.5</td>
<td>.2</td>
<td>0.054</td>
<td>0.059</td>
</tr>
<tr>
<td>5</td>
<td>.072</td>
<td>0.050</td>
<td>0.074</td>
</tr>
<tr>
<td>.8</td>
<td>.2</td>
<td>0.045</td>
<td>0.048</td>
</tr>
<tr>
<td>5</td>
<td>.088</td>
<td>0.068</td>
<td>0.056</td>
</tr>
</tbody>
</table>

For Park’s (2002) NNH stochastic volatility case, the conventional bootstrap intervals are subject to serious undercoverage, especially when $K = 1$. The type I errors on both sides are much larger than 0.05. As in the case of a shift in variance, asymptotic intervals have the largest coverage distortion among the three methods of constructing confidence intervals considered. The wild bootstrap confidence intervals lead to some improvements over other two methods. Actually they are very accurate when the sample size is 250. One particularly interesting thing is that the performance of wild bootstrap confidence interval is also quite good when $\sigma_{z\eta} \neq 0$, in which case Assumption 2.2 is violated.

© Royal Economic Society 2008
Table 3. Actual type I errors of confidence intervals based on asymptotic distribution (ASY), recursive design conventional bootstrap (CB), and wild bootstrap (WB), under non-stationary and non-linear heteroskedasticity of Park (2002) for $K = 0.5$, $\sigma_{zt} \in \{0, 0.1, 0.2, 0.3\}$, $\theta \in \{0.5, 0.9\}$ and the sample size $n \in \{50, 100, 250\}$.

<table>
<thead>
<tr>
<th>$\sigma_{zt}$</th>
<th>$n = 50$</th>
<th>$n = 100$</th>
<th>$n = 250$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><strong>ASY</strong></td>
<td><strong>CB</strong></td>
<td><strong>WB</strong></td>
</tr>
<tr>
<td></td>
<td>$L$</td>
<td>$R$</td>
<td>$L$</td>
</tr>
<tr>
<td>0</td>
<td>0.124</td>
<td>0.036</td>
<td>0.095</td>
</tr>
<tr>
<td>0.1</td>
<td>0.096</td>
<td>0.031</td>
<td>0.076</td>
</tr>
<tr>
<td>0.2</td>
<td>0.120</td>
<td>0.045</td>
<td>0.093</td>
</tr>
<tr>
<td>0.3</td>
<td>0.098</td>
<td>0.038</td>
<td>0.065</td>
</tr>
</tbody>
</table>

5. CONCLUSION AND POSSIBLE EXTENSIONS

In this paper, I establish a limit theory for the autoregressive model with stable autoregressive roots around a polynomial trend under non-stationary volatility in wide sense. The residual-based recursive wild bootstrap procedure is shown to be consistent when the OLS estimates are asymptotically mixed Gaussian. Simulations reveal that the wild bootstrap possesses better finite sample performance than inference procedures based on the asymptotic approximation under non-stationary volatility. Several extensions are possible.

First, it is natural to ask whether there exists a bootstrap method which recovers the correct limit distribution in general allowing for the dependence between $\{\sigma_t\}$ and $\{\eta_t\}$, viz. under Assumption 2.1 without imposing Assumption 2.2. However, even if such a bootstrap procedure exists, it may not produce asymptotic refinements since in the presence of non-stationary volatility the limit distribution is non-pivotal, depending on nuisance parameters involved in the volatility function $\sigma_t$. One possible solution for this complication is to first estimate the volatility function $\sigma_t$ non-parametrically (as in Hansen 1995; Xu and Phillips 2008) and then apply the conventional residual-based bootstrap method to the re-weighted regression equation.

Second, Chung and Park (2007) studied a specification of the volatility process for which least squares estimators are mixed Gaussian in the limit without imposing the independence assumption (like Assumption 2.2). They assumed the volatility function to be an integrable transformation of...
Table 4. Actual type I errors of confidence intervals based on asymptotic distribution (ASY), recursive design conventional bootstrap (CB), and wild bootstrap (WB), under non-stationary and nonlinear heteroskedasticity of Park (2002) for $K = 1$, $\sigma_z \eta \in \{0, 0.1, 0.2, 0.3\}$, $\theta \in \{0.5, 0.9\}$ and the sample size $n \in \{50, 100, 250\}$.

<table>
<thead>
<tr>
<th>$\sigma_z \eta$</th>
<th>$\theta = 0.5$</th>
<th>$\theta = 0.9$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ASY</td>
<td>CB</td>
</tr>
<tr>
<td>$n = 50$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>.159</td>
<td>.054</td>
</tr>
<tr>
<td>0.1</td>
<td>.138</td>
<td>.042</td>
</tr>
<tr>
<td>0.2</td>
<td>.140</td>
<td>.053</td>
</tr>
<tr>
<td>0.3</td>
<td>.124</td>
<td>.052</td>
</tr>
<tr>
<td>$n = 100$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>.153</td>
<td>.052</td>
</tr>
<tr>
<td>0.1</td>
<td>.200</td>
<td>.025</td>
</tr>
<tr>
<td>0.2</td>
<td>.129</td>
<td>.078</td>
</tr>
<tr>
<td>0.3</td>
<td>.131</td>
<td>.065</td>
</tr>
<tr>
<td>$n = 250$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>.144</td>
<td>.099</td>
</tr>
<tr>
<td>0.1</td>
<td>.125</td>
<td>.103</td>
</tr>
<tr>
<td>0.2</td>
<td>.148</td>
<td>.099</td>
</tr>
<tr>
<td>0.3</td>
<td>.119</td>
<td>.095</td>
</tr>
</tbody>
</table>

an integrated process. Asymptotic validity of the wild bootstrap considered here without imposing Assumption 2.2 is expected in this case but it needs further verification.

Third, I assume in this paper that the order of the autoregression $p$ is finite and known. Relaxation of this assumption is important and useful in many practical cases such as when we believe the data are generated from an ARMA model or the autoregressive order is selected from information criteria like AIC or BIC. Gonçalves and Kilian (2007) considered robust inference in AR($\infty$) under conditional heteroskedasticity of unknown form. However, the asymptotic properties of the various order-selection information criteria under non-stationary volatility are still unknown. But it seems likely that the framework in Ploberger and Phillips (1996) and Phillips (1996) is general enough to apply in this case.

Finally, it is also interesting and important to consider the robust inference under non-stationary volatility when we do not know whether the autoregressive roots are outside, on or inside the unit circle. All of these topics are left for future research.

**Acknowledgements**

This paper is based on the third chapter of my Ph.D. dissertation submitted to Yale University. I am grateful to my advisor, Peter C. B. Phillips, for his invaluable guidance and support. I also benefited from...
the comments by Editor, Richard J. Smith, Co-Editor, Frank Windmeijer, three anonymous referees, Donald W. K. Andrews, Yongmiao Hong, Paramaraj Jegenathan, Jiaoyang Li, Robert A. M. Taylor and other participants at the econometrics seminar at Yale University, the Singapore Econometric Study Group (SESG), July, 2005 and the International Symposium on Econometric Theory and Application (SETA) in Xiamen, April, 2006.

REFERENCES


© Royal Economic Society 2008
In APPENDIX A, we state the proofs of the main results presented in Section 3. For a \( d(d \geq 1) \)-dimensional vector \( x = (x_1, \ldots, x_d) \), its Euclidean norm is defined as \( |x| = (\sum_{i=1}^d x_i^2)^{1/2} \). For a random variable \( x \), its \( L^q(q \geq 1) \)-norm is defined as \( \|x\|_q = (E|x|^q)^{1/q} \). The following lemma is useful throughout the section.

**Lemma A1:** (Billingsley 1968, theorem 4.2, p. 25; see also Brockwell and Davis 1991, proposition 6.3.9) If \( \xi_n \) and \( \chi_{nL}(n \geq 1) \) are random vectors such that

(i) \( \chi_{nL} \Rightarrow \chi_L \) as \( n \to \infty \) for each \( L \geq 1 \),

(ii) \( \chi_{nL} \Rightarrow \chi_L \) as \( n \to \infty \), and

(iii) \( \lim_{L \to \infty} \limsup_{n \to \infty} P(|\xi_n - \chi_{nL}| > \epsilon) = 0 \) for every \( \epsilon > 0 \), then \( \xi_n \Rightarrow \chi \) as \( n \to \infty \).

**Proof of Lemma 3.1.** (i) Let \( \tilde{X}_t = (Z_{t1}', T_t')' \) where \( Z_t = (\tilde{Y}_{t-1}, \ldots, \tilde{Y}_{t-p})' \) and \( T_t = (1, t, \ldots, t^n)' \). Since \( \gamma(\alpha - \alpha) = (\gamma_n^{-2}Y^{-1} \sum_{i=1}^n \tilde{X}_i \tilde{X}_i'Y^{-1})^{-1}(\gamma_n^{-2}Y^{-1} \sum_{t=1}^n \tilde{X}_t \tilde{e}_t' \), Lemma 3.1 (i) follows from

\[
\gamma_n^{-2}Y^{-1} \sum_{i=1}^n \tilde{X}_i \tilde{X}_i'Y^{-1} \Rightarrow Q \tag{A.1}
\]

and \( \gamma_n^{-2}Y^{-1} \sum_{i=1}^n \tilde{X}_i \tilde{e}_i' \Rightarrow \int Vd\sigma_{p+m+1} \), where the first is from (a) (b) (c) and the second from (d) and (e) in the following:

(a) \( n^{-1} \gamma_n^{-2} \sum_{i=1}^n Z_t Z_t' \Rightarrow \Omega \int \sigma^2 \);  
(b) for \( i = 0, \ldots, m, n^{-1} \gamma_n^{-2} \sum_{i=1}^n i^p Z_t Z_t' \Rightarrow 0 \);  
(c) \( \sum_{t=1}^n T_t T_t' \Rightarrow R, \) where \( R = \text{diag}(n^{1/2}, n^{3/2}, \ldots, n^{(2m+1)/2}) \);  
(d) \( n^{-1/2} \gamma_n^{-2} \sum_{i=1}^n Z_t \tilde{e}_t \Rightarrow \int \sigma^2 d\sigma_p \), where \( \sigma_p(r) \) is a vector Brownian motion with the variance matrix \( \Omega \).

(e) \( \gamma_n^{-1} \sum_{t=1}^n T_t \tilde{e}_t \Rightarrow \int d\sigma(p, r \sigma, \ldots, r^m \sigma) dB_{m+1} \), where \( B_{m+1}(r) \) is a vector Brownian motion with the variance matrix \( u'r \), independent of \( B_p(r) \) in (d).

© Royal Economic Society 2008
We prove (a) first. Since \( Z_t = \sum_{j=1}^{\infty} b_j e_{t-j} \), we define
\[
\xi_n \defeq n^{-1} \gamma_n^{-2} \sum_{t=1}^{n} \sum_{i,j=1}^{\infty} b_i b_j' \left( n^{-1} \gamma_n^{-2} \sum_{r=1}^{n} e_{t-r} e_{t-j} \right).
\]

Now we show \( \xi_n \Rightarrow \Omega \int \sigma^2 \) using Lemma A1. First we establish that, for fixed \( i \) and \( j \),
\[
n^{-1} \gamma_n^{-2} \sum_{t=1}^{n} e_{t-i} e_{t-j} \Rightarrow \int \sigma^2, \quad \text{if } i = j, \quad \mathbb{P} > 0 \text{ if } i \neq j.
\]  \( \text{(A.2)} \)

If \( i = j \),
\[
\mathbb{E} \left( n^{-1} \gamma_n^{-2} \sum_{t=1}^{n} e_{t-i}^2 - n^{-1} \sum_{t=1}^{n} \gamma_n^{-2} \sigma_{t-i}^2 \right)^2
= n^{-2} \mathbb{E} \left[ \sum_{t=1}^{n} \gamma_n^{-2} \sigma_{t-i}^2 \left( \eta_{t-i}^2 - 1 \right) \right]^2
= n^{-2} \sum_{t=1}^{n} \left[ \gamma_n^{-4} \mathbb{E} \sigma_{t-i}^4 \right] \mathbb{E} (\eta_{t-i}^2 - 1)^2 + 2n^{-2} \gamma_n^{-4} \mathbb{E} \sigma_{t-i}^2 \mathbb{E} (\eta_{t-i}^2 - 1)^2 \mathbb{E} (\xi_{t-i}^2 - 1) = 0,
\]
by Assumption 2.1(ii). So by Markov inequality \( n^{-1} \gamma_n^{-2} \sum_{t=1}^{n} e_{t-i}^2 = n^{-1} \sum_{t=1}^{n} |\xi_{t-i}|^2 = o_P(1) \Rightarrow \int \sigma^2 \) by Assumption 2.1(i). If \( i \neq j \), without loss of generality, assume \( i < j \). Similarly, it can be shown
\[
n^{-1} \gamma_n^{-2} \sum_{t=1}^{n} e_{t-i} e_{t-j} = n^{-1} \gamma_n^{-2} \sum_{t=1}^{n} \mathbb{E} e_{t-i} e_{t-j} + o_P(1)
= n^{-1} \gamma_n^{-2} \sum_{t=1}^{n} \mathbb{E} \sigma_{t-i} \eta_{t-j} + o_P(1) = o_P(1).
\]

Thus (A.2) holds. Since \( Z_t = \sum_{j=1}^{\infty} b_j e_{t-j} \), we write
\[
n^{-1} \gamma_n^{-2} \sum_{t=1}^{n} Z_t Z_t' = \sum_{i,j=1}^{\infty} b_i b_j' \left( n^{-1} \gamma_n^{-2} \sum_{t=1}^{n} e_{t-i} e_{t-j} \right).
\]

For fixed \( L \), let \( \chi_{nL} \defeq \sum_{i,j=1}^{L} b_i b_j' (n^{-1} \gamma_n^{-2} \sum_{t=1}^{n} e_{t-i} e_{t-j}) \Rightarrow \sum_{i=1}^{L} b_i b_j' \int \sigma^2 \defeq \chi_L \) as \( n \to \infty \) by (A.2). As \( L \to \infty \), \( \chi_L \Rightarrow \sum_{i=1}^{\infty} b_i b_j' \int \sigma^2 = \Omega \int \sigma^2 \defeq \chi \). Further, we have
\[
\mathbb{P}(|\xi_n - \chi_{nL}| > \epsilon) \leq \epsilon^{-1} \mathbb{E} \left( \sum_{i,j,L+1} b_i b_j' \left( n^{-1} \gamma_n^{-2} \sum_{t=1}^{n} e_{t-i} e_{t-j} \right) \right)
\]
\[
\leq \epsilon^{-1} \sum_{i,j,L+1} |b_i| \cdot |b_j| \left( n^{-1} \sum_{t=1}^{n} \left[ \gamma_n^{-4} \mathbb{E} \sigma_{t-i}^4 \mathbb{E} \sigma_{t-j}^2 \right]^{1/2} \right)
\]
\[
\leq C \epsilon^{-1} \sum_{i,j,L+1} |b_i| \cdot |b_j|, \quad \text{for any } n \text{ and some constant } C
\]
\[
\to 0, \quad \text{as } L \to \infty.
\]  \( \text{(A.3)} \)
where exchange of the order of expectation and infinite sum follows from absolute summability (see, e.g., Rao 1973, p. 111). Thus (a) follows from Lemma A1.

(b) We write \( \xi_n \stackrel{\text{def}}{=} n^{-(i+1)} \gamma_n^{-1} \sum_{t=1}^n t' Z_t = \sum_{j=1}^n b_j [n^{-1} \sum_{t=1}^n (t/n)' \gamma_n^{-1} \xi_{t-j}] \). Let \( \chi_{nL} \stackrel{\text{def}}{=} \sum_{j=1}^L b_j [n^{-1} \sum_{t=1}^n (t/n)' \gamma_n^{-1} \xi_{t-j}] \). For fixed \( j \), \( (tn)' \gamma_n^{-1} \xi_{t-j} \) is a uniformly integrable \( L^1 \)-mixingale since \( \mathbb{E} [(tn)' \gamma_n^{-1} \xi_{t-j}]^2 \leq \gamma_n^{-2} E \xi_{t-j}^2 = \gamma_n^{-2} E \sigma_{t-j}^2 < \infty \). By the law of large numbers (Andrews 1988) \( n^{-1} \sum_{t=1}^n (t/n)' \gamma_n^{-1} \xi_{t-j} \overset{p}{\to} 0 \), and so \( \chi_{nL} \overset{p}{\to} 0 \) as \( n \to \infty \). By the similar arguments in (A.3) we have \( \mathbb{P} (|\xi_n - \chi_{nL}| > \epsilon) \to 0 \) as \( L \to \infty \). Thus (b) follows from Lemma A1.

(c). It follows from the fact that \( n^{-(i+1)} \sum_{t=1}^n t' \to (i + 1)^{-1} \) for \( i = 0, 1, \ldots \).

(d). For \( j = 0, 1, \ldots, \), define \( S_j' := n^{-1/2} \sum_{l=1}^n f_{l,j} \eta_l \) and \( dS_j := S_j' - S_{j-1}' = n^{-1/2} \sum_{l=1}^n f_{l,j} \eta_l \), where \( f_{l,j} \) is as in Assumption 2.1(i). Let \( \xi_n \stackrel{\text{def}}{=} n^{-1/2} \gamma_n^{-2} \sum_{t=1}^n Z_t \xi_t = \sum_{j=1}^L b_j n^{-1/2} \sum_{l=1}^n \gamma_n^{-2} \sigma_{t-j} \sigma_t \eta_{t-j} \eta_t \) and

\[
\chi_{nL} \stackrel{\text{def}}{=} \sum_{j=1}^L b_j n^{-1/2} \sum_{l=1}^n \gamma_n^{-2} \sigma_{t-j} \sigma_t \eta_{t-j} \eta_t = \sum_{j=1}^L b_j \sum_{l=1}^n \gamma_n^{-2} \sigma_{t-j} \sigma_t dS_j' \cdot
\]

For fixed \( j \), \( \sum_{t=1}^n \gamma_n^{-2} \sigma_{t-j} \sigma_t dS_j' \to \int \sigma^2 dW^j \) as \( n \to \infty \), by assumption \( A(i) \) and theorem 2.1 of Hansen (1992). So for fixed \( L \), \( \chi_{nL} \to \sum_{j=1}^L b_j \int \sigma^2 dW^j \) as \( n \to \infty \). Note that \( W_j \) is independent over \( j \). In particular, for \( j \geq 1 \), \( \mathbb{E} S_j' S_j' = n^{-1} \sum_{r,s} \mathbb{E} \eta_{r-j} \eta_{s-j} \eta_r \eta_s = n^{-1} \sum_{t,s} \mathbb{E} \eta_{r-j} \eta_{s-j} \eta_r \eta_s = 0 \), if \( r > s \); \( \mathbb{E} S_j' S_{j-1}' = n^{-1} \sum_{t,s} \mathbb{E} \eta_{r-j} \eta_{s-j} \eta_r \eta_s = 0 \), if \( r \leq s \). For \( i, j \geq 1 \), without loss of generality, assume \( i < j \). \( \mathbb{E} S_j' S_i' = n^{-1} \sum_{t,s} \mathbb{E} \eta_{r-i} \eta_{s-i} \eta_{s-j} \eta_r = 0 \), if \( r > s \); \( \mathbb{E} S_j' S_{j-i}' = n^{-1} \sum_{t,s} \mathbb{E} \eta_{r-i} \eta_{s-i} \eta_{s-j} \eta_r = 0 \), if \( r \leq s \). Let \( B_j = b_j W_j \) and \( B_p = \sum_{j=1}^p B_j \). As \( L \to \infty \),

\[
\chi_{nL} \to \int \sigma^2 dW^j = \int \sigma^2 dB_j \quad \text{and} \quad \chi_{n} \to \int \sigma^2 dB_p \quad \text{as} \quad n \to \infty,
\]

by dominated convergence theorem and noting \( |\chi_{nL}| \leq \sum_{j=1}^\infty |b_j| \cdot |\int \sigma^2 dW^j| \) and \( \mathbb{E} (\int \sigma^2 dW^j) = \mathbb{E} (\int \sigma^2) = \infty \). Also

\[
\mathbb{P} (|\xi_n - \chi_{nL}| > \epsilon) \leq n^{-1} \sum_{j=1}^\infty |b_j| \sum_{l=1}^n \gamma_n^{-2} \sigma_{t-j} \sigma_t \eta_{t-j} \eta_t \leq n^{-1} \sum_{j=1}^\infty |b_j| \left( n^{-1} \sum_{l=1}^n \mathbb{E} \eta_{t-j}^2 \left[ \gamma_n^{-2} \mathbb{E} \sigma_{t-j}^4 \right]^{1/2} \right) \to 0,
\]

as \( L \to \infty \) by Assumption 2.1 (ii). Thus (d) follows from Lemma A1.

(e) It follows from

\[
\gamma_n^{-1} \sum_{t=1}^n T_t \xi_t = \sum_{i=1}^n \sqrt{n^{-1} T_t' \cdot \gamma_n^{-1} \sigma_t} \cdot d \left( n^{-1/2} \sum_{t=1}^n \eta_t \right)
\]

\[
\to \int (\sigma, r \sigma, \ldots, \sigma^m \sigma) dW^0 = \int \text{diag}(\sigma, r \sigma, \ldots, \sigma^m \sigma) dW^0 = \int \text{diag}(\sigma, r \sigma, \ldots, \sigma^m \sigma) dB_{m+1}
\]

by assumption 2.1(i) and theorem 2.1 in Hansen (1992). Independence of \( B_{m+1}(r) \) in (e) and \( B_p(r) \) in (d) follows from the independence of \( W^0(s) \) and \( W^j(s) \) for \( j \geq 1 \).
So the proof of Lemma 3.1 (i) is complete.

(ii). Under Assumption 2.2, $V(r)$ is independent of $B_{p+m+1}(r)$. Then the stochastic integral in (i) reduces to a mixed Gaussian random variable. The (ii) follows immediately from (i).

\section*{Proof of Corollary 3.1}

If we establish

\begin{equation}
\left( \gamma^{-1} \sum_{i=1}^{n} \tilde{X}_i \tilde{X}_i' \right)^{-1} \left( \sum_{i=1}^{n} \tilde{X}_i \tilde{X}_i' \tilde{e}_i^2 \right) \left( \sum_{i=1}^{n} \tilde{X}_i \tilde{X}_i' \right)^{-1} = Q^{-1} \left( \int V \Lambda V \right) Q^{-1},
\end{equation}

then using $X_i = G^{-1} \tilde{X}_i$, we have

\[
\tilde{C} = \left( \gamma^{-1} \sum_{i=1}^{n} X_i X_i' \right)^{-1} \left( \sum_{i=1}^{n} X_i X_i' \tilde{e}_i \right) \left( \sum_{i=1}^{n} X_i X_i' \right)^{-1}
\]

\[
= \left( \gamma^{-1} G^{-1} \sum_{i=1}^{n} \tilde{X}_i \tilde{X}_i' G^{-1} \right)^{-1} \left( \gamma^{-1} G^{-1} \sum_{i=1}^{n} \tilde{X}_i \tilde{X}_i' G^{-1} \right)^{-1}.
\]

\[
= \gamma G' \left[ \sum_{i=1}^{n} \tilde{X}_i \tilde{X}_i' \right] \left( \sum_{i=1}^{n} X_i X_i' \tilde{e}_i \right) \left( \sum_{i=1}^{n} \tilde{X}_i \tilde{X}_i' \right)^{-1} (\gamma' \gamma^{-1})^{-1} \gamma' g_j
\]

\[
= \gamma' \left( \gamma, \gamma^{-1} \right)^{-1} \left( \gamma^{-1} \sum_{i=1}^{n} \tilde{X}_i \tilde{X}_i' \right)^{-1} \left( \sum_{i=1}^{n} \tilde{X}_i \tilde{X}_i' \tilde{e}_i \right) \left( \sum_{i=1}^{n} \tilde{X}_i \tilde{X}_i' \right)^{-1} (\gamma' \gamma^{-1})^{-1} \gamma' g_j
\]

For $1 \leq j \leq p$, $\sqrt{n}(\widehat{\theta}_j - \theta_j)/\sqrt{Q^{-1}(\int V \Lambda V) Q^{-1}} \Rightarrow \mathcal{N}(0, 1)$ and $\tilde{C}_{jj} \Rightarrow Q^{-1}(\int V \Lambda V) Q^{-1}$.

For $1 \leq i \leq m$, if $n^{-1/2} \gamma_n \rightarrow 0$, $\sqrt{n}(\widehat{\theta}_i - \delta_i)/\sqrt{Q^{-1}(\int V \Lambda V) Q^{-1}} \Rightarrow \mathcal{N}(0, 1)$ and $\tilde{C}_{p+i, p+i} \Rightarrow \mathcal{N}(0, 1)$ with $c = 0$. For $i = 0$, if $\gamma_n \rightarrow \infty$ and $n^{-1/2} \gamma_n \rightarrow 0$, then Corollary 3.1 holds. Now we prove (A.4), which follows from

\begin{equation}
\gamma_n^{-4} \gamma^{-1} \sum_{i=1}^{n} \tilde{X}_i \tilde{X}_i \tilde{e}_i^2 \gamma^{-1} \Rightarrow \int V \Lambda V.
\end{equation}

First we have

\begin{equation}
\gamma_n^{-4} \gamma^{-1} \sum_{i=1}^{n} \tilde{X}_i \tilde{X}_i \tilde{e}_i^2 \gamma^{-1} \Rightarrow \int V \Lambda V.
\end{equation}
(c). \( \gamma_n^{-2} \mathbf{1} T T' \mathbf{1}^{-1} \sum_{i=1}^{n} T_i T'_i \mathbf{1}^{-1} \Rightarrow f(\sigma, r_\sigma, \ldots, r_m \sigma)(\sigma, r_\sigma, \ldots, r_m \sigma); \)

We prove (a) first. Note that \( \gamma_n^{-2} \sum_{i=1}^{L} b_i b_i' \sum_{j=1}^{n} \gamma_j \gamma_{i-j}^2 \Rightarrow \sum_{i=1}^{L} b_i b_i' \int \sigma^4 \) as \( n \to \infty. \) Then (a) holds by Lemma A1 provided that

\[
P\left( \left| \sum_{i,j=L+1}^{\infty} b_i b_i' \frac{1}{n} \sum_{j=1}^{n} \gamma_j \gamma_{i-j}^2 \right| > \epsilon \right) \to 0,
\]

which follows from arguments similar to those in (A.3) since \( \max_{1 \leq t \leq n} \mathbb{E} \eta_t^4, \max_{1 \leq t \leq n} \mathbb{E} [\gamma_n^{-1} \sigma_t]^4 \leq C < \infty \) (Assumption 2.1 (ii)).

(b). Note that for fixed \( j \geq 1, \) \( \gamma_n^{-3} n^{-(i+1)} \sum_{i=1}^{n} t_i \epsilon_{i-j}^2 \Rightarrow 0. \) Indeed, \( \gamma_n^{-3} (t/n)^j (\epsilon_{i-j}^2 - \epsilon_{i-j}^2) \) is a uniformly integrable m.d. sequence. By a WLLN for m.d. sequences,

\[
\gamma_n^{-3} n^{-(i+1)} \sum_{i=1}^{n} t_i \epsilon_{i-j}^2 = n^{-1} \sum_{i=1}^{n} (t/n)^j \gamma_n^{-3} \epsilon_{i-j}^2 + o_p(1)
\]

\[
= n^{-1/2} \sum_{i=1}^{n} (t/n)^j \gamma_n^{-3} \sigma_i^2 d \left( n^{-1/2} \sum_{t=1}^{i} \eta_{i-t} \right) + o_p(1)
\]

\[
\Rightarrow n^{-1/2} \int r^s \sigma^2 dW^0 + o_p(1) = o_p(1).
\]

Thus \( \gamma_n^{-3} n^{-(i+1)} \sum_{i=1}^{n} t_i Z_i \epsilon_i^2 = \sum_{i=1}^{n} b_i n^{-1} \sum_{i=1}^{n} (t/n)^j \gamma_n^{-3} \epsilon_{i-j}^2 \Rightarrow 0 \) by Lemma A1 provided that

\[
P\left( \left| \sum_{j=L+1}^{\infty} b_i b_i' \frac{1}{n} \sum_{j=1}^{n} \gamma_j \gamma_{i-j}^2 \right| > \epsilon \right) \to 0, \quad \text{as } L \to \infty,
\]

which follows from the similar arguments in (A.3).

(c). By a WLLN for m.d. sequences

\[
\gamma_n^{-2} \mathbf{1} T T' \mathbf{1}^{-1} \sum_{i=1}^{n} T_i T'_i \mathbf{1}^{-1} \Rightarrow \gamma_n^{-2} \mathbf{1} T T' \mathbf{1}^{-1} + o_p(1)
\]

\[
= \sum_{i=1}^{n} \left( \gamma_n^{-1} \mathbf{1}^{-1} T_i \sigma_i \right) \left( \gamma_n^{-1} \mathbf{1}^{-1} T_i \sigma_i \right)' + o_p(1)
\]

\[
\Rightarrow \int (\sigma, r_\sigma, \ldots, r_m \sigma)' (\sigma, r_\sigma, \ldots, r_m \sigma).
\]

So (c) holds. Combining the results (a), (b) and (c) gives (A.6). Then

\[
\gamma_n^{-2} \mathbf{1} T T' \mathbf{1}^{-1} \sum_{i=1}^{n} \tilde{X}_i \tilde{X}_i \mathbf{1}^{-1} = \gamma_n^{-2} \mathbf{1} T T' \mathbf{1}^{-1} \sum_{i=1}^{n} \tilde{X}_i \tilde{X}_i \mathbf{1}^{-1} + o_p(1)
\]

\[
= \sum_{i=1}^{n} \left( \gamma_n^{-1} \mathbf{1}^{-1} \tilde{X}_i \hat{\sigma} \right) \left( \gamma_n^{-1} \mathbf{1}^{-1} \tilde{X}_i \hat{\sigma} \right)' + o_p(1)
\]

\[
\Rightarrow \int V \Lambda V,
\]

where the first equality follows from (A.1) and the fact that \( \gamma_n^{-2} (\hat{\epsilon}_i^2 - \epsilon_i^2) = O_p(n^{-1/2}), \) which holds since

\[
\hat{\epsilon}_i = \epsilon_i - X_i (\hat{\beta} - \beta) = \epsilon_i - \tilde{X}_i (\hat{\alpha} - \alpha) = \epsilon_i + O_p(1/\sqrt{n}). \quad (A.7)
\]

© Royal Economic Society 2008
Proof of Theorem 3.2. Define $\hat{Y}_{t-j} = Y_{t-j}^* - \sum_{i=0}^{m} \hat{\mu}_i (t-j)^i, 0 \leq j \leq p$. Let $\hat{\mu}_m$ and $\hat{\mu}_{m-i} (1 \leq i \leq m)$ be the right-hand side of (3.2) except using $\delta_i, \theta_i$ and $\hat{\mu}_i$ instead of $\delta_i, \theta_i$ and $\mu_i$. So (3.16) can rewritten as

$$Y_t^* = \sum_{i=0}^{m} \hat{\mu}_i t^i + \sum_{j=1}^{n} \hat{\theta}_j Y_{t-j}^* + \hat{\varepsilon}_t^* = \hat{X}_t^* \hat{\alpha} + \hat{\varepsilon}_t^*,$$

where $\hat{X}_t^* = (\hat{Y}_{t-n}^*, \ldots, \hat{Y}_{t-p}^*, 1, t, \ldots, t^m)$. Let $\hat{\alpha}^*$ be the OLS coefficient estimates of regressing $Y_t^*$ on $1, t, \ldots, t^m, \hat{Y}_{t-1}^*, \ldots, \hat{Y}_{t-p}^*$. To prove Theorem 3.2, it suffices to show

$$\mathcal{Y}(\hat{\alpha}^* - \alpha) = \left(\gamma_n^{-2} \mathcal{Y}^{-1} \sum_{j=1}^{n} \hat{X}_j^* \hat{X}_j^* \mathcal{Y}^{-1}\right)^{-1} \left(\gamma_n^{-2} \mathcal{Y}^{-1} \sum_{j=1}^{n} \hat{X}_j^* \hat{\varepsilon}_j^*\right) \xrightarrow{d_{\mathbb{P}}} \mathcal{MN}_{p+m+1} \left(0, Q^{-1} \left(\int \mathcal{V} \mathcal{V} \right) Q^{-1}\right),$$

which follows from results

(a) $n^{-1} \gamma_n^{-2} \sum_{j=1}^{n} Z_j^* Z_j^* \xrightarrow{d_{\mathbb{P}}} \Omega \mathcal{V} \sigma^2$, in probability, where $Z_j^* = (\hat{Y}_{t-j}^*, \ldots, \hat{Y}_{t-p}^*)';$

(b) for $i = 0, \ldots, m, n^{-1} \gamma_n^{-1} \sum_{j=1}^{n} t^i Z_j^* \xrightarrow{d_{\mathbb{P}}} 0$, in probability;

(c) $n^{-1/2} \gamma_n^{-2} \sum_{j=1}^{n} Z_j^* \hat{\varepsilon}_j^* \xrightarrow{d_{\mathbb{P}}} \mathcal{MN}_{p} (0, \Omega \mathcal{V} \sigma^2)$, in probability;

(d) $\gamma_n^{-1} \mathcal{Y}^{-1} \sum_{j=1}^{n} T_j^* \mathcal{V} \sigma \xrightarrow{d_{\mathbb{P}}} \mathcal{MN} (0, \int (\sigma, r \sigma, \ldots, r^m \sigma) (\sigma, r \sigma, \ldots, r^m \sigma), \mathcal{V} \sigma)$, in probability.

(e) $E_n^* \gamma_n^{-2} \sum_{j=1}^{n} Z_j^* \mathcal{Y}^{-1} \sum_{j=1}^{n} T_j^* \mathcal{V} \sigma = 0$.

Here by $\hat{\varepsilon}_n^* \xrightarrow{p} 0$ in probability we mean $\mathbb{P}^*(|\hat{\varepsilon}_n^*| > \epsilon) = o_p(1)$ for any $\epsilon > 0$. Note that $\hat{\varepsilon}_n^* \xrightarrow{p} 0$ implies $\hat{\varepsilon}_n^* \xrightarrow{d_{\mathbb{P}}} 0$.

Using the autoregression in (3.3), under Assumption 2.1 it can be shown to be strongly consistent for (converges $a.s - \mathbb{P}$ to) $\theta$ by establishing $n^{-1} \sum_{j=1}^{n} \gamma_n^{-2} \hat{Y}_{t-j}^* \varepsilon_{t-j} \xrightarrow{a.s - \mathbb{P}} 0$, which holds by the strong LLN for martingale differences (see, e.g. White 1999; theorem 3.76) provided that

$$\left\|\gamma_n^{-2} \hat{Y}_{t-j}^* \varepsilon_{t-j} \right\|_2 \leq \sum_{j=0}^{\infty} \left|\psi_j\right| \left\|\gamma_n^{-2} \varepsilon_{t-j} \right\|_2 \leq \sum_{j=0}^{\infty} \left|\psi_j\right| \left\|\gamma_n^{-2} \varepsilon_{t-j} \right\|_2 \leq C \sum_{j=0}^{\infty} \left|\psi_j\right| < \infty.$$

Here we make the simplifying assumption that the bootstrap DGP (3.16) is initialized at minus infinity, since the initial condition does not affect the limit distribution theory as long as it is bounded. Thus, for sufficiently large $n > \bar{n}$, $\hat{\theta} (L) = \hat{\theta}_0 + \hat{\theta}_1 L + \cdots + \hat{\theta}_p L^p$ is invertible, $a.s - \mathbb{P}$, and $\hat{Y}_t^* = \hat{\theta} (L)^{-1} \hat{\varepsilon}_t^* \xrightarrow{d_{\mathbb{P}}} \sum_{j=0}^{\infty} \psi_j \hat{\varepsilon}_{t-j}^*$ with $\sum_{j=0}^{\infty} \left|\psi_j\right| < \infty, a.s - \mathbb{P}$. Define $\hat{b}_j = (\hat{\psi}_{j-1}, \ldots, \hat{\psi}_{j-p})$ for $j \geq 1$, with $\hat{\psi}_j = 0$ for $j < 0$. Thus $Z_j^* = \sum_{j=1}^{\infty} \hat{b}_j \hat{\varepsilon}_{t-j}^*$. Denote $\hat{\Omega} = \sum_{j=1}^{\infty} \hat{b}_j \hat{b}_j^*$. We prove (a) by using Lemma A1 in the bootstrap world. First we show for fixed $i$ and $j$,

$$n^{-1} \sum_{j=1}^{n} \gamma_n^{-2} \hat{\varepsilon}_{t-i} \hat{\varepsilon}_{t-j} \xrightarrow{d_{\mathbb{P}}} \int \sigma^2, \text{ if } i = j, \xrightarrow{p} 0 \text{ if } i \neq j.$$  

(A.8)

© Royal Economic Society 2008
When $i = j$, we have $n^{-1} \sum_{t=1}^{n} \gamma_{n}^{-2} \tilde{e}_{t-i}^2 - n^{-1} \sum_{t=1}^{n} \gamma_{n}^{-2} \tilde{e}_{t-i}^2 \xrightarrow{p} 0$ provided that

$$E^* \left( n^{-1} \sum_{t=1}^{n} \gamma_{n}^{-2} \tilde{e}_{t-i}^2 - n^{-1} \sum_{t=1}^{n} \gamma_{n}^{-2} \tilde{e}_{t-i}^2 \right)^2$$

$$= n^{-2} \gamma_{n}^{-4} E^* \left( \sum_{t=1}^{n} \tilde{e}_{t-i}^2 (v_{t-i}^2 - 1) \right)^2$$

$$= n^{-2} \sum_{t=1}^{n} \left( \gamma_{n}^{-1} \tilde{e}_{t-i} \right)^4 E^* (v_{t-i}^2 - 1)^2$$

$$\leq C n^{-3} \sum_{t=1}^{n} \left( \gamma_{n}^{-1} \tilde{e}_{t-i} \right)^4 = o_p(1),$$

which holds since

$$\gamma_{n}^{-1} \tilde{e}_{t} = \gamma_{n}^{-1} \tilde{e}_{t} + o_p(1) = O_p(1) \quad (A.9)$$

by (A.7), provided that $\gamma_n / \sqrt{n} \rightarrow 0$. The case when $i \neq j$ in (A.8) can be proved similarly.

For fixed $L$, we have $\sum_{i,j=1}^{L} \tilde{b}_i \tilde{b}_j (n^{-1} \sum_{t=1}^{n} \gamma_{n}^{-2} \tilde{e}_{t-i} \tilde{e}_{t-j}) \xrightarrow{d} \sum_{i,j=1}^{L} b_i b_j \sigma^2$, as $n \rightarrow \infty$, by (3.5), consistent estimation of the impulse responses and the continuous mapping theorem, and $\sum_{i,j=1}^{L} b_i b_j \int \sigma^2 \xrightarrow{d} \sum_{i=1}^{n} b_i b_i \int \sigma^2$ when $L \rightarrow \infty$. Now (a) holds provided that

$$E^* \left( \sum_{i,j=L+1}^{\infty} \tilde{b}_i \tilde{b}_j \left( n^{-1} \sum_{t=1}^{n} \gamma_{n}^{-2} \tilde{e}_{t-i} \tilde{e}_{t-j} \right) \right) > \epsilon$$

$$\leq \epsilon^{-1} E^* \left( \sum_{i,j=L+1}^{\infty} \tilde{b}_i \tilde{b}_j \left( n^{-1} \sum_{t=1}^{n} \gamma_{n}^{-2} \tilde{e}_{t-i} \tilde{e}_{t-j} \right) \right)$$

$$\leq \epsilon^{-1} \sum_{i,j=L+1}^{\infty} \left| \tilde{b}_i \right| \cdot \left| \tilde{b}_j \right| \left( n^{-1} \sum_{t=1}^{n} \left( \gamma_{n}^{-2} \tilde{e}_{t-i} \tilde{e}_{t-j} \right)^{1/2} \right)^2 \rightarrow 0, \quad \text{as} \quad L \rightarrow \infty,$$

by (A.9) and $\sum_{i=1}^{n} \left| \tilde{b}_i \right| < \infty, \ a.s. = \mathbb{P}$.

(b) By the Markov inequality it suffices to show

$$E^* \left( n^{-i+1} \gamma_{n}^{-1} \sum_{t=1}^{n} t_i Z_i^* \right)^i \left( n^{-i+1} \gamma_{n}^{-1} \sum_{t=1}^{n} t_i Z_i^* \right) = o_p(1).$$

Indeed,

$$E^* \left( n^{-i+1} \gamma_{n}^{-1} \sum_{t=1}^{n} t_i Z_i^* \right)^i \left( n^{-i+1} \gamma_{n}^{-1} \sum_{t=1}^{n} t_i Z_i^* \right)$$

$$= n^{-2(i+1)} \gamma_{n}^{-2} \sum_{t=1}^{n} t_i^2 E^* Z_i^* Z_i^* \leq n^{-2} \sum_{t=1}^{n} \gamma_{n}^{-2} E^* Z_i^* Z_i^*.$$

Since $Z_i^* = \sum_{j=0}^{t-1} \tilde{b}_j \tilde{e}_{t-j}$, we have $\gamma_{n}^{-2} E^* Z_i^* Z_i^* = \gamma_{n}^{-2} \sum_{j=0}^{t-1} \tilde{b}_j \tilde{e}_{t-j} \tilde{e}_{t-j} = o_p(1)$ for $t \leq n$. So (b) holds.

(c) First we establish for any fixed $i \geq 1$,

$$n^{-1/2} \gamma_{n}^{-2} \sum_{t=1}^{n} \tilde{e}_{t-i} \tilde{e}_{t-i} \xrightarrow{d} \mathcal{N}_p \left( 0, \int \sigma^2 \right), \quad \text{in probability.} \quad (A.10)$$
Define $\mathcal{F}^*_t$ to be the sigma field generated by $\{v_s: s \leq t\}$, then $y_{n-2}^\gamma e_{t-1}^\gamma e_t^\gamma$ is an m.d. array with respect to $\mathcal{F}^*_t$. Note that

$$
\left\| n^{-1} \sum_{t=1}^n y_n^{-4} (e_{t-i}^\gamma e_t^\gamma - \sigma_{t-i}^\gamma \sigma_t^\gamma) \right\|_{1+a} \leq n^{-1-a} \left( \sum_{t=1}^n y_n^{-4} \left( \left\| e_{t-i}^\gamma e_t^\gamma \right\|_{1+a} + \left\| \sigma_{t-i}^\gamma \sigma_t^\gamma \right\|_{1+a} \right) \right)^{1+a} \to 0,
$$

since $y_n^{-4} \left\| e_{t-i}^\gamma e_t^\gamma \right\|_{1+a} = y_n^{-4} (E e_{t-i}^\gamma e_t^\gamma + 2a \sigma_{t-i}^\gamma \sigma_t^\gamma) \frac{1}{\gamma} \leq (y_n^{-8} E \sigma_{t-i}^\gamma \sigma_t^\gamma + 4a E \sigma_{t-i}^\gamma \sigma_t^\gamma + 4a E \eta_{t-i}^\gamma \eta_t^\gamma + 4a E \eta_{t-i}^\gamma \eta_t^\gamma) \frac{1}{\gamma^2} < \infty$ by Assumption 1 (i) and similarly $y_n^{-4} \left\| \sigma_{t-i}^\gamma \sigma_t^\gamma \right\|_{1+a} < \infty$. Thus $n^{-1} \sum_{t=1}^n y_n^{-4} \sum_{i=1}^n e_{t-i}^\gamma e_t^\gamma = n^{-1} \sum_{t=1}^n e_{t-i}^\gamma e_t^\gamma + o_p(1) = n^{-1} \sum_{t=1}^n y_n^{-4} \sigma_{t-i}^\gamma \sigma_t^\gamma + o_p(1) \Rightarrow \sigma^4$. Then (A.10) holds by the CLT for m.d. arrays (Davidson 1994) provided that

$$
n^{-1} y_n^{-4} \sum_{t=1}^n e_{t-i}^\gamma e_t^\gamma - n^{-1} \sum_{t=1}^n y_n^{-4} \sigma_{t-i}^\gamma \sigma_t^\gamma = o_p(1) \quad \text{(A.11)}
$$

and

$$
n^{-2} y_n^{-8} \sum_{t=1}^n e_{t-i}^\gamma e_t^\gamma = o_p(1). \quad \text{(A.12)}
$$

(A.11) holds since

$$
E^* \left[ n^{-1} y_n^{-4} \sum_{t=1}^n (e_{t-i}^\gamma e_t^\gamma - \sigma_{t-i}^\gamma \sigma_t^\gamma)^2 \right] = n^{-2} y_n^{-8} E^* \left[ \sum_{t=1}^n \sigma_{t-i}^\gamma \sigma_t^\gamma (v_{t-i}^\gamma v_t^\gamma - 1) \right] \leq C n^{-2} \sum_{t=1}^n y_n^{-8} \sigma_{t-i}^\gamma \sigma_t^\gamma = o_p(1),
$$

by (A.9), and (A.12) holds since

$$
n^{-2} y_n^{-8} \sum_{t=1}^n e_{t-i}^\gamma e_t^\gamma = n^{-2} y_n^{-8} \sum_{t=1}^n \sigma_{t-i}^\gamma \sigma_t^\gamma (v_{t-i}^\gamma v_t^\gamma - 1) \leq C n^{-2} \sum_{t=1}^n y_n^{-8} \sigma_{t-i}^\gamma \sigma_t^\gamma = o_p(1). \quad \text{(A.11)}
$$

Note that $n^{-1/2} y_n^{-2} \sum_{t=1}^n Z_t^\gamma e_t^\gamma = \sum_{t=1}^\infty \hat{b}_t (n^{-1/2} \sum_{t=1}^n y_n^{-2} e_{t-i}^\gamma e_t^\gamma)$. Thus (c) can be proved by using Lemma A1 provided that

$$
\mathbb{P}^* \left( \left| \sum_{t=L+1}^\infty \hat{b}_t \left( n^{-1/2} \sum_{t=1}^n y_n^{-2} e_{t-i}^\gamma e_t^\gamma \right) \right| < \epsilon \right) \leq e^{-2 \epsilon^2} \mathbb{E}^* \left( \left| \sum_{t=L+1}^\infty \hat{b}_t \left( n^{-1/2} \sum_{t=1}^n y_n^{-2} e_{t-i}^\gamma e_t^\gamma \right) \right| \right)^2 \leq e^{-2} \sum_{t=L+1}^\infty |\hat{b}_t|^2 \left( n^{-2} \sum_{t=1}^n (y_n^{-8} e_{t-i}^\gamma e_t^\gamma)^{1/2} \right)^2 = o_p(1), \quad \text{as } L \to \infty,
$$

by (A.9) and $\sum_{t=L+1}^\infty |\hat{b}_t| < \infty, \ a.s. - \mathbb{P}$.

(d). It suffices to show that for $\lambda = (\lambda_1, \ldots, \lambda_{m+1})' > 0$ with $\sum_{i=1}^{m+1} \lambda_i = 1$,

$$
\lambda' y_n^{-1} T_n^{\gamma-1} \sum_{t=1}^n T_t e_t^\gamma \Rightarrow \text{MAN} \left( 0, \int \lambda' (\sigma, r \sigma, \ldots, r^m \sigma)' (\sigma, r \sigma, \ldots, r^m \sigma) \lambda' \right), \text{ in probability}. \quad \text{(A.13)}
$$

© Royal Economic Society 2008
Note that $\lambda_n^{-1} n^{-1/2} \sum_{i=1}^{m+1} \lambda_i (t/n)^i \hat{e}_i = \gamma_n^{-1} \pi_i \hat{e}_i$, where $\pi_i = \sum_{i=1}^{m+1} \lambda_i (t/n)^i \leq 1$, is an m.d. array with respect to $\mathcal{F}_s = \sigma v_s : s \leq t$, with $\mathbb{E}^* \left[ \gamma_n^{-1} \pi_i \hat{e}_i \right]^2 = \gamma_n^{-2} \pi_i^2 \hat{e}_i$. Since

$$
\mathbb{E}^* \left[ n^{-1} \sum_{i=1}^{n} \gamma_n^{-2} \pi_i^2 \hat{e}_i^2 - n^{-1} \sum_{i=1}^{n} \gamma_n^{-2} \pi_i^2 \hat{e}_i^2 \right]^2 
= n^{-2} \mathbb{E}^* \left[ \sum_{i=1}^{n} \gamma_n^{-2} \pi_i^2 \hat{e}_i^2 (v_i^2 - 1) \right]^2 
\leq n^{-2} \mathbb{E}^* \left[ \sum_{i=1}^{n} \gamma_n^{-2} \pi_i^2 (v_i^2 - 1) \right]^2 
= n^{-2} \sum_{i=1}^{n} \gamma_n^{-4} \pi_i^4 \mathbb{E}^* (v_i^2 - 1)^2 
\leq C n^{-2} \sum_{i=1}^{n} \gamma_n^{-4} \pi_i^4 = o_P(1),
$$

by (A.9), then by Markov inequality and $\pi_i \to \sum_{i=1}^{m+1} \lambda_i r^i$,

$$
n^{-1} \sum_{i=1}^{n} \gamma_n^{-2} \pi_i^2 \hat{e}_i^2 = n^{-1} \sum_{i=1}^{n} \gamma_n^{-2} \pi_i^2 \hat{e}_i^2 + o_P(1) \Rightarrow \int \lambda'(\sigma, r \sigma, \ldots, r^m \sigma) (\sigma, r \sigma, \ldots, r^m \sigma) \lambda'.
$$

Further we have

$$
n^{-2} \sum_{i=1}^{n} \gamma_n^{-4} \pi_i^4 \mathbb{E}^* \hat{e}_i^4 = n^{-2} \sum_{i=1}^{n} \gamma_n^{-4} \pi_i^4 \mathbb{E}^* v_i^4 \leq C n^{-2} \sum_{i=1}^{n} \gamma_n^{-4} \pi_i^4 = o_P(1).
$$

Then (A.13) holds by the CLT for m.d. arrays.

(e). It follows from $\mathbb{E}^* Z_t^* \hat{e}_t \hat{e}_s = \mathbb{E}^* Z_t^* \mathbb{E}^* \hat{e}_t \hat{e}_s = 0$ if $s \geq t$, $\mathbb{E}^* \hat{e}_t \hat{e}_s = \mathbb{E}^* \hat{e}_t \mathbb{E}^* \hat{e}_s = 0$ if $s < t$. □

**Proof of Corollary 3.2.** The proof uses arguments similar to the proof of Corollary 3.1. □

© Royal Economic Society 2008