Harvard University
Harvard University Biostatistics Working Paper Series

Year 2004 *Paper* 16

A Robust Regression Model for a First-Order Autoregressive Time Series with Unequal Spacing: Technical Report

E. Andres Houseman*

*Harvard School of Public Health, ahousema@hsph.harvard.edu

A Robust Regression Model for a First-Order Autoregressive Time Series with Unequal Spacing: Technical Report

E. Andrés Houseman
Department of Biostatistics, Harvard School of Public Health,
655 Huntington Ave., Boston, Massachusetts 02115, U.S.A
ahousema@hsph.harvard.edu

October 14, 2004

Abstract

Time series arise often in environmental monitoring settings, which typically involve measuring processes repeatedly over time. In many such applications, observations are irregularly spaced and, additionally, are not distributed normally. We describe the technical details justfying a simple, robust approach for estimating regression parameters and a first-order autocorrelation parameter in a time series where the observations are irregularly spaced. Estimates are obtained from an estimating equation constructed as a linear combination of estimated innovation errors, suitably robustified by symmetric and possibly bounded functions. Under MCAR assumption and mild regularity conditions, the proposed estimating equation yields consistent and asymptotically normal estimates. Technical details are developed using the theory of mixingales described by Davidson (1994) and standard Z-estimation theory as described by van der Vaart (1998). In particular, the sequence of estimating function components can be shown to be an ergodic L_1 mixingale to which a weak law of large numbers and a central limit theorem apply. Lipschitz conditions ensure the functional convergence required to complete the proof.

1 Introduction

Time series arise often in environmental monitoring settings, which typically involve measuring processes repeatedly over time and/or space. Although the time series regression literature is vast, beginning with Durbin (1960), much of the development has occurred in the econometrics literature, where regular time series data are the norm. In the environmetric setting, the expense and difficulty of collecting data often leads to irregularity in the observed time series. In addition, outliers or heavy tailed distributions can occur. See, for example, Houseman et al. (2004). Time series in the environmental setting have been addressed by authors such as Brumback et al. (2000), who proposed a *transitional regression model* for Poisson outcomes. However, the paper did not address unequal spacing in the time series, and it addressed outliers primarily through overdispersion of the Poisson outcome. Robust methods for time series have been proposed by numerous authors, such as Denby & Martin (1979) and McDougall (1994). However, they have not addressed unequally spaced time series in a regression setting. On the other hand, Omori (2003) recently extended the methods of Zeger (1988) for unequally spaced observations; both of these papers are focused, however, on count data and the estimating equations are motivated by a Poisson assumption. Unlike Denby & Martin (1979) and McDougall (1994), they do not specifically address bounded-influence estimation.

This article provides theoretical details of a simple, robust approach for estimating regression parameters and a positive first-order autocorrelation parameter for a time series where the spacings between observations are unequal. In our approach, which is similar in spirit to Denby & Martin (1979), we construct a method-of-moments estimator (Mátyás, 1999), or Z-estimator (van der Vaart, 1998). The major technical references are Davidson (1994) and van der Vaart (1998). These will be referred to throughout the text simply as Davidson and Van der Vaart, respectively.

2 Robust Estimation for a Univariate First-Order Autoregressive Time Series

We now describe a simple robust estimating-equations approach to estimating regression coefficients and autocorrelation parameter in a first-order autoregressive time series. We first state several assumptions about our model, present some definitions, and then proceed to state two theorems. The theorems are proved in the Appendix.

Our first assumption describes a univariate first-order autoregressive time series regression model.

Assumption 1. Let $\mathcal{Y} = \{..., Y_{-1}, Y_0, Y_1, Y_2, ...\}$ be a sequence of univariate random variables from symmetric distributions having constant variance, and let $\mathcal{X} = \{..., X_{-1}, X_0, X_1, X_2, ...\}$ be a corresponding sequence of p-dimensional covariate vectors. Assume that $\mathrm{E}(Y_j|X_j) = X_j'\beta$ for each $j \in \mathcal{J} = \{..., -1, 0, 1, 2, ...\}$, and the process $Y_j - X_j'\beta$ is stationary. The outcomes follow a first-order autoregressive process, wherein $\mathrm{E}(Y_j|X_j, Y_{j-1}, X_{j-1}) = X_j'\beta + \rho(Y_{j-1} - X_{j-1}'\beta)$ for each $j \in \mathcal{J}$; in addition,

$$Y_j = X_j'\beta + \rho(Y_{j-1} - X_{j-1}'\beta) + U_j, \tag{1}$$

where $0 < \rho < 1$, and U_j is symmetric, independent of $Y_{j-1} - X_{j-1}'\beta$, has expectation zero, and has variance τ^2.

We also assume an open but bounded parameter space, which is plausible in most practical settings.

Assumption 2. Assume that $\sigma > 0$ is a known scale parameter and that the parameter $\theta = (\beta', \rho)'$ ranges over Θ, a convex open subset of a compact set of $\mathbb{R}^{p+1}$. In particular, there is an $\epsilon > 0$ such that $\epsilon < \rho < 1-\epsilon$ and $\sigma > \epsilon$; these conditions ensure that the AR-1 process described in Assumption 1 is well-behaved on the boundary of Θ. We denote $\theta_0 = (\beta_0', \rho_0')$ and σ_0 as the true values of the parameters.

We also assume regularity conditions for the covariate process $\{X_j\}$ and the mechanism which gives rise to the "gaps" in the time series. In essence, we require the missing data mechanism $\{\Delta_j\}$ to be *missing completely at random* (MCAR) in the sense of Little & Rubin (1987). We also assume sufficiently fast mixing that a weak law of large numbers and a central limit theorem apply. Theoretical details would be somewhat simplified if we assume stationarity. However, we would like to address the more general case in which $X_j'\beta$ contains terms for seasonal adjustment. It turns out that a sufficient condition for our methodology is that the expected values of the individual estimating function components converge to a constant for every $\theta \in \Theta$. This holds true in the more general situation where we assume periodic stationarity in the sense described as follows.

Assumption 3. Let $V_j = (X_j', \Delta_j, K_j)'$ and let $\{V_j\}$ be an ergodic process (in the sense of Davidson, Section 13.4, pages 199-203), independent of the innovation process $\{U_j\}$, such that Δ_j is Bernoulli, $K_j = (K_{j-1} + 1)(1 - \Delta_{j-1}) + \Delta_{j-1}$ is bounded, and $\{X_j\}$ is bounded and L_2 integrable. Moreover, assume that $\{V_j\}$ is periodically stationary in that there is an integer Q such that $\{V_{nQ+q}\}$ is stationary and ergodic for all $q \in \{0, 1, ..., Q-1\}$ and all integers n. Finally, assume that $\{V_{nQ+q}\}$ is α-mixing or ϕ-mixing of size $a > 1$ (in the sense of Davidson, Section 14.1, pages 209-211). For convenience, let $\{\mathcal{V}_h\}$ denote the corresponding adapted filtration.

Here, Δ_j is interpreted as an indicator of whether the outcome was observed or not and K_j is the gap between observed observations. Since we are interested in the behavior of $\mathcal{Y}$ conditional on $\{V_j\}$, the dependence of $\{V_j\}$ need be only sufficiently structured to support the technical details of the proofs. In particular, ergodicity is not required for consistency, only for asymptotic normality. This will make sense because the estimating equations defined below will essentially condition out the information in $\{\mathcal{V}_h\}$. However, the distribution of the estimator depends upon the specific realization $\{V_j\}$.

Next, we impose conditions on our *robustifying functions* and the distribution of the random portion of the time series.

Assumption 4. The functions $\lambda(\cdot)$ and $\omega(\cdot)$ are odd and have bounded and continuous first derivatives. In addition, $\lambda(\cdot)$ and $\omega(\cdot)$ have continuous second derivatives or, more generally, the derivatives $\dot{\lambda}(\cdot)$ and $\dot{\omega}(\cdot)$ satisfy Lipschitz conditions on Θ. In either case, there is a constant A such that $|\dot{\lambda}(\theta) - \dot{\lambda}(\tilde{\theta})| < A|\theta - \tilde{\theta}|$ for all $\theta, \tilde{\theta} \in \Theta$, and similarly for $\dot{\omega}(\cdot)$.

Assumption 4 is probably stronger than necessary, as we remark following the statement of Theorem 2. The next assumption is necessary for the application of a central limit theorem.

Assumption 5. Either $\lambda(\cdot)$ and $\omega(\cdot)$ are bounded or U_j possesses a fourth moment.

Finally, we define the components of our proposed estimating function.

Definition 1. Let $\mathcal{H}_j$ be the σ-field generated by $\{(U_h, V_h)\}_{h<j}$ and let $\{\mathcal{H}_j\}$ be the corresponding filtration, referred to as the *history*.

Definition 2. Let

$$\xi(\rho, k) = \left(\sum_{l=0}^{k-1} \rho^{2l}\right)^{1/2} = \left(\frac{1-\rho^{2k}}{1-\rho^2}\right)^{1/2}.$$

For convenience, denote $\xi(\rho, \infty) = (1-\rho^2)^{-1/2}$. In addition, reparametrize τ^2 as $\alpha^2\sigma^2$ for some known α^2; thus, σ^2 is a scale parameter that has a linear relationship with the variance of U_j.

Note that $\mathrm{E}[Y_j - X_j'\beta_0 - \rho_0^k(Y_{j-k} - X_{j-k}'\beta_0)|X_j, Y_{j-k}, X_{j-k}] = 0$ and $\mathrm{Var}[Y_j - X_j'\beta_0 - \rho_0^k(Y_{j-k} - X_{j-k}'\beta_0)|X_j, Y_{j-k}, X_{j-k}] = \xi(\rho_0, k)^2\tau^2$, and that the marginal variance of Y_j is $\xi(\rho_0, \infty)^2\tau^2$. These facts are useful in constructing zero-mean residuals that have variance suitably uniform for bounded influence estimation.

Definition 3. Let $N_n = \sum_{j=1}^n \Delta_j$,

$$\psi_j(\theta) = \lambda\left(\frac{Y_j - X_j\beta - \rho^{K_j}(Y_{j-K_j} - X_{j-K_j}\beta)}{\xi(\rho, k_i)\sigma}\right)\left[\begin{array}{c} X_j - \rho^{K_j}X_{j-K_j} \\ K_j\rho^{K_j-1}\omega[(Y_{j-K_j} - X_{j-K_j}\beta)/\xi(\rho,\infty)\sigma] \end{array}\right],$$

and $\Psi_n(\theta) = N_n^{-1}\sum_{j=1}^n \Delta_j\psi_j(\theta)$.

By the comment after Definition 2, it is easy to see that the arguments to the robustifying functions $\lambda(\cdot)$ and $\omega(\cdot)$ have stable variance and their tails are consequently downweighted uniformly. The specific form of the functions $\psi_j(\theta)$ is motivated by differentiating the arguments to $\lambda(\cdot)$ in the manner suggested by Wedderburn (1974), and robustifying potentially heavy-tailed coefficients as suggested by Denby & Martin (1979).

The main theorems of this paper concern the asymptotic behavior of the estimator $\widehat{\theta}_n$ obtained by solving $\Psi_n(\theta) = 0$. In summary, $\widehat{\theta}_n$ enjoys the usual attractive asymptotic properties, as described below.

Theorem 1 (Consistency). If $\widehat{\theta}_n$ solves $\Psi_n(\theta) = 0$ then $\widehat{\theta}_n \xrightarrow{\mathcal{P}} \theta_0$.

Theorem 2 (Asymptotic Normality). If for all θ in a neighborhood of θ_0, $\Psi_n(\theta)$ converges in probability to a continuous function with nonsingular derivative, and if $\widehat{\theta}_n$ solves $\Psi_n(\theta) = 0$, then there is a matrix Ω such that $N_n^{1/2}\Omega^{-1/2}(\widehat{\theta}_n - \theta_0) \xrightarrow{\mathcal{L}} N(0, I)$.

The proofs of Theorems 1 and 2 appear in the Appendix. The proof of Theorem 2 of suggests an estimator of Ω, which we propose formally as another theorem.

Theorem 3 (Variance Estimation). Let

$$\widehat{\Omega}_n = \dot{\Psi}_n(\widehat{\theta}_n)^{-1} \left(N_n^{-1/2} \sum_{j=1}^{n} \Delta_j \psi_j(\widehat{\theta}_n)\psi_j(\widehat{\theta}_n)' \right) \dot{\Psi}_n(\widehat{\theta}_n)^{-T},$$

where $\dot{\Psi}_n(\theta)$ is the first derivative matrix of $\Psi_n(\theta)$. Provided the conditions of Theorem 2 hold, $\widehat{\Omega}_n$ is a consistent estimator of Ω.

Theorems 2 and 3 require that the estimating function converges to a deterministic function that has a continuous first derivative near θ_0, a condition that is difficult to verify in general since it depends on the behavior of $\mathcal{X}$. However, in practical settings the existence of such a limit can usually be assumed.

Results can be extended to include the variance estimating function $\Psi_n^{(s)}$,

$$\psi_j^{(s)}(\theta) = \left(\frac{Y_j - X_j'\beta - \rho^{K_j}(Y_{j-K_j} - X_{j-K_j}\beta)}{\xi(\rho, K_j)} \right)^2 - \alpha^2\sigma^2,$$

provided U_j is bounded in $L_{4+\delta}$ for some $\delta > 0$. This slightly stronger condition is needed to ensure that $\psi_i^{(s)}(\theta)^2$ obeys a central limit theorem.

As suggested by the results in Denby & Martin (1979) for regular time series with mean zero, it should, in principle, be possible to extend Theorem 2 to the case where $\lambda(\cdot)$ and $\omega(\cdot)$ are simply continuous, provided some additional regularity conditions are assumed. One approach might be to approximate $\Psi_n(\theta)$ by its expectation, or by twice-differentiable functions that closely approximate $\Psi_n(\theta)$. Since our interest is primarily from an applied perspective, we have not fully investigated these approaches.

References

Brumback, B. A., Ryan, L. M., Schwartz, J. D., Neas, L. M., Stark, P. C. & Burge, H. A. (2000). Transitional regression models, with application to environmental time series. *Journal of the American Statistical Association* **95**, 16–27.

Davidson, J. (1994). *Stochastic Limit Theory.* New York: Oxford University Press.

Denby, L. & Martin, R. D. (1979). Robust estimation of the first-order autoregressive parameter. *Journal of the American Statistical Association* **74**, 140–146.

Durbin, J. (1960). Estimation of parameters in time-series regression models. *Journal of the Royal Statistical Society Series B-Statistical Methodology* **22**, 139–153.

Houseman, E. A., Ryan, L. M. & Coull, B. A. (2004). Cholesky residuals for assessing normal errors in a linear model with correlated outcomes. *Journal of American Statistical Association* **99**, 383–394.

Little, R. J. A. & Rubin, D. B. (1987). *Statistical Analysis with Missing Data.* New York: John Wiley & Sons.

Mátyás, L. (1999). *Generalized Method of Moments Estimation.* New York: Cambridge University Press.

McDougall, A. J. (1994). Robust methods for recursive autoregressive moving average estimation. *Journal of the Royal Statistical Society, Series B* **56**, 189–207.

Omori, Y. (2003). Estimation for unequally spaced time series of counts with serially correlated random effects. *Statistics and Probability Letters* **63**, 1–12.

van der Vaart, A. W. (1998). *Asymptotic Statistics.* New York: Cambridge University Press.

Wedderburn, R. W. M. (1974). Quasi-likelihood functions, generalized linear models, and the gauss-newton method. *Biometrika* **61**, 439–447.

Zeger, S. L. (1988). A regression model for time series of counts. *Biometrika* **75**, 621–629.

Appendix

In this appendix we shall prove Theorems 1, 2, and 3. Proofs follow from the mixingale law or large numbers (Davidson, Theorem 19.11, page 302) and a central limit theorem for ergodic mixingale (Davidson, Theorem 24.5, page 385). Of course, in order to make use of the theorems it is first necessary to prove that the estimating functions of Definition 3 form ergodic mixingales. We accomplish this indirectly by proving first that they form a sequence that is *near-epoch dependent* on the sequence $\{\mathcal{V}_j\}$. We review the relevant definitions as formulated by Davidson (Chapters 16 and 17). First, we repeat Davidson's definition of the concept of an L_p *mixingale* for $p \geq 1$.

Definition 4. Let $(\Omega, \mathcal{F}, P)$ be a probability space, let $\{\mathcal{F}_t\}$ be an increasing sequence of σ-subfields of $\mathcal{F}$, and let X_t be integrable random variables on $\mathcal{F}_t$. The sequence of pairs $\{X_t, \mathcal{F}_t\}_{-\infty}^{\infty}$, is called an L_p mixingale with respect to a sequence of non-negative constants $\{c_t\}_{-\infty}^{\infty}$ if there exists a sequence of non-negative constants $\{\zeta_t\}_0^{\infty}$ such that

$\zeta_m \to 0$ as $m \to \infty$ and

$$\|\mathrm{E}(X_t|\mathcal{F}_{t-m})\|_p \leq c_t \zeta_m \qquad \text{(Davidson, Condition 16.1)}$$

$$\|\mathrm{E}(X_t|\mathcal{F}_{t+m})\|_p \leq c_t \zeta_m \qquad \text{(Davidson, Condition 16.2)}$$

hold for all t and all $m \geq 0$. If $\zeta_m = O(m^{-s})$ for $s > r$, then the sequence is said to be of *size* $-r$.

Associated with the mixingale concept is the idea of *near-epoch dependence.*

Definition 5. Let $p > 0$. For a stochastic sequence $\{A_j\}_{-\infty}^{+\infty}$, possibly vector valued, on a probability space $(\Omega, \mathcal{F}, P)$, let $\mathcal{F}_{j-m}^{j+m} = \sigma(A_{j-m}, ..., A_{j+m})$ such that $\{\mathcal{F}_{j-m}^{j+m}\}_{m=0}^{\infty}$ is an increasing sequence of σ-fields. If a sequence of integrable random variables $\{Q_j\}_{-\infty}^{+\infty}$ satisfies

$$\|Q_j - \mathrm{E}(Q_j|\mathcal{F}_{j-m}^{j+m})\|_p \leq d_j \nu_m \qquad (2)$$

where $\nu_m \to 0$ and $\{d_j\}_{-\infty}^{+\infty}$ is a sequence of positive constants, Q_j will be said to be near-epoch dependent in L_p-norm (L_p-NED) on $\{A_j\}_{-\infty}^{+\infty}$. If $\nu_m = O(m^{-s})$ for $s > r$, then the dependence is said to be of *size* $-r$. It will be convenient to note when there is an M such that $\nu_m = 0$ for all $m > M$, in which case we will say that $\{Q_j\}$ is *M-dependent with respect to* $\{A_j\}$.

We try the reader's patience with another definition, which will pay off in notational convenience:

Definition 6. Define the scaled residuals R_j and S_j as

$$R_j = \frac{Y_j - X_j'\beta - \rho^{K_j}(Y_{j-K_j} - X_{j-K_j}'\beta)}{\xi(\rho, K_j)\sigma},$$

$$S_j = \frac{Y_j - X_j'\beta}{\xi(\rho,\infty)\sigma}.$$

We are now in a position to state and prove several lemmas. Lemmas 1 and 2 are technical lemmas needed to prove Lemma 3, which establishes the mixingale property required for a law of large numbers and a central limit theorem to apply.

Lemma 1. Fix $\theta \in \Theta$. Let $\{W_j\}$ be a bounded process, independent of the innovations $\{U_j\}$, and L_1-NED on $\{V_j\}$. Moreover, assume that $\{W_j\}$ is M-dependent with respect to $\{V_j\}$. Then $\{W_j\lambda(R_j) - \mathrm{E}[W_j\lambda(R_j)], \mathcal{H}_j\}$ is an L_1 mixingale of size -1 with respect to a bounded sequence of constants $\{c_j\}$.

Proof. Fix j, k, and $h < j$. Simple algebraic manipulation shows that

$$\begin{aligned} Z_{j+k}(\beta) - \rho^k Z_j(\beta) &= \sum_{l=0}^{j+k-h} \rho_0^l U_{j+k-l} - \sum_{l=0}^{j-h} \rho^k \rho_0^l U_{j-l} + (\rho^k X_j - X_{j+k})'(\beta - \beta_0) + \\ &\quad \rho_0^{j+k-h-1} Z_{h-1}(\beta_0) - \rho^k \rho_0^{j-h-1} Z_{h-1}(\beta_0), \end{aligned}$$

where $Z_l(\beta) = Y_l - X_l'\beta$. Let $C_k = (\rho_0^k - \rho^k)\xi(\rho,k)^{-1}\sigma^{-1}$,

$$\tilde{U}_{hjk} = \xi(\rho,k)^{-1}\sigma^{-1}\left[\sum_{l=0}^{j+k-h} \rho_0^l U_{j+k-l} - \sum_{l=0}^{j-h} \rho^k \rho_0^l U_{j-l}\right],$$

and

$$\tilde{X}_{jk} = \xi(\rho,k)^{-1}\sigma^{-1}(\rho^k X_j - X_{j+k})'(\beta - \beta_0).$$

Thus $R_j = \tilde{U}_{hjK_j} + \tilde{X}_{jK_j} + \rho_0^{j-h-1} C_{K_j} Z_{h-1}(\beta_0)$. By Assumption 4 and the mean value theorem,

$$W_j\lambda(R_j) = W_j\lambda(\tilde{U}_{hjK_j} + \tilde{X}_{jK_j}) + \rho_0^{j-h-1} D_{hj} W_j C_{K_j} Z_{h-1}(\beta_0), \tag{3}$$

where D_{hj} is a bounded random variable. Define $\mathcal{F}_{j-m}^{j+m}$ as in Definition 5, with $A_j = (U_j, V_j)$. Since $\tilde{U}_{hjK_j}$, W_j and $\tilde{X}_{jK_j}$ are M-dependent with respect to $\{A_j\}$, it follows that there is an m' such that

$$W_j\lambda(\tilde{U}_{hjK_j}+\tilde{X}_{jK_j}) - \mathrm{E}\left[W_j\lambda(\tilde{U}_{hjK_j}+\tilde{X}_{jK_j})\,\middle|\,\mathcal{F}_{j+(j-h)}^{j-(j-h)}\right] = 0 \tag{4}$$

for $h < j - m'$. For all such h,

$$\begin{aligned}\mathrm{E}\left(\rho_0^{j-h-1}D_{hj}W_jC_{K_j}Z_{h-1}(\beta_0)\,\middle|\,\mathcal{F}_{j+(j-h)}^{j-(j-h)}\right) &= \mathrm{E}[Z_{h-1}(\beta_0)]\mathrm{E}\left(\rho_0^{j-h-1}D_{hj}W_jC_{K_j}\,\middle|\,\mathcal{F}_{j+(j-h)}^{j-(j-h)}\right)\\ &= 0 \qquad (5)\end{aligned}$$

and

$$\|\rho_0^{j-h-1}D_{hj}W_jC_{K_j}Z_{h-1}(\beta_0)\|_1 \le \rho_0^{j-h-1}\|D_{hj}W_jC_{K_j}\|_1\|Z_{h-1}(\beta_0)\|_1. \tag{6}$$

By (3) - (6),

$$\left\|W_j\lambda(R_j) - \mathrm{E}\left(W_j\lambda(R_j)\,\middle|\,\mathcal{F}_{j+(j-h)}^{j-(j-h)}\right)\right\|_1 \le \rho_0^{j-h-1}\|D_{hj}W_jC_{K_j}\|_1\|Z_{h-1}(\beta_0)\|_1,$$

which satisfies (2). Thus, $\{W_j\lambda(R_j) - \mathrm{E}\left[W_j\lambda(R_j)\right]\}$ is L_1-NED on $\{A_j\}$. By Theorem 17.5 of Davidson (page 264), $\{W_j\lambda(R_j) - \mathrm{E}[W_j\lambda(R_j)], \mathcal{H}_j\}$ is an L_1 mixingale of size -1 with respect to a sequence of constants $\{c_j\}$, where

$$c_j \le \max\{\|W_j\lambda(R_j)\|_1, \|D_{hj}W_jC_{K_j}\|_1\|Z_{h-1}(\beta_0)\|_1\}.$$

Smoothness conditions (Assumption 4) and moment conditions (Assumption 5) ensure that $\{c_j\}$ is bounded. □

Lemma 2. Fix $\theta \in \Theta$. Let $\{W_j\}$ be a bounded process, independent of the innovations $\{U_j\}$, and L_1-NED on $\{V_j\}$. Moreover, assume that $\{W_j\}$ is M-dependent with respect to $\{V_j\}$. Then $\left(\{W_j\omega(S_{j-K_j})\lambda(R_j) - \mathrm{E}[W_j\omega(S_{j-K_j})\lambda(R_j)]\}, \mathcal{H}_j\right)$ is an L_1 mixingale of size -1 with respect to a bounded sequence of constants $\{c_j\}$.

Proof. As in the proof of Lemma 1, assume $h < j$ and let

$$S_j = \tilde{U}^*_{hj} + \tilde{X}^*_j + \rho_0^{j-h-1} C^*_{K_j} Z_{h-1}(\beta_0),$$

where $Z_l(\beta) = Y_l - X_l'\beta$, $C^*_k = \rho^{k-1}\xi(\rho,\infty)^{-1}\sigma^{-1}$, $\tilde{U}^*_{hj} = \xi(\rho,\infty)^{-1}\sigma^{-1}\left[\sum_{l=0}^{j-h}\rho_0^l U_{j-l}\right]$, and $\tilde{X}^*_j = -\xi(\rho,\infty)^{-1}\sigma^{-1}X_j'(\beta-\beta_0)$. Also as in the proof of Lemma 1, $\omega(S_j) = \omega(\tilde{U}^*_{hj} + \tilde{X}^*_j) + D^*_{hj}\rho_0^{j-h}C^*_{K_j}Z_{h-1}(\beta_0)$, where D^*_{hj} is a bounded random variable. Using techniques similar to the proof of Lemma 1, it is straightforward to show that (2) is satisfied by the individual components $W_j\lambda(\tilde{U}_{hjK_j}+\tilde{X}_{jK_j})\omega(\tilde{U}^*_{hj}+\tilde{X}^*_j)$, $W_j\rho_0^{j-h-1}D_{hj}C_{K_j}Z_{h-1}(\beta_0)\omega(\tilde{U}^*_{hj}+\tilde{X}^*_j)$, $W_j\lambda(\tilde{U}_{hjK_j}+\tilde{X}_{jK_j})D^*_{hj}\rho_0^{j-h-1}C^*_{K_j}Z_{h-1}(\beta_0)$, and $W_j\rho_0^{2(j-h-1)}D_{hj}D^*_{hj}C_{K_j}C^*_{K_j}Z_{h-1}(\beta_0)^2$. It follows that $\{W_j\omega(S_{j-K_j})\lambda(R_j) - \mathrm{E}[W_j\omega(S_{j-K_j})\lambda(R_j)]\}$ is L_1-NED on $\{A_j\}$, and the proof is completed by an application of Theorem 17.5 of Davidson (page 264). □

Lemma 3. The process $\{\Delta_j\psi_j(\theta) - \mathrm{E}[\Delta_j\psi_j(\theta)]\}_j$ together with $\mathcal{H}_j$ forms an L_1 mixingale of size -1 with respect to a bounded sequence of constants $\{c_j\}$.

Proof. The statement follows easily from Lemmas 1 and 2. □

Lemmas 5 and 6 assert pointwise and global (respectively) convergence of the estimating functions. However, first we must prove a technical lemma, Lemma 4, in order to relax stationarity assumptions to the weaker requirement of periodic stationarity.

Lemma 4. Fix $\theta \in \Theta$.Then

$$N_n^{-1}\sum_{j=1}^n \mathrm{E}[\Delta_j\lambda(R_j)(X_j - \rho X_{j-K_j})]$$

and and

$$N_n^{-1}\sum_{j=1}^n \mathrm{E}[\Delta_j K_j \rho^{K_j}\omega(S_j)\lambda(R_j)]$$

converge to constants.

Proof. Let $W_j = \Delta_j \lambda(R_j)(X_j - \rho X_{j-K_j})$. The distribution of W_j is periodic in the sense of Assumption 3; consequently, let $\mu_q = \mathrm{E}(W_{Ql+q})$. It follows that for each $q = 0, ..., Q-1$, $n^{-1}\sum_{l=1}^{n} \mathrm{E}(W_{Ql+q}) = \mu_q$. Let $[a]$ denote the integer portion of a and $a\%Q$ the modulus of a over Q. Then as $n \to \infty$,

$$\begin{aligned} n^{-1}\sum_{j=1}^{n} \mathrm{E}(Q_j) &= n^{-1}\sum_{l=1}^{[n/Q]}\sum_{q=0}^{Q-1} \mathrm{E}(W_{Ql+q}) + n^{-1}\sum_{q=0}^{n\%Q} \mathrm{E}(W_{Ql+q}) \\ &= \frac{[n/Q]}{Q[n/Q] + n\%Q}\sum_{q=0}^{Q-1}[n/Q]^{-1}\sum_{l=1}^{[n/Q]} \mathrm{E}(W_{Ql+q}) + n^{-1}\sum_{q=0}^{n\%Q} \mathrm{E}(W_{Ql+q}) \\ &\to Q^{-1}\sum_{q=0}^{Q-1}\mu_q + 0 = \mu, \end{aligned}$$

where $\mu = Q^{-1}\sum_{q=0}^{Q-1}\mu_q$. By Assumption 3, for each $q \in \{0, ..., Q-1\}$, there is a π_q such that $n^{-1}\sum_{m=1}^{n} \Delta_{mQ+q} \xrightarrow{\mathcal{P}} \pi_q$; consequently, $n^{-1}N_n \xrightarrow{\mathcal{P}} Q^{-1}\sum_{q=0}^{Q-1}\pi_q \equiv \bar{\pi}$. It follows that $N_n^{-1}\sum_{j=1}^{n} \mathrm{E}(W_j) = (n^{-1}N_n)^{-1} n^{-1}\sum_{j=1}^{n} \mathrm{E}(W_j) \xrightarrow{\mathcal{P}} \bar{\pi}^{-1}\mu$. A similar proof establishes the result for $N_n^{-1}\sum_{j=1}^{n} \mathrm{E}[\Delta_j K_j \rho^{K_j}\omega(S_j)\lambda(R_j)]$. □

Lemma 5. Fix $\theta \in \Theta$. Then there is a constant $\Psi(\theta)$ such that $N_n^{-1}\sum_{j=1}^{n} \Delta_j \psi_j(\theta) \xrightarrow{\mathcal{P}} \Psi(\theta)$.

Proof. As in the proof of Lemma 4, there is a constant π_0 such that $n^{-1}N_n \xrightarrow{\mathcal{P}} \pi_0 > 0$. It follows that $nN_n^{-1} \xrightarrow{\mathcal{P}} \pi_0^{-1}$, thus $\|nN_n^{-1}\|_1 < \infty$. Therefore, $\|nN_n^{-1}\Delta\psi_j(\theta)\|_1 < \infty$, and there is a constant C such that $\|N_n^{-1}\Delta\psi_j(\theta) - \mathrm{E}[N_n^{-1}\Delta\psi_j(\theta)]\|_1 \le n^{-1}C$ for all j.

Clearly $nC^{-1}N_n^{-1}\Delta\psi_j(\theta)$ is uniformly integrable. Also, $\sum_{j=1}^{n} Cn^{-1} = C < \infty$ and $\sum_{j=1}^{n} C^2 n^{-2} = n^{-1}C^2 \to 0$ as $n \to \infty$. By the mixingale law of large numbers (Davidson, Theorem 19.11, page 302),

$$N_n^{-1}\sum_{j=1}^{n} \Delta_j\psi_j(\theta) - \mathrm{E}\left[N_n^{-1}\sum_{j=1}^{n} \Delta_j\psi_j(\theta)\right] \xrightarrow{\mathcal{P}} 0$$

The conclusion follows from Lemma 4. □

Lemma 6. $\sup_{\theta \in \Theta} |N_n^{-1} \sum_{j=1}^n \Delta_j \psi_j(\theta) - \Psi(\theta)| \xrightarrow{\mathcal{P}} 0$

Proof. The functions $\lambda(\cdot)$ and $\omega(\cdot)$ possess continuous first derivatives, so for any $\theta, \tilde{\theta} \in \Theta$,

$$\begin{aligned}
\left| N_n^{-1} \sum_{j=1}^n \Delta_j \psi_j(\theta) - N_n^{-1} \sum_{j=1}^n \Delta_j \psi_j(\tilde{\theta}) \right| &= \left| N_n^{-1} \sum_{j=1}^n \Delta_j [\psi_j(\theta) - \psi_j(\tilde{\theta})] \right| \\
&= \left| N_n^{-1} \sum_{j=1}^n \Delta_j \dot{\psi}(\bar{\theta})(\theta - \tilde{\theta}) \right| \\
&\leq \left\| N_n^{-1} \sum_{j=1}^n \Delta_j \dot{\psi}(\bar{\theta}) \right\|_1 |\theta - \tilde{\theta}| \\
&\leq \left\| N_n^{-1} \sum_{j=1}^n \Delta_j \right\|_1 D|\theta - \tilde{\theta}| \\
&= D|\theta - \tilde{\theta}|,
\end{aligned}$$

where $\bar{\theta}$ lies between θ and $\tilde{\theta}$, and D is an upper bound of the matrix norms $\|\dot{\psi}(\bar{\theta})\|_1$ as $\bar{\theta}$ ranges over the compact closure of Θ. By Theorem 21.10 of Davidson (page 339), the collection of functions $\left\{ N_n^{-1} \sum_{j=1}^n \Delta_j \psi_j(\theta) \right\}$ is stochastically equicontinuous. The result follows from Lemma 5 of this technical report and Theorem 21.9 of Davidson (page 337). □

We are now in a position to address the major theorems of this paper. First, we state and prove Lemma 7, which asserts the kernel inspiration for the estimating equation: that the estimation function components are independent and have expectation zero when evaluated at θ_0.

Lemma 7. Evaluated at the true parameter values, $\mathrm{E}(\psi_j) = 0$ and $\mathrm{E}(\psi_h \psi_j') = 0$ for all $j \neq h$.

Proof. Evaluated at the true parameter values, each residual R_j is symmetric with expectation zero. Since $\lambda(\cdot)$ is odd, iterated expectation (conditioning on $\mathcal{X}$ and $\mathcal{H}_j$)

demonstrates that $\mathrm{E}(\psi_j) = 0$. Assume without loss of generality that $h < j$, so that ψ_h is a constant with respect to $\mathcal{H}_j$. Then

$$\mathrm{E}(\psi_h \psi_j') = \mathrm{E}[\mathrm{E}(\psi_h \psi_j' | \mathcal{H}_j)] = \mathrm{E}[\psi_h \mathrm{E}(\psi_j' | \mathcal{H}_j)] = \mathrm{E}[\psi_h \cdot 0] = 0.$$

□

Proof of Theorem 1. The result follows from Lemmas 6, 7, and Theorem 5.9 of van der Vaart (1998, page 46). □

Proof of Theorem 2. Let $\dot{\psi}_j(\theta)$ be the continuous first derivative matrix of $\psi_j(\theta)$, and let $\bar{D}_n(\theta) = N_n^{-1} \sum_{j=1}^n \Delta_j \dot{\psi}_j(\theta)$. Then for $\theta, \tilde{\theta} \in \Theta$

$$\begin{aligned} |\bar{D}_n(\theta) - \bar{D}_n(\tilde{\theta})| &\leq N_n^{-1} \sum_{j=1}^n |\dot{\psi}_j(\theta) - \dot{\psi}_j(\tilde{\theta})| \\ &\leq N_n^{-1} A |\theta - \tilde{\theta}| \\ &= O_p(1) |\theta - \tilde{\theta}|, \end{aligned}$$

by Assumption 4, the Lipschitz conditions on $\dot{\lambda}(\cdot)$ and $\dot{\omega}(\cdot)$. By Davidson (Theorem 21.10, page 339), the collection of functions $\{\bar{D}_n(\theta) - \bar{D}_n(\theta_0)\}$ is stochastically equicontinuous. Consequently, $|\bar{D}_n(\widehat{\theta}_n) - \bar{D}_n(\theta_0)| \xrightarrow{\mathcal{P}} 0$ by Theorem 1 of this technical report, Davidson (Theorem 21.9, page 337), and van der Vaart (Theorem 5.9, page 46).

By differentiation of $\psi_j(\theta)$ and application of Lemmas 1 and 2, it is straightforward (but tedious) to show that $\{\Delta_j \dot{\psi}_j(\theta) - \mathrm{E}[\Delta_j \dot{\psi}_j(\theta)], \mathcal{H}_j\}$ is an L_1 mixingale. Thus, the mixingale law of large numbers (Davidson, Theorem 19.11, page 302) implies that there is a function $\dot{\Psi}(\theta)$ such that $N_n^{-1} \sum_{j=1}^n \Delta_j \dot{\psi}_j(\theta) \xrightarrow{\mathcal{P}} \dot{\Psi}(\theta)$. The uniform continuity of the functions $\dot{\Psi}(\theta)$ over Θ ensures that $\dot{\Psi}(\theta_0)$ is in fact the first derivative of $\Psi(\theta_0)$, which by assumption is nonsingular in a neighborhood of θ_0. Now,

$$0 = \dot{\Psi}(\theta_0)^{-1} \sum_{j=1}^{n} \Delta_j \psi_j(\widehat{\theta}_n) = \sum_{j=1}^{n} \Delta_j \dot{\Psi}(\theta_0)^{-1} \psi_j(\theta_0) + \sum_{j=1}^{n} \Delta_j \dot{\Psi}(\theta_0)^{-1} \dot{\psi}_j(\bar{\theta})(\widehat{\theta}_n - \theta_0),$$

where $\bar{\theta}$ lies between $\widehat{\theta}_n$ and θ_0. Thus,

$$\begin{aligned} N_n^{1/2}(\widehat{\theta}_n - \theta_0) &= -N_n^{-1/2} \left[N_n^{-1} \sum_{j=1}^{n} \Delta_j \dot{\Psi}(\theta_0)^{-1} \dot{\psi}_j(\bar{\theta}) \right]^{-1} \sum_{j=1}^{n} \Delta_j \dot{\Psi}(\theta_0)^{-1} \psi_j(\theta_0) \\ &= -N_n^{-1/2} \Xi_n \sum_{j=1}^{n} \Delta_j \dot{\Psi}(\theta_0)^{-1} \psi_j(\theta_0), \end{aligned}$$

where Ξ_n converges in probability to the identity matrix. The moment conditions of Assumption 5 ensure that

$$\limsup_{n\to\infty} N_n^{-1/2} \mathrm{E} \left[\sum_{j=1}^{n} \Delta_j \dot{\Psi}(\theta_0)^{-1} \psi_j(\theta_0) \right] < \infty.$$

The sequence $\{\Delta_j \Psi(\theta_0)^{-1} \dot{\psi}_j(\theta)\}$ is ergodic, since it involves functions of terms of an ergodic sequence and an independent (therefore ergodic) sequence. By Lemma 3, the sequence is also an L_1 mixingale of size -1. A central limit theorem for stationary ergodic mixingales (Davidson, Theorem 24.5, page 385), applied to linear functions of each of the finite number of stationary subsequences, combined with the Cramer-Wold Theorem (Davidson, Theorem 25.5, page 405), proves that

$$N_n^{-1/2} \sum_{j=1}^{n} \Delta_j \dot{\Psi}(\theta_0)^{-1} \psi_j(\theta_0) \xrightarrow{\mathcal{L}} N(0, \Omega),$$

where $\Omega = \dot{\Psi}(\theta_0)^{-1} \Sigma_0 \dot{\Psi}(\theta_0)^{-T}$ and Σ_0 is the limiting variance of $N_n^{-1/2} \sum_{j=1}^{n} \Delta_j \psi_j(\theta_0)$. From an application of the standard Slutzky theorem, $N_n^{1/2}(\widehat{\theta}_n - \theta_0) \xrightarrow{\mathcal{L}} N(0, \Omega)$. □

Proof of Theorem 3. Following the notation in the proof of Theorem 2, it is necessary only to show that Σ_0 is consistently estimated by

$$\widehat{\Sigma}_n = N_n^{-1/2} \sum_{j=1}^{n} \Delta_j \psi_j(\widehat{\theta}_n) \psi_j(\widehat{\theta}_n)'.$$

Conditional on $\mathcal{H}_j$, $\psi_j(\theta_0)$ is independent of $\psi_h(\theta_0)$ for $h < j$. (See Lemma 7.) Consequently, if

$$\Sigma_n(\theta) \equiv N_n^{-1/2} \sum_{j=1}^{n} \Delta_j \psi_j(\theta) \psi_j(\theta)',$$

then $\Sigma_n(\theta_0) \xrightarrow{\mathcal{P}} \Sigma_0$. An approach similar to that used to prove consistency of $\widehat{\theta}_n$ can be used to show that in fact $\widehat{\Sigma}_n = \Sigma_n(\widehat{\theta}_n) \xrightarrow{\mathcal{P}} \Sigma_0$. We review the approach by sketching the proof.

Using an approach similar to the proofs of Lemmas 1, 2, 3, a mixingale property can be shown to apply to the terms $\Delta_j \psi_j(\theta) \psi_j(\theta)'$. Via Theorem 21.10 of Davidson, Assumption 4 can be used to demonstrate the stochastic equicontinuity necessary for an application of Theorem 21.9 of Davidson, by which we conclude that $\sup_{\theta \in \Theta} \|\Sigma_n(\theta) - \Sigma(\theta)\| \xrightarrow{\mathcal{P}} 0$ for some continuous matrix function $\Sigma(\theta)$ such that $\Sigma(\theta_0) = \Sigma_0$. Combining Theorem 1 of this paper and Theorem 5.9 of van der Vaart prove the necessary result. □