

Bent-cable regression with autoregressive noise

Grace S. CHIU and Richard A. LOCKHART

Key words and phrases: Change points; conditional least squares; maximum likelihood; segmented regression; time series.

MSC 2010: Primary 62M10; secondary 62E20.

Abstract: Motivated by time series of atmospheric concentrations of certain pollutants the authors develop bent-cable regression for autocorrelated errors. Bent-cable regression extends the popular piecewise linear (broken-stick) model, allowing for a smooth change region of any non-negative width. Here the authors consider autoregressive noise added to a bent-cable mean structure, with unknown regression and time series parameters. They develop asymptotic theory for conditional least-squares estimation in a triangular array framework, wherein each segment of the bent cable contains an increasing number of observations while the autoregressive order remains constant as the sample size grows. They explore the theory in a simulation study, develop implementation details, apply the methodology to the motivating pollutant dataset, and provide a scientific interpretation of the bent-cable change point not discussed previously.

1. INTRODUCTION

The breadth of applications of the bent-cable regression methodology has been demonstrated by Chiu, Lockhart & Routledge (2006). The bent cable is a continuous segmented function with three phases: the incoming and outgoing linear phases, joined smoothly by a quadratic bend of non-negative width. Smoothness is a result of imposing continuity on the first derivative of the bent-cable function, with respect to the univariate argument of the function, or to any of the five function parameters (regression coefficients); they are the incoming intercept and slope, β_0 and β_1 , respectively, the change in slope between the linear phases, β_2 , the centre of the bend, τ , and the half-width of the bend, γ . Taking $\boldsymbol{\theta} = (\beta_0, \beta_1, \beta_2, \tau, \gamma)'$ and $q(x; \tau, \gamma) = (x - \tau + \gamma)^2 / (4\gamma) \mathbf{1}\{|x - \tau| \leq \gamma\} + (x - \tau) \mathbf{1}\{x > \tau + \gamma\}$, the bent-cable function is

$$f(x; \boldsymbol{\theta}) = \beta_0 + \beta_1 x + \beta_2 q(x; \tau, \gamma). \quad (1)$$

To model the dependence of response Y_t on covariate x_t by bent-cable regression, we write

$$Y_t = f(x_t; \boldsymbol{\theta}) + W_t \quad (2)$$

where W_t 's are response errors, or noise, with mean 0.

Chiu, Lockhart & Routledge (2006) give an extensive list of work using conventional continuous change-point models. A popular one is the piecewise-linear (broken-stick) model, which, as discussed by these authors, can be inappropriate in the absence of clear scientific theory or evidence that supports an abrupt change as presumed by the model's kink. In contrast, the bend of the

bent cable allows for a smooth transition from one regime to the other, without ruling out abruptness that may be indicated by the data. This extra flexibility makes the bent cable an appealing alternative to kinked models in many practical settings.

Chiu, Lockhart & Routledge (2006) develop least-squares (LS) estimation theory for bent-cable regression when W_t 's are independent and identically distributed (iid). Despite intrinsic irregularity of the estimation problem due to a non-differentiable score function, they show that standard asymptotic results (consistency and asymptotic normality for parameter estimators, and asymptotic chi-squared distribution for deviance statistics) apply to LS estimation (equivalent to maximum likelihood (ML) estimation when W_t 's are normal) of the regression coefficients, θ . However, many practical applications of bent-cable regression may involve time-series data. One example is the pattern of atmospheric concentrations of chlorofluorocarbons (CFCs) over recent decades. Figure 1(a) shows 152 monthly mean measurements of CFC-11 (a common type of CFC), recorded in parts-per-trillion (ppt), starting with January 1988 as the zeroth month. (Data source: National Oceanic and Atmospheric Administration Earth System Research Laboratory (NOAA/ESRL), <ftp://140.172.192.211/hats/cfcs/cfc11/insituGCs/RITS/monthly/fl1mlomo.dat>.) The data appear to follow a bent cable, and clearly exhibit serial correlation. Thus, proper analysis of these data calls for the modelling of the autocorrelation alongside the bent-cable mean structure.

Existing works, such as Hallin, Taniguchi, Serroukh & Choy (1999), discuss ML asymptotics for estimates of both regression and time-series parameters, but only in the context of linear regression. Moreover, the bent cable is an irregular model not readily handled by standard asymptotic approaches. Evidently, there is a need for asymptotic theory to be specially developed for bent-cable regression involving serially correlated W_t 's. For the CFC data, the autocorrelation function (ACF) of the *detrended* series, i.e. $\{Y_t - f(x_t; \hat{\theta})\}$ where $\hat{\theta}$ is the estimated bent-cable coefficient assuming white noise, is given in Figure 1(b); the decay pattern suggests a roughly autoregressive (AR) structure. Since the AR model is one of the most common in time-series, we focus on the AR noise structure for bent-cable regression of serial data.

Details of the extended methodology to account for AR noise appear in Sections 2 and 3. Section 3 develops large-sample theory in a triangular array asymptotic framework. Implementation algorithms are proposed in Section 4. In Section 5, we apply our method to describe structural characteristics of the CFC-11 data of Figure 1. We use these data in Section 6 to illustrate the notion of the *critical time point*, a bent-cable “change point” previously not discussed by Chiu, Lockhart & Routledge (2006). Simulations presented in Section 7 exemplify the performance of the method for modelling finite time series in practice. We conclude with discussion in Section 8.

Note that we do not consider correlation structures other than AR. In many practical situations, the functional form of the regression model, i.e. the bent cable here, is of primary interest. When serial correlation is non-negligible, we acknowledge this temporal dependency as a statistical nuisance by assuming the stationary AR model as an approximation, then estimating the primary bent-cable coefficients and secondary AR parameters given a reasonably chosen AR order. Of course, alternative methods should be considered if the assumption of a stationary AR structure and/or bent-cable model is deemed inappropriate.

2. BENT-CABLE MODEL WITH $AR(p)$ NOISE

We focus on time-series contexts, in which a continuous response, Y_t , is regressed on time, $x_t=t$, as the covariate. We consider equally spaced observations at times $t=0, 1, \dots, T$. Particularly, we assume that the finite dataset $\mathbf{Y}=(Y_0, Y_1, \dots, Y_T)'$ is generated by (2), where

$$W_t = \phi_1 W_{t-1} + \phi_2 W_{t-2} + \dots + \phi_p W_{t-p} + \varepsilon_t \quad (3)$$

is a stationary zero-mean $AR(p)$ series, i.e. ε_t are iid with $E(\varepsilon_t)=0$ and $\text{Var}(\varepsilon_t)=\sigma^2$, and $\phi \in \mathcal{S} \equiv \{(\phi_1, \dots, \phi_p)' : \phi_p \neq 0, \sum_{i=1}^p \phi_i z^i \neq 1 \text{ for all complex } z \text{ such that } |z| \leq 1\}$. To be estimated are the

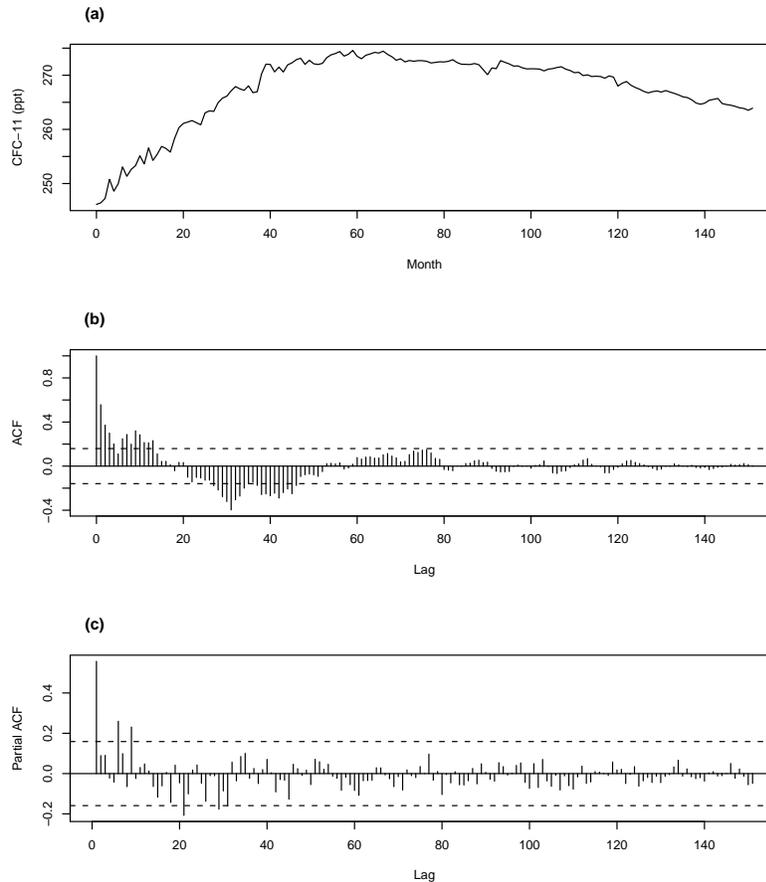


Figure 1: (a) CFC-11 monthly means from January 1988 to August 2000, recorded at the NOAA/ESRL station in Mauna Loa; (b) ACF plot for the detrended data (i.e. observed data minus bent-cable ML fit assuming white noise); (c) PACF plot for the detrended data.

bent-cable parameter, θ , the AR parameter, ϕ , and the innovation variance, σ^2 ; the latter two are regarded as nuisance.

Note that despite having discrete covariate values, the underlying bent cable, f , is continuous in t . Consequently, the transition parameters τ and γ are not restricted to being integer values. Furthermore, we consider $0 < \tau - \gamma < \tau + \gamma < T$ and $\beta_2 \neq 0$, so that all three phases of the bent cable are present and covered by the observed data.

3. CONDITIONAL MAXIMUM LIKELIHOOD ESTIMATION

For the purpose of defining the likelihood function, we consider normally distributed innovations, ε_t 's, although normality is not necessary in practice (see Section 3.2). As we explain below, our method can be referred to as conditional least-squares estimation irrespective of normality. However, for convenience, we adhere to the “likelihood” nomenclature throughout this article.

To develop estimation theory, we require the distinction between the “true” model parameter, $(\theta_0, \phi_0, \sigma_0^2)$, and a candidate value, (θ, ϕ, σ^2) . As is common for AR(p) models, our estimation of $(\theta_0, \phi_0, \sigma_0^2)$ is based on the conditional log-likelihood function $\ell_T(\theta, \phi, \sigma^2)$, with the first p

observations y_0, y_1, \dots, y_{p-1} regarded as given. Let $P(\cdot)$ denote a density function, and write $f_t(\boldsymbol{\theta}) = f(t; \boldsymbol{\theta})$. Then, ignoring irrelevant constants, $\ell_T(\boldsymbol{\theta}, \boldsymbol{\phi}, \sigma^2) = \ln P_{\boldsymbol{\theta}, \boldsymbol{\phi}, \sigma}(Y_p, Y_{p+1}, \dots, Y_T | y_0, y_1, \dots, y_{p-1}) = -(1/2)\sigma^{-2}S_T(\boldsymbol{\theta}, \boldsymbol{\phi}) - (T+1-p)\ln\sigma$, where

$$S_T(\boldsymbol{\theta}, \boldsymbol{\phi}) = \sum_{t=p}^T \left[(Y_t - f_t(\boldsymbol{\theta})) - \sum_{i=1}^p \phi_i (Y_{t-i} - f_{t-i}(\boldsymbol{\theta})) \right]^2 \quad (4)$$

is referred to as the conditional sum-of-squares error (CSSE). The conditional ML estimator (CMLE) of $(\boldsymbol{\theta}_0, \boldsymbol{\phi}_0, \sigma_0^2)$ is the argument that maximizes ℓ_T over the parameter space. The estimation of $\boldsymbol{\eta}_0 \equiv (\boldsymbol{\theta}_0, \boldsymbol{\phi}_0)$ does not depend on that of σ_0^2 . Hence, we can first take σ^2 as a constant, compute the CMLE, $\hat{\boldsymbol{\eta}}$, then evaluate the CMLE for σ_0^2 by

$$\hat{\sigma}^2 = \bar{S}_T(\hat{\boldsymbol{\eta}}) \equiv \frac{S_T(\hat{\boldsymbol{\eta}})}{T+1-p}. \quad (5)$$

The purpose of our current work is to extend the existing ML / LS theory for bent-cable regression of Chiu, Lockhart & Routledge (2006) who assume white noise. A direct extension to the case of AR(p) noise is CML theory, since $\hat{\boldsymbol{\eta}}$ maximizes ℓ_T , and hence, minimizes the CSSE, S_T (i.e. conditional LS estimation). We do not consider theory for full ML where $P_{\boldsymbol{\theta}, \boldsymbol{\phi}, \sigma}(\mathbf{Y})$ is maximized. While ML and CML are different approaches when $p > 0$, they are asymptotically equivalent for finite values of p .

In standard applications of order- (p, d, q) autoregressive integrated moving average (ARIMA) models, none of p , d , or q is part of statistical inference (e.g. Brockwell & Davis 2002, 2006). We follow this convention and develop our estimation theory for a given p . In practice, model diagnostics can suggest a suitable choice of p , as illustrated in Section 5.

3.1 CML large-sample theory

Although ML asymptotics exist for linear models with AR noise and for bent cables with white noise, extending these results to our current context is not straightforward. The reason is as follows. The number of data points in the incoming phase is, up to rounding, $\tau_0 - \gamma_0$. These points provide the information about the incoming slope. As a result, useful large-sample approximations must require $\tau_0 - \gamma_0$ to be large. Similarly, the number of points in the bend is essentially $2\gamma_0$, and in the outgoing phase, $T - (\tau_0 + \gamma_0)$; both quantities, too, must be large. These three conditions cannot be met simultaneously if we derive large-sample approximations by taking limits in which $\boldsymbol{\theta}_0$ is fixed. Instead, we take a limit in which the true parameter value, written as $\boldsymbol{\theta}_{0,T}$, depends on T in such a way that all three quantities above become large. The mathematics is simplest if the proportion of observations in each phase stays away from 0 as T grows.

Thus, for large-sample theory, we regard data from (2) as the T th row in a triangular array:

$$Y_{t,T} = f_t(\boldsymbol{\theta}_{0,T}) + W_t \quad (= f_t(\boldsymbol{\theta}_0) + W_t = Y_t). \quad (6)$$

We let

$$\tilde{\boldsymbol{\theta}} \equiv (\tilde{\beta}_0, \tilde{\beta}_1, \tilde{\beta}_2, \tilde{\tau}, \tilde{\gamma})', \quad \mathbb{M}_T = \text{diag}\{1, 1/T, 1/T, T, T\}, \quad \boldsymbol{\theta} \equiv \boldsymbol{\theta}_T = \mathbb{M}_T \tilde{\boldsymbol{\theta}} \quad (7)$$

so that $\boldsymbol{\theta}_0 \equiv \boldsymbol{\theta}_{0,T} = \mathbb{M}_T \tilde{\boldsymbol{\theta}}_0$. We introduce the notation

$$\tilde{f}_{t,T}(\tilde{\boldsymbol{\theta}}_0) = \tilde{\beta}_{0,0} + \tilde{\beta}_{1,0} t/T + \tilde{\beta}_{2,0} q(t/T; \tilde{\tau}_0, \tilde{\gamma}_0).$$

Figure 2 illustrates the fact that due to (7),

$$\tilde{f}_{t,T}(\tilde{\boldsymbol{\theta}}) = f_{t/T}(\tilde{\boldsymbol{\theta}}) = f_t(\boldsymbol{\theta}_T). \quad (8)$$

Hence, (6) can be expressed in terms of $\tilde{\theta}_0$ as

$$Y_{t,T} = \tilde{f}_{t,T}(\tilde{\theta}_0) + W_t. \quad (9)$$

We develop asymptotics for CML estimation of $\tilde{\eta}_0 \equiv (\tilde{\theta}'_0, \phi'_0)'$ based on minimization of

$$\tilde{S}_T(\tilde{\theta}, \phi) = \sum_{t=p}^T \left[(Y_{t,T} - \tilde{f}_{t,T}(\tilde{\theta})) - \sum_{i=1}^p \phi_i (Y_{t-i,T} - \tilde{f}_{t-i,T}(\tilde{\theta})) \right]^2 \quad (10)$$

which is equivalent to (4) due to (7). Note that when applying bent-cable regression in practice, one is never required to deal with (9), (10), or $\tilde{\eta}_0$; the CMLE is $\hat{\eta} = ([\mathbb{M}_T^{-1} \hat{\theta}]', \hat{\phi}')'$. The triangular array is a mere technical device for establishing large-sample approximations. This device is also employed in other works on change-point asymptotics, e.g. Fotopoulos, Jandhyala & Tan (2009).

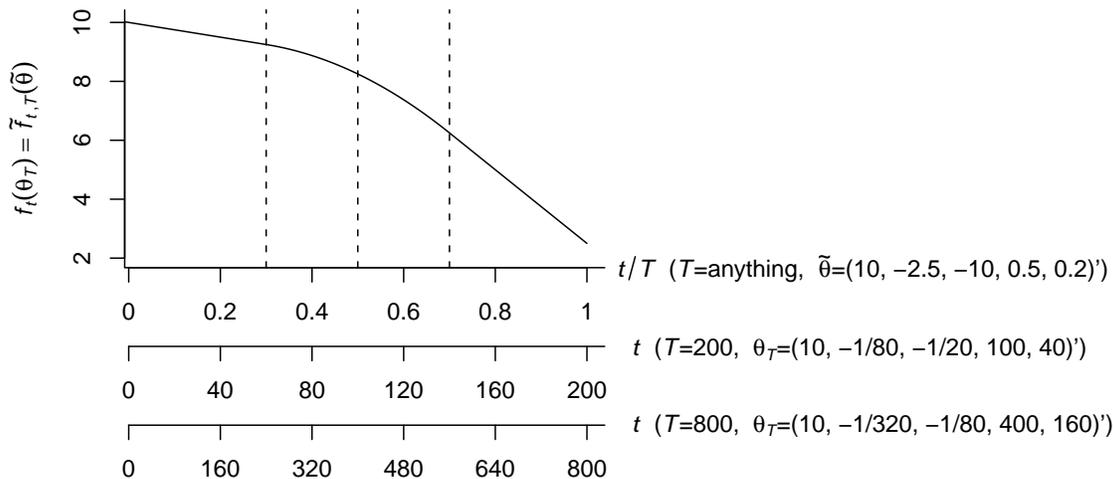


Figure 2: Two elements ($T=200$ and $T=800$) in the sequence of bent-cable parameters $\{\theta_T\}$, and the corresponding parameter θ in the triangular array.

We emphasize that although the triangular array in (9) involves W_t , the AR noise series itself is T -free. The Associate Editor suggests the following. Our asymptotic framework here can be contrasted with an alternative framework that is based also on (9) but coupled with a different AR series, $W_{t,T}$, that depends on T in such a way that $\text{Corr}(W_{t,T}, W_{t+k,T}) \rightarrow 1$ as $T \rightarrow \infty$, i.e. the time gap between consecutive observations decreases as $T \rightarrow \infty$. Under (3) and (9), the effective sample size (ESS) (Thiébaux & Zwiers 1984) of $\{W_t, t/T \in A \subset [0,1]\}$ approaches ∞ as $T \rightarrow \infty$, and so consistent estimators are possible. However, under the alternative framework, the ESS of $\{W_{t,T}, t=0, \dots, T\}$ remains bounded as $T \rightarrow \infty$, and thus consistent estimators are not possible.

3.2 Statement of theorems

The main results of this article are given as theorems below. Proofs appear in the Appendix.

THEOREM 1 (CONSISTENCY). *Given is the model comprising (3) and (9), where ε_t 's are i.i.d. with mean 0 and finite fourth moment. Consider estimating $\tilde{\eta}_0$ by minimizing (10) over some compact*

set $\mathcal{K} \subset \Omega = \{\tilde{\boldsymbol{\theta}}: \tilde{\beta}_2 \neq 0, 0 < \tilde{\tau} - \tilde{\gamma} < \tilde{\tau} + \tilde{\gamma} < 1\} \times \{\boldsymbol{\phi}: \boldsymbol{\phi} \in \mathcal{S}\}$, and estimating σ_0^2 by (5). Take $T \rightarrow \infty$. Then, $\hat{\boldsymbol{\eta}}$ and $\hat{\sigma}^2$ (referred to as the CMLEs) are consistent estimators of $\boldsymbol{\eta}_0$ and σ_0^2 , respectively.

As usual, asymptotic normality of $\hat{\boldsymbol{\eta}}$ follows from its consistency. Since the map from $\hat{\boldsymbol{\eta}}$ to $\hat{\boldsymbol{\eta}}$ is linear, asymptotic normality of $\hat{\boldsymbol{\eta}}$ is equivalent to that of $\hat{\boldsymbol{\eta}}$. However, due to the irregularity of the bent-cable model, some care is required to define the Hessian and Fisher information. Again we take $\sigma = \sigma_0$ to be constant, then consider quantities derived from S_T in (4) but omit the term *conditional* and the dependence on T in the notation to avoid clutter.

First, the score function is $\mathbf{U}_\sigma(\boldsymbol{\eta}) = -(1/2)\sigma^{-2}\nabla S_T(\boldsymbol{\eta})$. Although continuous, \mathbf{U}_σ is cusped (hence, non-differentiable) along the $T+1$ pairs of hyper-rays $\mathcal{R} = \{\boldsymbol{\eta} : \tau \pm \gamma = t \text{ for } t=0,1,\dots,T\}$; also see Chiu, Lockhart & Routledge (2006). Here, we adopt their definition of \mathbb{V}_σ^+ , the directional Hessian, by taking a directional gradient of \mathbf{U}_σ . Hence, \mathbb{V}_σ^+ is well-defined on Ω and is identical to $\nabla \mathbf{U}_\sigma(\boldsymbol{\eta})$ over $\Omega \setminus \mathcal{R}$. Finally, we define

$$\mathbb{I}_\sigma(\boldsymbol{\eta}_0) = -E[\mathbb{V}_\sigma^+(\boldsymbol{\eta}_0)]. \quad (11)$$

We refer to $\mathbb{I}_\sigma(\boldsymbol{\eta}_0)$ as the ‘‘Fisher information’’ matrix, since, for normal ε_t ’s, it has the usual definition apart from directional derivatives. (For convenience of nomenclature, we use the term ‘‘Fisher information’’ even when normality is not assumed.) We derive \mathbb{V}_σ^+ in the Appendix.

THEOREM 2 (ASYMPTOTIC NORMALITY). *Under the conditions of Theorem 1,*

1. *the matrix $T^{-1}\mathbb{I}_\sigma(\boldsymbol{\eta}_0)$ is positive definite for all sufficiently large T , and similarly, $P_{\boldsymbol{\eta}_0}\{T^{-1}\mathbb{I}_\sigma(\hat{\boldsymbol{\eta}}) \text{ is positive definite}\} \rightarrow 1$ as $T \rightarrow \infty$;*
2. *whenever $T^{-1}\mathbb{I}_\sigma(\boldsymbol{\eta}_0)$ is positive definite, it has a unique lower triangular square-root with positive diagonal entries, denoted by $[T^{-1}\mathbb{I}_\sigma(\boldsymbol{\eta}_0)]^{1/2}$, and $\sqrt{T}[T^{-1}\mathbb{I}_\sigma(\boldsymbol{\eta}_0)]^{1/2}(\hat{\boldsymbol{\eta}} - \boldsymbol{\eta}_0)$ converges in distribution to a standard $(5+p)$ -variate normal random variable; the statement holds true with $\mathbb{I}_\sigma(\hat{\boldsymbol{\eta}})$ replacing $\mathbb{I}_\sigma(\boldsymbol{\eta}_0)$;*
3. *Assertions (1) and (2) hold true when $\hat{\sigma}^2$ replaces σ^2 in the expression of $\mathbb{I}_\sigma(\boldsymbol{\eta}_0)$ or $\mathbb{I}_\sigma(\hat{\boldsymbol{\eta}})$.*

4. COMPUTING THE CMLE AND FISHER INFORMATION

Again, quantities with a *tilde* (‘‘ \sim ’’) are not considered in practice. For computations, we first propose three algorithms in Sections 4.1 to 4.3 for obtaining the CMLEs $\hat{\boldsymbol{\eta}}$ and $\hat{\sigma}^2$; the subscript ‘‘0’’ is dropped from this discussion to reduce clutter. The latter two algorithms involve some technicality that is discussed in Section 4.4. A description of the computation of \mathbb{I}_σ follows in Section 4.5. An implementation of all five sections is available to the user of our methodology through the `bentcableAR` package publicly distributed by the R Project for Statistical Computing.

4.1 Unconstrained conditional least-squares

The theory of this article concerns the minimization of (4) over the $\boldsymbol{\theta}$ and $\boldsymbol{\phi}$ spaces simultaneously. In practice, constrained optimization to yield weak stationarity (i.e. $\hat{\boldsymbol{\phi}} \in \mathcal{S}$) is a non-trivial exercise. While many time-series analysis programs have such constraints built-in for fitting linear regression with AR noise, we are unaware of ones that readily handle the bent-cable counterpart. Here, we propose a simple algorithm for obtaining $\hat{\boldsymbol{\eta}}$ over the unconstrained \mathbb{R}^{5+p} instead of Ω , although the long-run behaviour of $\hat{\boldsymbol{\phi}}$ from this algorithm is unclear. In the case that it results in $\hat{\boldsymbol{\phi}} \notin \mathcal{S}$ but stationarity is insisted upon, alternative algorithms in Sections 4.2 and 4.3 may be considered.

Minimization of the objective function S_T over an unconstrained parameter space can be easily handled by standard optimization software. An initial estimate $\hat{\boldsymbol{\eta}}^{(0)}$ is required due to non-linearity; for this, consider the *profile deviance surface* from Chiu, Lockhart & Routledge (2006) for $p=0$, where $\boldsymbol{\beta}=(\beta_0, \beta_1, \beta_2)'$ is profiled out. As $p>0$ in the current estimation problem, we additionally profile out $\boldsymbol{\phi}$, and evaluate this surface, D_{prof} , at each point on a fine (τ, γ) -grid. Note that our estimation problem is linear in $\boldsymbol{\beta}$ and $\boldsymbol{\phi}$. Hence, D_{prof}^j , the profile deviance evaluated at the j th grid point, is a function of the CMLEs $\hat{\boldsymbol{\beta}}_j$ and $\hat{\boldsymbol{\phi}}_j$, both of which correspond to having the j th grid point as the given value of the underlying transition parameters. Consequently, all of D_{prof}^j , $\hat{\boldsymbol{\beta}}_j$, and $\hat{\boldsymbol{\phi}}_j$ can be readily obtained for each j using a standard software that produces CML fits for an AR(p) model with a linear trend. Next, define $(\hat{\tau}^{(0)}, \hat{\gamma}^{(0)}) = \arg \max_j D_{\text{prof}}^j$, and $\hat{\boldsymbol{\beta}}^{(0)}$ and $\hat{\boldsymbol{\phi}}^{(0)}$ to be the associated CMLEs given $(\hat{\tau}^{(0)}, \hat{\gamma}^{(0)})$. Finally, write $\hat{\boldsymbol{\theta}}^{(0)}=(\hat{\boldsymbol{\beta}}^{(0)}, \hat{\tau}^{(0)}, \hat{\gamma}^{(0)})'$ and $\hat{\boldsymbol{\eta}}^{(0)}=(\hat{\boldsymbol{\theta}}^{(0)}, \hat{\boldsymbol{\phi}}^{(0)})'$.

The quality of such $\hat{\boldsymbol{\eta}}^{(0)}$ as an initial estimate naturally depends on the grid resolution. As computational burden increases with resolution, a less intensive alternative (that is possibly less precise) is to take $\hat{\boldsymbol{\theta}}^{(0)}$ to be the MLE of $\boldsymbol{\theta}$ assuming $p=0$ (see Chiu, Lockhart & Routledge 2006), then set $\hat{\boldsymbol{\phi}}^{(0)}$ to be the MLE of the AR coefficients for the residual series $\{Y_t - f_t(\hat{\boldsymbol{\theta}}^{(0)})\}$. In practice, however, we observe that for some datasets, such initial values often *mislead* the optimization software to converge on a local minimum of S_T that may be some distance from the global minimum. In either case, substitute the converged $\hat{\boldsymbol{\eta}}$ into (5) to obtain $\hat{\sigma}^2$.

4.2 Iterative conditional / full maximum likelihood hybrid algorithm

This algorithm addresses the case in which unconstrained optimization of Section 4.1 leads to $\hat{\boldsymbol{\phi}} \notin \mathcal{S}$. However, it does not address the case of $\hat{\beta}_2=0$ and/or $\hat{\tau} \leq 0$ and/or $\hat{\gamma} \leq 0$. Nevertheless, such may not be a practical issue, since most optimization programs allow bounded parameter spaces, which may be applied to τ and γ . For example, the R function `optim()` allows quasi-Newton optimization with optional box constraints for each parameter, although from our experience, unconstrained quasi-Newton is generally faster and converges more readily. Finally, a CMLE consisting of boundary values for any of these three parameters may simply indicate that either the underlying model does not satisfy the non-degeneracy assumptions, or more data are needed for proper parameter estimation. Of course, the former is a case of model misspecification, and is beyond the scope of this article. For the latter, large-sample inference from Section 3.2 may be inappropriate.

For our hybrid algorithm here, note that \mathbf{Y} has a multivariate normal distribution with mean $(f_0(\boldsymbol{\theta}_0), \dots, f_T(\boldsymbol{\theta}_0))'$ and covariance matrix $c_0 \boldsymbol{\Sigma}_0$, where

$$\boldsymbol{\Sigma}_0 \equiv \boldsymbol{\Sigma}(\phi_0) = \begin{bmatrix} 1 & r_1 & r_2 & \dots & \dots & r_T \\ r_1 & 1 & r_1 & r_2 & \dots & r_{T-1} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ r_T & r_{T-1} & \dots & \dots & r_1 & 1 \end{bmatrix}, \quad r_h = \frac{c_h}{c_0} \quad \forall h = 0, 1, \dots, T$$

and $c_h \equiv c_h(\phi_0, \sigma_0)$ is the lag- h autocovariance for $\{W_t\}$. Also note that $\boldsymbol{\Sigma}_0$ does not depend on σ_0 since r_h is the autocorrelation. Now, the full log-likelihood is $\ell^{\text{full}}(\boldsymbol{\eta}, \sigma^2) \equiv \ln P_{\boldsymbol{\eta}, \sigma}(\mathbf{Y}) = \ell_T(\boldsymbol{\eta}, \sigma^2) + \ell^e(\boldsymbol{\eta}, \sigma^2)$, where ℓ_T can be regarded as the *working* component, and $\ell^e(\boldsymbol{\eta}, \sigma^2) = \ln P_{\boldsymbol{\eta}, \sigma}(Y_0, Y_1, \dots, Y_{p-1})$ is the *residual* component. This setup resembles that for maximization-by-parts by Song, Fan & Kalbfleisch (2005). Specifically, ℓ_T dominates ℓ^e in the information it contains about ℓ^{full} in applications where T is much greater than p . (Full) ML estimation via maximization-by-parts requires iteratively solving (A) $\nabla \ell_T(\boldsymbol{\eta}, \sigma^2) = \mathbf{0}$ — call the solution $\boldsymbol{\psi}^*$ — and (B) $\nabla \ell_T(\boldsymbol{\eta}, \sigma^2) = -\nabla \ell^e(\boldsymbol{\psi}^*)$. Note that Step (A) alone is CML. Here, for the purpose of restricting $\hat{\boldsymbol{\phi}} \in \mathcal{S}$, we modify Step (B) as (B*) to incorporate some information from ℓ^e to obtain Step (A) estimates at each iteration.

Our iterative approach is as follows. For Step (A), maximize ℓ_T over $\boldsymbol{\theta}$ while holding $\boldsymbol{\phi}$ fixed. This can be broken down into linear and non-linear components: (i) given (τ, γ) and $\boldsymbol{\phi}$, computing $\boldsymbol{\Sigma}(\boldsymbol{\phi})$ is routine (see Section 4.4), so that one may exploit the simple closed form generalized LS (GLS) linear solution of $\hat{\boldsymbol{\beta}}$ that maximizes ℓ^{full} (thereby incorporating ℓ^e in the iterations of Step (A)); (ii) given $\boldsymbol{\beta}$ and $\boldsymbol{\phi}$, maximizing ℓ_T reduces to non-linear LS estimation of (τ, γ) , which is a two-dimensional problem that can be easily implemented with standard optimization software. For Step (B*), maximize ℓ^{full} over $\boldsymbol{\phi}$ for $\boldsymbol{\theta}$ fixed, i.e. obtain the AR coefficient MLEs for the mean-zero series $\{Y_t - f_t(\boldsymbol{\theta})\}$. Such estimates that also satisfy stationarity constraints are readily available by using a standard time-series analysis software, which may additionally provide an estimate for σ . Our algorithm iterates between (A) and (B*) until convergence. Thus, the overall algorithm solves $\nabla \ell^{\text{full}}(\boldsymbol{\eta}) = \mathbf{0}$ over $\boldsymbol{\beta}$ via (A)(i), $\nabla \ell_T(\boldsymbol{\eta}) = \mathbf{0}$ over (τ, γ) via (A)(ii), and $\nabla \ell^{\text{full}}(\boldsymbol{\eta}) = \mathbf{0}$ over $\boldsymbol{\phi}$ via (B*). In essence, this is an iterative full ML procedure except for (A)(ii). For it to be full ML, $\nabla \ell^{\text{full}}$ would replace $\nabla \ell_T$ in (A)(ii), although the extra complexity due to ℓ^e may not be worthwhile if one's purpose is to merely restrict $\hat{\boldsymbol{\phi}} \in \mathcal{S}$.

Note that with this hybrid algorithm, one has the choice of taking the $\hat{\sigma}^2$ associated with the last iteration from Step (B*), or substituting the converged $\hat{\boldsymbol{\eta}}$ into (5) to obtain $\hat{\sigma}^2$.

4.3 Iterative CML-ML-method-of-moments hybrid algorithm

This algorithm is identical to that of Section 4.2, with the exception that the computation of $\hat{\boldsymbol{\phi}}^{(i)}$ at iteration i is based on solving the Yule-Walker (YW) equations that involve sample autocovariance values expressed as functions of $\hat{\boldsymbol{\theta}}^{(i)}$. Standard time-series analysis software provide this computation to yield the method-of-moments (MM) estimate $\hat{\boldsymbol{\phi}}^{(i)}$ that is guaranteed to fall inside \mathcal{S} for all i (Brockwell & Davis 2002).

4.4 Autocovariance as a function of $\boldsymbol{\phi}$ and σ

The previous two algorithms involve the computation of a GLS solution $\hat{\boldsymbol{\beta}}_{\boldsymbol{\phi}, \tau, \gamma}$ based on the autocorrelation matrix $\boldsymbol{\Sigma}(\boldsymbol{\phi})$. To this end, first consider the autocovariance c_h expressed in terms of the true values $\boldsymbol{\phi}_0$ and σ_0 . For an AR(p) process $\{W_t\}$, it can be shown that

$$\text{Var}(W_t) = c_0(\boldsymbol{\phi}_0, \sigma_0) = \sigma_0^2 \left\{ 1 - \sum_{i=1}^p \phi_{0,i}^2 - 2 \sum_{i < j} \phi_{0,i} \phi_{0,j} r_{j-i}(\boldsymbol{\phi}_0) \right\}^{-1} \quad (12)$$

$i, j = 1, \dots, p$

for all t . Stationarity of $\{W_t\}$ ensures that (12) is well defined. At lag $h > 0$, one can verify that the autocorrelation r_h satisfies the system

$$\begin{aligned} 0 = & \phi_{0,h} + (\phi_{0,2h} - 1) r_h(\boldsymbol{\phi}_0) + \mathbf{1}\{h > 1\} (\phi_{0,1} + \phi_{0,2h-1}) r_{h-1}(\boldsymbol{\phi}_0) \\ & + \mathbf{1}\{h > 2\} [(\phi_{0,h-1} + \phi_{0,h+1}) r_1(\boldsymbol{\phi}_0) + (\phi_{0,h-2} + \phi_{0,h+2}) r_2(\boldsymbol{\phi}_0) + \dots + \\ & (\phi_{0,2} + \phi_{0,2h-2}) r_{h-2}(\boldsymbol{\phi}_0)] + \mathbf{1}\{2h < p\} [\phi_{0,2h+1} r_{h+1}(\boldsymbol{\phi}_0) + \phi_{0,2h+2} r_{h+2}(\boldsymbol{\phi}_0) \\ & + \dots + \phi_{0,p} r_{p-h}(\boldsymbol{\phi}_0)] \quad \text{for } 1 \leq h < p, \end{aligned} \quad (13)$$

$$r_h(\boldsymbol{\phi}_0) = \phi_{0,1} r_{h-1}(\boldsymbol{\phi}_0) + \phi_{0,2} r_{h-2}(\boldsymbol{\phi}_0) + \dots + \phi_{0,p} r_{h-p}(\boldsymbol{\phi}_0) \quad \text{for } h \geq p. \quad (14)$$

Note that (13)–(14) for $1 \leq h \leq p$ form the system of YW equations. For general $\boldsymbol{\phi}$ and σ , we replace all instances of $\boldsymbol{\phi}_0$ in (12)–(14) with $\boldsymbol{\phi}$ (i.e. $\phi_{0,i}$ with ϕ_i) and of σ_0 with σ . These equations can be put in the context of our hybrid algorithms, as follows.

Given $\boldsymbol{\phi}$ at any iteration, we solve (13) for $r_h(\boldsymbol{\phi})$, where $h=1, \dots, p-1$. Substitute them in (14) to obtain $r_h(\boldsymbol{\phi})$ for $h=p, \dots, T$. This completes the computation of $\boldsymbol{\Sigma}(\boldsymbol{\phi})$. Note that our usage of

the YW equations here may be unorthodox, as observed moments of c_h are typically substituted in them to solve for MM estimates of $\phi_{0,i}$.

4.5 Fisher information as a function of $\boldsymbol{\eta}$

In practice, only Assertion (3) of Theorem 2 is directly consequential, and it requires the computation of $\mathbb{I}_{\hat{\sigma}}(\hat{\boldsymbol{\eta}})$. Thus, it is necessary to consider \mathbb{I}_{σ} as a function of $\boldsymbol{\eta}$. To do so, we again fix $\sigma=\sigma_0$, and expand the expression on the right-hand-side of (11), then replace $\boldsymbol{\eta}_0$ by $\boldsymbol{\eta}$ to yield $\mathbb{I}_{\sigma}(\boldsymbol{\eta})$. Next, partition $\mathbb{I}_{\sigma}(\boldsymbol{\eta})$ into an upper-left 5×5 block, a lower-right $p\times p$ block, and the $5\times p$ and $p\times 5$ off-diagonal blocks, as follows:

$$\mathbb{I}_{\sigma}(\boldsymbol{\eta}) = \left[\begin{array}{c|c} \mathbb{I}_{\sigma}^{\boldsymbol{\theta}}(\boldsymbol{\eta}) & \mathbb{I}_{\sigma}^{\boldsymbol{\eta}}(\boldsymbol{\eta}) \\ \hline -\frac{\mathbb{I}_{\sigma}^{\boldsymbol{\theta}}(\boldsymbol{\eta})}{[\mathbb{I}_{\sigma}^{\boldsymbol{\eta}}(\boldsymbol{\eta})]'} & \mathbb{I}_{\sigma}^{\phi}(\boldsymbol{\eta}) \end{array} \right].$$

Note that \mathbb{I}_{σ} is made up of $T+1-p$ summands over $t=p, p+1, \dots, T$. In the Appendix, we show that the t th summand of $\mathbb{I}_{\sigma}^{\boldsymbol{\theta}}(\boldsymbol{\eta})$ has (j, k) th element equal to $\sigma^{-2}[\nabla_j D_t(\boldsymbol{\eta})][\nabla_k D_t(\boldsymbol{\eta})]$, where $D_t(\boldsymbol{\eta}) = \sum_{i=1}^p \phi_i [Y_{t-i} - f_{t-i}(\boldsymbol{\theta})] - [Y_t - f_t(\boldsymbol{\theta})]$, with $D_t(\boldsymbol{\eta}_0) = -\varepsilon_t$, and ∇_j denotes the partial derivative with respect to the j th element of $\boldsymbol{\theta}$. Indeed, $\mathbb{I}_{\sigma}^{\boldsymbol{\theta}}$ for $p=0$ corresponds to the Fisher information of Chiu, Lockhart & Routledge (2006). We also derive in the Appendix that $\mathbb{I}_{\sigma}^{\boldsymbol{\eta}}(\boldsymbol{\eta}) \equiv \mathbb{O}$ (matrix of zeros), and that the t th summand of $\mathbb{I}_{\sigma}^{\phi}(\boldsymbol{\eta}_0)$ has (j, k) th element equal to $\sigma^{-2}c_{|j-k|}(\phi_0, \sigma)$. Take $\sigma^{-2}c_{|j-k|}(\phi, \sigma)$ over j, k , and t to form $\mathbb{I}_{\sigma}^{\phi}(\boldsymbol{\eta})$, which is $\boldsymbol{\theta}$ -free.

Finally, $\hat{\boldsymbol{\eta}}$ and $\hat{\sigma}$ are substituted into the above expressions to yield $\mathbb{I}_{\hat{\sigma}}(\hat{\boldsymbol{\eta}})$. Note that obtaining $\mathbb{I}_{\hat{\sigma}}^{\phi}(\hat{\boldsymbol{\eta}})$ may be computationally involved for $p>1$, due to the solving of (12)–(14).

5. ANALYSIS OF CFC-11 DATA

The NOAA/ESRL Global Monitoring Division (GMD) records hourly measurements of atmospheric concentration of CFCs from numerous monitoring stations around the globe. These data date as far back as the 1970's, and are publicly available on the GMD website. We apply our methodology to analyze the CFC-11 monthly means made at each of several stations using the Radiatively Important Trace Species (RITS) system; in this article, we only discuss the Mauna Loa (MLO) station from July 1987 to August 2000. Subsequent measurements at MLO were made using the newer Chromatograph for Atmospheric Trace Species system. We restrict our attention to the RITS data to avoid the possible non-stationarity element due to a switch in measuring devices.

Online GMD data are occasionally updated, and the version that we analyze was last updated on May 4, 2001. Note that each monthly mean is computed based on a different number of measurements whose sample standard deviation varies from month to month. Also, five means are missing (January 1992, and June to September 1995). As our current method cannot handle missing data, they are replaced with the corresponding measurements made by the Halocarbons and other Atmospheric Trace Species Group Flask Sampling Program: we use the 1992 flask data (last updated on August 25, 1999; ftp://140.172.192.211/hats/cfcs/cfc11/flasks/monthly/oldgc/f11mon.dat) and the 1995 data (last updated on November 24, 2004; ftp://140.172.192.211/hats/cfcs/cfc11/flasks/monthly/otto/mlomon_f11.txt). We expect the introduction of a different sampling scheme for 5 of more than 150 observations to be a negligible source of non-stationarity.

Preliminary visual inspection of the complete time series suggests that the observations from 1987 may violate the linearity of the bent cable's incoming phase. Consequently, we drop these values and analyze the remaining 152 observations shown in Figure 1. This minor truncation is justified, in that our main interest is in the structure of the change of CFC concentration over time, from being an increasing trend in earlier years to a decreasing one more recently. With the emphasis placed on the smoothly bent phase, Chiu, Lockhart & Routledge (2006) have demonstrated that

such types of changes often can be effectively described by the bend joining two linear phases to form a bent cable.

The corresponding partial ACF (PACF) is given in Figure 1(c). As the horizontal lines delimit $\pm 1.96/\sqrt{(T+1)}$, they can be regarded as crude guidelines for determining p . Thus, we see that the lag-1 autocorrelation heavily dominates, although that at lags 6, 9, and 21 is potentially non-negligible. Nevertheless, the main objective is to understand the response mean structure while accounting for non-trivial autocorrelation. We first focus on $p=1$, and only consider higher values of p for possibly improved fits.

5.1 AR(1)

With $p=1$, the default CML algorithm of Section 4.1 produces the parameter estimates in Table 1. The fitted bent cable appears in Figure 3(a) in bold gray, with the estimates $\hat{\tau}$ and $\hat{\tau} \pm \hat{\gamma}$ delimited by bold gray vertical lines. As $\hat{\phi} \in \mathcal{S}$ here, the alternative hybrid algorithms of Sections 4.2 and 4.3 are unnecessary. The fitted residuals, \widehat{W}_t , and fitted innovations, $\widehat{\varepsilon}_t$, appear in Figure 3(b) in gray and black, respectively. Some non-random pattern remains, as well as heteroskedasticity between the earlier and latter months. Moreover, the P/ACF of $\widehat{\varepsilon}_t$ (not shown) exhibits potentially non-negligible autocorrelation at various lags as high as 31. Altogether, these suggest that a stationary AR(1) process may not be fully adequate to describe these data properly.

Table 1: Bent-cable regression estimates for the CFC-11 data.

	AR(1)	AR(6)
$(\widehat{\beta}_0, \widehat{\beta}_1, \widehat{\beta}_2)$	(247.29, 0.64, -0.75)	(249.34, 0.55, -0.67)
$(\widehat{\tau}, \widehat{\gamma})$	(46.35, 24.77)	(49.00, 21.03)
$\widehat{\phi}$	0.56	(0.57, 0.03, 0.07, -0.04, -0.16, 0.28)
$\widehat{\sigma}^2$	0.56	0.46
approx. 95% CI for \mathcal{T}_θ	63.47 ± 3.53	62.71 ± 4.93

Despite this, a drastic advantage of the AR(1) model over the “naive” fit assuming white noise is evident in the assessment of uncertainty associated with bent-cable parameters. The difference is seen in Figure 4, where profile deviance contours for the AR(1) and naive fits are shown. The profiling technique mentioned in Section 4.1 was employed to produce these plots. The innermost contour with value -6 for each plot may be regarded as an approximate 95% confidence region (CR) for (τ_0, γ_0) computed assuming the corresponding noise structure. (This approximation is based on the asymptotic χ_2^2 distribution, in light of Theorem 2, of the (conditional) profile deviance drop as a function of (τ, γ) . The tactics used by Chiu, Lockhart & Routledge 2006 can be employed here for its proof). The naive fit produces a smooth, rounded CR that is wrongly optimistic in the uncertainty associated with the transition parameters. In contrast, by acknowledging a positive and substantial lag-1 autocorrelation, we reveal the extent of this uncertainty (in the form of an unbounded CR) and the irregularity of the estimation problem (in the form of an infinite diagonal ridge), despite a large T . (Refer to Chiu, Lockhart & Routledge 2006 for an in-depth discussion on interpreting the profile deviance surface).

5.2 Other AR(p) fits

As the AR(1) model is not fully adequate, one may consider a general ARIMA structure and/or non-stationarity for $\{W_t\}$, but such is beyond the scope of our methodology. Instead, we consider a larger p as a compromise, while keeping the fit somewhat parsimonious. For example, an AR(6) CML

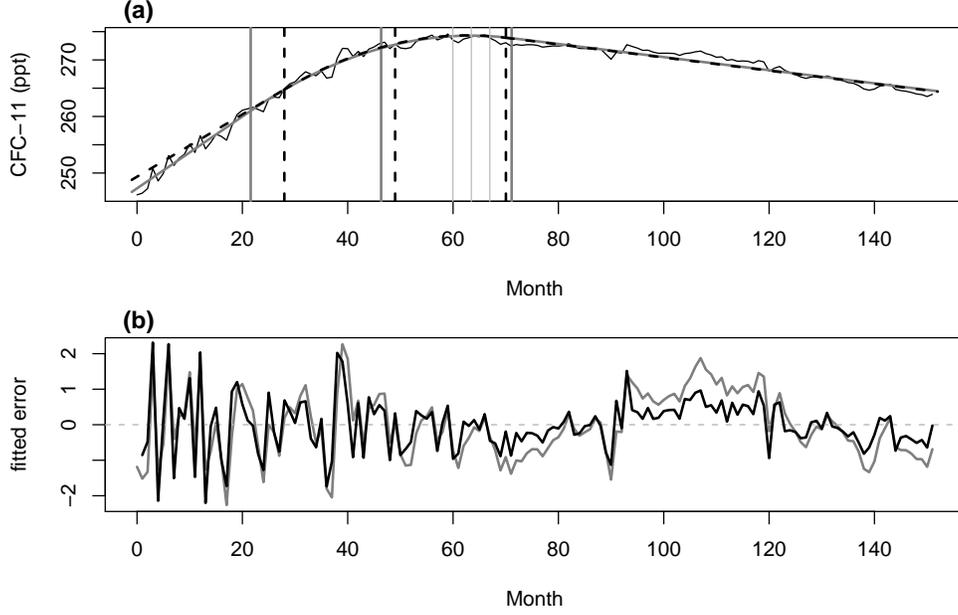


Figure 3: (a) Fitted bent cable with transition region assuming AR(1) (—) and AR(6) (- - -) noise; an approximate 95% CI for the CTP corresponding to the AR(1) fit is also shown (—); (b) fitted residuals (—) and fitted innovations (—) from the AR(1) fit.

fit appears in Figure 3(a) in dashed lines. Aside from a milder incoming slope for AR(6), both estimated cables virtually coincide. Thus, the characteristics of \widehat{W}_t (not shown for $p=6$) are highly comparable between fits. For $\widehat{\varepsilon}_t$'s (also not shown), the larger p does not improve heteroskedasticity, although the corresponding P/ACF shows no substantial deviation from that for white noise.

To quantitatively compare the two models, consider an AR(6) model subject to the null constraint $H_0 : \phi_{0,2} = \dots = \phi_{0,6} = 0$. Due to the nature of CML estimation, one can easily show that this reduced AR(6) fit is equivalent to an AR(1) fit on the reduced dataset $\{y_6, y_7, \dots, y_T\}$. Now, denote by ℓ_T^m the maximized conditional log-likelihood for the full fit, and ℓ_T^{mr} for the reduced. Let S_T^m and S_T^{mr} denote the corresponding values of the CSSE, both of which are sums of $T^* = T + 1 - 6$ terms. Then, the conditional deviance statistic is

$$2[\ell_T^m - \ell_T^{mr}] = \frac{S_T^{mr}}{S_T^{mr}/T^*} + T^* \ln \frac{S_T^{mr}}{T^*} - \frac{S_T^m}{S_T^m/T^*} - T^* \ln \frac{S_T^m}{T^*} = T^* [\ln S_T^{mr} - \ln S_T^m] = 4.3.$$

When compared to the χ_5^2 distribution, it suggests little evidence against H_0 , i.e. having $p=1$ is no less adequate than $p=6$ here.

There are other potential shortcomings with the use of a large p . Note that the earlier half of the AR(6) incoming phase appears to be positively biased. Given virtually identical outgoing slopes for both AR fits, first-order differentiability of the bent cable forces the AR(6) fit to exhibit a narrower bend. Hence, the duration of this bend may also be biased. Indeed, the bias of the incoming phase may be expected to increase with p , as the randomness of the first p data points (out of a finite sample) is entirely unaccounted for by the CML method, possibly leading to overfitting. For this, raising p simply to compensate for the inadequacy of an AR model when the data show a more general ARIMA structure and/or non-stationarity is not recommended for CML inference.

To further investigate the notion of overfitting, we obtained an unconstrained AR(9) CML fit

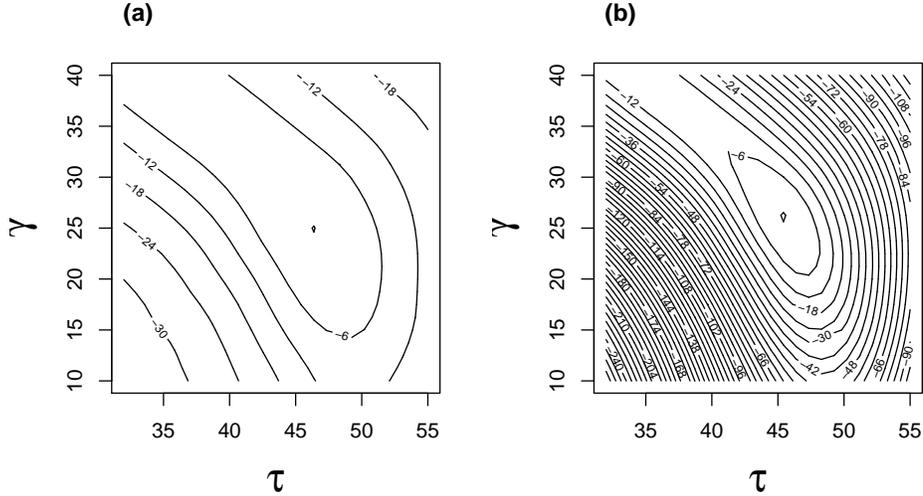


Figure 4: Profile deviance contours for the (a) AR(1) fit and (b) naive (AR(0)) fit.

(not shown) for these data. A notable peculiarity of this fit is its near non-stationarity: $1 - \sum_{i=1}^p \hat{\phi}_i = 0.06$ (≈ 0), as opposed to 0.44 and 0.25 for $p=1$ and 6, respectively. However, the striking feature is that the fit hovers entirely above the observed data. Interestingly, such a non-sensical fit exhibits a very reasonable bend that ranges from 29 to 64. Moreover, while \widehat{W}_t here is uniformly negative, $\widehat{\varepsilon}_t$ much better resembles white noise than when $p=1$ or 6, with substantially reduced heteroskedasticity. Nevertheless, having discarded information on the variability of 6% (9 out of 152) of the time series and nearly 1/3 of the incoming phase has drastically impaired CML estimation for the model as a whole.

In contrast, the CML-ML-MM hybrid algorithm results in an alternative AR(9) fit (also not shown) that highly resembles the AR(1) CML fit, and hence, is much more sensible than the AR(9) CML counterpart. Thus, it is evident that accounting for the variability of all data in part of the estimation procedure helps to counter the effects of potential overfitting. However, since the conditional likelihood methodology of this article handles CML estimation only, we shall not pursue further inference based on the hybrid fit.

6. ESTIMATING THE CRITICAL TIME POINT

Let us again drop the subscript “0” for true parameter values throughout this section.

In some applications, it may be of interest to locate the point at which the bent cable’s slope changes sign. In a temporal context, it is the *critical time point* (CTP) at which the response mean structure takes either an upturn from a decreasing trend, or a downturn from an increasing trend. We denote it by \mathcal{T}_θ . It parallels the break point of the conventional broken stick (the bent cable’s limiting case as $\gamma \downarrow 0$) when the mean response begins to move in a different direction. For this, the idea of a CTP is inapplicable to any bent cable whose slope does not exhibit a sign change, including that whose either linear phase is flat. (A similar idea to the CTP in such cases may be τ , the centre of the bend, although care is needed to distinguish between τ and \mathcal{T}_θ otherwise: see Section 6.1.) Hence, we restrict our attention to the bend region $[\tau - \gamma, \tau + \gamma]$, and define $\mathcal{T}_\theta = \arg_t \{ \partial f_t(\theta) / \partial t = 0 \}$. Then, one can verify that $\mathcal{T}_\theta = \tau - \gamma - 2\beta_1\gamma / \beta_2$, which is estimated by $\widehat{\mathcal{T}} \equiv \widehat{\mathcal{T}}_\theta$.

As \mathcal{T}_θ is non-linear in the vector components of θ , we employ a first-order Taylor expansion about $\hat{\theta}$ for the large-sample distribution of \hat{T} . That is,

$$\hat{T} \approx \mathcal{T}_\theta + \xi'_\theta(\hat{\theta} - \theta) \quad \text{where} \quad \xi_\theta = \left(0, -\frac{2\gamma}{\beta_2}, \frac{2\beta_1\gamma}{\beta_2^2}, 1, -\frac{2\beta_1 + \beta_2}{\beta_2} \right)'.$$

By Theorem 2, an approximate $100(1-\delta)\%$ Wald confidence interval (CI) for \mathcal{T}_θ is therefore

$$\hat{T} \pm z_{\delta/2} \sqrt{\hat{\xi}' [\mathbb{I}_\sigma^\theta(\hat{\eta})]^{-1} \hat{\xi}} \quad (15)$$

where $\hat{\xi} \equiv \xi_{\hat{\theta}}$ and z_q is the $(1-q)$ th quantile of the standard normal distribution.

6.1 Example: CFC-11 data revisited

Since January 1, 1989, many countries around the world have adopted stringent policies to control the use and production of CFC-releasing substances, thanks to the Montréal Protocol. Earlier non-participating countries have recently followed suit. As a result, atmospheric concentrations of CFCs have shown a steady decline for more than a decade. Of course, the decline did not start immediately following the CFC ban; any impact on the atmosphere would take time to develop. Therefore, it may be of interest to estimate the CTP at which the decline began to take place. However, as monitoring stations are situated in geographical locations that are vastly spread out, CTPs deduced from various stations may be very different. This is evident in the graphs of monthly mean CFC readings found at the GMD website. This difference causes ambiguity in the interpretation of individual CTPs. To properly handle this, Khan, Chiu & Dubin (2009) develop, in a Bayesian framework, bent-cable regression methodology for longitudinal data that arise from multiple observational units. Instead, we focus on CFC-11 detection by the MLO station only, and make classical inference on the corresponding \mathcal{T}_θ .

We apply (15) to the AR(1) and AR(6) CML fits from Section 5. The two resulting CIs appear in Table 1, the former of which is also shown in Figure 3(a) as light gray vertical lines. In both instances, \mathcal{T}_θ is estimated to take place in around the 63rd month (April 1993); apparently, it took only about 4 years for the Montréal Protocol to show its success. The key feature here is a corresponding CI that resides entirely above $\hat{\tau}$ (the upper 95% confidence limit for τ is less than 56 for both AR(1) and AR(6), computed based on the respective profile deviance as a function of τ only, and compared to the χ_1^2 distribution). This phenomenon has significant implications. The conventional change-point technique of broken-stick fitting would direct one's attention to the estimate of the kink, which is perceivably around $\hat{\tau}$ of a bent-cable fit. However, if the data possibly exhibit a gradual transition such as the bend of a cable, then the "change point" in the form of a CTP could be a substantial distance from τ . Therefore, if the start of a downturn from an upward trend (or *vice versa*) is of scientific interest, then it is vital for the focus to be placed on \mathcal{T}_θ instead of τ . In our example, mistaking $\hat{\tau} < 56$ (September 1992) for the change point could result in an unwarranted early declaration of the Protocol's success.

7. SIMULATIONS

We now discuss simulations in the AR(1) case to exemplify the large-sample behaviour of the CMLEs, $\hat{\eta}$ and $\hat{\sigma}^2$, given finite samples. Again, we use subscript "0" for true parameter values.

7.1 Performance of large-sample approximations for CFC-11 analysis

First, we illustrate that large-sample approximations from our theory are valid for the CFC analysis. To do so, we conducted simulations, using the CML AR(1) fitted cable from Section 5.1 (see Table

1) as the underlying structure. Innovations ε_t 's were generated from a t -distribution with 10 degrees of freedom (df), but scaled to have variance $\sigma_0^2=0.56$. The choice of df was to impose heavier tails than a normal distribution. Taking $\rho_0\equiv\phi_{0,1}=0.56$, noise W_t 's were generated by (3) with a starting value of 0, and a burn-in of 1,000 values that were subsequently discarded. Responses y_t 's resulted from substituting $\boldsymbol{\theta}_{0,T}=(247.29, 0.64, -0.75, 46.35, 24.77)'$ and W_t 's into (6); 1,000 such $\{y_t\}$ series were generated. Unconstrained CMLEs assuming normality were computed for all 1,000 time series, with the knowledge of a stationary AR(1) noise structure. Since all $\hat{\rho}$ values fell inside $(-1, 1)$, hybrid algorithms were unnecessary. However, one simulation resulted in $\hat{\gamma}<0$, for which conditional least-squares was repeated but optimized over $\gamma\geq 0$. We focus on the parameters of main interest: the transition, $\boldsymbol{\alpha}_0\equiv(\tau_0, \gamma_0)$, and the CTP, $\mathcal{T}_{\boldsymbol{\theta}_0}$. Empirical coverage of nominal 95% CRs/CIs from these 1,000 datasets appears in the first row of Table 2.

Table 2: Empirical coverage (%) out of 1,000 simulated bent-cable AR($p=1$) time series with non-normal innovations. CRs for (τ_0, γ_0) and CIs for $\mathcal{T}_{\boldsymbol{\theta}_0}$ have a nominal 95% confidence level, based on critical values from χ^2 , F , and normal distributions. For the bottom four rows, $\boldsymbol{\eta}_{0,T}=(\boldsymbol{\theta}'_{0,T}, \rho_0)'$, with $\sigma_0^2=0.01$ and $\tilde{\boldsymbol{\theta}}_0=(10, -2.5, -10, 0.5, 0.2)'$ as appears in Figure 2.

truth	T	(τ_0, γ_0)				$\mathcal{T}_{\boldsymbol{\theta}_0}$
		(A) Deviance		(B) Wald		Wald
		χ^2	F	χ^2	F	
CFC fit from Table 1	151	91.6	91.8	87.9	88.2	89.7
(1)(a): $\boldsymbol{\eta}_{0,T} = (10, -1/80, -1/20, 100, 40, 0.8)'$	200	87.2	88.1	93.9	94.1	—
(1)(b): $\boldsymbol{\eta}_{0,T} = (10, -1/80, -1/20, 100, 40, 0.2)'$	200	94.1	94.2	94.0	94.2	—
(2)(a): $\boldsymbol{\eta}_{0,T} = (10, -1/320, -1/80, 400, 160, 0.8)'$	800	94.2	94.3	95.1	95.1	—
(2)(b): $\boldsymbol{\eta}_{0,T} = (10, -1/320, -1/80, 400, 160, 0.2)'$	800	96.2	96.2	97.4	97.5	—

Wald CIs for $\mathcal{T}_{\boldsymbol{\theta}_0}$ were computed using (15). For $\boldsymbol{\alpha}_0$, we follow Chiu, Lockhart & Routledge (2006) and consider (A) deviance-based CRs and (B) Wald CRs, where actual CRs needed not be computed: for (A), a (conditional) profile deviance drop between 0 and $-\chi_2^2(0.05)=-5.99$ or $-2F_{2,(T+1-p)-2}(0.05)=-6.1136$ implied coverage; likewise, $(\hat{\boldsymbol{\alpha}}-\boldsymbol{\alpha}_0)'\mathbb{S}^{-1}(\hat{\boldsymbol{\alpha}}-\boldsymbol{\alpha}_0)<5.99$ or 6.1136 implied coverage for (B), where \mathbb{S} is the 2×2 block of $[\mathbb{I}_{\hat{\sigma}}(\hat{\boldsymbol{\eta}})]^{-1}$ that corresponds to $\boldsymbol{\alpha}$. As expected for finite samples, observed coverage for either $\boldsymbol{\alpha}_0$ or $\mathcal{T}_{\boldsymbol{\theta}_0}$ was somewhat less than the nominal 95% (from 88% to 92%), with increasing coverage using Wald, χ^2 , and F critical values, respectively. Qualitative features of the fits would have been impractical to assess visually on a case-by-case basis, as simulations were fully automated. However, in the absence of model misspecification aside from non-normality, qualitative anomalies were not expected. Overall, results here suggest that our asymptotic framework provided proper inference for atmospheric CFC-11 concentrations.

7.2 Large-sample approximations for a bent cable with no CTP

Next, we ran simulations featuring true bent cables whose slope did not change signs; we used the cable from Figure 2. Two values of T were investigated: (1) $T=200$ and (2) $T=800$. For each, two values of the AR(1) parameter were examined: (a) $\rho_0=0.8$ and (b) $\rho_0=0.2$. Aside from these parameters and $\sigma_0^2=0.01$, all simulation and estimation procedures were identical to Section 7.1.

Parameter constraints proved to be unnecessary here. Empirical coverage from the 4,000 fits appear in Table 2, bottom four rows, in ascending order of the effective information available from the data, a qualitative notion related to the ESS. Here, we are interested in a qualitative assessment

of the performance of normal approximations based on our asymptotic framework. Hence, formal computation of the ESS is not pursued, although we use the term “ESS” in the discussion, in the sense that for a fixed sample size, the more correlated the data, the less information they contain, or the smaller the ESS.

For our simulations with the smaller three ESSs, moderate to no under-coverage (from 87% to 95%) was observed, with F -based coverage higher than the χ^2 -based by a small margin that approached 0 as the ESS increased, roughly speaking. Curiously for these three, Wald coverage was no less or even higher than deviance-based coverage, the latter of which is typically considered more reliable (Harrell 2001). The same peculiarity was true for the largest ESS (97+% vs. 96%), in addition to both types of coverage being higher than nominal; though, deviance-based coverage was closer to nominal, and hence, superior in this case. Considering all four sets of simulations, Wald coverage appeared closer to nominal when ρ_0 was closer to 1; the advantage was more noticeable for a smaller T . All this perhaps was attributable to the non-trivial underestimation of the nuisance parameters ρ_0 and σ_0 for small ESSs (see Figure 5), leading to a false sense of high precision through the use of $\mathbb{I}_{\hat{\sigma}}(\hat{\boldsymbol{\eta}})$ for Wald CRs. Specifically, each graph in Figure 5 comprises boxplots for Sets (1)(a), (1)(b), (2)(a), and (2)(b) by increasing ESS / decreasing negative bias from left to right. (All these distributions are roughly symmetric; hence, the median and mean are similar.) Thus, one cannot determine which of deviance-based and Wald CRs is generally preferable in bent-cable time series regression, in light of how the negative bias for $\rho_0 > 0$ and σ_0 can compensate for the theoretical inferiority of Wald CRs apparent in most types of regression settings. Finally, note that Table 2 and Figure 5 together reflect consistency of $\hat{\boldsymbol{\eta}}$ and $\hat{\sigma}^2$.

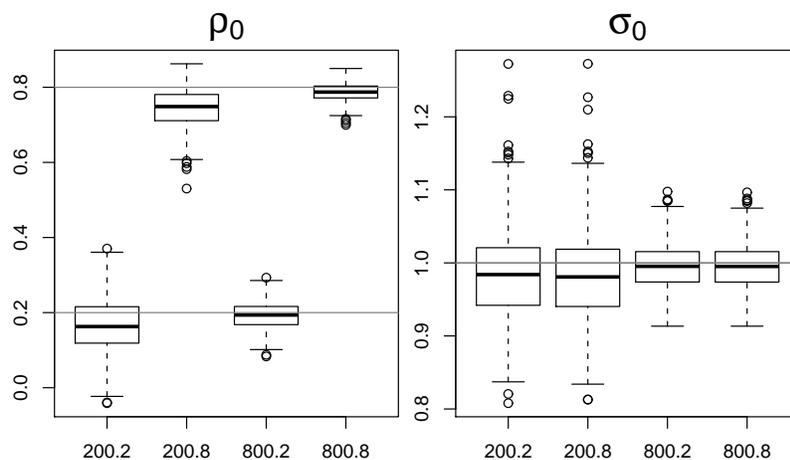


Figure 5: Boxplots for CMLs of covariance parameters: each tickmark on the x -axes indicates the values of T and ρ_0 for the 1,000 simulated datasets that produced the boxplot, e.g. “800.2” refers to $T=800$ and $\rho_0=0.2$; horizontal lines in gray indicate the actual parameter values that generated the data.

Indeed, negative bias in the estimate of ρ_0 for AR(1) time series with a linear trend has been well documented in work such as Marriott & Pope (1954). To conduct our own investigation for $\hat{\rho}$ and $\hat{\sigma}$, we ran separate sets of simulations (not shown) involving a response mean that is (i) horizontal and ranging from 0 to an unknown constant, (ii) a straight line with non-trivial slope, and (iii) a bent cable. Our results indicate that for fixed T , ρ_0 , and σ_0 , bias increases as more unknown parameters in the mean structure are being estimated. Bias corrections for ρ_0 of a mean-zero AR(1) series appear in Mudelsee (2001) and White (1961), and those for AR(p) coefficients in linear regression are in Cheang & Reinsel (2000). However, bias corrections for nuisance parameter

estimates in bent-cable regression is beyond the scope of our current article.

We return to the simulations of Table 2 and Figure 5. In addition to consistency, we also investigated asymptotic normality for all model parameters except σ_0 . Histograms for simulated estimates (not shown) exhibited a general “bell shape.” Exceptions to symmetry occurred for the distribution of $\hat{\gamma}_{T=200}$ which showed mild skewness for both $\rho_0=0.8$ and 0.2 . For the former, the left tail was truncated at 0 because of a large variability due to a small ESS and a lower γ -bound of 0. However, for $T=800$, all parameter estimates exhibited normality.

8. DISCUSSION

Previous articles (Chiu, Lockhart & Routledge 2006 and related works) have discussed the potential advantage of fitting bent cables instead of broken sticks for describing phenomena that are traditionally considered piecewise-linear in nature. Bent-cable regression for different types of correlated data is an area of ongoing research (Khan, Chiu & Dubin submitted; Reynolds & Chiu submitted). In this article, we have restricted our attention to bent-cable regression in contexts involving a single time series, where the notion of a change point is often of practical significance. For bent cables, we have referred to it as the CTP, and demonstrated its distinct interpretation from the centre of the cable’s transition region, the latter being a time point which may be easily confused as the change point due to its close relation to the break of the conventional broken stick.

For temporally correlated data, both inference for and implementation of bent-cable regression are substantially more complex than the case of independent data. Specifically, classical inference is impossible for this type of regression under the standard asymptotic framework in which model parameters are invariant to an ever-increasing number of observations over time. Instead, we have employed a framework that involves an AR noise series indexed only by time (t), coupled with a triangular array of bent-cable models indexed by t and the number of observations (T); all elements of the triangular array share a common parameter that is T -free and linearly related to the bent-cable parameter at hand (dependent on T) that corresponds to the observed time scale. This non-standard framework is used to provide useful distributional approximations for the practitioner of the bent-cable methodology when confidence limits on the observed time scale are concerned. Assuming a stationary AR temporal structure, we have developed inference procedures and implementation of our methodology using conditional least-squares estimation. Data analyses and simulations have demonstrated the practicality of our asymptotic framework and methodology.

Our inference theory is developed, as is common, for a known AR order, p . However, in very few practical situations is p predetermined; in general, p must be estimated or *selected* (sometimes multiple times) according to certain criteria, and the subsequent fit(s) diagnosed. Any formal estimation of p is expected to influence the behaviour of other parameter estimators, but this topic is beyond the scope of our article. Here, estimation of the bent-cable parameters is our primary interest; when serial correlation is non-negligible, we acknowledge the dependence over time by applying the stationary AR model as an approximation, without excessive concern over the underlying value of p . This approach is similar to that of Brockwell & Davis (2002), page 141. Our modelling framework has shown promise in practical settings of serially correlated data. Of course, some situations may call for models other than the bent cable or $\text{AR}(p)$. As with any modelling exercise in practice, alternative methodologies should be considered if the current modelling framework performs inadequately.

Finally, readers interested in applying the methodology of this paper are encouraged to consider the `bentcableAR` package, publicly available through the Comprehensive R Archive Network.

APPENDIX

We derive the conditional score, directional Hessian, and Fisher information after establishing some notational conventions. We drop the subscript σ on \mathbf{U} , V^+ and \mathbb{I} to reduce clutter. The

various derivatives and right derivatives involve several indicators for which notation is useful: $J_{1t}(\tau, \gamma) = \mathbb{1}\{t > \tau + \gamma\}$, $J_{2t}(\tau, \gamma) = \mathbb{1}\{|t - \tau| \leq \gamma\}$, and $J_{3t}(\tau, \gamma) = \mathbb{1}\{\tau - \gamma < t \leq \tau + \gamma\}$. To shorten expressions below we suppress τ and γ from the J 's. Throughout this appendix we use $s = s(t) = t/T$ as a shorthand to relabel quantities in the triangular array when necessary. Both \sum_s and \sum_t extend over $t = p, \dots, T$, and is equivalent to ranging s over $\Psi(T) = \{p/T, (p+1)/T, \dots, 1\}$. We will need the fact that for continuous functions g and h on the unit interval and integers $1 \leq i, j \leq p$, we have

$$\lim_{T \rightarrow \infty} \frac{1}{T} \sum_s g(s - i/T) h(s - j/T) = \int_0^1 g(u) h(u) du; \quad (16)$$

the convergence is uniform in any family of Lipschitz functions g, h with a uniform bound on the functions and on their Lipschitz constants.

Now, let $Z_t(\boldsymbol{\theta}) = Y_t - f_t(\boldsymbol{\theta}) = W_t + f_t(\boldsymbol{\theta}_0) - f_t(\boldsymbol{\theta})$, and let $\varphi_0 = -1$ and $\varphi_i = \phi_i$ for $i = 1, \dots, p$. Thus, D_t from Section 4.5 can be rewritten as $D_t(\boldsymbol{\eta}) = \sum_{i=0}^p \varphi_i Z_{t-i}(\boldsymbol{\theta})$, so that $S_t(\boldsymbol{\eta}) = \sum_t D_t^2(\boldsymbol{\eta})$ and the t th summand of $\mathbf{U}(\boldsymbol{\eta})$ is $\mathbf{U}_t(\boldsymbol{\eta}) = -\sigma^{-2} [D_t(\boldsymbol{\eta})][\nabla D_t(\boldsymbol{\eta})]$. One can show that the ϕ -derivatives are $\nabla_{\phi_j} D_t(\boldsymbol{\eta}) = Z_{t-j}(\boldsymbol{\theta})$ for $j=1, \dots, p$. For $\boldsymbol{\theta}$ -derivatives,

$$\nabla_{\beta_0} D_t(\boldsymbol{\eta}) = -\sum_{i=0}^p \varphi_i, \quad \nabla_{\beta_1} D_t(\boldsymbol{\eta}) = -\sum_{i=0}^p \varphi_i (t - i), \quad (17)$$

$$\nabla_{\beta_2} D_t(\boldsymbol{\eta}) = -\sum_{i=0}^p \varphi_i (\alpha_{1,t-i} \alpha_{4,t-i} + \gamma \alpha_{2,t-i}^2), \quad (18)$$

$$\nabla_{\tau} D_t(\boldsymbol{\eta}) = \beta_2 \sum_{i=0}^p \varphi_i (\alpha_{1,t-i} + \alpha_{2,t-i}), \quad \nabla_{\gamma} D_t(\boldsymbol{\eta}) = -\beta_2 \sum_{i=0}^p \varphi_i \alpha_{3,t-i}, \quad (19)$$

where $\alpha_{1t} = J_{1t}$, $\alpha_{2t} = [t - (\tau - \gamma)] / (2\gamma) J_{2t}$, $\alpha_{3t} = (1/4)[1 - (t - \tau)^2 / \gamma^2] J_{2t}$, and $\alpha_{4t} = t - \tau$. Assemble to obtain \mathbf{U}_t .

We denote by $V_{j,k,t}^+(\boldsymbol{\eta})$ the derivative from the right of the k th vector component of $\mathbf{U}_t(\boldsymbol{\eta})$ with respect to η_j . Apply (11) to define $I_{j,k,t}(\boldsymbol{\eta})$, the (j, k) th entry of the t th summand of $\mathbb{I}(\boldsymbol{\eta})$. For $p=0$, all expressions of \mathbf{U} , \mathbb{V}^+ , and \mathbb{I} reduce to those of Chiu, Lockhart & Routledge (2006). For a general p , with the argument $\boldsymbol{\eta}$ suppressed from the notation, some tedious algebra can show that the (j, k) th entry of \mathbb{I}^θ is the expectation of $-V_{j,k,t}^+ = [(\nabla_j D_t)(\nabla_k D_t) - Q_{j,k,t}] / \sigma^2$, where

$$Q_{j,k,t} = \begin{cases} -D_t \sum_{i=0}^p \varphi_i (\alpha_{1,t-i} + \alpha_{2,t-i}) & j, k \in \{\beta_2, \tau\} \\ D_t \sum_{i=0}^p \varphi_i \alpha_{3,t-i} & j, k \in \{\beta_2, \gamma\} \\ (2\gamma)^{-1} \beta_2 D_t \sum_{i=0}^p \varphi_i J_{3,t-i} & j = k = \tau \\ (2\gamma^2)^{-1} \beta_2 D_t \sum_{i=0}^p \varphi_i \alpha_{4,t-i} J_{3,t-i} & j = \tau, k = \gamma \\ (2\gamma^2)^{-1} \beta_2 D_t \sum_{i=0}^p \varphi_i \alpha_{4,t-i} J_{2,t-i} & j = \gamma, k = \tau \\ (2\gamma^3)^{-1} \beta_2 D_t \sum_{i=0}^p \varphi_i \alpha_{4,t-i}^2 J_{2,t-i} & j = k = \gamma \\ 0 & \text{otherwise} \end{cases}.$$

Note that $E[Q_{j,k,t}(\boldsymbol{\eta}_0)] = 0$ since $E[D_t(\boldsymbol{\eta}_0)] = 0$. By (11), $I_{j,k,t}^\theta(\boldsymbol{\eta}) = \sigma^{-2} [\nabla_j D_t(\boldsymbol{\eta})][\nabla_k D_t(\boldsymbol{\eta})]$.

For \mathbb{I}^θ , one can show that

$$V_{\beta_0 \phi_j, t}^+ = V_{\phi_j \beta_0, t}^+ = \frac{1}{\sigma^2} \left[D_t + Z_{t-j} \sum_{i=0}^p \varphi_i \right], \quad (20)$$

$$V_{\beta_1 \phi_j, t}^+ = V_{\phi_j \beta_1, t}^+ = \frac{1}{\sigma^2} \left[D_t (t - j) + Z_{t-j} \sum_{i=0}^p \varphi_i (t - i) \right], \quad (21)$$

$$V_{\beta_2\phi_j,t}^+ = V_{\phi_j\beta_2,t}^+ = \frac{1}{\sigma^2} \left[D_t(\alpha_{1,t-j}\alpha_{4,t-j} + \gamma\alpha_{2,t-j}^2) + Z_{t-j} \sum_{i=0}^p \varphi_i(\alpha_{1,t-i}\alpha_{4,t-i} + \gamma\alpha_{2,t-i}^2) \right], \quad (22)$$

$$V_{\tau\phi_j,t}^+ = V_{\phi_j\tau,t}^+ = -\frac{\beta_2}{\sigma^2} \left[D_t(\alpha_{1,t-j} + \alpha_{2,t-j}) + Z_{t-j} \sum_{i=0}^p \varphi_i(\alpha_{1,t-i} + \alpha_{2,t-i}) \right], \quad (23)$$

$$V_{\gamma\phi_j,t}^+ = V_{\phi_j\gamma,t}^+ = \frac{\beta_2}{\sigma^2} \left[D_t\alpha_{3,t-j} + Z_{t-j} \sum_{i=0}^p \varphi_i\alpha_{3,t-i} \right]. \quad (24)$$

Evaluated at $\boldsymbol{\eta}_0$, all (20)–(24) have mean 0. Hence, by (11), $\mathbb{I}^\eta(\boldsymbol{\eta}) \equiv \mathbb{O}$.

Finally, for \mathbb{I}^ϕ , it can be shown that $V_{\phi_j\phi_k,t}^+(\boldsymbol{\eta}_0) = -\sigma^{-2}[Z_{t-j}(\boldsymbol{\theta}_0)][Z_{t-k}(\boldsymbol{\theta}_0)] = -\sigma^{-2}W_{t-j}W_{t-k}$ for all $j, k = 1, \dots, p$. Hence, by (11), $I_{\phi_j\phi_k,t}(\boldsymbol{\eta}) = \sigma^{-2}c_{|j-k|}(\boldsymbol{\phi}, \sigma)$.

For the following proofs, recall $\Psi(T)$, (7), (8), and that $\tilde{\boldsymbol{\eta}} = (\tilde{\boldsymbol{\theta}}', \boldsymbol{\phi}')'$. Regularity of the estimation problem hinges on the $\tilde{\boldsymbol{\eta}}$ parametrization or Ψ scale, which ensures that the design points satisfy the design conditions of Chiu, Lockhart & Routledge (2006).

Proof of Theorem 1. Consider $\tilde{S}_T(\tilde{\boldsymbol{\eta}})$ from (10) and

$$H(\tilde{\boldsymbol{\eta}}) \equiv \sigma_0^2 + (\boldsymbol{\phi} - \boldsymbol{\phi}_0)' \mathbb{I}^* (\boldsymbol{\phi} - \boldsymbol{\phi}_0) + \left(1 - \sum_{i=1}^p \phi_i\right)^2 \int_0^1 [f_s(\tilde{\boldsymbol{\theta}}) - f_s(\tilde{\boldsymbol{\theta}}_0)]^2 ds,$$

where $\mathbb{I}^* = c_0 \boldsymbol{\Sigma}_0 = \text{Cov}(W_{t-1}, \dots, W_{t-p})$ for all t . Then, we have the following lemma.

LEMMA 1. As $T \rightarrow \infty$,

1. the function $T^{-1}\tilde{S}_T(\tilde{\boldsymbol{\eta}})$ converges to $H(\tilde{\boldsymbol{\eta}})$ uniformly on \mathcal{K} , in probability;
2. on Ω , the function H is continuous and uniquely minimized by $\tilde{\boldsymbol{\eta}}_0$, i.e. $H(\tilde{\boldsymbol{\eta}}) > H(\tilde{\boldsymbol{\eta}}_0)$ for all $\tilde{\boldsymbol{\eta}} \in \Omega \setminus \tilde{\boldsymbol{\eta}}_0$; and
3. the minimizer of \tilde{S}_T over \mathcal{K} , denoted by $\hat{\tilde{\boldsymbol{\eta}}}_{\mathcal{K}}$, converges to $\tilde{\boldsymbol{\eta}}_0$ in probability.

Assertion (3) is consistency of the CMLE $\hat{\tilde{\boldsymbol{\eta}}}$. Assertion (1) of the lemma and consistency of $\hat{\tilde{\boldsymbol{\eta}}}$ imply that $\hat{\sigma}^2 = T^{-1}S_T(\hat{\tilde{\boldsymbol{\eta}}}_T) = T^{-1}\tilde{S}_T(\hat{\tilde{\boldsymbol{\eta}}}) \xrightarrow{P} H(\tilde{\boldsymbol{\eta}}_0) = \sigma_0^2$. That is, $\hat{\sigma}^2$ is consistent. The proof of Lemma 1 below completes the proof of Theorem 1.

Proof of Lemma 1. Assertion (1). Let $\mathbf{a} = \boldsymbol{\phi} - \boldsymbol{\phi}_0$, $\mathbf{X}_s = (W_{s-1/T}, \dots, W_{s-p/T})'$, and $d_s = \sum_{i=0}^p \varphi_i [f_{s-i/T}(\tilde{\boldsymbol{\theta}}) - f_{s-i/T}(\tilde{\boldsymbol{\theta}}_0)]$. Write $\tilde{S}_T = \sum_s (\varepsilon_s - \mathbf{a}'\mathbf{X}_s + d_s)^2 = T[A_T + B_T + C_T + 2(\Gamma_T + \Delta_T + \Lambda_T)]$ where $A_T = T^{-1} \sum_s \varepsilon_s^2$, $B_T = \mathbf{a}'[T^{-1} \sum_s \mathbf{X}_s \mathbf{X}_s']\mathbf{a}$, $C_T = \sum_{i=0}^p \sum_{j=0}^p \varphi_i \varphi_j C_{ijT}^*$, $C_{ijT}^* = T^{-1} \sum_s [f_{s-i/T}(\tilde{\boldsymbol{\theta}}) - f_{s-i/T}(\tilde{\boldsymbol{\theta}}_0)][f_{s-j/T}(\tilde{\boldsymbol{\theta}}) - f_{s-j/T}(\tilde{\boldsymbol{\theta}}_0)]$, $\Gamma_T = -\sum_{i=1}^p (\phi_i - \phi_{0,i})T^{-1} \times \sum_s \varepsilon_s W_{s-i/T}$, $\Delta_T = T^{-1} \sum_s \varepsilon_s d_s$, and $\Lambda_T = -\sum_{i=1}^p (\phi_i - \phi_{0,i})T^{-1} \sum_s W_{s-i/T} d_s$.

By ergodicity, we have $\tilde{\boldsymbol{\eta}}$ -free and a.s. convergence for (i) $A_T \rightarrow \sigma_0^2$, (ii) $T^{-1} \sum_s \mathbf{X}_s \mathbf{X}_s' \rightarrow \mathbb{I}^*$, and (iii) $T^{-1} \sum_s \varepsilon_s W_{s-i/T} \rightarrow 0$ for all i . By (ii), $B_T(\boldsymbol{\phi})$ converges to $\mathbf{a}'\mathbb{I}^*\mathbf{a}$ uniformly on \mathcal{K} . By compactness, $\exists M_{\mathcal{K}} < \infty$ such that $|\varphi_i|$, $|\phi_i - \phi_{0,i}| < M_{\mathcal{K}}$ for all i . This and (iii) imply that

$$\sup_{\mathcal{K}} |\Gamma_T(\boldsymbol{\phi})| \leq \sum_{i=1}^p M_{\mathcal{K}} |T^{-1} \sum_s \varepsilon_s W_{s-i/T}| \xrightarrow{\text{a.s.}} 0.$$

As $T \rightarrow \infty$, we have $C_{ijT}^* \rightarrow g(\tilde{\boldsymbol{\eta}}) \equiv \int_0^1 [f_s(\tilde{\boldsymbol{\eta}}) - f_s(\tilde{\boldsymbol{\eta}}_0)]^2 ds$ uniformly on \mathcal{K} , since f_s is continuously differentiable and \mathcal{K} is compact. Since $\{W_t\}$ is stationary, we have

$$\sum_{i=0}^p \sum_{j=0}^p \varphi_i \varphi_j = \left(\sum_{i=0}^p \varphi_i \right)^2 = \left(1 - \sum_{i=1}^p \phi_i \right)^2 > 0 \quad (25)$$

$$\implies \sup_{\mathcal{K}} \left| \sum_{i=0}^p \sum_{j=0}^p \varphi_i \varphi_j C_T^*(\tilde{\boldsymbol{\eta}}) - \left(1 - \sum_{i=1}^p \phi_i \right)^2 g(\tilde{\boldsymbol{\eta}}) \right| \leq \sum_{i=0}^p \sum_{j=0}^p M_{\mathcal{K}}^2 [\sup_{\mathcal{K}} |C_{ijT}^*(\tilde{\boldsymbol{\eta}}) - g(\tilde{\boldsymbol{\eta}})|] \rightarrow 0.$$

Hence, C_T converges to $(1 - \sum_{i=1}^p \phi_i)^2 g$ uniformly on \mathcal{K} .

Finally, by compactness of \mathcal{K} and a gridding argument such as in Chiu (2002), one can show that $\sup_{\mathcal{K}} |\Delta_T(\tilde{\boldsymbol{\eta}})|$, $\sup_{\mathcal{K}} |T^{-1} \sum_s W_{s-i/T} d_s(\tilde{\boldsymbol{\eta}})| \xrightarrow{P} 0$. Hence,

$$\sup_{\mathcal{K}} |\Lambda_T(\tilde{\boldsymbol{\eta}})| \leq \sum_{i=1}^p M_{\mathcal{K}} (\sup_{\mathcal{K}} |T^{-1} \sum_s W_{s-i/T} d_s(\tilde{\boldsymbol{\eta}})|) \xrightarrow{P} 0.$$

Uniform convergence of $T^{-1} \tilde{S}_T$ to H follows.

Assertion (2). Continuity of H is elementary. The rest of the assertion is also straightforward, since (i) $H(\tilde{\boldsymbol{\eta}}_0) = \sigma_0^2$ and (ii) for $\tilde{\boldsymbol{\eta}} \neq \tilde{\boldsymbol{\eta}}_0$, positive definiteness of \mathbb{I}^* and (25) imply $H(\tilde{\boldsymbol{\eta}}) > \sigma_0^2$.

Assertion (3). This is a standard consequence of Assertions (1) and (2). \square

Proof of Theorem 2. Note that the use of T in the statement of the theorem is unnecessary, but it allows the theorem to correspond directly to its counterpart in Chiu, Lockhart & Routledge (2006). For Assertions (1) and (2), take $\sigma_0^2 = 1$ without loss.

Assertion (1). The map $\boldsymbol{\eta} \mapsto \tilde{\boldsymbol{\eta}}$ is linear with a diagonal transformation matrix which we denote by \mathbb{A} . In fact, \mathbb{A} is block diagonal with \mathbb{M}_T as the $\boldsymbol{\theta}$ -block and the identity matrix as the $\boldsymbol{\phi}$ -block. Thus, $\tilde{\mathbb{I}} = \mathbb{A}[\mathbb{I}(\boldsymbol{\eta})]\mathbb{A}$ is the Fisher information for the model parametrized by $\tilde{\boldsymbol{\eta}}$. Since $\tilde{\mathbb{I}}$ is positive definite (pd) if and only if \mathbb{I} is pd, it suffices to consider $\tilde{\mathbb{I}}^{\boldsymbol{\phi}} = \mathbb{I}^{\boldsymbol{\phi}}$ and $\tilde{\mathbb{I}}^{\boldsymbol{\theta}}$ separately.

For the former, recall that \mathbb{I}^* is pd and T -free. Since $\mathbb{I}^{\boldsymbol{\phi}}(\boldsymbol{\eta}_0) = (T+1-p)\mathbb{I}^*$, we have that $\mathbb{I}^{\boldsymbol{\phi}}(\boldsymbol{\eta}_0)$ is pd for all T and $T^{-1}\mathbb{I}^{\boldsymbol{\phi}}(\boldsymbol{\eta}_0) \rightarrow \mathbb{I}^*$. For $\tilde{\mathbb{I}}^{\boldsymbol{\theta}}$, some algebra yields $\tilde{\mathbb{I}}^{\boldsymbol{\theta}}(\tilde{\boldsymbol{\eta}}_0) = \sum_0^p \sum_0^p \varphi_{0,i} \varphi_{0,j} \times \sum_s [\tilde{\boldsymbol{\nabla}}^{\boldsymbol{\theta}} f_{s-i/T}(\tilde{\boldsymbol{\theta}}_0)] [\tilde{\boldsymbol{\nabla}}^{\boldsymbol{\theta}} f_{s-j/T}(\tilde{\boldsymbol{\theta}}_0)]'$, where the superscript for $\tilde{\boldsymbol{\nabla}}$ denotes the corresponding vector elements. By (16), $T^{-1}\tilde{\mathbb{I}}^{\boldsymbol{\theta}}(\tilde{\boldsymbol{\eta}}_0) \rightarrow \tilde{\mathbb{I}}_{\infty}^{\boldsymbol{\theta}}$ whose (j, k) th entry is $(1 - \sum_1^p \phi_{0,i})^2 \int_0^1 [\tilde{\boldsymbol{\nabla}}_{\tilde{\theta}_j} f_s(\tilde{\boldsymbol{\theta}}_0)] [\tilde{\boldsymbol{\nabla}}_{\tilde{\theta}_k} f_s(\tilde{\boldsymbol{\theta}}_0)] ds$. The arguments surrounding identifiability in Chiu, Lockhart & Routledge (2006) imply that $\boldsymbol{v}' \tilde{\mathbb{I}}_{\infty}^{\boldsymbol{\theta}} \boldsymbol{v} > 0$ for all non-zero $\boldsymbol{v} \in \mathbb{R}^5$. Hence, $T^{-1}\tilde{\mathbb{I}}^{\boldsymbol{\theta}}(\tilde{\boldsymbol{\eta}}_0)$ is asymptotically pd. The rest of Assertion (1) follows from a uniform convergence argument such as that for Theorem 1.

Assertion (2). On the Ψ scale the score function is $\tilde{\boldsymbol{U}} = \mathbb{A}\boldsymbol{U}$, although \mathbb{A} cancels out of the quantity in Assertion (2). The usual tactic here is to establish unbiasedness and asymptotic normality of the score $\tilde{\boldsymbol{U}}(\tilde{\boldsymbol{\eta}}_0)$. To this end, we apply the following lemma.

LEMMA 2. *Under the conditions of Theorems 1 and 2,*

$$\frac{T^{-1/2} \boldsymbol{v}' [\tilde{\boldsymbol{U}}(\tilde{\boldsymbol{\eta}}_0)]}{\sqrt{T^{-1} \boldsymbol{v}' [\tilde{\mathbb{I}}(\tilde{\boldsymbol{\eta}}_0)] \boldsymbol{v}}} \xrightarrow{\mathcal{L}} N(0, 1) \quad \text{as } T \rightarrow \infty \text{ for all non-zero } \boldsymbol{v} \in \mathbb{R}^{5+p}.$$

The proof of Lemma 2 appears at the end of this appendix. The rest of the proof for Assertion (2) hinges on (i) a one-term Taylor-type expansion of $\tilde{\boldsymbol{U}}(\hat{\boldsymbol{\eta}})$ about $\tilde{\boldsymbol{\eta}}_0$, involving $\tilde{\boldsymbol{V}}^+(\hat{\boldsymbol{\eta}})$; (ii) the uniqueness in probability of $\hat{\boldsymbol{\eta}}$ as the minimizer of $\tilde{S}_T(\hat{\boldsymbol{\eta}})$ over a neighbourhood of $\tilde{\boldsymbol{\eta}}_0$; and (iii) the uniform closeness in probability of $\tilde{\boldsymbol{V}}^+(\hat{\boldsymbol{\eta}})$ to $\tilde{\mathbb{I}}(\tilde{\boldsymbol{\eta}}_0)$ on this neighbourhood. As in Chiu, Lockhart & Routledge (2006), the directional Hessian for (i) plays the role of a well-defined gradient of the score function in a standard Taylor expansion in a regular estimation problem. Here, our problem

of estimating $\tilde{\boldsymbol{\eta}}_0$ would have been regular, if not for the non-differentiability of $\tilde{\boldsymbol{U}}$ with respect to $\tilde{\boldsymbol{\theta}}$ along isolated hyper-rays. Since the estimation of $\boldsymbol{\phi}$ yields no additional irregularity, the argument for (i) is essentially identical to that for Lemma 2 from Chiu, Lockhart & Routledge (2006), except for an extended parameter space due to $\boldsymbol{\phi}$. Similarly, the argument for (ii) is a direct extension of Step 1 in the proof of their Theorem 2, Assertion 2, coupled with our Theorem 1 above. For (iii), the assertion can be verified by some algebra. Assemble and rescale from $\hat{\boldsymbol{\eta}}$ to $\tilde{\boldsymbol{\eta}}$ to conclude asymptotic normality of $[\mathbb{I}(\boldsymbol{\eta}_0)]^{1/2}(\hat{\boldsymbol{\eta}} - \boldsymbol{\eta}_0)$. That of $\sqrt{T}[T^{-1}\mathbb{I}(\tilde{\boldsymbol{\eta}})]^{1/2}(\tilde{\boldsymbol{\eta}} - \boldsymbol{\eta}_0)$ follows from Theorem 1.

Assertion (3). This is standard likelihood theory from the consistency of $\hat{\sigma}^2$ and the properties of $\tilde{\mathbb{V}}^+(\tilde{\boldsymbol{\eta}})$ used to prove Assertion (2).

Proof of Lemma 2. We apply the martingale central limit theorem (MCLT) (Theorem 3.2 and Corollary 3.1 of Hall & Heyde 1980). Let $X_{t,T} \equiv T^{-1/2}\mathbf{v}'\tilde{\boldsymbol{U}}_t(\tilde{\boldsymbol{\eta}}_0)$ and $M_{t,T} = \sum_{u=p}^t X_{u,T}$. We claim that $M_{t,T}$ is a martingale array for each T relative to the sigma fields \mathcal{F}_t generated by $\{W_0, \dots, W_{p-1}, \varepsilon_p, \varepsilon_{p+1}, \dots, \varepsilon_t\}$ (or equivalently by $\{W_0, \dots, W_t\}$). The lemma would then follow if we verify the two conditions of the MCLT.

We write $\mathbf{V}_t = (W_{t-1}, \dots, W_{t-p})'$ and let $\mathbf{v}_1 = (v_1, \dots, v_5)$ and $\mathbf{v}_2 = (v_6, \dots, v_{5+p})$ so that $\mathbf{v}' = (\mathbf{v}_1, \mathbf{v}_2)$. Note from (17)–(19) that $X_{t,T} = T^{-1/2}\varepsilon_t[\mathbf{v}_1\mathbf{a}_t + \mathbf{v}_2\mathbf{V}_t]$ where

$$\mathbf{a}_t = \mathbb{M}_T \sum_{j=0}^p \varphi_j \nabla^{\boldsymbol{\theta}} f(t-i; \boldsymbol{\theta}) = \sum_{j=0}^p \varphi_j \tilde{\nabla}^{\tilde{\boldsymbol{\theta}}} f(s-i/T; \tilde{\boldsymbol{\theta}}).$$

Since each W_{t-j} is \mathcal{F}_t -measurable and ε_t has mean 0, it is easy to check that $M_{t,T}$ is a martingale.

The first condition of the MCLT which must be checked is that for each $\delta > 0$, we have

$$\sum_{t=p}^T E [X_{t,T}^2 \mathbf{1}\{|X_{t,T}| > \delta\} | \mathcal{F}_{t-1}] = \frac{1}{T} \sum_{t=p}^T (\mathbf{v}_1\mathbf{a}_t + \mathbf{v}_2\mathbf{V}_t)^2 E [\varepsilon_t^2 \mathbf{1}\{|X_{t,T}| > \delta\} | \mathcal{F}_{t-1}] \xrightarrow{P} 0.$$

To this end, we let $Z_t = \delta\sqrt{T}(|\mathbf{v}_1\mathbf{a}_t + \mathbf{v}_2\mathbf{V}_t|)^{-1}$. For $\delta > 0$, we have

$$E [\varepsilon_t^2 \mathbf{1}\{|X_{t,T}| > \delta\} | \mathcal{F}_{t-1}] = E [\varepsilon_t^2 \mathbf{1}\{|\varepsilon_t| > Z_t\} | \mathcal{F}_{t-1}] \leq \frac{E [\varepsilon_t^4 \mathbf{1}\{|\varepsilon_t| > Z_t\} | \mathcal{F}_{t-1}]}{Z_t^2} \leq \frac{E(\varepsilon_t^4)}{Z_t^2}.$$

Assembling, and using $(a+b)^4 \leq 8(a^4 + b^4)$, we have

$$\sum_{t=p}^T E [X_{t,T}^2 \mathbf{1}\{|X_{t,T}| > \delta\} | \mathcal{F}_{t-1}] \leq \frac{8E(\varepsilon_t^4)}{\delta^2 T} \cdot \frac{1}{T} \sum_{t=p}^T \{(\mathbf{v}_1\mathbf{a}_t)^4 + (\mathbf{v}_2\mathbf{V}_t)^4\}. \quad (26)$$

The terms involving $(\mathbf{v}_2\mathbf{V}_t)^4$ go to 0 by ergodicity of $\{W_t\}$. Components of \mathbf{a}_t may be computed from (17)–(19) by replacing each occurrence of β_2 , γ , and τ by $\tilde{\beta}_2$, $\tilde{\gamma}$, and $\tilde{\tau}$, respectively, and by replacing $t-i$ by $s-i/T$. It follows that there is a constant C not depending on t such that $|\mathbf{v}_1\mathbf{a}_t| \leq C$; the right hand side of (26) therefore converges to 0.

In the proof of Theorem 2 we showed that $T^{-1}\tilde{\mathbb{I}}$ converges to a positive definite matrix, $\tilde{\mathbb{I}}_\infty$. It therefore remains only to establish the second condition of the MCLT which becomes

$$\sum_{t=p}^T E[X_{t,T}^2 | \mathcal{F}_{t-1}] = \frac{1}{T} \sum_{t=p}^T \sigma^2(\mathbf{v}_1\mathbf{a}_t + \mathbf{v}_2\mathbf{V}_t)^2 \rightarrow \mathbf{v}'\tilde{\mathbb{I}}_\infty\mathbf{v}. \quad (27)$$

Convergence of $T^{-1}\sum_t(\mathbf{v}_1\mathbf{a}_t)^2$ to $\mathbf{v}_1\tilde{\mathbb{I}}_\infty\mathbf{v}_1'$ follows from the application of (16). Convergence of the cross product term $T^{-1}\sum_t(\mathbf{v}_1\mathbf{a}_t)(\mathbf{v}_2\mathbf{V}_t)$ to 0 is a moment calculation, using the exponential decay of the autocovariance function of the $\{W_t\}$ and the previous bound on $|\mathbf{a}_t|$. Finally, convergence

of $T^{-1} \sum_t (\mathbf{v}_1 \mathbf{a}_t)^2$ to $\mathbf{v}_2 \tilde{\mathbb{I}}_\infty^\phi \mathbf{v}_2' = \mathbf{v}_2 \tilde{\mathbb{I}}^* \mathbf{v}_2'$ uses ergodicity and the variance-covariance of $\{W_t\}$. This establishes (27) and completes the proof of Lemma 2. \square

ACKNOWLEDGEMENTS

This research has been funded by the Natural Sciences and Engineering Research Council of Canada through Discovery Grants to G.S. Chiu and R.A. Lockhart. We thank the Editor, Associate Editor, and referees for their valuable suggestions; Drs. James Elkins and Thayne Thompson, Global Monitoring Division, National Oceanic and Atmospheric Administration Earth System Research Laboratory, for clarifications of their data used in this article; and Prof. Richard Routledge, Simon Fraser University, for suggesting the use of these data.

REFERENCES

- P. J. Brockwell & R. A. Davis (2002). *Introduction to Time Series and Forecasting*, Second Edition. Springer, New York.
- P. J. Brockwell & R. A. Davis (2006). *Time Series: Theory and Methods*, Second Edition. Springer, New York.
- W.-K. Cheang & G. C. Reinsel (2000). Bias reduction of autoregressive estimates in time series regression model through restricted maximum likelihood. *Journal of the American Statistical Association*, 95, 1173–1184.
- G. S. Chiu (2002). *Bent-Cable Regression for Assessing Abruptness of Change*. Unpublished doctoral dissertation, Simon Fraser University, Burnaby, Canada.
- G. Chiu, R. Lockhart & R. Routledge (2006). Bent-cable regression theory and applications. *Journal of the American Statistical Association*, 101, 542–553.
- S. B. Fotopoulos, V. K. Jandhyala & L. Tan (2009). Asymptotic study of the change-point mle in multivariate Gaussian families under contiguous alternatives. *Journal of Statistical Planning and Inference*, 139, 1190–1202.
- P. Hall & C. C. Heyde (1980). *Martingale Theory and Its Applications*. Academic Press, New York.
- M. Hallin, M. Taniguchi, A. Serroukh & K. Choy (1999). Local asymptotic normality for regression models with long-memory disturbance. *The Annals of Statistics*, 27, 2054–2080.
- F. E. Harrell (2001). *Regression Modeling Strategies: with Applications to Linear Models, Logistic Regression, and Survival Analysis*. Springer, New York.
- S. A. Khan, G. Chiu & J. A. Dubin (2009). Atmospheric concentration of chlorofluorocarbons: addressing the global concern with the longitudinal bent-cable model. *Chance*, 22(3), 8–17.
- S. A. Khan, G. S. Chiu & J. A. Dubin (submitted). Flexible bent-cable models for mixture longitudinal data.
- F. H. C. Marriott & J. A. Pope (1954). Bias in the estimation of autocorrelations. *Biometrika*, 41, 390–402.
- M. Mudelsee (2001). Note on the bias in the estimation of the serial correlation coefficient of AR(1) processes. *Statistical Papers*, 42, 517–527.
- P. S. Reynolds & G. S. Chiu (submitted). Early warning of hemorrhage-induced hypothermia: estimating the thermoregulatory transition using bent cable regression.
- P. X.-K. Song, Y. Fan & J. D. Kalbfleisch (2005). Maximization by parts in likelihood inference. *Journal of the American Statistical Association*, 100, 1145–1167.
- H. J. Thiébaux and F. W. Zwiers (1984). The interpretation and estimation of effective sample size. *Journal of Climate and Applied Meteorology*, 23, 800–811.
- J. S. White (1961). Asymptotic expansions for the mean and variance of the serial correlation coefficient. *Biometrika*, 48, 85–94.

Received 27 February 2009
Accepted 2 March 2010

Grace S. CHIU: grace.chiu@csiro.au
CSIRO Mathematics, Informatics and Statistics
Canberra, ACT
Australia, 2601
and
Department of Statistics and Actuarial Science
University of Waterloo
Waterloo, ON
Canada, N2L 3G1

Richard A. LOCKHART: lockhart@stat.sfu.ca
Department of Statistics and Actuarial Science
Simon Fraser University
Burnaby, BC
Canada, V5A 1S6