Econometric Theory, **0**, 2026, 1–47.
doi:10.1017/S0266466626100449

LOCAL POLYNOMIAL ESTIMATION OF TIME-VARYING PARAMETERS IN NONLINEAR MODELS

DENNIS KRISTENSEN
University College London, United Kingdom

YOUNG JUN LEE
College of Economics and Finance, Hanyang University, Republic of Korea

We develop a novel asymptotic theory for local polynomial extremum estimators of time-varying parameters in a broad class of nonlinear time-series models, including discrete-valued ones. We show that the proposed estimators are asymptotically normally distributed under weak conditions. We also provide a precise characterization of the leading bias term due to smoothing. We demonstrate the usefulness of our general results by establishing primitive conditions for local (quasi-)maximum-likelihood estimators of time-varying models threshold autoregressions, ARCH models and Poisson autoregressions with exogenous covariates, to be normally distributed in large samples and characterize their leading biases. An empirical study of U.S. corporate default counts demonstrates the applicability of the proposed local linear estimator for Poisson autoregressions, shedding new light on the dynamic properties of U.S. corporate defaults.

1. INTRODUCTION

There is ample empirical evidence of time-varying parameters in many econometric models (see, e.g., Ghysels and Hall, 1990; Inoue and Rossi, 2011; Caldara et al., 2012; Christoffersen, Jacobs, and Ornthanalai, 2012; Giacomini and Rossi, 2016). Most studies aiming at accommodating this feature assume a fully parametric model for this time variation; one example of this is structural break models. This has the advantage that the time-varying version of a given model stays parametric and can be estimated using existing methods. The disadvantage is that the researcher runs the risk of choosing a misspecified model for the time variation.

To reduce this risk, methods that treat the problem of time-varying parameters as a structured nonparametric one have been developed: They assume that the sequence of time-varying parameters arise as values of an underlying function which is then estimated nonparametrically; one popular class of estimators that

The authors would like to thank Sokbae Lee and participants at various seminars and conferences for valuable comments and suggestions. Address correspondence to Dennis Kristensen, University College London, United Kingdom, email: d.kristensen@ucl.ac.uk.

© The Author(s), 2026. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (https://creativecommons.org/licenses/by/4.0), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

https://doi.org/10.1017/S0266466626100449 Published online by Cambridge University Press

falls in this category are local estimators which includes the so-called rolling-window estimator. However, the existing literature has mostly focused on local constant (Nadaraya–Watson) kernel estimators of the time-varying parameters (see, e.g., Dahlhaus, Richter, and Wu, 2019; Bardet, Doukhan, and Wintenberger, 2022).

We here propose to estimate the time-varying parameters using local polynomial estimators since these are known to have a number of attractive properties compared to the local constant one (see, e.g., Fan, Heckman, and Wand, 1995). Under very weak restrictions on the model being estimated and the time-series data being used, we develop an asymptotic theory for the estimators. Specifically, we show that they are normally distributed in large samples and provide a complete characterization of the leading variance and bias components.

The class of estimators includes as special cases the local constant estimator and the local linear estimator. We find that the local constant estimator requires stronger regularity conditions to be well-behaved in large samples and will generally suffer from additional biases in the interior of the domain compared to the local linear estimator. These additional biases involve the so-called time derivative process of the stationary approximation to data which is not present in the biases of the local linear one. Moreover, the local linear estimator enjoys the well-known automatic boundary adjustment property: At the beginning and end of the sample, it will perform better than the local constant one. This feature is important since often the main interest is on the values that the time-varying parameters take at the end of the sample.

The two most closely related papers to ours are Dahlhaus et al. (2019) and Bardet et al. (2022) who develop a general theory for local constant estimators of time-varying parameters in Markov models and infinite memory models, respectively. Our general framework encompasses theirs as special cases and we consider a broader class of estimators than they do. Moreover, our proof techniques are different and, when specializing to local constant estimators, allow us to arrive at the same results as they do under weaker conditions: First, our theory imposes minimum requirements on the data-generating process (DGP) with the main assumption being that it is locally stationary. Second, it imposes weaker restrictions on the bandwidth used in the estimation; in particular, we can allow for the bandwidth being chosen using standard bandwidth selection rules, such as cross-validation, which is not the case in Dahlhaus et al. (2019) and Bardet et al. (2022). Second, we characterize the leading bias term of the estimators, which is in contrast to Dahlhaus et al. (2019) and Bardet et al. (2022). To demonstrate these attractive features of our general theory, we apply it to a class of Markov models with exogenous covariates.

Another important feature of our theory is that it also applies to discrete-valued time-series models, such as Poisson autoregressions. Such models are not covered by the theories of Dahlhaus et al. (2019) and Bardet et al. (2022) since these require the DGP to be smooth w.r.t. time which is violated when the time series is discrete-valued. In contrast, our theory for local linear estimators impose very

https://doi.org/10.1017/S0266466626100449 Published online by Cambridge University Press

weak smoothness conditions on the DGP due to our new proof techniques and so applies to discrete-valued time-series models without any modifications. For the local constant estimator, we combine the ideas and techniques of Truquet (2019) with our main result to obtain a complete analysis of this estimator.

We also contribute to the literature on asymptotic analysis of local polynomial estimators of varying-coefficient models by extending existing results (see, e.g., Fan et al., 1995; Loader, 2006) to cover situations where the objective functions are non-concave. This proves to be a non-trivial extension, but at the same time an important one since the log-likelihood functions of many non-linear models are non-concave.

As an empirical application, we revisit the empirical study of Agosto et al. (2016), where a Poisson autoregressive model with additional covariates was used to model and analyze U.S. defaults. Using the proposed methodology, we find substantial time variation in the model parameters that the original study was unable to capture. In particular, we find that the ability of macroeconomic and financial variables to predict defaults have varied substantially over time. A battery of informal tests of the time-varying model against the time-invariant version finds strong support for the former.

The remainder of the article is organized as follows. Framework and estimators are introduced in Section 2. Section 3 presents the asymptotic theory of the estimators. Section 4 provides examples of the theory when applied to particular models. We present the results of two simulation studies and the empirical application in Sections 5 and 6, respectively. We conclude in Section 7, where we discuss how our results potentially could be applied to time-varying infinite-memory models. All proofs have been relegated to the Appendix.

2. FRAMEWORK

We are given n observations, $Z_{n,t} \in (\mathcal{Z}, \|\cdot\|)$, $t = 1, \ldots, n$, where $(\mathcal{Z}, \|\cdot\|)$ is a Banach space, from a time-series model characterized by a finite-dimensional vector of unknown parameters $\theta \in \Theta \subset \mathbb{R}^{d_\theta}$ to be estimated. In most applications, $Z_{n,t}$ will also be finite-dimensional but our theory allows for, for example, functional data as well. We take as given an objective function $\ell_{n,t}(\theta) = \ell\left(\mathcal{Z}_{n,t}; \theta\right) \in \mathbb{R}$, where $\mathcal{Z}_{n,t} = \left(Z_{n,t}, \ldots Z_{n,0}, Z_{n,-1}, \ldots\right)$. Since we do not observe the process before $t = 1$, we here initialize the process at deterministic values chosen by us, $Z_{n,-t} = z_{-t}$, $t \geq 1$. Under regularity conditions stated below, the effect of the initial values will vanish asymptotically.

The objective function is assumed to identify the data-generating parameter as its maximizer, $\theta = \arg\max_{\theta' \in \Theta} \mathbb{E}\left[\ell_{n,t}\left(\theta'\right)\right]$, if θ indeed was time-invariant, and $\ell_{n,t}\left(\theta'\right)$ was stationary and ergodic. In this case, the natural estimator is to replace population expectations by sample ones and estimate θ by $\hat{\theta} = \arg\max_{\theta' \in \Theta} \frac{1}{n} \sum_{t=1}^{n} \ell_{n,t}\left(\theta'\right)$.

Suppose now that in fact θ is varying over time so that $\mathcal{Z}_{n,t}$ is generated by $\theta_{n,t} = \theta(t/n)$, $t = 1, \ldots, n$, where $\theta : [0,1] \mapsto \Theta$ is an unknown function that

https://doi.org/10.1017/S0266466626100449 Published online by Cambridge University Press

characterizes the time variation in the parameters.[1] At the same time, the objective function is still assumed to identify the parameter in the sense that $\theta(t/n) = \arg\max_{\theta' \in \Theta} \mathbb{E}\left[\ell_{n,t}(\theta')\right]$. We then propose to estimate $\theta(u)$ at any given value $u \in [0,1]$ using local polynomial estimators: First, for t/n in a neighborhood of u, we approximate $\theta(t/n)$ by the following polynomial of order $m \geq 0$:

$$\theta^*_{u,\beta}(t/n) := \beta_1 + \beta_2(t/n-u) + \cdots + \beta_{m+1}(t/n-u)^m/m! = D_m(t/n-u)\beta, \quad \textbf{(1)}$$

where $\beta = \left(\beta_1', \ldots, \beta_{m+1}'\right)' \in \mathbb{R}^{(m+1)d_\theta}$ with $\beta_{i+1} = \theta^{(i)}(u) = \partial^i \theta(u)/\partial u^i \in \mathbb{R}^{d_\theta}$ and $D_m(u) = \left(1, u, u^2/2 \ldots, u^m/m!\right) \otimes I_{d_\theta} \in \mathbb{R}^{d_\theta \times (m+1)d_\theta}$.

Next, to control the approximation error, $\theta(t/n) - \theta^*_{u,\beta}(t/n)$, we introduce a kernel weighted version of the "global" objective function evaluated at the polynomial approximation:

$$L_n(\beta|u) = \frac{1}{n}\sum_{t=1}^{n} K_b(t/n-u)\,\ell_{n,t}\left(\theta^*_{u,\beta}(t/n)\right),$$

where $K_b(\cdot) = K(\cdot/b)/b$, $K : \mathbb{R} \mapsto \mathbb{R}$ is a kernel function, and $b = b_n > 0$ is a bandwidth. The kernel weights ensure that when $t/n - u$ is "large," the corresponding observations are down weighted in the estimation, thereby controlling for the aforementioned approximation error. We then estimate the polynomial coefficients by

$$\hat{\beta}(u) = \arg\max_{\beta \in \mathcal{B}} L_n(\beta|u), \quad \textbf{(2)}$$

where

$$\mathcal{B} = \left\{\beta \in \mathbb{R}^{(m+1)d_\theta} : \theta^*_{u,\beta}(v) \in \Theta \text{ for } v \in [0,1]\right\}.$$

The estimated β coefficients are used as estimates of $\theta(u)$ and its first m derivatives, $\hat{\theta}^{(i)}(u) = \hat{\beta}_{i+1}(u)$, $i = 0, \ldots, m$. When $m = 0$, we recover the standard local-constant estimator. The above class of estimators is similar to the ones considered in Fan et al. (1995) for so-called varying-coefficient models, except that we consider time-series models with the "regressor" that we smooth over being normalized time, t/n, and do not restrict $\theta \mapsto \ell_{n,t}(\theta)$ to be convex.

The choice of the order of the polynomial, m, should reflect the degree of smoothness that we are willing to assume $u \mapsto \theta(u)$ has. If $\theta(u)$ is m times differentiable, then we should use this m in the estimation for optimal control of the bias in the nonparametric estimation. On the other hand, increasing the order of the polynomial tends to increase the variability of the resulting estimator, since more local parameters are introduced in the estimation. For a further discussion of this issue, we refer the reader to Section 3.3 of Fan and Gijbels (2018).

Our framework includes Markov processes and stochastic processes with infinite memory (see, e.g., Doukhan and Wintenberger, 2008; Bardet et al., 2022) as

[1] Data points now depend on sample size n through $\theta(t/n)$, which is why we write $Z_{n,t}$ instead of of Z_t.

https://doi.org/10.1017/S0266466626100449 Published online by Cambridge University Press

special cases. In Section 4, we apply our general theory to the following class of Markov models with exogenous covariates:

$$Y_{n,t} = G\left(Y_{n,t-1}, \ldots, Y_{n,t-q}, X_{n,t-1}, \varepsilon_t; \theta\left(t/n\right)\right), \tag{3}$$

where $G : \mathcal{Y}^q \times \mathcal{X} \times \mathcal{E} \times \Theta$ is a known function, $X_{n,t-1}$ is a Markov process of exogenous covariates, and ε_t is a sequence of errors. For a given specification of G and the distribution of ε_t, we can then derive the corresponding log-likelihood for the model with time-invariant parameters, $\ell_{n,t}(\theta) = \ell\left(Z_{n,t}, X_{n,t-1}; \theta\right)$, where $Z_{n,t} = \left(Y_{n,t}, X_{n,t}\right)$. Below, we provide three examples of models that our theory applies to.

Example 1. Time-varying threshold autoregressive with exogenous covariates (tv-TAR-X) model with two regimes:

$$Y_{n,t} = \omega + \sum_{i=1}^{q} \alpha_{1,i}\left(t/n\right) Y_{n,t-i}^{+} + \sum_{i=1}^{q} \alpha_{2,i}\left(t/n\right) Y_{n,t-1}^{-} + \gamma\left(t/n\right)' X_{n,t-1} + \varepsilon_t, \tag{4}$$

where $y^+ := \max\{y, 0\}$ and $y^- := \min\{y, 0\}$. Here, $X_{n,t-1}$ contains additional predictors, and ε_t is i.i.d. with $\mathbb{E}[\varepsilon_t] = 0$ and $\mathbb{E}\left[\varepsilon_t^2\right] < \infty$. A natural estimator of $\theta = \left(\omega, \alpha_1', \alpha_2', \gamma'\right)'$ is the least-squares one so that $\ell_{n,t}(\theta) = -\left(Y_{n,t} - \omega - \sum_{i=1}^{q} \alpha_{1,i} Y_{n,t-1}^{+} - \sum_{i=1}^{q} \alpha_{2,i} Y_{n,t-1}^{-} - \gamma' X_{n,t-1}^{-}\right)^2$.

Example 2. Time-varying ARCH model with covariates (tv-ARCH-X):

$$Y_{n,t} = \lambda_{n,t}\left(\theta\left(t/n\right)\right)\varepsilon_t^2, \quad \lambda_{n,t}(\theta) = \omega + \sum_{i=1}^{q} \alpha_i Y_{n,t-1} + \gamma' X_{n,t-1}, \tag{5}$$

where $\theta = \left(\omega, \alpha', \gamma'\right)'$, ε_t is i.i.d. with $\mathbb{E}\left[\varepsilon_t^2\right] = 1$, and $X_{n,t-1}$ contains additional predictors. The Gaussian log-likelihood function takes the form $\ell_{n,t}(\theta) = -Y_{n,t}/\lambda_{n,t}(\theta) - \log\left(\lambda_{n,t}(\theta)\right)$, where $\theta = \left(\omega, \alpha', \gamma'\right)'$.

Example 3. Time-varying Poisson autoregression with exogenous covariates (tv-PARX):

$$Y_{n,t} | \mathcal{F}_{n,t-1} \sim \text{Poisson}\left(\lambda_{n,t}\left(\theta\left(t/n\right)\right)\right), \quad \lambda_{n,t}(\theta) = \omega + \sum_{i=1}^{q} \alpha_i Y_{n,t-1} + \gamma' X_{n,t-1}, \tag{6}$$

where $\text{Poisson}(\lambda)$ denotes the Poisson distribution with intensity parameter λ, $\theta = \left(\omega, \alpha', \gamma'\right)'$, and $X_{n,t-1}$ contains additional predictors. The log-likelihood function takes the form $\ell_{n,t}(\theta) = Y_{n,t} \log\left(\lambda_{n,t}(\theta)\right) - \lambda_{n,t}(\theta)$.

3. ASYMPTOTIC THEORY

We here provide an asymptotic theory for $\hat{\beta}$. One complication of this analysis is that the components of $\hat{\beta}$ converge with different rates. We follow the existing literature and handle this issue by introducing a re-scaled version of $\hat{\beta}$ (see,

https://doi.org/10.1017/S0266466626100449 Published online by Cambridge University Press

e.g., Han and Kristensen, 2014, for a similar approach): Define $\hat{\alpha} = U_n \hat{\beta} = (\hat{\theta}(u)', b\hat{\theta}^{(1)}(u)', \ldots, b^m \hat{\theta}^{(m)}(u)')'$, where

$$U_n = diag\{1, b, \ldots, b^m\} \otimes I_{d_\theta} \in \mathbb{R}^{(m+1)d_\theta \times (m+1)d_\theta}$$

is a weighting matrix containing their relative convergence rates, Given that U_n is non-singular, the estimation problem (2) is equivalent to solving

$$\hat{\alpha} = \arg\max_{\alpha \in \mathcal{A}} Q_n(\alpha|u), \quad Q_n(\alpha|u) = \frac{1}{n} \sum_{t=1}^{n} K_b(t/n - u)\, \ell_{n,t}\left(D_{m,b}(t/n - u)\, \alpha\right), \tag{7}$$

where $D_{m,b}(u) = D_m(u/b)$ and

$$\mathcal{A} = \left\{\alpha \in \mathbb{R}^{(m+1)d_\theta} : D_m(v)\, \alpha \in \Theta,\ \forall v \in \mathcal{K}\right\} \tag{8}$$

with $\mathcal{K}$ denoting the support of K. We will then analyze the properties of $\hat{\alpha}$.

Due to the time-varying parameters, $Z_{n,t}$ will generally be non-stationary. To develop an asymptotic theory that allows for this feature, we will rely on the concept of local stationarity as introduced by Dahlhaus (1997); see also Dahlhaus and Subba Rao (2006) and Dahlhaus et al. (2019). We first generalize this concept to sequences of random functions.

DEFINITION 1. *A triangular family of random sequences $W_{n,t}(\theta)$, $\theta \in \Theta$, $t = 1, 2, \ldots, n$ and $n \geq 1$, is uniformly locally stationary on Θ (ULS (p, r, Θ)) for some $p, r > 0$ if there exists a family of processes $W_t^*(\theta|u)$, $u \in [0,1]$, such that: (i) the process $\{W_t^*(\theta|u)\}$ is stationary and ergodic for all $(\theta, u) \in \Theta \times [0,1]$ and (ii) for some $C < \infty$ and $\rho < 1$,*

$$\mathbb{E}\left[\sup_{\theta \in \Theta} \left\| W_{n,t}(\theta) - W_t^*(\theta|u) \right\|^p\right]^{1/p} \leq C\left(\left|\frac{t}{n} - u\right|^r + \frac{1}{n^r} + \rho^t\right).$$

If $W_{n,t}(\theta) = W_{n,t}$ does not depend on any parameters, we write LS (p, r). The above condition states that $W_{n,t}(\theta)$ may be non-stationary, but it is locally in time well-approximated by a stationary version $W_t^*(\theta|u)$. Compared to existing definitions of local stationarity, we allow for an additional term ρ^t to appear in the approximation error. This is needed in order to allow for the initial value of $W_{n,t}(\theta)$ to be chosen arbitrarily. In contrast, by not including ρ^t in their definitions, most of the existing literature implicitly assumes that $W_{n,t}(\theta)$ has been initialized at $W_{n,0}(\theta) = W_0^*(\theta|u)$. When used in the analysis of local estimators, this latter definition implicitly requires that the DGP changes as the researcher varies u in the local log-likelihood which is a rather peculiar assumption. In contrast, the above definition allows for $W_{n,0}(\theta)$ to be initialized at a given fixed value—as long as the impact of this dies out with rate ρ.

For an example of how the additional error term appears in autoregressive models, we refer the reader to the proof of Lemma 3 in Section 4 which allows for arbitrary initialization of the DGP. The additional error term due to different initializations is here assumed to decay geometrically and so our definition rules

https://doi.org/10.1017/S0266466626100449 Published online by Cambridge University Press

out long-memory-type processes. This is mostly for simplicity and we expect that most of our results can be generalized to allow for slower decay rates.

We will then require that $\ell_{n,t}(\theta)$ is ULS (p, r, Θ) with stationary approximation $\ell_t^*(\theta|u) = \ell\left(\mathcal{Z}_t^*(u), \theta\right)$, where $\mathcal{Z}_t^*(u) = \left(Z_t^*(u), Z_{t-1}^*(u), \dots\right)$ is the stationary solution to the model being estimated when $\theta_{n,t} = \theta(u)$ is constant. To illustrate, consider again (3). Under regularity conditions on G and ε_t (see Section 4 for details), the stationary solution will in this case take the form

$$Y_t^*(u) = G\left(Y_{t-1}^*(u), \dots, Y_{t-q}^*(u), X_{t-1}^*(u), \varepsilon_t; \theta(u)\right),$$

where we impose the high-level condition that the exogenous covariates are locally stationary. If the DGP is locally stationary, it follows under great generality that the likelihood and its derivatives are also locally stationary (cf. Section 4).

The next step in our proof is to establish a uniform law of large numbers (ULLN) for the stationary approximation of $Q_n(\alpha|u)$, $Q_n^*(\alpha|u) = \frac{1}{n}\sum_{t=1}^n K_b(t/n - u)\, \ell_t^*(D_b(t/n-u)\alpha|u)$. A sufficient condition for a ULLN to hold is that $\theta \mapsto \ell_{n,t}^*(\theta|u)$ is *L_p-continuous*.

DEFINITION 2. *A stationary process $W_t^*(\theta|u)$ is said to be L_p-continuous w.r.t. θ if $\mathbb{E}\left[\left\|W_t^*(\theta|u)\right\|^p\right] < \infty$ for all $\theta \in \Theta$ and*

$$\forall \epsilon > 0 \exists \delta > 0 : \mathbb{E}\left[\sup_{\theta' : \|\theta - \theta'\| < \delta} \left\|W_t^*(\theta|u) - W_t^*\left(\theta'|u\right)\right\|^p\right]^{1/p} < \epsilon.$$

Imposing L_p-continuity w.r.t. θ is weaker than almost surely continuity: If $\theta \mapsto W_t^*(\theta|u)$ is almost surely continuous with $\mathbb{E}\left[\sup_{\theta\in\Theta}\left\|W_t^*(\theta|u)\right\|^p\right] < \infty$, the process is also L_p-continuous since $DW_t(\delta) = \sup_{\|\theta-\theta'\|\le\delta}\left\|W_t^*(\theta|u) - W_t^*\left(\theta'|u\right)\right\|^p$, $\delta > 0$, will then satisfy $\lim_{\delta\to 0} DW_t(\delta) = 0$ almost surely and so, by dominated convergence, $\lim_{\delta\to 0}\mathbb{E}[DW_t(\delta)] = 0$. It is easily verified that L_p-continuity w.r.t. θ implies stochastic equicontinuity of $Q_n^*(\alpha|u)$ and so a ULLN holds (cf. Lemma A(i) in Appendix A.1).

We are now ready to state the regularity conditions used to show consistency.

Assumption 1. (i) $K(\cdot) \ge 0$ has compact support $\mathcal{K}$ and $\int_{-\infty}^{+\infty} K(v)\, dv = 1$; (ii) for some $\Lambda < \infty$, $|K(v) - K(\tilde{v})| \le \Lambda |v - \tilde{v}|$, $v, \tilde{v} \in \mathbb{R}$; and (iii) $v \mapsto \theta(v)$ is continuous at u.

Assumption 2. (i) Θ is compact and the true value $\theta(u) \in \Theta$; (ii) $\theta \mapsto \ell_t^*(\theta|u)$ is L_1-continuous; and (iii) $\theta \mapsto \mathbb{E}\left[\ell_t^*(\theta|u)\right]$ has a unique maximum at $\theta(u) \in \Theta$.

Assumption 3. $\ell_{n,t}(\theta)$ is ULS $(1, r, \Theta)$ for some $r > 0$ with stationary approximation $\ell_t^*(\theta|u)$.

Assumption 1(i) imposes stronger than usual assumptions on K and excludes, among others, the Gaussian kernel and higher-order kernels. It includes, on the other hand, the Epanechnikov and the triangular kernel. The restriction that

https://doi.org/10.1017/S0266466626100449 Published online by Cambridge University Press

$K(\cdot) \geq 0$ is used to ensure the identification of the parameters when $m > 0$; without this, identification is not necessarily guaranteed; see below for further discussion. For the analysis of the local constant estimator ($m = 0$), all subsequent results will go through with K having full support and taking negative values.

The compact support assumption greatly simplifies our analysis of local polynomial estimation of non-concave models: In order to establish uniform convergence of the likelihood, we require Θ to be compact as is standard in the literature. But under this restriction, it is easily checked that $D_{m,b}(v)\alpha \notin \Theta$ as $b \to 0$ for any given $\alpha = (\alpha_1, \ldots, \alpha_{m+1})$ with $\alpha_i \neq 0$ for some $i \geq 1$ and any $v \neq 0$. Thus, to allow for kernels with unbounded support, we would generally need the parameter space $\mathcal{A}$, as defined in (8), to collapse at $\{(\alpha_1, 0, \ldots, 0) : \alpha_1 \in \Theta\}$ as $b \to 0$. Such shrinking behavior in turn means that a formal Taylor expansion of $\ell_{n,t}\left(D_{m,b}(v)\alpha\right)$ w.r.t. α is difficult to obtain and so standard arguments to establish asymptotic normality of $\hat{\alpha}$ cannot be applied. On the other hand, by restricting the support $\mathcal{K}$ to be compact, $K_b(v)\,\ell_{n,t}\left(D_{m,b}(v)\alpha\right)$ is well-defined for all $\alpha \in \mathcal{A}$ and $v \in \mathbb{R}$ (where we set $K_b(v)\,\ell_{n,t}\left(D_{m,b}(v)\alpha\right) = 0$ for $v/b \notin \mathcal{K}$). Moreover, $(\alpha_1, 0, \ldots, 0)$ is an interior point of $\mathcal{A}$, and so in our analysis of $\hat{\alpha}$, we can employ standard arguments involving a Taylor expansion of the score function around this point.

Assumption 2 is standard in the analysis of "global" extremum estimators of stationary models on the form $\tilde{\theta}(u) = \arg\max_{\theta\in\Theta} \sum_{t=1}^{n} \ell_t^*(\theta|u)/n$. In particular, for a given time-series model, we can import existing results for verification of Assumption 2 (ii) and (iii); see Section 4 for more details. Assumption 2 in conjunction with $K(\cdot) \geq 0$ ensures that the local polynomial estimator identifies $\theta(u)$. If we allow for kernels that take negative values, we have to replace Assumption 2(iii) with the following more abstract identification condition: The function $Q^*(\alpha|u) = \int K(v)\,\mathbb{E}\left[\ell_t^*(D_m(v)\alpha|u)\right]dv$ satisfies $Q^*(\alpha|u) < Q^*((\theta(u), 0, \ldots, 0)|u)$ for any $\alpha \neq (\theta(u), 0, \ldots, 0)$. We have not been able to provide primitive conditions for this to hold when K can take negative values and so instead impose the positivity constraint on K.

If the objective function $\theta \mapsto \ell_{n,t}(\theta)$ is concave and Θ is convex, we can replace Assumption 3 with the following pointwise versions: For any $\theta \in \Theta$, $\ell_{n,t}(\theta)$ is locally stationary and $\mathbb{E}\left[|\ell_t^*(\theta|u)|\right] < \infty$ (see Theorem 2.7 in Newey and McFadden, 1994). Under the above assumptions, the following consistency result holds.

THEOREM 1. *Let Assumptions 1–3 hold. Then, as* $b \to 0$ *and* $nb \to \infty$, $\hat{\alpha} \to^p (\theta(u), 0, \ldots, 0)'$. *In particular,* $\hat{\theta}(u) \to^p \theta(u)$.

Note that the above theorem only shows the consistency of $\hat{\theta}(u)$, and so, at this stage, we cannot make any statements regarding $\hat{\theta}^{(i)}(u)$, $i = 1, \ldots, m$. This is similar to other results for nonlinear extremum estimators that converge with different rates (see, e.g., Theorem 9 in Han and Kristensen, 2014, where a global consistency result is only provided for the component with the fastest rate).

https://doi.org/10.1017/S0266466626100449 Published online by Cambridge University Press

However, under additional regularity conditions on the quasi-likelihood function, we can provide a more precise analysis of the estimators, including local consistency of $\hat{\theta}^{(k)}(u)$, $1 \leq k \leq m$. With $s_{n,t}(\theta) = \partial \ell_{n,t}(\theta) / (\partial \theta) \in \mathbb{R}^{d_\theta}$ and $h_{n,t}(\theta) = \partial^2 \ell_{n,t}(\theta) / (\partial \theta \partial \theta') \in \mathbb{R}^{d_\theta \times d_\theta}$, $D_{n,t}(u) = D_{m,b}(t/n - u)$ and $K_{n,t}(u) = K_b(t/n - u)$, the score and Hessian of $Q_n(\alpha|u)$ are given by

$$S_n(\alpha|u) = \frac{\partial Q_n(\alpha|u)}{\partial \alpha} = \frac{1}{n} \sum_{t=1}^{n} K_{n,t}(u) D_{n,t}(u)' s_{n,t}\left(D_{n,t}(u)\alpha\right),$$

$$H_n(\alpha|u) = \frac{\partial^2 Q_n(\alpha|u)}{\partial \alpha \partial \alpha'} = \frac{1}{n} \sum_{t=1}^{n} K_{n,t}(u) D_{n,t}(u)' h_{n,t}\left(D_{n,t}(u)\alpha\right) D_{n,t}(u).$$

It is easily checked that $\alpha_0 := U_n \beta_0$, where $\beta_0 = (\theta(u)', \theta^{(1)}(u)', \ldots, \theta^{(m)}(u)')'$, belongs to the interior of $\mathcal{A}$ for all n large enough due to Assumption 4(ii) below in conjunction with Assumption 2. Due to the consistency result, so will $\hat{\alpha}$ w.p.a. 1. Thus, $\hat{\alpha}$ will satisfy the first-order condition of (7) which combined with the mean-value theorem yields

$$0 = S_n\left(\hat{\alpha}|u\right) = S_n(\alpha_0|u) + H_n(\bar{\alpha}|u)\left(\hat{\alpha} - \alpha_0\right) = S_n(\alpha_0|u) + H_n(\bar{\alpha}|u)\, U_n(\hat{\beta} - \beta_0), \tag{9}$$

where $\bar{\alpha}$ is situated on the line segment connecting $\hat{\alpha}$ and α_0. We then decompose the score function into a bias and variance component, $S_n(\alpha_0|u) = B_n(u) + S_n(u)$, where

$$B_n(u) = \frac{1}{n} \sum_{t=1}^{n} K_{n,t}(u) D_{n,t}(u)' b_{n,t}, \quad b_{n,t} = s_{n,t}\left(\theta_u^*(t/n)\right) - s_{n,t}(\theta(t/n)) \tag{10}$$

$$S_n(u) = \frac{1}{n} \sum_{t=1}^{n} K_{n,t}(u) D_{n,t}(u)' s_{n,t}, \quad s_{n,t} = s_{n,t}(\theta(t/n)) \tag{11}$$

and $\theta_u^*(t/n)$ was defined in eq. (1). This decomposition is different from the one usually employed in the analysis of kernel estimators of time-varying coefficients, where $s_{n,t}$ is replaced by the stationary version of the score function evaluated at $\theta(u)$, $s_t^*(\theta(u)|u)$ (see, e.g., Dahlhaus and Subba Rao, 2006; Dahlhaus et al., 2019). This "usual" choice has the consequence that the corresponding bias term generally involves the time derivative process of the score function and so the resulting analysis tends to impose stronger regularity conditions on the model being estimated. By instead centering the analysis around $s_{n,t}$, we can obtain the leading term of the bias $B_n(u)$ through a standard Taylor expansion w.r.t. θ:

$$b_{n,t} \cong h_{n,t}\left(\theta_u^*(t/n)\right)\left\{\theta_u^*(t/n) - \theta(t/n)\right\} \cong -h_{n,t}(\theta(u)) \frac{\theta^{(m+1)}(u)}{(m+1)!} \{t/n - u\}^{m+1}. \tag{12}$$

Thus, our approach allows for a simpler derivation of the leading bias and variance terms under the following weak regularity conditions.

https://doi.org/10.1017/S0266466626100449 Published online by Cambridge University Press

Assumption 4. (i) $\theta \mapsto \ell_{n,t}(\theta)$ is twice continuously differentiable; (ii) $\theta(u)$ lies in the interior of Θ and is $m+1$ times continuously differentiable; and (iii) $s_{n,t}$ is a martingale difference (MGD) array w.r.t. $\mathcal{F}_{n,t} = \mathcal{F}\left\{Z_{n,t}, Z_{n,t-1}, \ldots\right\}$ and LS $(2,r)$ with stationary approximation $s_t^*(\theta(u)|u)$ for some $r > 0$.

Assumption 5. (i) $h_{n,t}(\theta)$ is ULS $(1,r,\mathcal{N}(u,\epsilon))$ with continuous stationary approximation $h_t^*(\theta(u))$ for some $r > 0$, where $\mathcal{N}(u,\epsilon) := \{\theta \in \Theta : \|\theta - \theta(u)\| < \epsilon\}$ for some arbitrarily small $\epsilon > 0$ and (ii) $H(u) \equiv \mathbb{E}\left[h_t^*(\theta(u)|u)\right]$ is non-singular.

Similar to Assumption 2, Assumption 4 contains standard regularity conditions used in the analysis of regular parametric estimators on the form $\tilde{\theta}(u) = \arg\max_{\theta\in\Theta} \sum_{t=1}^n \ell_t^*(\theta|u)$. At the same time, Assumption 4(iii) is non-standard compared to the existing literature (as discussed above) and allows us to apply a novel martingale central limit theorem for locally stationary sequences to $S_n(u)$:

$$\sqrt{nb}S_n(u) \to^d N(0, \mathbb{K}_2 \otimes \Omega(u)), \quad \Omega(u) = \mathbb{E}\left[s_t^*(\theta(u)|u)\, s_t^*(\theta(u)|u)'\right] \tag{13}$$

(see Lemma A(iii) in Appendix A.1). This result can be seen as a generalization of the standard CLT for stationary and ergodic MGDs that allows for locally stationary processes. The MGD assumption amounts to assuming that the time-varying model is correctly specified and has to be verified on a case-by-case basis (see Section 4.2 for examples of this).

Finally, Assumption 5 together with the expansion in eq. (12) is used to derive the limits of $B_n(u)$ and $H_n(\bar{\alpha}|u)$:

$$H_n(\bar{\alpha}|u) \to^p \mathbb{K}_1 \otimes H(u), \quad B_n(u) = -b^{m+1}\mu_1 \otimes H(u)\frac{\theta^{(m+1)}(u)}{(m+1)!} + o_P\left(b^{m+1}\right), \tag{14}$$

where $\mu_i = \int_{\mathbb{R}} K(v)\, v^{m+i} D_m(v)\, dv$ and $\mathbb{K}_i = \int_{\mathbb{R}} K^i(v)\, D_m(v)\, D_m(v)'\, dv$, $i \geq 1$. Combining (9), (13), and (14), we obtain the following theorem.

THEOREM 2. *Suppose that Assumptions 1–5 hold. Then, as $b \to 0$ and $nb \to \infty$,*

$$\sqrt{nb}U_n\left\{\hat{\beta} - \beta_0 - R_n Bias(u)\right\} \to^d N\left(0, \mathbb{K}_1^{-1}\mathbb{K}_2\mathbb{K}_1^{-1} \otimes V(u)\right),$$

where $R_n = diag\left\{b^{m+1}, b^m, \ldots, b\right\} \otimes I_{d_\theta}$, $V(u) = H(u)^{-1}\Omega(u)H(u)^{-1}$, and $Bias(u) = \mathbb{K}_1^{-1}\mu_1 \otimes \frac{\theta^{(m+1)}(u)}{(m+1)!}$.

In particular, for $i = 0, 1, \ldots, m$,

$$\sqrt{nb^{2i+1}}\left\{\hat{\theta}^{(i)}(u) - \theta^{(i)}(u) - b^{m+1-i} Bias_i(u)\right\} \to^d N\left(0, \kappa_{2,i} V(u)\right), \tag{15}$$

where $Bias_i(u) = \kappa_{1,i}\frac{\theta^{(m+1)}(u)}{(m+1)!}$ and $\kappa_{1,i}$ and $\kappa_{2,i}$ denote the ith element of $\mathbb{K}_1^{-1}\mu_1$ and (i,i)th element of $\mathbb{K}_1^{-1}\mathbb{K}_2\mathbb{K}_1^{-1}$, respectively.

https://doi.org/10.1017/S0266466626100449 Published online by Cambridge University Press

Similar to existing results for local polynomial estimators in a cross-sectional setting, the leading bias term in (15) only depends on $\theta^{(m+1)}(u)$ and so the estimators adapt to the curvature of $\theta(u)$. The asymptotic variance in Theorem 2 can be estimated using plug-in methods: It follows from the proof of Theorem 2 that $H_n\left(\hat{\alpha}|u\right) \to^p \mathbb{K}_1 \otimes H(u)$ and

$$\hat{W}(u) = \frac{1}{n}\sum_{t=1}^{n} K_{n,t}^2(u) D_{n,t}(u)' s_{n,t}\left(D_{n,t}(u)\hat{\alpha}\right) s_{n,t}\left(D_{n,t}(u)\hat{\alpha}\right)' D_{n,t}(u)$$
$$\to^p \mathbb{K}_2 \otimes \Omega(u).$$

Compared to most existing asymptotic results for the local constant estimator, such as Dahlhaus et al. (2019), the above result with $m \geq 1$ imposes much weaker restrictions on the bandwidth. In particular, standard bandwidth selection rules can be employed here but not under most of the existing theories since their conditions require undersmoothing (i.e., $b \to 0$ at a faster rate than the optimal one). This is due to the fact that these theories do not provide a complete characterization of the leading bias term. The few papers that do characterize the leading bias term, such as Dahlhaus and Subba Rao (2006), require the so-called time derivatives of the stationary score function to exist and be well-behaved since these enter their bias expressions. Our conditions and results, on the other hand, do not require these and are analogous to the ones found in the literature on local polynomial likelihood estimators (see, e.g., Theorem 1b of Fan et al., 1995).

Equation (15) holds for any value of $m \geq 0$ and $i = 0, \ldots, m$. However, if K is symmetric, then $\kappa_{1,i} = 0$ when $m - i$ is even. In particular, for the local constant estimator ($m = i = 0$), Theorem 2 only informs us that the bias component of $\hat{\theta}(u)$ is $o_p(b)$ which is not a sharp rate. To obtain the leading bias term in the cases where $m - i$ is even, a higher-order expansion of $b_{n,t}$ in eq. (10) is necessary. This expansion requires additional assumptions involving aforementioned time derivatives and standard derivatives w.r.t. θ of $h_t^*(\theta(u)|u)$. To present these, we need the following additional concept.

DEFINITION 3. *A stationary process $W_t^*(\theta|u)$ is said to be L_p-differentiable w.r.t. u if there exists a stationary and ergodic process $\partial_u W_t^*(\theta|u)$ with $\mathbb{E}\left[\left\|\partial_u W_t^*(\theta|u)\right\|^p\right] < \infty$ such that*

$$\mathbb{E}\left[\left\|W_t^*(\theta|u+\Delta) - W_t^*(\theta|u) - \partial_u W_t^*(\theta|u)\Delta\right\|^p\right]^{1/p} = o(\Delta),\ \Delta \to 0.$$

Our definition of time differentiability is slightly weaker compared to the one found in Dahlhaus et al. (2019) and other papers where differentiability w.r.t. u has to hold almost surely. With this definition in hand, we are ready to introduce the following additional regularity conditions in order to derive the leading bias term when $m - i$ is even.

Assumption 6. $\partial h_{n,t}(\theta)/(\partial\theta_i)$ exists and is ULS $(1, r, \mathcal{N}(u, \epsilon))$ with L_1-continuous stationary approximation $\partial h_t^*(\theta|u)/(\partial\theta_i)$, $i = 1, \ldots, d_\theta$.

https://doi.org/10.1017/S0266466626100449 Published online by Cambridge University Press

Assumption 7. $h_t^*(\theta|u)$ is L_1 -differentiable w.r.t. u at $\theta = \theta(u)$ with time-derivative $\partial_u h_t^*(\theta(u)|u) = \partial h_t^*(\theta|u)/(\partial u)|_{\theta=\theta(u)} \in \mathbb{R}^{d_\theta \times d_\theta}$.

Assumption 8. $\sum_{t=1}^{\infty} \left|\mathrm{Cov}\left(h_{ij,0}^*(\theta(u)|u), h_{ij,t}^*(\theta(u)|u)\right)\right| < \infty$, $i,j = 1,\ldots,d_\theta$.

The time-derivative $\partial_u h_t^*(\theta|u)$ will generally involve time derivatives of the underlying stationary approximation of data: If $h_t^*(\theta|u) = h\left(\mathcal{Z}_t^*(u);\theta\right)$ for some function h which is differentiable w.r.t. $\mathcal{Z}_t^*(u)$, then it takes the form $\partial_u h_t^*(\theta|u) = \sum_{i=0}^{\infty} \partial h(z_0, z_1, z_2, \ldots;\theta)/(\partial z_i)|_{z=\mathcal{Z}_t^*(u)} \times \partial_u Z_{t-i}^*(u)$, where $\partial_u Z_{i,t}^*(u)$ is the time derivative of $Z_t^*(u)$. The short memory condition imposed in Assumption 8 is used to control the variance component of the first-order bias term derived in Theorem 2. In Section 4, we use the concept of τ—weak dependence (Doukhan and Wintenberger, 2008) to verify Assumption 8.

Under the above additional assumptions, we obtain the following higher-order expansion of the bias component.

THEOREM 3. *Suppose Assumptions 1–8 hold and $\theta(\cdot)$ is $m+2$ times continuously differentiable. Then, as $b \to 0$ and $nb \to \infty$, the bias $B_n(u)$ defined in (10) satisfies, with r given in Assumption 6,*

$$B_n(u) = -b^{m+1}\left[\mu_1 \otimes H(u)\frac{\theta^{(m+1)}(u)}{(m+1)!} + o_P(b)\right] - b^{m+2}\left[\mu_2 \otimes B_1(u) + o_p(1)\right] \tag{16}$$
$$- b^{2m+2}\left[\mu_2 \otimes B_2(u) + o_p(1)\right] + O_P\left(1/n^{\min\{1,r\}}\right) + o_p\left(\frac{1}{\sqrt{nb}}\right),$$

where, with $\partial_u H(u) = \mathbb{E}\left[\partial_u h_t^*(\theta|u)\right]_{\theta=\theta(u)}$ and $\partial_{\theta_i} H(u) = \mathbb{E}\left[\partial h_t^*(\theta|u)/(\partial\theta_i)\right]_{\theta=\theta(u)}$,

$$B_1(u) = H(u)\frac{\theta^{(m+2)}(u)}{(m+2)!} + \left(\partial_u H(u) + \sum_{i=1}^{d_\theta} \theta_i^{(1)}(u)\,\partial_{\theta_i} H(u)\right)\frac{\theta^{(m+1)}(u)}{(m+1)!},$$

$$B_2(u) = \frac{-1}{2\{(m+1)!\}^2}\sum_{i=1}^{d_\theta} \theta_i^{(m+1)}(u)\,\partial_{\theta_i} H(u)\,\theta^{(m+1)}(u).$$

As a special case, we obtain the following result for the local constant estimator.

COROLLARY 1. *Under the assumptions of Theorem 3 with $m = 0$ together with $\int_{\mathbb{R}} K(v)\,v dv = 0$, the local constant estimator satisfies, as $b \to 0$, $nb^3 \to \infty$, and $n^{\min\{r,1\}}b \to \infty$,*

https://doi.org/10.1017/S0266466626100449 Published online by Cambridge University Press

$$\sqrt{nb}\left\{\hat{\theta}(u)-\theta(u)-b^2\mu_2 Bias_0(u)\right\}\rightarrow^d N\left(0,\kappa_{2,0}V(u)\right),$$

where $Bias_0(u)=H^{-1}(u)[B_1(u)+B_2(u)]$ *and* $B_1(u)$ *and* $B_2(u)$ *are given in Theorem 3.*

Two equivalent representations of $B_1(u)+B_2(u)$ *are, with* $s_t^*(\theta|u)$ *denoting the stationary approximation of* $s_{n,t}(\theta)$*,*

$$\begin{aligned}B_1(u)+B_2(u)&=\frac{1}{2}\frac{\partial^2}{\partial v^2}\mathbb{E}\left[s_t^*(\theta(v)|u)\right]_{v=u}+\frac{\partial^2}{\partial u\partial v}\mathbb{E}\left[s_t^*(\theta(v)|u)\right]_{v=u}\\&=-\frac{1}{2}\frac{\partial^2}{\partial u^2}\mathbb{E}\left[s_t^*(\theta|u)\right]_{\theta=\theta(u)},\end{aligned}$$

where the second representation is only well-defined if $s_t^*(\theta|u)$ *is twice* L_1*-differentiable w.r.t. u.*

To our knowledge, this is the first complete characterization of the leading bias term of local constant estimators in general time-varying parameter models. The final characterization of the bias, $B_1(u)+B_2(u)=-\frac{1}{2}\partial_u^2\mathbb{E}\left[s_t^*(\theta(v)|u)\right]_{v=u}$, corresponds to the one obtained in Dahlhaus and Subba Rao (2006) for the time-varying ARCH model. This characterization, however, requires $s_t^*(\theta|u)$ to be twice differentiable w.r.t. u, while our characterization only requires $h_t^*(\theta|u)$ to be once differentiable w.r.t. u.

Comparing Theorems 2 and 3, we see that the local linear and local constant estimators share convergence rate and asymptotic variance, but that the latter suffers from additional biases. This is consistent with the theory found for local constant and local linear estimators in a cross-sectional setting (see, e.g., Fan et al., 1995).

3.1. Discrete-Valued Time Series

The above theory for the local constant estimator does not cover discrete-valued time-series models. Specifically, Theorem 3 requires $h_t^*(\theta|u)$ to be differentiable w.r.t. u (cf. Assumption 7). This property rarely holds when $h_t^*(\theta|u)$ is a function of discrete-valued random variables since these are generally not smooth functions of the underlying parameters of the model (see Truquet, 2019 for more details). Thus, Theorem 3 does not apply to, for example, the Poisson autoregressive model.

But Theorem 2 still applies. We therefore combine the ideas of Truquet (2019, 2020) with Theorem 2 to obtain a theory for the local constant estimator that also covers models with discrete-valued outcomes. This is achieved by replacing Assumptions 7 and 8 with the following ones.

Assumption 9. $v\mapsto\mathbb{E}\left[h_t^*(\theta(u)|v)\right]$ is continuously differentiable at u.

https://doi.org/10.1017/S0266466626100449 Published online by Cambridge University Press

Assumption 10. (i) $\bar{V}_{ijkl}(v_1, v_2) := \sum_{t=1}^{\infty} \mathrm{Cov}\left(h^*_{ij,0}(\theta(u)|v_1), h^*_{kl,t}(\theta(u)|v_2)\right)$ exists for all (v_1, v_2) in a neighborhood of (u, u) and all (i, j, k, l) and (ii) $\bar{V}_{ijkl}(v_1, v_2)$ is continuously differentiable at $(v_1, v_2) = (u, u)$.

If $v \mapsto h^*_t(\theta(u)|v)$ is L_1-differentiable w.r.t. v at u, then $\partial_v \mathbb{E}\left[h^*_t(\theta(u)|v)\right]_{v=u} = \mathbb{E}\left[\partial_v h^*_t(\theta(u)|v)\right]_{v=u}$. Thus, Assumption 9 is weaker than Assumption 7 and is satisfied as long as the cumulative distribution function of $h^*_t(\theta(u)|v)$ is differentiable w.r.t. v (cf. Lemma 2 below). This property holds for many discrete-valued models, including Poisson autoregressions and dynamic discrete choice models (cf. Truquet, 2019, 2020).

Assumption 10, on the other hand, is stronger than Assumption 8. However, similar to Assumption 8, part (i) is satisfied if $h^*_t(\theta(u)|v)$ is τ-weakly dependent for v in a neighborhood of u since this in turn implies that the joint process $\left(h^*_t(\theta(u)|v_1), h^*_t(\theta(u)|v_2)\right)$ is weakly dependent for (v_1, v_2) in a neighborhood of (u, u). Moreover, part (ii) will hold under the same conditions that ensure Assumption 9 holds, namely, that the joint distribution function of $\left(h^*_0(\theta(u)|v_1), h^*_t(\theta(u)|v_2)\right)$ is differentiable.

The following result shows that the results for the local constant estimator remain essentially the same under Assumptions 9 and 10 in place of Assumptions 7 and 8.

THEOREM 4. *Suppose Assumptions 1–6, 9, and 10 hold and $\theta(\cdot)$ is $m+2$ times continuously differentiable. Then the conclusions of Theorem 3 and Corollary 1 still hold, except that $\partial_u H(u)$ in the expression of $B_1(u)$ is now defined as $\partial_u H(u) = \partial_v E\left[h^*_t(\theta(u)|v)\right]_{v=u}$.*

3.2. Behavior at Boundary

We have already seen that the local linear estimator has smaller biases than the local constant one in the interior of its domain, $u \in (0, 1)$. Another well-known advantage of the local linear estimators in a cross-sectional setting is that they exhibit automatic boundary carpentering. This property also holds in our setting where the boundaries are $u = 0$ and $u = 1$. Since the results for $u = 1$ are similar, we here only analyze the properties of the local constant ($m = 0$) and the local linear ($m = 1$) estimators at $u = cb$ for some constant $c > 0$. Combining the intermediate bias–variance analysis carried out in the proofs of Theorem 2 and 3 with the arguments of Fan et al. (1995), we find that the two estimators remain asymptotically normally distributed but their asymptotic biases and variances take different forms.

COROLLARY 2. *Let $\hat{\theta}_0(u)$ and $\hat{\theta}_1(u)$ be the local constant and local linear estimators, respectively, of $\theta(u)$. Under the assumptions of Theorem 3 with $m = 1$ together with $\mu_1 = \int_{\mathbb{R}} K(v) v dv = 0$, as $b \to 0$, $nb^3 \to \infty$, and $n^{\min\{r,1\}} b \to \infty$,*

https://doi.org/10.1017/S0266466626100449 Published online by Cambridge University Press

and for $m \in \{0,1\}$,

$$\sqrt{nb}\left(\hat{\theta}_m(cb) - \theta(cb) - b^{m+1}Bias_m\right) \xrightarrow{d} N(0, a_m V(0_+)),$$

where $V(0_+) = \lim_{u \downarrow 0} V(u)$ *and*

$$Bias_0 = \left(\kappa^c_{1,0}\right)^{-1}\kappa^c_{1,1}\theta^{(1)}(0_+); \quad Bias_1 = \frac{1}{2}\frac{\left(\kappa^c_{1,2}\right)^2 - \kappa^c_{1,1}\kappa^c_{1,3}}{\kappa^c_{1,0}\kappa^c_{1,2} - \left(\kappa^c_{1,1}\right)^2}\theta^{(2)}(0_+);$$

$$a_0 = \kappa^c_{2,0}/\left(\kappa^c_{1,0}\right)^2;$$

$$a_1 = \left[\left(\kappa^c_{1,2}\right)^2\kappa^c_{2,0} - 2\kappa^c_{1,1}\kappa^c_{1,2}\kappa^c_{2,1} + \left(\kappa_{1,1}\right)^2\kappa^c_{2,2}\right]/\left[\kappa^c_{1,0}\kappa^c_{1,2} - \left(\kappa^c_{1,1}\right)^2\right]^2.$$

We refer to Fan et al. (1995) for the precise expressions of the constants $\kappa^c_{i,j}$, $i = 1,2$ and $j = 0,1,2$. At the boundary, both the biases and variances of the local constant and local linear estimators are now different. While the difference between two asymptotic variances is a constant scale, compare a_0 and a_1 above, the biases are now of a different order: The local linear estimator still enjoys a bias of order $O\left(b^2\right)$ while the bias of the local constant one blows up and becomes of order $O(b)$. Thus, the local constant estimator will generally suffer from significantly larger biases at the boundary compared to the local linear one.

4. APPLICATIONS

In this section, we illustrate how our general results can be verified for a class of Markov models with exogenous covariates and provide concrete examples within this framework. Application to infinite-memory models, as considered in Bardet et al. (2022), is discussed in Section 7 as a direction for future work.

By inspection of our Assumptions 1–5, we observe that, for a given parametric model and estimator, most of the assumptions are easily verifiable using standard arguments known from the literature on regular parametric estimators. The only ones that are non-standard are Assumptions 3 and 5, which require the researcher to show that $\ell_{n,t}(\theta)$ and its first two derivatives are ULS. Similarly, Assumption 6 is also a ULS requirement, while Assumption 8 and/or 10 will hold if $h^*_t(\theta(u)|u)$ is weakly dependent.

The following lemma provides sufficient conditions for a given transformation of a time series to be ULS if the underlying time series is ULS. It also shows that if the stationary version is τ-weakly dependent (see Doukhan and Wintenberger, 2008 for the definition), then so is the transformation.

LEMMA 1. *Let* $(\mathcal{Z}, \|\cdot\|)$ *be a Banach space and* $f : (\mathcal{Z}, \|\cdot\|)^\infty \times \Theta \mapsto \mathbb{R}^d$, $d < \infty$, *be a given mapping. Then the following hold, where* $\mathcal{Z}_{n,t}(\theta) := \left(Z_{n,t}(\theta), Z_{n,t-1}(\theta), \ldots\right)$ *and* $\mathcal{Z}^*_t(\theta|u) := \left(Z^*_t(\theta|u), Z^*_{t-1}(\theta|u), \ldots\right)$ *is its stationary approximation:*

https://doi.org/10.1017/S0266466626100449 Published online by Cambridge University Press

1. *Suppose that (i) for some* $g:(\mathcal{Z},\|\cdot\|)^{\infty}\times\Theta\mapsto\mathbb{R}_{+}$ *and* $\{a_i\}_{i=0}^{\infty}$ *with* $\sum_{i=0}^{\infty}a_i<\infty$ *so that*

$$\|f(z_1;\theta)-f(z_2;\theta)\|\leq\left\{\sum_{k=1}^{2}g(z_k)\right\}\sum_{i=0}^{\infty}a_i\left\|z_{1,i}-z_{2,i}\right\|, \tag{17}$$

for all $\theta\in\Theta$ *and* $z_k=\left(z_{k,0},z_{k,1},\ldots\right)\in\mathcal{Z}^{\infty}$, $k=1,2$ *and (ii) for some* $p_1,p_2\geq 1$ *and* $r>0$, $Z_{n,t}(\theta)\in\mathcal{Z}$ *is ULS* (p_1,r,Θ) *with* $\mathbb{E}\left[\sup_{\theta\in\Theta}\left\|Z_{n,t}(\theta)\right\|^{p_1}\right]<\infty$, $\mathbb{E}\left[\sup_{\theta\in\Theta}g\left(\mathcal{Z}_t^*(\theta|u)\right)^{p_2}\right]<\infty$ *and* $\mathbb{E}\left[\sup_{\theta\in\Theta}g\left(\mathcal{Z}_{n,t}(\theta)\right)^{p_2}\right]<\infty$. *Then* $f\left(\mathcal{Z}_{n,t}(\theta);\theta\right)$ *is ULS* (p,r,Θ), *where* $p=(1/p_1+1/p_2)^{-1}$, *and* $\mathbb{E}\left[\sup_{\theta\in\Theta}\left\|f\left(\mathcal{Z}_t^*(\theta);\theta\right)\right\|^{p}\right]<\infty$
2. *Suppose that: (i) in part 1. of the lemma holds with (ii')* $g(z)\leq C(1+\sum_{i=0}^{\infty}b_i\|z_i\|^{s})$ *for some* $s\geq 0$ *and* $\{b_i\}_{i=0}^{\infty}$ *with* $\sum_{i=0}^{\infty}b_i<\infty$, $k=1,2$*; (iii')* $Z_{n,t}(\theta)\in\mathcal{Z}$ *is ULS* (p_1,r,Θ) *with* $\mathbb{E}\left[\sup_{\theta\in\Theta}\left\|Z_{n,t}(\theta)\right\|^{p_1}\right]<\infty$ *and* $\mathbb{E}\left[\sup_{\theta\in\Theta}\left\|Z_t^*(\theta|u)\right\|^{p_1}\right]<\infty$, *for some* $p_1\geq s$. *Then* $f\left(\mathcal{Z}_{n,t}(\theta);\theta\right)$ *is ULS* (p,r,Θ) *with* $\mathbb{E}\left[\sup_{\theta\in\Theta}\left\|f\left(\mathcal{Z}_t^*(\theta|u);\theta\right)\right\|^{p}\right]<\infty$ *and* $p=p_1/(s+1)$.
3. *Suppose furthermore that* $Z_t^*(\theta|u)$ *is* τ*-weakly dependent with weak dependence coefficients* $\tau_Z(k)$. *Then* $f\left(\mathcal{Z}_t^*(\theta|u);\theta\right)$ *is also* τ*-weakly dependent with* $\mathbb{E}\left[\left\|f\left(\mathcal{Z}_t^*(\theta|u);\theta\right)\right\|^{p}\right]$ *whose weak dependence coefficients vanish at the same rate as* $\tau_Z(k)$. *In particular, if* $p>2$ *and* $\tau_Z(k)$ *vanish at geometric rate then* $\sum_{t=1}^{\infty}\left|\mathrm{Cov}\left(f\left(\mathcal{Z}_0^*(\theta|u);\theta\right),f\left(\mathcal{Z}_t^*(\theta|u);\theta\right)\right)\right|<\infty$.

Next, we here provide more primitive conditions for Assumption 9 that can be applied to discrete-valued random variables.

LEMMA 2. *Suppose that the distribution function* $F_Z(z;u)$ *of* $Z_t^*(u)$ *satisfies* $dF_Z(z;u)=p(z;u)\,d\mu(z)$ *for some measure* μ, *where* $u\mapsto p(z;u)$ *is differentiable w.r.t.* μ*-a.s. Then for any function* $f(z)$ *with* $\mathbb{E}\left[\left\|f\left(Z_t^*(u);u\right)\right\|\right]<\infty$, $\partial_u\mathbb{E}\left[f\left(Z_t^*(u)\right)\right]=\int f(z)\,\partial_u p(z;u)\,d\mu(z)$ *assuming the integral is well defined.*

A sufficient condition for $\int|f(z)\,\partial_u p(zu)|\,d\mu(z)<\infty$ is $\mathbb{E}[\left|f\left(Z_t^*(u)\right)\right|\left|s\left(Z_t^*(u)\right)\right|]<\infty$, where $s=\partial_u p/p$ is the score function. For example, if $|s(z;u)|\leq C\left(1+\|z\|^{\delta}\right)$, $\delta\geq 0$, then the result will hold under $E\left[\left\|Z_t^*(\theta|u)\right\|^{\delta}\left\|f\left(Z_t^*(\theta|u);u\right)\right\|\right]<\infty$. For Markov models, the conditions of Lemma 2 can be verified by importing results from Truquet (2020) and Vazquez-Abad and Kushner (1992); see proof of Corollary 6 for an example of this.

We summarize our findings in the following corollary.

COROLLARY 3. *Suppose that (i) K satisfies Assumption 1; (ii) the stationary version of the model, as summarized by* $\theta\mapsto\ell_t^*(\theta|u)$, *satisfies Assumptions 2 and 5(ii); (iii)* $s_{n,t}$ *is an MGD; (iv)* $\ell_{n,t}(\theta)$, $s_{n,t}$, *and* $h_n,t(\theta)$ *satisfy the conditions imposed on f in Lemma 1; (v)* $Z_{n,t}$ *is LS* (p,r) *and* $\mathbb{E}\left[\left\|Z_t^*(u)\right\|^{p}\right]<\infty$. *Then, the conclusions of Theorem 2 hold.*

https://doi.org/10.1017/S0266466626100449 Published online by Cambridge University Press

Suppose that, in addition, (vi) the stationary distribution satisfies either Assumption 7 or 9; (vii) $Z_t^*(u)$ is τ-weakly dependent with $\tau_Z(k)$ vanishing at geometric rate; and (viii) $h_t^*(\theta|u)$ satisfies (17) with $\delta \geq 0$ so that $p/(\delta+1) > 2$ and the conditions of Lemma 2. Then, the conclusions of Theorem 4 hold.

If the model and estimator of interest are regular in the sense that its stationary version satisfies the above assumptions, then all that remains to be shown is that $Z_{n,t}$ is locally stationary with $Z_t^*(u)$ being weakly dependent. The next section provides primitive conditions for this to hold for Markov models with exogenous covariates, while the second section revisits the examples of Section 2 and provides primitive conditions under which our main results apply to these.

4.1. Markov Models with Exogenous Covariates

Consider the following class of q-Markov models with exogenous covariates:

$$Y_{n,t} = G\left(Y_{n,t-1}, \ldots, Y_{n,t-q}, X_{n,t-1}, \varepsilon_t; \theta(t/n)\right), \tag{18}$$

$$X_{n,t} = H\left(X_{n,t-1}, \ldots, X_{n,t-q}, \eta_t, t/n\right). \tag{19}$$

We here restrict $X_{n,t}$ to be a Markov process. We could allow for $X_{n,t}$ to exhibit richer dynamics, for example, allow for $X_{n,t}$ to depend on lags of $Y_{n,t}$. However, this would lead to more complicated assumptions and so we here maintain (19) for simplicity. We impose the following assumptions on (18) and (19).

Assumption 11. (i) (ε_t, η_t) is i.i.d. over time; (ii) for some $\rho_x < 1$, $p \geq 1$, and $x_0 \in \mathcal{X}^q$, $\mathbb{E}\left[\|H(x;\eta_t,u) - H(\tilde{x};\eta_t,u)\|^p\right] \leq \rho_x \|x-\tilde{x}\|^p$ and $\mathbb{E}\left[\|H(x_0;\eta_t,u)\|^p\right] < \infty$ for all $x, x' \in \mathcal{X}^q$ and $u \in [0,1]$; (iii) for some $(x_0, y_0) \in \mathcal{X}^q \times \mathcal{Y}^q$, $L < \infty$ and $\rho_y < 1$, $\sup_{\theta\in\Theta} \mathbb{E}[\|G(y_0, x_0, \varepsilon_t; \theta)\|^p] < \infty$ and

$$\mathbb{E}\left[\|G(y,x,\varepsilon_t;\theta) - G\left(y',x',\varepsilon_t;\theta\right)\|^p\right]^{1/p} \leq L\|x-x'\| + \rho_y\|y-y'\|$$

for all (x,y), $\left(x',y'\right) \in \mathcal{X}^q \times \mathcal{Y}^q$; (iv) for some $\tilde{p} \geq 1$, $r > 0$, and $\delta \geq 0$ and all $\theta, \theta' \in \Theta$,

$$\mathbb{E}\left[\|G(y,x,\varepsilon_t;\theta) - G\left(y,x,\varepsilon_t;\theta'\right)\|^{\tilde{p}}\right]^{1/\tilde{p}} \leq C\left(1 + \|x\|^\delta + \|y\|^\delta\right)\|\theta - \theta'||^r;$$

$$\mathbb{E}\left[\|H(x,\eta_t;u) - H(x,\eta_t;v)\|^{\tilde{p}}\right]^{1/\tilde{p}} \leq C\left(1 + \|x\|^\delta\right)|u-v|^r;$$

(v) $\mathbb{E}\left[\|Y_{n,0}\|^{\tilde{p}}\right] < \infty$ and $\mathbb{E}\left[\|X_{n,0}\|^{\tilde{p}}\right] < \infty$.

This assumption is similar to the one found in Dahlhaus et al. (2019). However, in contrast to this article, we do not require the process to be initialized at the stationary distribution, $(Y_{n,0}, X_{n,0}) = (Y_0^*(u), X_0^*(u))$, and instead allow it to be initialized at any given value or distribution.

https://doi.org/10.1017/S0266466626100449 Published online by Cambridge University Press

LEMMA 3. *Under Assumption 11(i)–(iii), for any* $u \in [0,1]$, *there exists a stationary and* τ *-weakly dependent solution* $\left\{Y_t^*(u), X_t^*(u)\right\}$ *to*

$$Y_t^*(u) = G\left(Y_{t-1}^*(u), X_{t-1}^*(u), \varepsilon_t; \theta(u)\right), \tag{20}$$

$$X_t^*(u) = H\left(X_{t-1}^*(u), \eta_t, u\right), \tag{21}$$

with $\tau_{Y,X}(k) \leq O\left(\rho^k\right)$. *If furthermore Assumption 11(iv) and (v) hold,* $\sup_{u\in[0,1]} \mathbb{E}\left[\|X_t^*(u)\|^{\tilde{p}\delta}\right]$, $\sup_{u\in[0,1]} \mathbb{E}\left[\|Y_t^*(u)\|^{\tilde{p}\delta}\right] < \infty$ *and* $u \mapsto \theta(u)$ *and* $u \mapsto H(x, \eta_t, u)$ *is continuously differentiable, then* $\left(Y_{n,t}, X_{n,t}\right)$ *is LS* $(\tilde{p}, r)$ *with* $\sup_{n,t} \mathbb{E}\left[\|Y_{n,t}\|^{\tilde{p}}\right] < \infty$ *and* $\sup_{n,t} \mathbb{E}\left[\|X_{n,t}\|^{\tilde{p}}\right] < \infty$ *for any given initial values* $\left(Y_{n,0}, X_{n,0}\right)$ *with* $\mathbb{E}\left[\|Y_{n,0}\|^p\right] < \infty$ *and* $\mathbb{E}\left[\|X_{n,0}\|^p\right] < \infty$.

Combining Corollary 3 with Lemma 3, we obtain a set of easily verifiable primitive conditions for local M-estimators of time-varying parameters in Markov models to satisfy Theorems 2 and 4 (see Section 4.2 for examples of the verification procedure).

4.2. Specific Examples

We here apply Corollary 3 and Lemma 3 to Examples 1–3. Note here that the first and third examples cannot be analyzed using some of the existing theories since these rely on differentiability of the DGP. Moreover, none of the existing theories allow for covariates to be included in the model. As such, the results below are new to the literature.

Throughout this section, the following assumption is maintained where $\theta(u)$ and Θ are specified in each of the following examples.

Assumption 12. (i) The kernel K satisfies Assumption 1; (ii) $\theta(u) \in \text{Int}(\Theta)$ and $\theta(\cdot)$ is twice continuously differentiable; (iii) the additional predictors $X_{n,t}$ of the model solve (19), where $H(x, \eta_\tau; u)$ satisfies the conditions in Assumption 11 with $p = \tilde{p} = 2$; and (iv) ε_t is i.i.d.

For the time-varying TAR-X model in Example 1, we obtain the following result.

COROLLARY 4. *Let* $\Theta = \mathbb{R}^{2q+d_X+1}$ *and suppose that Assumption 12 holds with* $\mathbb{E}\left[\varepsilon_t^2\right] < \infty$ *and*

$$\sup_{u\in[0,1]} \sum_{i=1}^{q} \max\left\{\left|\alpha_{1,i}(u)\right|, \left|\alpha_{2,i}(u)\right|\right\} < 1. \tag{22}$$

Then, the results of Theorem 2 apply to the local linear estimators of the tv-TAR (1) *model (4) with* $H(u) = \mathbb{E}\left[\tilde{X}_t^*(u)\tilde{X}_t^*(u)'\right]$, $\Omega(u) = \mathbb{E}\left[\varepsilon_t^2\right] H(u)$, *and*

$$\tilde{X}_t^*(u) = \left(Y_{t-1}^{*+}(u), \ldots, Y_{t-q}^{*+}(u), Y_{t-1}^{*-}(u), \ldots, Y_{t-q}^{*-}(u), X_{t-1}^*(u)\right).$$

https://doi.org/10.1017/S0266466626100449 Published online by Cambridge University Press

If, in addition, $\mathbb{E}\left[\varepsilon_t^4\right] < \infty$, ε_t *has a differentiable density and* $H(x, \eta_t, u)$ *satisfies Assumption 4.1(L3) and (L4) in Dahlhaus et al. (2019), then the results of Theorem 3 hold for the local constant estimators with* $\partial_\theta H(u) = 0$ *and* $\partial_u H(u) = \partial H(\theta|u) / (\partial u)|_{\theta=\theta(u)}$.

Note here that we do not need to restrict Θ to be compact since here $\ell_{n,t}(\theta)$ is concave. Equation (22) ensures that the model indeed has a locally stationary solution and is τ-weakly dependent. The additional restrictions imposed in the second part of the corollary are used to show that the time derivative of $\left(Y_t^*(u), X_t^*(u)\right)$ exists so that Assumption 7 holds.

Next, consider the local Gaussian QMLE of the tv-ARCH-X model given in Example 2. Under eq. (23) below, there exists a locally stationary solution to the model which takes the form

$$Y_t^*(u) = \lambda_t^*(u)\varepsilon_t^2, \quad \lambda_t^*(u) = \theta(u)'\tilde{X}_t^*(u),$$

where $\tilde{X}_t^*(u) = \left(1, Y_{t-1}^*(u), \ldots, Y_{t-q}^*(u), X_{t-1}^*(u)'\right)'$.

COROLLARY 5. *Assume that Assumption 12 holds,* $\Theta \subseteq \mathbb{R}_+^{1+q+d_X}$ *is compact with* $\omega \geq \underline{\omega} > 0$ *for all* $\theta \in \Theta$, $X_{n,t-1} \in \mathbb{R}_+^{d_X}$, $\mathbb{E}\left[\varepsilon_t^4\right] < \infty$, $\inf_{u\in[0,1]} \omega(u) > 0$, *and*

$$\sup_{u\in[0,1]} \sum_{i=1}^{q} \alpha_i(u) < 1. \tag{23}$$

Then, the results of Theorem 2 apply to the local linear estimators of the tv-ARCH-X model (5) with

$$H(u) = -\mathbb{E}\left[\frac{\partial_\theta \lambda_t^*(u)\left(\partial_\theta \lambda_t^*(u)\right)'}{\lambda_t^*(u)^2}\right], \quad \Omega(u) = -\mathrm{Var}\left(\varepsilon_t^2\right) H(u).$$

If, in addition, $\mathbb{E}\left[\varepsilon_t^{4+\delta}\right] < \infty$, *for some* $\delta > 0$, *and* $H(x, \eta_t, u)$ *satisfies Assumption 4.1(L3) and (L4) in Dahlhaus et al. (2019), then the results of Theorem 3 apply to the local constant estimators of the tv-ARCH-X model with* $\partial_u H(u) = \partial H(\theta|u) / (\partial u)|_{\theta=\theta(u)}$ *and*

$$\partial_{\theta_i} H(u) = 2\mathbb{E}\left[\frac{\partial_{\theta_i}\lambda_t^*(u)\,\partial_\theta \lambda_t^*(u)\left(\partial_\theta \lambda_t^*(u)\right)'}{\lambda_t^*(u)^3}\right].$$

Our conditions on the model are more or less identical to Bardet et al. (2022) but allows for exogenous regressors to be included. Our conditions are substantially weaker compared to Dahlhaus and Subba Rao (2006) and Inoue, Jin, and Pelletier (2019) who impose much stronger moment conditions.

Finally, we apply our results to the local MLE of the tv-PARX model in Example 3. Under (24) below, the tv-PARX process is locally stationary with

https://doi.org/10.1017/S0266466626100449 Published online by Cambridge University Press

stationary solution $Y_t^*(u)|\mathcal{F}_{t-1}^*(u) \sim \text{Poisson}\left(\lambda_t^*(u)\right)$, where $\lambda_t^*(u) = \theta(u)' \tilde{X}_t^*(u)$ and $\tilde{X}_t^*(u) = \left(1, Y_{t-1}^*(u), \ldots, Y_{t-q}^*(u), X_{t-1}^*(u)'\right)'$.

COROLLARY 6. *Suppose that Assumption 12 holds, $\Theta \subseteq \mathbb{R}_+^{1+q+d_X}$ is compact with $\omega \geq \underline{\omega} > 0$ for all $\theta \in \Theta$, $X_{n,t-1} \in \mathbb{R}_+^{d_X}$, $\inf_{u\in[0,1]} \omega(u) > 0$, and*

$$\sup_{u\in[0,1]} \sum_{i=1}^{q} \alpha_i(u) < 1. \tag{24}$$

Then, the results of Theorems 2 apply to the local linear estimators of the tv-PARX model (6) with

$$\Omega(u) = \mathbb{E}\left[\frac{\left(\partial_\theta \lambda_t^*(u)\right)\left(\partial_\theta \lambda_t^*(u)\right)'}{\lambda_t^*(u)}\right] = -H(u), \quad \partial_\theta \lambda_t^*(u) = \tilde{X}_t^*(u).$$

Suppose furthermore that $X_t^*(u)$ satisfies the following: Its Markov transition kernel is continuous w.r.t. the Lebesgue measure with density $p_X(x|x_0;u) > 0$ which is differentiable w.r.t. u for all (x,x_0); $|\partial_u p_X(x|x_0;u)|/p_X(x|x_0;u) \leq C(1+\|x\|+\|x_0\|)$; $p_X(x|x_0;u) \leq \bar{p}_X^{(0)}(x|x_0;u)$ and $|\partial_u p_X(x|x_0;u)| \leq \left|\bar{p}_X^{(1)}(x|x_0)\right|$, where $\int \|x\|^2 \bar{p}_X^{(k)}(x|x_0;u)\,dx < \infty$, $k = 0, 1$. Moreover, $\mathbb{E}\left[1 + X_t^*(u)^2 | X_{t-1}^*(u) = x\right] \leq \rho(u)\left(1+\|x\|^2\right) + b(u)$ for some $\rho(u) < 1$ and $b(u) < \infty$. Then the local constant estimator satisfies Theorem 3 with

$$\partial_{\theta_i} H(u) = 2\mathbb{E}\left[\frac{\partial_{\theta_i}\lambda_t^*(u)\,\partial_\theta \lambda_t^*(u)\left(\partial_\theta \lambda_t^*(u)\right)'}{\lambda_t^*(u)^3}\right].$$

5. SIMULATION STUDY

In this section, we examine the finite-sample performances of the local constant and local linear estimators. All reported results are based on 1,000 simulated data sets. The overall performance of the estimators is evaluated using the mean absolute deviation error (MADE), $MADE_i := \frac{1}{n}\sum_{t=1}^{n} \mathbb{E}\left[|\hat{\theta}_i(t/n) - \theta_i(t/n)|\right]$, as well as their integrated bias, variance, and mean squared error. All results are based on the Epanechnikov kernel and with the bandwidth chosen using the cross-validation method proposed in Richter and Dahlhaus (2019).

5.1. Time-Varying ARCH

We first consider the time-varying ARCH(1) in eq. (5), where $\varepsilon \sim i.i.d.N(0,1)$, $\omega(u) = 0.7 - 0.5\sin(4\pi u)$, and $\alpha(u) = 0.45 + 0.4\sin(4\pi u)$. We estimate $\omega(u)$ and $\alpha(u)$ using both Gaussian log-likelihood and the WLS method of Fryzlewicz, Sapatinas, and Subba Rao (2008). Table 1 reports the performance of the estimators. For all sample sizes, the local MLEs perform better than the local WLS estimators.

https://doi.org/10.1017/S0266466626100449 Published online by Cambridge University Press

TABLE 1. Performance of the local constant (LC) and local linear (LL) estimators for tvARCH model: Integrated squared bias (IBias2), integrated variance (IVar), integrated mean squared errors (IMSE), and MADE.

		$\omega(u)$				$\alpha(u)$			
		WLS		ML		WLS		ML	
n		LC	LL	LC	LL	LC	LL	LC	LL
250	IBias2	0.0212	0.0222	0.0202	0.0204	0.0217	0.0279	0.0178	0.0213
	IVar	0.0647	0.0787	0.0549	0.0578	0.0734	0.0848	0.0771	0.0720
	IMSE	0.0859	0.1009	0.0752	0.0782	0.0951	0.1126	0.0948	0.0933
	MADE	0.2223	0.2386	0.2136	0.2141	0.2462	0.2643	0.2414	0.2379
	IBias2BD	0.0397	0.0173	0.0310	0.0215	0.0226	0.0208	0.0181	0.0144
500	IBias2	0.0116	0.0118	0.0091	0.0091	0.0107	0.0140	0.0076	0.0096
	IVar	0.0346	0.0390	0.0313	0.0322	0.0448	0.0484	0.0464	0.0442
	IMSE	0.0461	0.0508	0.0405	0.0413	0.0555	0.0624	0.0541	0.0537
	MADE	0.1651	0.1743	0.1555	0.1556	0.1878	0.1977	0.1813	0.1794
	IBias2BD	0.0383	0.0071	0.0346	0.0128	0.0206	0.0092	0.0207	0.0134
1,000	IBias2	0.0067	0.0060	0.0045	0.0043	0.0060	0.0072	0.0038	0.0049
	IVar	0.0205	0.0225	0.0182	0.0183	0.0278	0.0297	0.0274	0.0257
	IMSE	0.0272	0.0285	0.0228	0.0226	0.0337	0.0369	0.0312	0.0306
	MADE	0.1261	0.1301	0.1152	0.1148	0.1463	0.1515	0.1374	0.1353
	IBias2BD	0.0322	0.0017	0.0274	0.0049	0.0183	0.0039	0.0164	0.0072

For sample sizes of $n = 250$ and $n = 500$, the local constant MLE performs as well as the local linear MLE in terms of the global measures, but the latter performs best in terms of IMSE and MADE for $n = 1,000$. Thus, the overall superiority of the local linear estimator indicated by the theory appears to be a large-sample property.

To compare the performance of the estimators near the end of the sample, we also evaluate the bias of the estimators for the first and last 2.5% of time periods corresponding to $u \in [0, 0.025] \cup [0.975, 1]$. This is reported as IBias2BD in Table 1. As predicted by the theory, we find that, relative to the local constant versions, the local linear WLS and ML estimators enjoy significantly smaller biases near the boundaries of $[0, 1]$ for all sample sizes.

5.2. Time-Varying Poisson Autoregression with Exogenous Covariates (PARX)

We here report simulation results for the local linear MLE of the following PARX(1) model with an additional exogenous regressor $X_{n,t}$:

$$\lambda_{n,t} = \omega(t/n) + \alpha(t/n) Y_{n,t-1} + \gamma(t/n) X_{n,t-1}^2,$$

https://doi.org/10.1017/S0266466626100449 Published online by Cambridge University Press

TABLE 2. Performance of the local linear estimators for tvPARX models: Integrated squared bias (IBias2), integrated variance (IVar), integrated mean squared errors (IMSE), and median of MADE.

		DGP1			DGP2		
n		$\omega(u)$	$\alpha(u)$	$\gamma(u)$	$\omega(u)$	$\alpha(u)$	$\gamma(u)$
250	IBias2	0.0330	0.0037	0.0004	0.0339	0.0035	0.0003
	IVar	0.1550	0.0078	0.0422	0.1409	0.0077	0.0383
	IMSE	0.1880	0.0115	0.0426	0.1748	0.0111	0.0386
	MADE	0.2407	0.0839	0.1463	0.2371	0.0832	0.1447
500	IBias2	0.0111	0.0016	0.0002	0.0137	0.0016	0.0002
	IVar	0.0528	0.0039	0.0216	0.0648	0.0042	0.0208
	IMSE	0.0639	0.0055	0.0218	0.0784	0.0057	0.0209
	MADE	0.1645	0.0586	0.1080	0.1717	0.0593	0.1079
1,000	IBias2	0.0044	0.0007	0.0001	0.0044	0.0008	0.0001
	IVar	0.0266	0.0023	0.0115	0.0286	0.0022	0.0111
	IMSE	0.0310	0.0030	0.0116	0.0330	0.0030	0.0112
	MADE	0.1216	0.0426	0.0798	0.1199	0.0428	0.0798

where $\omega(u) = 0.6 - 0.3u + 0.3\sin(2\pi u)$, $\alpha(u) = 0.3 + 0.3u - 0.3\sin(2\pi u)$, and $\gamma(u) = 1 - 0.5\cos(\pi u)$. The dynamics of the exogenous regressor was chosen as

$$X_{n,t} = \sqrt{\sigma(t/n) + \beta(t/n) X_{n,t-1}^2}\varepsilon_t,\ \varepsilon_t \sim \text{i.i.d.}N(0,1).$$

where either

- DGP1: $\sigma(u) = 0.8$ and $\beta(u) = 0.4$ so that $X_{n,t} = X_t$ is strictly stationary.
- DGP2: $\sigma(u) = 0.8 + 0.4\cos(2\pi u)$ and $\beta(u) = 0.4 - 0.2\sin(2\pi u)$ so $X_{n,t}$ is locally stationary.

Table 2 reports the overall performance of the estimators in terms of integrated squared bias, variance, MSE, and MADE. The table shows that, in all sample sizes, the local linear estimator behaves well for both DGP1 and DGP2. All bias, variance, and MADE decrease as the sample size increases. Finally, similar to the case of the tvARCH model, the local linear estimator performs well near the boundaries; we leave out these results since they are similar to the ones reported for the tv-ARCH model above.

6. EMPIRICAL APPLICATION

We here revisit the empirical analysis of U.S. corporate defaults carried out in Agosto et al. (2016) with the aim of examining whether there is evidence of structural instability in the time series. The data set consists of monthly number of bankruptcies among Moody's rated industrial firms in the United States for the

https://doi.org/10.1017/S0266466626100449 Published online by Cambridge University Press

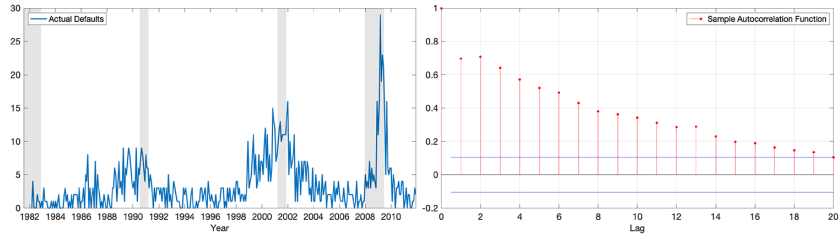


FIGURE 1. Left: Number of defaults per month among Moody's rated U.S. industrial firms in the period 1982–2011; Right: autocorrelation function of defaults.

period 1982–2011 ($n = 360$ observations), collected from Moody's Credit Risk Calculator (CRC). Figure 1 shows the time series of default counts together with its sample autocorrelation function, which reveals high temporal dependence in default counts and existence of default clusters over time.

With $Y_{n,t} \in \{0, 1, 2, \ldots\}$, $t \geq 1$ denoting the number of defaults in a given month, we use a log-PARX model to gain better understanding of the dynamics of $Y_{n,t}$ as a function of its own past, $Y_{n,t-m}$, $m \geq 1$, but also in terms of additional covariates $X_{n,t} \in \mathbb{R}^{d_x}$, which include relevant macroeconomic and financial factors as considered in Agosto et al. (2016). We model $Y_{n,t}$ as a conditional Poisson distribution $Y_{n,t}|\mathcal{F}_{n,t-1} \sim \text{Poisson}\left(\lambda_{n,t}\left(\theta\left(t/n\right)\right)\right), t = 1, 2, \ldots, n$, where the intensity $\lambda_{n,t}\left(\theta\left(t/n\right)\right)$ depends on past counts, covariates $X_{n,t-1}$. and a vector of time-varying parameters $\theta\left(t/n\right)$. Our favored specification of $\lambda_{n,t}\left(\theta\right)$ is the log-PARX model of Fokianos and Tjøstheim (2011), which is here augmented by the chosen set of exogenous variables, $X_{n,t-1}$,

$$\log \lambda_{n,t}\left(\theta\right) = \omega + \sum_{i=1}^{p} \alpha_i \log\left(Y_{n,t-i} + 1\right) + \gamma' X_{n,t-1}, \tag{25}$$

so that $\theta = \left(\omega, \alpha_1, \ldots, \alpha_p, \gamma'\right)'$.

We here deviate from Agosto et al. (2016) that specifies $\lambda_{n,t}$ to be a linear function of past counts and factors. The reasons for us favoring the above log-specification over the linear one are three-fold: First, (25) do not impose positivity constraints on the parameters which facilitates the numerical computation of the estimators; second, it allows us to include any predictors we wish without the need of first transforming them to ensure that each component of the resulting $X_{n,t-1}$ is positive; third, when we estimate both models using the U.S. default data, we found that the log-specification delivers a better fit.

Similar to Agosto et al. (2016), we use $\sum_{i=1}^{p} \alpha_i\left(t/n\right)$ as a measure of contagion in the financial markets: If $\sum_{i=1}^{p} \alpha_i\left(t/n\right)$ is large, then firms defaulting today will lead to a large increase in the risk of other firms defaulting next period everything else equal.

https://doi.org/10.1017/S0266466626100449 Published online by Cambridge University Press

As exogenous covariates, we consider the same financial, credit market, and macroeconomic variables as in Agosto et al. (2016): Realized volatility (RV) computed using daily squared return on the S&P 500 index, the Leading Index released by the Federal Reserve (LI), year-to-year change in Industrial Production Index (IP), one-year return on the S&P 500 index (SPX), the three-month Treasury bill rate ($TB3$), and BAA Moody's rated to 10-year Treasury spread (SP).[2] Agosto et al. (2016) decompose LI and IP into their positive and negative parts to deal with above-mentioned issue of $X_{n,t-1}$ having to be positive in the linear specification. In contrast, no such transformations are needed for our log-specification. Moreover, as well as RV, we also consider the logarithm of RV, $\log(RV)$, to evaluate if the latter is a better predictor; again, this would not be possible in the linear specification of Agosto et al. (2016).

The lag length p is chosen using BIC where, for a given model, the log-likelihood is evaluated at the estimated time-varying parameters.[3] According to this version of BIC, the preferred specification is $p = 2$. Importantly, by allowing for the parameters to be time-varying, a much more parsimonious model is selected by BIC: If we do model selection where for each model we restrict the estimated parameters to be constant over time, the preferred model is $p = 6$. Moreover, the persistence, or contagion, of the estimated time-invariant version, as measured by $\sum_{i=1}^{p} \hat{\alpha}_i$, is substantially higher than for the time-varying version, as measured by $\sum_{i=1}^{p} \hat{\alpha}_i(u)$, $u \in [0,1]$. This is consistent with the findings reported in Hillebrand (2005) for GARCH models: Neglecting structural changes in parameters causes the estimates of these to be suffering from a strong upward bias which in turn leads BIC to selecting a bigger model.

As benchmark, we first estimate the time-invariant version of (25) with $p = 2$. Table 3 shows the estimation results for four different specifications of the time-invariant LLPARX(2) model: Column (1) contains the results when only (RV, LI) are included; column (2) when only $(\log RV, LI)$ are included; column (3) when all covariates are included except for $\log(RV)$; and column (4) when all covariates except RV are included. As in Agosto et al. (2016), once we control for the information contained in RV and LI, none of the other four covariates are found to be relevant in predicting future defaults. We also observe that the two specifications using $\log(RV)$ appear to perform better than the ones using RV. Based on these results, our favored specification of the time-varying version is to use LI and $\log(RV)$ as exogenous variables:

$$\begin{aligned}\log \lambda_{n,t} = \omega(t/n) + \alpha_1(t/n) \log\left(Y_{n,t-1}+1\right) + \alpha_2(t/n) \log\left(Y_{n,t-2}+1\right) \\ + \beta_{RV}(t/n) \log\left(RV_{n,t-1}\right) + \beta_{LI}(t/n)\, LI_{n,t-1}.\end{aligned} \tag{26}$$

[2] These covariates are tested for the existence of default covariates in Das et al. (2007), Duffie, Saita, and Wang (2007), and Lando and Nielsen (2010).

[3] As pointed out by a referee, we are unable to provide a formal justification for using BIC to choose the lag length in our time-varying parameter version of the PARX model. So the application of this model selection criterion to our setting is very ad hoc.

https://doi.org/10.1017/S0266466626100449 Published online by Cambridge University Press

Table 3. Estimation results of different LLPARX models.

	(1)	(2)	(3)	(4)
ω	0.2186	1.0425	0.1003	1.3015
	(0.1022)	(0.2491)	(0.2118)	(0.4188)
α_1	0.3040	0.2777	0.3093	0.2836
	(0.0492)	(0.0496)	(0.0497)	(0.0501)
α_2	0.5033	0.4876	0.5082	0.4956
	(0.0491)	(0.0486)	(0.0499)	(0.0492)
RV	5.8496		5.4864	
	(3.9738)		(4.4365)	
$\log(RV)$		0.1230		0.1484
		(0.0373)		(0.0437)
LI	−0.1767	−0.1482	−0.1864	−0.1860
	(0.0310)	(0.0308)	(0.0490)	(0.0487)
IP			0.0029	−0.0167
			(0.0480)	(0.0481)
SPX			0.2074	0.2412
			(0.2408)	(0.2413)
*TB*3			0.0065	0.0007
			(0.0134)	(0.0136)
SP			0.0300	−0.0426
			(0.0585)	(0.0615)

Note: Standard errors are in parentheses.

Figure 2 shows the time series of the local linear estimates of the time-varying parameters of (26) together with the time-invariant estimates reported in column (2) of Table 3. Pointwise confidence bands are computed based on the asymptotic distribution derived in Theorem 3. The Leading Index is pointwise significant for most of the sample period which highlights the link between macroeconomic activity and corporate defaults also found in Agosto et al. (2016). At the same time, this link exhibits substantial time variation, in particular at the end of the sample during the Great Recession. The link between realized volatility and defaults of industrial firms is less significant but also appears to be changing over time. Similar to Agosto et al. (2016), the realized volatility and the Leading Index are strong explanatory variables during the Great Recession (2007–2011). However, differently, both of them are still relevant in the late 1980s and early 1990s. Especially, the effect of $\log(RV)$ tends to be negative in this period which cannot be captured by the aforementioned linear version of the PARX model.

https://doi.org/10.1017/S0266466626100449 Published online by Cambridge University Press

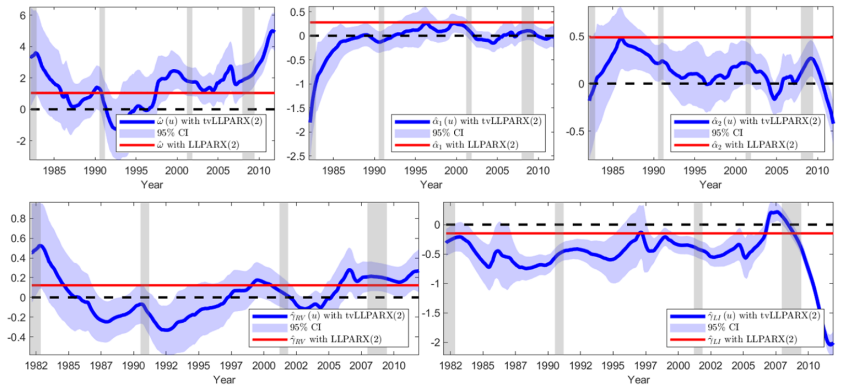


FIGURE 2. Local linear estimate of time-varying parameter in eq. (26): Shaded areas are the 95% confidence intervals.

Finally, judging from $\hat{\alpha}_1(t/n) + \hat{\alpha}_2(t/n)$, there is very little contagion in the financial markets from the 1990s onward, which is somewhat consistent with the findings in Agosto et al. (2016). However, at the end of the sample, we actually find a negative effect of today's log-defaults on tomorrow's default risk. Also note that the time-varying estimates, $\hat{\alpha}_1(t/n)$ and $\hat{\alpha}_2(t/n)$, remain below the corresponding time-invariant ones, $\hat{\alpha}_1$ and $\hat{\alpha}_2$ as marked by the horizontal red lines, throughout the sample period. Again, this seems to indicate that ignoring time variation in the parameters of PARX models lead to over estimation of the level of persistence/contagion.

To assess in-sample fit and whether the reported time variation in the parameters is statistically significant, we carry out an array of graphical and quantitative diagnostic tools for time series. First, we plot in the left panel of Figure 3 the actual default counts together with the predicted defaults $\hat{Y}_{n,t} := \hat{\lambda}_{n,t} = \lambda_{n,t}\left(\hat{\theta}(t/n)\right)$. As can be seen from this plot, the time-varying LLPARX model captures the default counts dynamics well. In the right panel of Figure 3, the sample autocorrelation function of the standardized Pearson residuals $\hat{e}_{n,t} = \hat{\lambda}_{n,t}^{-1/2}\left(Y_{n,t} - \hat{\lambda}_{n,t}\right)$ is plotted. Under correct specification, $e_{n,t}$ should be white noise—the plotted sample autocorrelation function supports this.

We also evaluate the adequacy of fit using the probability integral transform (PIT). We follow Davis and Liu (2016) and compute the PITs by

$$\hat{u}_{n,t} := F_{n,t}\left(Y_{n,t} - 1\right) + \nu_t\left[F_{n,t}\left(Y_{n,t}\right) - F_{n,t}\left(Y_{n,t} - 1\right)\right],$$

where $\{\nu_t\}$ is a sequence of i.i.d. random variables from a standard uniform distribution, and $F_{n,t}$ is the CDF of a Poisson($\hat{\lambda}_{n,t}$) distribution. Under correct model specification, $\hat{u}_{n,t}$ is a sequence of i.i.d. random variables from the standard uniform distribution. Figure 4 depicts the histogram of PIT which show that the

https://doi.org/10.1017/S0266466626100449 Published online by Cambridge University Press

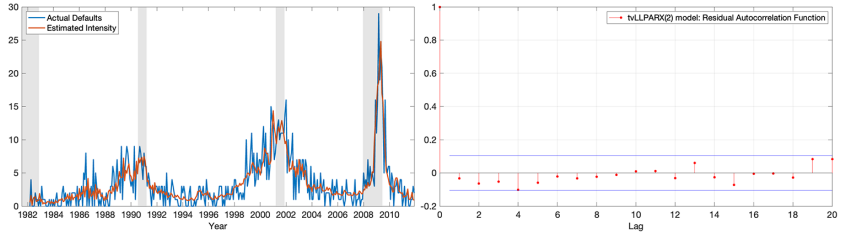


FIGURE 3. Left: Actual number of defaults (blue) and estimated intensity (red); Right: Sample autocorrelation function of Pearson residuals.

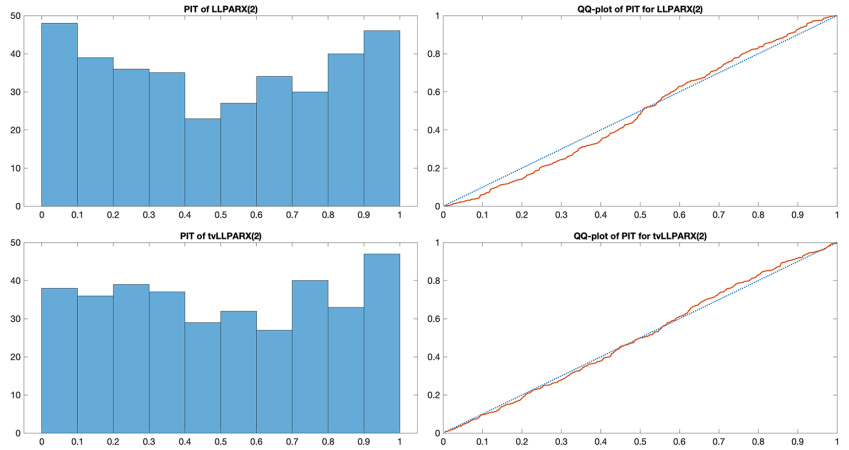


FIGURE 4. Left: Histograms of randomized PITs for log-linear PARX(2) and time-varying log-linear PARX(2) models fitted to the U.S. default data; Right: QQ-plots of the randomized PIT against standard uniform distribution for the corresponding models.

tvLLPARX(2) model provides a better in-sample fit than the corresponding time-invariant model.

Finally, in Table 4, we report the log-likelihood, AIC, and BIC values and the *p*-value from a Kolmogorov–Smirnov test of the PITs being uniformly distributed of each of four specifications in columns (1)–(4) of Table 3 together with the time-varying versions of (2) and (4), labeled tv (2) and tv(4), respectively. From these, we see that allowing for time-varying parameters increase the in-sample fit dramatically. While this is not a formal statistical test of time variation, it provides strong informal evidence of such. As mentioned earlier, we also see that specification (2) is favored over (1), (3), and (4) in the time-invariant case, and that (2) is favored over (4) in the time-varying case.

We complete the analysis by conducting a final sensitivity analysis of (26). This is done by including the remaining exogenous covariates in addition to realized

https://doi.org/10.1017/S0266466626100449 Published online by Cambridge University Press

Table 4. In-sample fit of time-invariant and time-varying LLPARX models.

	(1)	(2)	tv (2)	(3)	(4)	tv (4)
logL	696.3	700.6	777.7	696.8	701.8	778.6
AIC	−1,382.5	−1,391.2	−1,526.6	−1,375.6	−1,385.5	−1,539.3
BIC	−1,363.1	−1,371.7	−1,507.2	−1,340.6	−1,350.5	−1,504.3
p-value of PIT	0.0726	0.1319	0.5432	0.0588	0.1544	0.6445

volatility and Leading Index in the time-varying version of the model:

$$\begin{aligned}\log \lambda_{n,t} = {} & \omega(t/n) + \alpha_1(t/n)\log\left(Y_{n,t-1}+1\right) + \alpha_2(t/n)\log\left(Y_{n,t-2}+1\right) \\ & + \gamma_{RV}(t/n)\log\left(RV_{n,t-1}\right) + \gamma_{LI}(t/n)\,LI_{n,t-1} + \gamma_{IP}(t/n)\,IP_{n,t-1} \\ & + \gamma_{SPX}(t/n)\,SPX_{n,t-1} + \gamma_{TB3}(t/n)\,TB3_{n,t-1} + \gamma_{SP}(t/n)\,SP_{n,t-1}. \end{aligned} \tag{27}$$

The estimation results are provided in Figure 5. The estimates, except for the realized volatility, are consistent with our baseline findings. The Leading Index remains highly significant, which shows that macroeconomic factors are relevant in predicting future defaults. The link between short-term interest rates and defaults of industrial firms changes over time. During the late 1980s and the Great Recession (2007–2011), interest rates play a role in determining the interest expense of firms. In the 1990s, similar to the finding in Duffie et al. (2007), the sign of the coefficient for the short-term rate is consistent with the fact that the U.S. Federal Reserve often increases the short-term rates to control business expansions. Controlling for *LI* and *TB*3, other covariates are estimated to be insignificant for most of the sample period except for the period of the Great Recession (2007–2011), in which financial, credit market, and macroeconomic variables are significant explanators of the default intensity. These are novel findings that the original analysis of Agosto et al. (2016) did not reveal.

7. CONCLUSION AND FUTURE WORK

We have here developed a general theory for kernel regression-type estimators of time-varying parameters and used this to provide primitive conditions for the estimators in Markov models. Another interesting application would be infinite memory models of the form

$$Y_{n,t} = G\left(Y_{n,t-1}, Y_{n,t-2}, \ldots, \varepsilon_t; \theta(t/n)\right).$$

The local constant estimator of $\theta(u)$ was analyzed in Bardet et al. (2022) using techniques different from ours. Specifically, they provide primitive conditions

https://doi.org/10.1017/S0266466626100449 Published online by Cambridge University Press

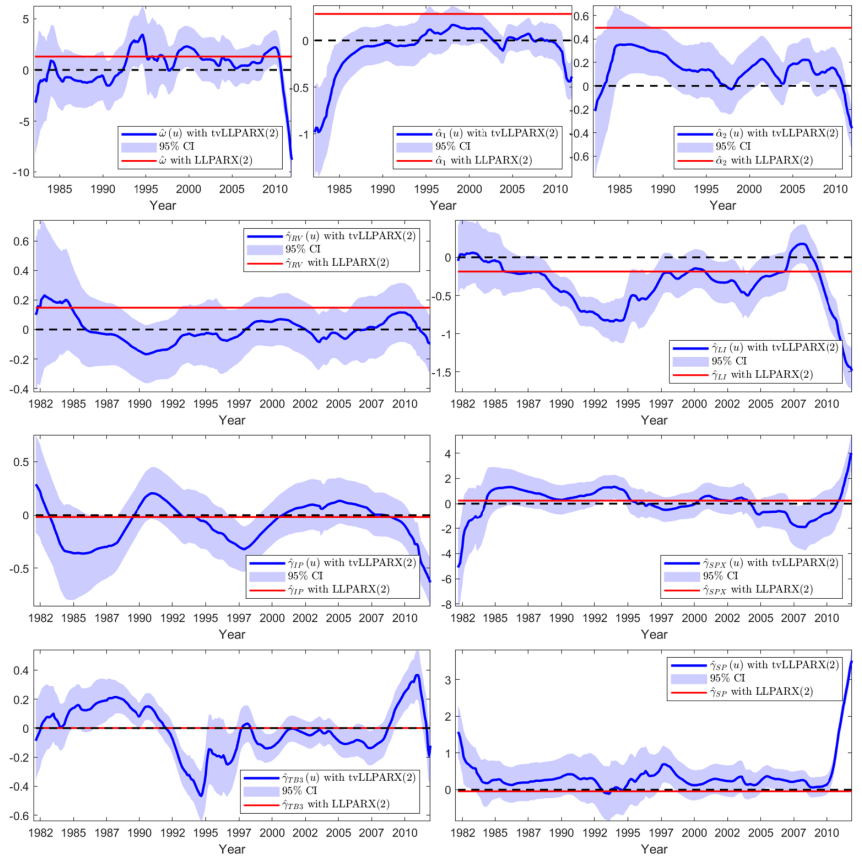


FIGURE 5. Local linear estimate of time-varying parameter in eq. (27): Shaded areas are the 95% confidence intervals.

under which $Y_{n,t}$ satisfies

$$\sup_{[n(u-cb)]\leq t\leq [n(u-cb)]+2cnb} E\left[\left\|Y_{n,t}-Y_t^*(u)\right\|^p\right]^{1/p} \leq C\left(b^\rho+\lambda_{[n(u-cb)]+t}\right),$$

where $Y_t^*(u)$ is the stationary solution to $Y_t^*(u)=G\left(Y_{t-1}^*(u),Y_{t-2}^*(u),\ldots,\varepsilon_t;\right.$ $\left.\theta(u)\right)$. But this result unfortunately does not suffice for our theory since this requires $Y_{n,t}$ to satisfy our version of local stationarity as given in Definition 1. At the same time, Bardet et al. (2022) do provide primitive conditions for the assumptions on the stationary version of Corollary 3 to hold. Thus, under the high-level assumption that $Y_{n,t}$ satisfies our version of local stationarity, we can combine the results of Bardet et al. (2022) with our Corollary 3 to obtain a theory for local polynomial estimators of infinite memory models. Under this high-level

https://doi.org/10.1017/S0266466626100449 Published online by Cambridge University Press

condition, we are also able to provide a precise characterization of the leading bias term of their estimator, something which was missing in the analysis of Bardet et al. (2022). We leave it to future research to establish primitive conditions under which infinite memory models satisfy our Definition 1.

A. Appendix

A.1. Auxiliary Results

In the following, assume that L satisfies: (i) $L(\cdot)$ has a compact support and (ii) for some $\Lambda < \infty$, $\left|L(v) - L(v')\right| \leq \Lambda \left|v - v'\right|$, $v, v' \in \mathbb{R}$. We denote $L_b(\cdot) := L(\cdot/b)/b$.

LEMMA A1. *The following hold as $b \to 0$ and $nb \to \infty$:*

(i) Suppose $\{W_{n,t}(\theta)\}$ is ULS (p, r, Θ) with its stationary approximation $\{W_t^*(\theta|u)\}$ being L_p continuous for some $p \geq 1, r > 0$, and Θ is compact. Then, with $\mathcal{A}$ defined in Assumption 2 and for any $u \in (0,1)$,

$$\sup_{\alpha \in \mathcal{A}} \left\| \frac{1}{n} \sum_{t=1}^{n} L_b(t/n - u) W_{n,t}\left(D_{m,b}(t/n - u)\alpha\right) - \int L(v) \mathbb{E}\left[W_t^*(D_m(v)\alpha|u)\right] dv \right\| = o_p(1).$$

(ii) Suppose $\{W_{n,t}, \mathcal{F}_{n,t}\}$ is an MGD array and LS (p, r) for some $p \geq 2$ and $r > 0$ with its stationary approximation $W_t^*(u)$ being L_p-continuous. Then, for any $u \in (0,1)$,

$$\sqrt{\frac{b}{n}} \sum_{t=1}^{n} L_b(t/n - u) W_{n,t} \to^d N\left(0, \int L^2(v)\, dv \times \mathbb{E}\left[W_t^*(u)^2\right]\right);$$

$$\sqrt{\frac{b}{n}} \sum_{t=1}^{n} L_b(t/n - cb) W_{n,t} \to^d N\left(0, \int_{-c}^{+\infty} L^2(v)\, dv \times \mathbb{E}\left[W_t^*(u)^2\right]\right).$$

(iii) Suppose $W_t^*(u)$ is stationary and ergodic with $\sum_{s=0}^{\infty} \left|\mathrm{Cov}\left(W_t^*(u), W_{t+s}^*(u)\right)\right| < \infty$. Then,

$$\left| \frac{1}{n} \sum_{t=1}^{n} L_b(t/n - u) W_t^*(u) - \int L(v)\, dv \times \mathbb{E}\left[W_t^*(u)\right] \right| = o_p\left(1/\sqrt{nb}\right).$$

Proof of (i). We first show that for all $\theta \in \Theta$,

$$\frac{1}{n} \sum_{t=1}^{n} L_b(t/n - u) W_{n,t}(\theta) \to^p \int L(v)\, dv \times \mathbb{E}\left[W_t^*(\theta|u)\right].$$

Note that $L(v) = 0$ for $|v| \geq \bar{v}$ for some $\bar{v} > 0$. Then, Minkowski's inequality implies that

$$\mathbb{E}\left[\left\| \frac{1}{n} \sum_{t=1}^{n} L_b(t/n - u) \left\{W_{n,t}(\theta) - W_t^*(\theta|u)\right\} \right\|^p \right]^{1/p}$$
$$\leq \frac{1}{n} \sum_{t=1}^{n} |L_b(t/n - u)| \, \mathbb{E}\left[\left\| W_{n,t}(\theta) - W_t^*(\theta|u) \right\|^p\right]^{1/p}$$

https://doi.org/10.1017/S0266466626100449 Published online by Cambridge University Press

$$\leq \frac{C}{n}\sum_{t=1}^{n}|L_b(t/n-u)|\left(b^r\left|\frac{t/n-u}{b}\right|^r+1/n^r+\rho^t\right)$$
$$\leq \frac{C}{n}\sum_{t=1}^{n}|L_b(t/n-u)|\times\left(b^r\bar{v}^r+1/n^r+\rho^t\right)=O\left(b^r\right)+O\left(n^{-r}\right)+O\left(\frac{1}{\sqrt{nb}}\right),$$

where we have used that

$$\frac{1}{n}\sum_{t=1}^{n}|L_b(t/n-u)|\rho^{rt}\leq\frac{1}{\sqrt{nb}}\sqrt{\frac{1}{n}\sum_{t=1}^{n}\left(L^2\right)_b(t/n-u)}\sqrt{\sum_{t=1}^{n}\rho^{2rt}}=O\left(\frac{1}{\sqrt{nb}}\right).$$

Next, with $\bar{W}_t=W_t^*(\theta|u)-\mathbb{E}\left[W_t^*(\theta|u)\right]$, $\frac{1}{n}\sum_{t=1}^{n}L_b(t/n-u)\bar{W}_t=\frac{1}{nb}\sum_{t=\underline{t}}^{\bar{t}}L_b(t/n-u)\bar{W}_t$ for sufficiently large n, where $\bar{t}=[n(u+\bar{v}b)]$ and $\underline{t}=[n(u-\bar{v}b)]$. Here, $[x]$ denotes the integer part of any real number x. By summation by parts, we have, with $S_{n,t}=\sum_{j=\underline{t}}^{t}\bar{W}_j$,

$$\frac{1}{n}\sum_{t=\underline{t}}^{\bar{t}}L_b(t/n-u)\bar{W}_t=\frac{1}{n}\sum_{t=\underline{t}}^{\bar{t}}L_b(t/n-u)\left(S_{n,t}-S_{n,t-1}\right)$$
$$=\frac{1}{n}\sum_{t=\underline{t}}^{\bar{t}-1}\left[L_b(t/n-u)-L_b((t+1)/n-u)\right]S_{n,t}+\frac{1}{n}L_b\left(\bar{t}/n-u\right)S_{n,\bar{t}}.$$

Since $\left\{\bar{W}_t\right\}$ is stationary, $S_{n,t}$ has the same distribution as $\tilde{S}_{n,t}=\sum_{j=1}^{t-\underline{t}+1}\bar{W}_j$. Thus, for some constant M, $\left|\frac{1}{n}\sum_{t=1}^{n}L_b(t/n-u)\bar{W}_t\right|\leq\frac{M}{nb}\sup_{t\leq\bar{t}-\underline{t}+1}\left|\tilde{S}_{n,t}\right|$. The ergodic theorem yields $\tilde{S}_{n,t}/t\to 0$ which in turn implies that $\frac{1}{n}\sum_{t=1}^{n}L_b(t/n-u)\bar{W}_t$ tends to zero almost surely. Finally, using the mean-value theorem, there exists $v_{n,t}\in\left[\frac{t-1}{n},\frac{t}{n}\right]$ so that with $\bar{L}=\sup_v L(v)$,

$$\left|\frac{1}{n}\sum_{t=1}^{n}L_b(t/n-u)-\int L_b(x-u)\,dx\right|=\left|\frac{1}{nb}\sum_{t=1}^{n}L_b(t/n-u)-\sum_{t=1}^{n}\int_{(t-1)/n}^{t/n}L_b(x-u)\,dx\right|$$
$$\leq\frac{1}{nb}\sum_{t=1}^{n}\left|L_b(t/n-u)-L_b\left(v_{n,t}-u\right)\right|$$
$$\leq\frac{1}{nb}\sum_{t=1}^{n}\Lambda\left|\frac{t/n-v_{n,t}}{b}\right|=O\left(\frac{1}{nb}\right),$$

which shows that $\frac{1}{n}\sum_{t=1}^{n}L_b(t/n-u)\mathbb{E}\left[W_t^*(\theta|u)\right]=\int L_b(x-u)\,dx\mathbb{E}\left[W_t^*(\theta|u)\right]+O(1/(nb))$.

For the uniform convergence, we note that by definition of $\mathcal{A}$, $D_b(v-u)\alpha\in\Theta$ for all $v\in\operatorname{supp}(L)$ and $\alpha\in\mathcal{A}$. Thus, $\frac{1}{n}\sum_{t=1}^{n}K_b(t/n-u)W_{n,t}\left(D_{n,t}(u)\alpha\right)$, where $D_{n,t}(u)=D_b(t/n-u)$, is well-defined for $\alpha\in\mathcal{A}$, and

$$\mathbb{E}\left[\sup_{\alpha\in\mathcal{A}}\left\|W_{n,t}\left(D_{n,t}(u)\alpha\right)-W_t^*\left(D_{n,t}(u)\alpha|u\right)\right\|^p\right]\leq\mathbb{E}\left[\sup_{\theta\in\Theta}\left\|W_{n,t}(\theta)-W_t^*(\theta|u)\right\|^p\right]$$
$$\leq C\left(|t/n-u|^r+1/n^r+\rho^t\right)^p.$$

https://doi.org/10.1017/S0266466626100449 Published online by Cambridge University Press

Using Hölder's and Minkowski's inequalities,

$$\begin{aligned}
&\mathbb{E}\left[\sup_{\alpha\in\mathcal{A}}\left\|\frac{1}{n}\sum_{t=1}^{n}L_b\left(t/n-u\right)\left\{W_{n,t}\left(D_{n,t}\left(u\right)\alpha\right)-W_t^*\left(D_{n,t}\left(u\right)\alpha|u\right)\right\}\right\|\right]\\
&\quad\leq\frac{1}{n}\sum_{t=1}^{n}\left|L_b\left(t/n-u\right)\right|\mathbb{E}\left[\sup_{\alpha\in\mathcal{A}}\left\|W_{n,t}\left(D_{n,t}\left(u\right)\alpha\right)-W_t^*\left(D_{n,t}\left(u\right)\alpha|u\right)\right\|^p\right]^{1/p}\\
&\quad\leq C\frac{b^r}{n}\sum_{t=1}^{n}\left|L_b\left(t/n-u\right)\right|\left(\left|\frac{t/n-u}{b}\right|^r+1/n^r+\rho^t\right)=O\left(b^r\right).
\end{aligned}$$

Next,

$$\begin{aligned}
&\sup_{\alpha\in\mathcal{A}}\left\|\frac{1}{n}\sum_{t=1}^{n}L_b\left(t/n-u\right)\left\{W_t^*\left(D_{n,t}\left(u\right)\alpha|u\right)-\mathbb{E}\left[W_t^*\left(D_{n,t}\left(u\right)\alpha|u\right)\right]\right\}\right\|\\
&\quad\leq\sup_{\theta\in\Theta}\left\|\frac{1}{n}\sum_{t=1}^{n}L_b\left(t/n-u\right)\left\{W_t^*\left(\theta|u\right)-\mathbb{E}\left[W_t^*\left(\theta|u\right)\right]\right\}\right\|+o_p\left(1\right),
\end{aligned}$$

where $\frac{1}{n}\sum_{t=1}^{n}L_b\left(t/n-u\right)\left\{W_t^*\left(\theta|u\right)-\mathbb{E}\left[W_t^*\left(\theta|u\right)\right]\right\}=o_P\left(1\right)$ for all $\theta\in\Theta$. Thus, the result will follow if we can show stochastic equicontinuity of $\theta\mapsto\frac{1}{n}\sum_{t=1}^{n}L_b\left(t/n-u\right)W_t^*\left(\theta|u\right)$ but this follows from the assumption of $\theta\mapsto W_t^*\left(\theta|u\right)$ being L_p continuous: For a given $\theta\in\Theta$ and $\epsilon>0$, there exists $\delta>0$ so that

$$\begin{aligned}
&\mathbb{E}\left[\sup_{\theta':\|\theta-\theta'\|<\delta}\left\|\frac{1}{n}\sum_{t=1}^{n}L_b\left(t/n-u\right)W_t^*\left(\theta|u\right)-\frac{1}{n}\sum_{t=1}^{n}L_b\left(t/n-u\right)W_t^*\left(\theta'|u\right)\right\|\right]\\
&\quad\leq\frac{1}{n}\sum_{t=1}^{n}\left|L_b\left(t/n-u\right)\right|\mathbb{E}\left[\sup_{\theta':\|\theta-\theta'\|<\delta}\left\|W_t^*\left(\theta|u\right)-W_t^*\left(\theta'|u\right)\right\|\right]\\
&\quad=\frac{\epsilon}{n}\sum_{t=1}^{n}\left|L_b\left(t/n-u\right)\right|=O\left(\epsilon\right).
\end{aligned}$$

Proof of (ii). Observe that $\sqrt{b/n}\sum_{t=1}^{n}L_b\left(t/n-u\right)W_{n,t}$ is a martingale with quadratic variation $Q_n=\frac{b}{n}\sum_{t=1}^{n}L_b^2\left(t/n-u\right)W_{n,t}^2$. Since $W_{n,t}$ is LS $(2,r)$, it follows from part (i) that $Q_n\to^p\int L^2\left(x\right)dx\mathbb{E}\left[W_t^*\left(u\right)^2\right]$. The result will then follow if the Lindeberg condition is satisfied (cf. Brown, 1971). With $m_{n,t}=\sqrt{b/n}L_b\left(t/n-u\right)W_{n,t}$ and $m_t^*\left(u\right)=\sqrt{b/n}L_b\left(t/n-u\right)W_t^*\left(u\right)$,

$$\begin{aligned}
\sum_{t=1}^{n}\left\|m_{n,t}\right\|^2 1\left(\left\|m_{n,t}\right\|>\varepsilon\right)\leq&\sum_{t=1}^{n}\left(\left\|m_{n,t}\right\|^2-\left\|m_t^*\left(u\right)\right\|^2\right)1\left(\left\|m_{n,t}\right\|>\varepsilon\right)\\
&+\sum_{t=1}^{n}\left\|m_t^*\left(u\right)\right\|^2 1\left(\left\|m_{n,t}\right\|>\varepsilon,\left\|m_t^*\left(u\right)\right\|\leq\varepsilon/\sqrt{2}\right)\\
&+\sum_{t=1}^{n}\left\|m_t^*\left(u\right)\right\|^2 1\left(\left\|m_t^*\left(u\right)\right\|>\varepsilon/\sqrt{2}\right).
\end{aligned}$$

https://doi.org/10.1017/S0266466626100449 Published online by Cambridge University Press

Recycling the arguments used in the proof of part (i), the first and third terms are $o_p(1)$. For the second term, employ the following inequality and Markov's inequality:

$$\sum_{t=1}^{n} \left\| m_t^*(u) \right\|^2 1\left(\left\| m_{n,t} \right\| > \varepsilon, \left\| m_t^*(u) \right\| \leq \varepsilon/\sqrt{2} \right)$$
$$\leq \frac{\varepsilon^2}{2} \sum_{t=1}^{n} 1\left(\left\| m_{n,t} \right\|^2 - \left\| m_t^*(u) \right\|^2 > \varepsilon^2/2 \right).$$

Proof of (iii). Assuming w.l.o.g. that $\mathbb{E}\left[W_t^*\right] = 0$,

$$\operatorname{Var}\left(\frac{1}{n} \sum_{t=1}^{n} L_b(t/n - u) W_t^*(u) \right) \leq \frac{1}{n^2} \sum_{t_1, t_2 = 1}^{n} \left| L_b(t_1/n - u) \right| \left| L_b(t_2/n - u) \right| \left| \mathbb{E}\left[W_{t_1}^* W_{t_2}^* \right] \right|$$
$$\leq \frac{\bar{L}}{(nb)^2} \sum_{t_1, t_2 = 1}^{n} \left| L\left(\frac{t_1/n - u}{b} \right) \right| \left| \mathbb{E}\left[W_{t_1}^* W_{t_2}^* \right] \right| = O\left(\frac{1}{nb} \right).$$

□

A.2. Proofs of Results in Section 3

Proof of Theorem 1. By Lemma A(i), $\sup_{\alpha \in \mathcal{A}} \left| Q_n(\alpha|u) - Q^*(\alpha|u) \right| = o_P(1)$, where $Q^*(\alpha|u) = \int_{\mathbb{R}} K(v) \mathbb{E}\left[\ell_t^*(D_m(v)\alpha|u) \right] dv$. Now, observe that for any $\alpha = (\alpha_1, \ldots, \alpha_{m+1})$ with $\alpha_i \neq 0$ for some $i \geq 2$, the polynomial $v \mapsto D_m(v)\alpha$ is non-constant almost everywhere. Thus, for any $\alpha \neq \alpha^* = (\theta(u), 0, \ldots, 0)$, $D_m(v)\alpha \neq \theta(u) = D_m(v)\alpha^*$ for almost all $v \in [0,1]$ and so by Assumption 2(iii) $\mathbb{E}\left[\ell_t^*(D_m(v)\alpha|u) \right] < \mathbb{E}\left[\ell_t^*(\theta(u)|u) \right] = \mathbb{E}\left[\ell_t^*\left(D_m(v)\alpha^*|u \right) \right]$ for almost every v. Since $K(\cdot) \geq 0$, this in turn implies that $Q^*(\alpha|u) \leq Q^*\left(\alpha^*|u\right)$. Finally, by the dominated convergence theorem together with Assumption 2(ii), $\alpha \mapsto Q^*(\alpha|u)$ is continuous. This proves $\hat{\alpha} \to^p \alpha^*$ (cf. Theorem 2.1 in Newey and McFadden, 1994). □

Proof of Theorem 2. From Theorem 1, we know that $\hat{\alpha} \to^p \alpha^* := (\theta(u), 0, \ldots, 0)$. It is easily checked that the limit is situated in the interior of $\mathcal{A}$ and so w.p.a.1. so will $\hat{\alpha}$. As a consequence, $\hat{\alpha}$ will satisfy (9) w.p.a.1. Adding and subtracting $S_n(u)$ and then rearranging yields

$$0 = \sqrt{nb} S_n(u) + H_n(\bar{\alpha}|u) \sqrt{nb} \left(\hat{\alpha} - \alpha_0 - H_n^{-1}(\bar{\alpha}|u) \{ S_n(\alpha_0|u) - S_n(u) \} \right).$$

Here, $H_n^{-1}(\bar{\alpha}|u)$ is well-defined w.p.a.1 since, as shown below, it converges toward an invertible matrix. The claimed asymptotic result now follows if we can verify the claims of eqs. (13) and (14).

Proof of eq. (13). With $L(u) = K(u) D_m(u)$, $\sqrt{nb} S_n(u) = \sqrt{\frac{b}{n}} \sum_{t=1}^{n} L_b(t/n - u) \otimes s_{n,t}$. The result now follows from Lemma A(ii) under Assumption 4.

Proof of first claim of eq. (14). With $L(u) = K(u) D_m(u) D_m(u)'$, we can write $H_n(\beta|u) = \frac{1}{n} \sum_{t=1}^{n} L_b(t/n - u) \otimes h_{n,t}\left(D_{n,t}(u)\beta \right)$. It follows from Lemma A(i) and Assumption 5 that

$$\sup_{\alpha \in \mathcal{B}(\epsilon)} \left\| H_n(\alpha|u) - \int K(v) D_m(v) D_m(v)' \otimes H(D_m(v)\alpha \mid u) dv \right\| = o_p(1),$$

https://doi.org/10.1017/S0266466626100449 Published online by Cambridge University Press

where $H(\theta|u) = \mathbb{E}\left[h_t^*(\theta|u)\right]$ is continuous in θ and $\mathcal{B}(\epsilon) = \{\alpha : \|\alpha - \alpha^*\| < \epsilon\}$. Since $\bar{\alpha} \to_p \alpha^*$ and $D_m(v)\alpha^* = \theta(u)$ for all v, continuity of $H(\cdot \mid u)$ yields

$$H_n(\bar{\alpha}|u) \to_p \mathcal{K}_1 \otimes H(\theta(u)|u) = \mathcal{K}_1 \otimes H(u).$$

Finally, $\mathbb{K}_1 = \int K(v) D_m(v) D_m(v)' dv$ is positive definite under standard kernel conditions, and $H(u)$ is invertible by assumption.

Proof of second claim of eq. (14). First observe that $D_{n,t}(u)\alpha^* = \theta_u^*(t/n)$, where $\theta_u^*(t/n)$ was defined in eq. (1). Now, employ the mean-value theorem to obtain that, for some $\bar{\theta}_{n,t}$ lying between $\theta_u^*(t/n)$ and $\theta(t/n)$ and some $u_{n,t} \in [t/n, u]$,

$$\begin{aligned}
b_{n,t} &= s_{n,t}\left(\theta_u^*(t/n)\right) - s_{n,t} \\
&= h_{n,t}\left(\bar{\theta}_{n,t}\right)\left\{\theta_u^*(t/n) - \theta(t/n)\right\} = -h_{n,t}\left(\bar{\theta}_{n,t}\right)\frac{\theta^{(m+1)}\left(u_{n,t}\right)}{(m+1)!}(t/n-u)^{m+1} \\
&= -h_{n,t}(\theta(t/n))\frac{\theta^{(m+1)}(t/n)}{(m+1)!}(t/n-u)^{m+1} \\
&\quad + \left\{h_{n,t}(\theta(t/n))\theta^{(m+1)}(t/n) - h_{n,t}\left(\bar{\theta}_{n,t}\right)\theta^{(m+1)}\left(u_{n,t}\right)\right\}\frac{(t/n-u)^{m+1}}{(m+1)!} \\
&=: b_{n,t}^{(1)} + b_{n,t}^{(2)}.
\end{aligned}$$

The first term is locally stationary and so by the same arguments as in the proof of Lemma A(ii):

$$\begin{aligned}
\frac{1}{n}\sum_{t=1}^{n} K_{n,t}(u) D_{n,t}(u)' b_{n,t}^{(1)} &= -\frac{b^{m+1}}{n}\sum_{t=1}^{n} K_{n,t}(u) D_{n,t}(u)' \left(\frac{t/n-u}{b}\right)^{m+1} \\
&\qquad h_{n,t}(\theta(t/n))\frac{\theta^{(m+1)}(t/n)}{(m+1)!} \\
&= -b^{m+1}\left\{\mu_1 \otimes H(u)\frac{\theta^{(m+1)}(u)}{(m+1)!} + o_p(1)\right\}.
\end{aligned}$$

For the second term, observe that for $|t/n-u| \leq Cb$, $\|\bar{\theta}_{n,t} - \theta(t/n)\| \leq \|\theta_u^*(t/n) - \theta(t/n)\| \leq \tilde{C}b^{m+1}$, and so, using the ULS property of $h_{n,t}(\theta)$,

$$\begin{aligned}
&\sup_{n,t}\mathbb{E}\left[\|h_{n,t}(\theta(t/n)) - h_{n,t}\left(\bar{\theta}_{n,t}\right)\|\right] \\
&\leq C\left(b^r\left|\frac{t/n-u}{b}\right|^r + 1/n^r\right) + \sup_{\|\theta-\theta^*\|\leq\tilde{C}b^{m+1}}\mathbb{E}\left[\|h_t^*(\theta|u) - h_t^*\left(\theta'|u\right)\|\right] \to 0,
\end{aligned}$$

as $n \to \infty$ and $b \to 0$. Similarly, $\sup_{n,t}\|\theta^{(m+1)}(t/n) - \theta^{(m+1)}\left(u_{n,t}\right)\| \to 0$ as $n \to \infty$. These two results combined show that the $\frac{b^{m+1}}{n}\sum_{t=1}^{n} K_{n,t}(u) D_{n,t}(u)' \left(\frac{t/n-u}{b}\right)^{m+1} b_{n,t}^{(2)} = o_p(1)$. □

Proof of Theorem 3. Write

$$b_{n,t} = s_{n,t}\left(\theta_u^*(t/n)\right) - s_{n,t} = s_t^*\left(\theta_u^*(t/n), t/n\right) - s_t^*(\theta(t/n), t/n) + O_P\left(n^{-r}\right).$$

https://doi.org/10.1017/S0266466626100449 Published online by Cambridge University Press

The proof proceeds exactly as the one of Theorem 2 except that the first-order expansion of $b_{n,t}$ is replaced by a second-order expansion of $s_{n,t}(\theta)$ w.r.t. θ combined with

$$\theta(t/n) = \theta_u^*(t/n) + \frac{\theta^{(m+1)}(u)}{(m+1)!}\{t/n-u\}^{m+1} + \frac{\theta^{(m+2)}(\bar{u}_{n,t})}{(m+2)!}\{t/n-u\}^{m+2},$$

where $u_{n,t}$ lies between u and t/n. This yields

$$\begin{aligned}
b_{n,t} &= h_{n,t}(\theta(t/n))\left\{\theta_u^*(t/n) - \theta(t/n)\right\} \\
&\quad + \frac{1}{2}\sum_{i=1}^{d_\theta}\left\{\theta_{u,i}^*(t/n) - \theta_i(t/n)\right\}\frac{\partial h_{n,t}(\bar{\theta}_{n,t})}{\partial\theta_i}\left\{\theta_u^*(t/n) - \theta(t/n)\right\} \\
&= -h_{n,t}(\theta(t/n))\left[\frac{\theta^{(m+1)}(u)}{(m+1)!}\{t/n-u\}^{m+1} + \frac{\theta^{(m+2)}(u_{n,t})}{(m+2)!}\{t/n-u\}^{m+2}\right] \\
&\quad + \frac{1}{2}\sum_{i=1}^{d_\theta}\frac{\theta_i^{(m+1)}(u_{n,t})}{(m+1)!}\frac{\partial h_{n,t}(\bar{\theta}_{n,t})}{\partial\theta_i}\frac{\theta^{(m+1)}(u_{n,t})}{(m+1)!}\{t/n-u\}^{2m+2}.
\end{aligned}$$

Thus, with $L_{n,t}^{(m)}(u) = K_{n,t}(u)D_{n,t}(u)'\left(\frac{t/n-u}{b}\right)^m$,

$$\begin{aligned}
B_n(u) &= \frac{1}{n}\sum_{t=1}^{n}K_{n,t}(u)D_{n,t}(u)'b_{n,t} \\
&= -b^{m+1}\left\{\frac{1}{n}\sum_{t=1}^{n}L_{n,t}^{(m+1)}(u)h_{n,t}(\theta(t/n))\right\}\frac{\theta^{(m+1)}(u)}{(m+1)!} \\
&\quad - b^{m+2}\left\{\frac{1}{n}\sum_{t=1}^{n}L_{n,t}^{(m+2)}(u)h_{n,t}(\theta(t/n))\theta^{(m+2)}(u_{n,t})\right\}\frac{1}{(m+2)!} \\
&\quad + \frac{b^{2m+2}}{2\{(m+1)!\}^2}\sum_{i=1}^{d_\theta}\left\{\frac{1}{n}\sum_{t=1}^{n}\theta_i^{(m+1)}(u_{n,t})L_{n,t}^{(2m+2)}(u)\frac{\partial h_{n,t}(\bar{\theta}_{n,t})}{\partial\theta_i}\theta^{(m+1)}(u_{n,t})\right\}.
\end{aligned}$$

We now analyze the three bracketed terms. First,

$$\begin{aligned}
&\frac{1}{n}\sum_{t=1}^{n}L_{n,t}^{(m+1)}(u)h_{n,t}(\theta(t/n)) \\
&= \frac{1}{n}\sum_{t=1}^{n}L_{n,t}^{(m+1)}(u)\left[h_{n,t}(\theta(t/n)) - h_{n,t}(\theta(u))\right] \\
&\quad + \frac{1}{n}\sum_{t=1}^{n}L_{n,t}^{(m+1)}(u)h_t^*(\theta(u)|u) + \frac{b}{n}\sum_{t=1}^{n}L_{n,t}^{(m+2)}(u)\partial_u h_t^*(\theta(u)|u) \\
&\quad + \frac{1}{n}\sum_{t=1}^{n}L_{n,t}^{(m+1)}(u)\left[h_{n,t}(\theta(u)) - h_t^*(\theta(u)|u) - \partial_u h_t^*(\theta(u)|u)\{t/n-u\}\right].
\end{aligned}$$

https://doi.org/10.1017/S0266466626100449 Published online by Cambridge University Press

By Lemma A(iii) together with Assumption 8 and Lemma A(i),

$$\frac{1}{n}\sum_{t=1}^{n} L_{n,t}^{(m+1)}(u)\left[h_{n,t}\left(\theta\left(t/n\right)\right)-h_{n,t}\left(\theta\left(u\right)\right)\right]=b\mu_{m+2}\sum_{i=1}^{d_\theta}\theta_i^{(1)}\left(u\right)\partial_{\theta_i}H\left(u\right)+o_P\left(b\right),$$

$$\frac{1}{n}\sum_{t=1}^{n} L_{n,t}^{(m+1)}(u)h_t^*\left(\theta\left(u\right)|u\right)=\mu_{m+1}H\left(u\right)+o_P\left(1/\sqrt{nb}\right), \tag{A.1}$$

$$\frac{1}{n}\sum_{t=1}^{n} L_{n,t}^{(m+2)}(u)\partial_u h_t^*\left(\theta\left(u\right)|u\right)=\mu_{m+2}\partial_u H\left(u\right)+o_P\left(1\right),$$

while, using Assumption 7, with $h_{n,t}\left(u\right)=h_{n,t}\left(\theta\left(u\right)\right)$ and $h_t^*\left(u\right)=h_t^*\left(\theta\left(u\right)|u\right)$,

$$\begin{aligned}
&\frac{1}{n}\sum_{t=1}^{n}\|L_{n,t}^{(m+1)}(u)\|\mathbb{E}\left[\|h_{n,t}\left(u\right)-h_t^*\left(u\right)-\partial_u h_t^*\left(u\right)\left\{t/n-u\right\}\|\right]\\
&\quad\leq\frac{1}{n}\sum_{t=1}^{n}\|L_{n,t}^{(m+1)}(u)\|C\left(1/n^r+\rho^t\right)\\
&\qquad+\frac{1}{n}\sum_{t=1}^{n}\|L_{n,t}^{(m+1)}(u)\|\mathbb{E}\left[\|h_t^*\left(t/n\right)-h_t^*\left(u\right)-\partial_u h_t^*\left(u\right)\left\{t/n-u\right\}\|\right]\\
&\quad=O_P\left(n^{-r}\right)+O_P\left(1/\sqrt{nb}\right)+o_P(b).
\end{aligned}$$

For the second and third terms, similar to the proof of eq. (14) and using Assumption 6,

$$\begin{aligned}
&\sum_{t=1}^{n} L_{n,t}^{(m+2)}\left(u\right)h_{n,t}\left(\theta\left(t/n\right)\right)\theta^{(m+2)}\left(u_{n,t}\right)\to^p\mu_{m+2}H\left(u\right)\theta^{(m+2)}\left(u\right),\\
&\sum_{t=1}^{n} L_{n,t}^{(2m+2)}\left(u\right)\theta_i^{(m+1)}\left(u_{n,t}\right)\frac{\partial h_{n,t}\left(\bar{\theta}_{n,t}\right)}{\partial\theta_i}\theta^{(m+1)}\left(u_{n,t}\right)\\
&\quad\to^p\mu_{2m+2}\theta_i^{(m+1)}\left(u\right)\partial_{\theta_i}H\left(u\right)\theta^{(m+1)}\left(u\right).
\end{aligned}$$

Collecting terms now yields the claimed result. □

Proof of Corollary 1. The first part is a direct consequence of Theorem 3. To show the second part, first observe that

$$B_1\left(u\right)+B_2\left(u\right)=H\left(u\right)\frac{\theta^{(2)}\left(u\right)}{2}+\left(\partial_u H\left(u\right)+\frac{1}{2}\sum_{i=1}^{d_\theta}\theta_i^{(1)}\left(u\right)\partial_{\theta_i}H\left(u\right)\right)\theta^{(1)}\left(u\right).$$

Next, using that $\mathbb{E}\left[s_t^*\left(\theta\left(v\right)|v\right)\right]=0$ for all v,

$$\begin{aligned}
0&=\frac{\partial^2}{\partial v^2}\mathbb{E}\left[s_t^*\left(\theta\left(v\right)|v\right)\right]_{v=u}\\
&=\frac{\partial^2}{\partial u^2}\mathbb{E}\left[s_t^*\left(\theta\left(v\right)|u\right)\right]_{v=u}+2\frac{\partial^2}{\partial u\partial v}\mathbb{E}\left[s_t^*\left(\theta\left(v\right)|u\right)\right]_{v=u}+\frac{\partial^2}{\partial v^2}\mathbb{E}\left[s_t^*\left(\theta\left(v\right)|u\right)\right]_{v=u},
\end{aligned}$$

https://doi.org/10.1017/S0266466626100449 Published online by Cambridge University Press

where it is easily checked that

$$2\frac{\partial^2}{\partial u\partial v}\mathbb{E}\left[s_t^*\left(\theta\left(v\right)|u\right)\right]_{v=u}+\frac{\partial^2}{\partial v^2}\mathbb{E}\left[s_t^*\left(\theta\left(v\right)|u\right)\right]_{v=u}=2\left\{B_1\left(u\right)+B_2\left(u\right)\right\}.$$ □

Proof of Theorem 4. Inspecting the proof of Theorem 3, we observe that Assumptions 7 and 8 are only invoked in the analysis of $\frac{1}{n}\sum_{t=1}^{n}L_{n,t}^{(m+1)}(u)h_{n,t}\left(\theta\left(t/n\right)\right)$. We here re-analyze this term under Assumptions 9 and 10: Write

$$\begin{aligned}\frac{1}{n}\sum_{t=1}^{n}L_{n,t}^{(m+1)}(u)h_{n,t}\left(\theta\left(t/n\right)\right)&=\frac{1}{n}\sum_{t=1}^{n}L_{n,t}^{(m+1)}(u)\left[h_{n,t}\left(\theta\left(t/n\right)\right)-h_{n,t}\left(\theta\left(u\right)\right)\right]\\&\quad+\frac{1}{n}\sum_{t=1}^{n}L_{n,t}^{(m+1)}(u)h_{n,t}\left(\theta\left(u\right)\right),\end{aligned}$$

where the first term satisfies eq. (A.1). Now, with $h_{n,t}:=h_{n,t}\left(\theta\left(u\right)\right)$ and $h_t^*\left(v\right):=h_t^*\left(\theta\left(u\right)|v\right)$,

$$\begin{aligned}\frac{1}{n}\sum_{t=1}^{n}L_{n,t}^{(m+1)}(u)h_{n,t}&=\frac{1}{n}\sum_{t=1}^{n}L_{n,t}^{(m+1)}(u)\left\{h_{n,t}-h_t^*\left(t/n\right)\right\}\\&\quad+\frac{1}{n}\sum_{t=1}^{n}L_{n,t}^{(m+1)}(u)\left\{h_t^*\left(t/n\right)-E\left[h_t^*\left(t/n\right)\right]\right\}\\&\quad+\frac{1}{n}\sum_{t=1}^{n}L_{n,t}^{(m+1)}(u)E\left[h_t^*\left(t/n\right)\right]\\&=:A_1+A_2+A_3.\end{aligned}$$

By Assumption 5(ii),

$$\mathbb{E}\left[\|A_1\|\right]\leq\frac{1}{n}\sum_{t=1}^{n}\left|L_{n,t}^{(m+1)}(u)\right|\mathbb{E}\left[\left\|h_{n,t}-h_t^*\left(t/n\right)\right\|\right]\leq\frac{C}{n^{1+r}}\sum_{t=1}^{n}\left|L_{n,t}^{(m+1)}(u)\right|=O\left(n^{-r}\right);$$

for any element $A_{2,i,j}$ of A_2, by Assumption 10 and with $V_{|t_1-t_2|}^{(i,j)}(v_1,v_2)=\mathrm{Cov}\left(h_{t_1,i,j}^*(v_1),h_{t_2,i,j}^*\left(\theta\left(v_2\right)|u\right)\right)$,

$$\begin{aligned}\mathbb{E}\left[A_{2,i,j}^2\right]&=\frac{1}{n^2}\sum_{t_1,t_2=1}^{n}L_{n,t_1}^{(m+1)}(u)L_{n,t_2}^{(m+1)}(u)V_{|t_1-t_2|}^{(i,j)}\left(t_1/n,t_2/n\right)\\&=\frac{1}{n^2}\sum_{t_1,t_2=1}^{n}L_{n,t_1}^{(m+1)}(u)L_{n,t_2}^{(m+1)}(u)\left\{V_{|t_1-t_2|}^{(i,j)}\left(t_1/n,t_2/n\right)-V_{|t_1-t_2|}^{(i,j)}\left(u,u\right)\right\}\\&\quad+\frac{1}{n^2}\sum_{t_1,t_2=1}^{n}L_{n,t_1}^{(m+1)}(u)L_{n,t_2}^{(m+1)}(u)V_{|t_1-t_2|}^{(i,j)}\left(u,u\right)\\&=\frac{1}{n^2}\sum_{t_1,t_2=1}^{n}L_{n,t_1}^{(m+1)}(u)L_{n,t_2}^{(m+1)}(u)\left\{V_{|t_1-t_2|}^{(i,j)}\left(t_1/n,t_2/n\right)-V_{|t_1-t_2|}^{(i,j)}\left(u,u\right)\right\}\end{aligned}$$

https://doi.org/10.1017/S0266466626100449 Published online by Cambridge University Press

$$+O\left(\frac{1}{nb}\right),$$

where, with

$$\begin{aligned}
\Delta V_{t_1,t_2}^{(i,j)} &:= V_{|t_1-t_2|}^{(i,j)}(t_1/n,t_2/n) - V_{|t_1-t_2|}^{(i,j)}(u,u) - \frac{\partial V_{|t_1-t_2|}^{(i,j)}(u,u)}{\partial u}\left[(t_1/n-u)+(t_2/n-u)\right] \\
&= O\left((t_1/n-u)^2+(t_2/n-u)^2\right), \\
&\left|\frac{1}{n^2}\sum_{t_1,t_2=1}^{n} L_{n,t_1}^{(m+1)}(u)L_{n,t_2}^{(m+1)}(u)\left\{V_{|t_1-t_2|}^{(i,j)}(t_1/n,t_2/n) - V_{|t_1-t_2|}^{(i,j)}(u,u)\right\}\right| \\
&\leq \frac{1}{n^2}\sum_{t_1,t_2=1}^{n}\left|L_{n,t_1}^{(m+1)}(u)\right|\left|L_{n,t_2}^{(m+1)}(u)\right|\left|\Delta V_{t_1,t_2}^{(i,j)}\right| \\
&\quad + \frac{1}{n^2}\sum_{t_1,t_2=1}^{n}\left|L_{n,t_1}^{(m+1)}(u)\right|\left|L_{n,t_2}^{(m+1)}(u)\right|\left|\frac{\partial V_{|t_1-t_2|}(u,u)}{\partial u}\right|\left[|t_1/n-u|+|t_2/n-u|\right] \\
&\leq \frac{C}{n^2}\sum_{t_1,t_2=1}^{n}\left|L_{n,t_1}^{(m+1)}(u)\right|\left|L_{n,t_2}^{(m+1)}(u)\right|\left\{(t_1/n-u)^2+(t_2/n-u)^2\right\} \\
&\quad + \frac{\bar{L}}{(nb)^2}\sum_{t_1,t_2=1}^{n}\left|L\left(\frac{t_1/n-u}{b}\right)\right|\left|\frac{\partial V_{|t_1-t_2|}(u,u)}{\partial u}\left[|t_1/n-u|+|t_2/n-u|\right]\right| \\
&= O\left(\frac{1}{nb}\right).
\end{aligned}$$

Finally, by Assumption 9,

$$\begin{aligned}
A_3 &= \frac{1}{n}\sum_{t=1}^{n} L_{n,t}^{(m+1)}(u)\mathbb{E}\left[h_t^*(u)\right] - b\times\frac{1}{n}\sum_{t=1}^{n} L_{n,t}^{(m+2)}(u)\partial_u\mathbb{E}\left[h_t^*(u)\right] \\
&\quad + \frac{1}{n}\sum_{t=1}^{n} L_{n,t}^{(m+1)}(u)\left\{\mathbb{E}\left[h_t^*(t/n)\right] - \mathbb{E}\left[h_t^*(u)\right] - \partial_u\mathbb{E}\left[h_t^*(u)\right](t/n-u)\right\} \\
&= \int_{\mathbb{R}} L^{(m+1)}(v)\,dv\times\mathbb{E}\left[h_t^*(u)\right] + b\int_{\mathbb{R}} L^{(m+1)}(v)\,v dv\times\partial_u\mathbb{E}\left[h_t^*(u)\right] + o(b).
\end{aligned}$$

□

A.3. Proofs of Results in Section 4

Proof of Lemma 1. First, for any $p>0$, with $\|\mathcal{Z}_{n,t}(\theta)\|_{a,p} := \left(\sum_{i=1}^{\infty} a_i\|Z_{n,t+1-i}(\theta)\|^p\right)^{1/p}$,

$$\begin{aligned}
\mathbb{E}\left[\sup_{\theta\in\Theta}\|f\left(\mathcal{Z}_t^*(\theta|u);\theta\right)\|^p\right]^{1/p} &\leq \mathbb{E}\left[\sup_{\theta\in\Theta}\|f\left(\mathcal{Z}_t^*(\theta|u);\theta\right) - f(0,0,\ldots;\theta)\|^p\right]^{1/p} \\
&\quad + \|f(0,0,\ldots;\theta)\| \\
&\leq \mathbb{E}\left[\sup_{\theta\in\Theta}\left\{g\left(\mathcal{Z}_t^*(\theta|u)\right) + g(0,0,\ldots)\right\}^p \|\mathcal{Z}_t^*(\theta|u)\|_{a,1}^p\right]^{1/p}
\end{aligned}$$

https://doi.org/10.1017/S0266466626100449 Published online by Cambridge University Press

$$+\|f(0,0,\ldots;\theta)\|,$$

where, for any q_1, q_2 satisfying $1/q_1 + 1/q_2 = 1$,

$$\mathbb{E}\left[\sup_{\theta\in\Theta} g\left(\mathcal{Z}_t^*(\theta|u)\right)^p \|\mathcal{Z}_{n,t}(\theta)\|_{a,1}^p\right]^{1/p} \leq \mathbb{E}\left[\sup_{\theta\in\Theta} g\left(\mathcal{Z}_t^*(\theta|u)\right)^{pq_1}\right]^{1/(pq_1)}$$
$$\mathbb{E}\left[\sup_{\theta\in\Theta} \left\|\mathcal{Z}_{t-i}^*(\theta|u)\right\|_{a,1}^{pq_2}\right]^{1/(pq_2)}.$$

For the right-hand side to be finite, we need $pq_1 = p_2$, $pq_2 = p_1$ and $1/q_1 + 1/q_2 = 1$; that is, $q_1 = p_2/p$ and $q_2 = p_1/p$ and $1/q_1 + 1/q_2 = p\{1/p_1 + 1/p_2\} = 1 \Leftrightarrow p = \{1/p_1 + 1/p_2\}^{-1}$. Thus, $f\left(\mathcal{Z}_t^*(\theta|u);\theta\right)$ is well-defined in the L_p-sense. With our choices of q_1, q_2, and p,

$$\begin{aligned}
&\mathbb{E}\left[\sup_{\theta\in\Theta} \|f\left(\mathcal{Z}_{n,t}(\theta);\theta\right) - f\left(\mathcal{Z}_t^*(\theta|t/n);\theta\right)\|^p\right]^{1/p} \\
&\quad\leq C\mathbb{E}\left[\sup_{\theta\in\Theta}\left\{g\left(\mathcal{Z}_{n,t}(\theta)\right) + g\left(\mathcal{Z}_t^*(\theta|u)\right)\right\}^p \|\mathcal{Z}_{n,t}(\theta) - \mathcal{Z}_t^*(\theta|t/n)\|_{a,1}^p\right]^{1/p} \\
&\quad\leq C\left\{\mathbb{E}\left[\sup_{\theta\in\Theta} g\left(\mathcal{Z}_{n,t}(\theta)\right)^{p_2}\right]^{1/p_2} + \mathbb{E}\left[\sup_{\theta\in\Theta} g\left(\mathcal{Z}_t^*(\theta|u)\right)^{p_2}\right]^{1/p_2}\right\} \\
&\qquad\times \sum_{i=1}^{\infty} a_i \mathbb{E}\left[\sup_{\theta\in\Theta} \|Z_{n,t-i}(\theta) - Z_{t+1-i}^*(\theta|t/n)\|^{p_1}\right]^{1/p_1} \\
&\quad\leq \frac{C}{n^r},
\end{aligned}$$

and, similarly, $\mathbb{E}\left[\sup_{\theta\in\Theta}\|f\left(\mathcal{Z}_t^*(\theta|u);\theta\right) - f\left(\mathcal{Z}_t^*(\theta|v);\theta\right)\|^p\right]^{1/p} \leq C|u-v|^r$.

Under (ii') and (iii'), condition (ii) of the first part of the lemma is satisfied with $p_2 = p_1/s \geq 1$:

$$\begin{aligned}
\mathbb{E}\left[\sup_{\theta\in\Theta}\|g\left(\mathcal{Z}_t^*(\theta|u)\right)\|^{p_2}\right] &\leq C\left(1 + \mathbb{E}\left[\sup_{\theta\in\Theta}\|\mathcal{Z}_t^*(\theta|u)\|_{b,s}^{sp_2}\right]\right) \\
&\leq C\left(1 + \sum_{i=0}^{\infty} b_i \mathbb{E}\left[\sup_{\theta\in\Theta}\|Z_{t-i}^*(\theta|u)\|^{p_1}\right]\right) < \infty,
\end{aligned}$$

and similarly for $\mathbb{E}\left[\sup_{\theta\in\Theta}\|g\left(\mathcal{Z}_{n,t}(\theta)\right)\|^{p_2}\right]$. Thus, the conclusions of the first part hold with $p = (1/p_1 + 1/(p_1/s))^{-1} = p_1/(s+1)$.

To show the third part, let $\tilde{\mathcal{Z}}_t^*(\theta|u)$ be an independent copy of $\mathcal{Z}_t^*(\theta|u)$. Applying the same inequalities as in the first part,

$$\mathbb{E}\left[\|f\left(\tilde{\mathcal{Z}}_t^*(\theta|u);\theta\right) - f\left(\mathcal{Z}_t^*(\theta|u);\theta\right)\|^p\right]^{1/p} \leq C\mathbb{E}\left[\|\tilde{\mathcal{Z}}_t^*(\theta|u) - \mathcal{Z}_t^*(\theta|u)\|^{p_1}\right]^{1/p_1}.$$

The claimed result now follows from the definition of τ-weak dependence (cf. p. 1999 of Doukhan and Wintenberger, 2008). This in turn implies that the short-memory condition is satisfied (cf. Doukhan and Wintenberger, 2008). □

Before proving Lemma 3, first consider q-Markov models without covariates on the form

$$Y_{n,t} = G\left(Y_{n,t-1},\ldots,Y_{n,t-q},\varepsilon_t,\theta(t/n)\right), \quad t = 1,\ldots,n,$$

https://doi.org/10.1017/S0266466626100449 Published online by Cambridge University Press

where $G : \mathcal{Y}^q \times \mathcal{E} \times \Theta \mapsto \mathcal{Y}$ is some known mapping, $\varepsilon_t \in \mathcal{E} \subseteq \mathbb{R}^{d_\varepsilon}$ is a sequence of i.i.d. errors, and $\theta(\cdot) \in \Theta$. Importantly, the initial value $Y_{n,0}$ can be arbitrarily chosen which is in contrast to most of the existing literature. Under regularity conditions, its corresponding stationary approximation $Y_t^*(u)$ will solve

$$Y_t^*(u) = G\left(Y_{t-1}^*(u) \ldots, Y_{t-q}^*(u), \varepsilon_t; \theta(u)\right), \quad u \in [0,1]. \tag{A.2}$$

We impose the following assumptions.

Assumption 13. (i) $\sup_{\theta\in\Theta} \mathbb{E}\left[\|G(y_0, \varepsilon_t; \theta)\|^p\right] < \infty$ for some $y_0 \in \mathcal{Y}^q$ and $p > 0$; (ii) there exists $\rho < 1$ so that for all $y, y' \in \mathcal{Y}^q$, $\mathbb{E}\left[\|G(y, \varepsilon_t; \theta) - G\left(y', \varepsilon_t; \theta\right)\|^p\right]^{1/p} \leq \rho\|y - y'\|$; (iii) there exist $\tilde{p} \geq 1$, $r > 0$, and $\delta \geq 0$ so that for all $\theta, \theta' \in \Theta$, $\mathbb{E}\left[\|G(y, \varepsilon_t; \theta) - G\left(y, \varepsilon_t; \theta'\right)\|^{\tilde{p}}\right]^{1/\tilde{p}} \leq C\left(1 + \|y\|^\delta\right)\|\theta - \theta'||^r$; and (iv) $\mathbb{E}\left[\|Y_{n,0}\|^{\tilde{p}}\right] < \infty$ and $\mathbb{E}\left[\|G(y, \varepsilon_t; \theta) - G\left(y, \varepsilon_t; \theta'\right)\|^{\tilde{p}}\right]^{1/\tilde{p}} \leq C\left(1 + \|y\|^\delta\right)\|\theta - \theta'||^r$.

LEMMA 5. *Under Assumption 13(i) and (ii), there exists a stationary and τ-weakly dependent solution, $\left\{Y_t^*(u)\right\}$ to (A.2) with $\sup_{u\in[0,1]} \mathbb{E}\left[\|Y_t^*(u)\|^p\right] < \infty$ with $\tau_Y(k) \leq O\left(\rho^k\right)$. If furthermore Assumption 13 (iii) and (iv) holds, $\sup_{u\in[0,1]} \mathbb{E}\left[\|Y_t^*(u)\|^{\tilde{p}\delta}\right] < \infty$ and $\theta(\cdot) \in \Theta$ is continuously differentiable, then $Y_{n,t}$ is LS $(\tilde{p}, r)$ with $\sup_{n,t} \mathbb{E}\left[\|Y_{n,t}\|^{\tilde{p}}\right] < \infty$ for any given initial value $Y_{n,0}$ with $\mathbb{E}\left[\|Y_{n,0}\|^p\right] < \infty$.*

Proof of Lemma 5. The first part of the result follows from Corollary 3.1 of Doukhan and Wintenberger (2008).

For the second part, define the stacked state vectors

$$\mathbf{Y}_{n,t} := (Y_{n,t}, Y_{n,t-1}, \ldots, Y_{n,t-q+1}) \in \mathcal{Y}^q, \qquad \mathbf{Y}_t^*(u) := (Y_t^*(u), Y_{t-1}^*(u), \ldots, Y_{t-q+1}^*(u)),$$

and define the associated Markov(1) update map $F : \mathcal{Y}^q \times \mathcal{E} \times \Theta \to \mathcal{Y}^q$ by

$$F((y_1, \ldots, y_q), \varepsilon, \theta) := \left(G(y_1, \ldots, y_q, \varepsilon, \theta), y_1, \ldots, y_{q-1}\right).$$

Then the q-Markov recursion is equivalent to the Markov(1) recursion on the enlarged state space:

$$\mathbf{Y}_{n,t} = F(\mathbf{Y}_{n,t-1}, \varepsilon_t, \theta(t/n)), \qquad \mathbf{Y}_t^*(u) = F(\mathbf{Y}_{t-1}^*(u), \varepsilon_t, \theta(u)).$$

Therefore, it suffices to prove the desired bounds for the Markov(1) process $\{\mathbf{Y}_{n,t}\}$; the corresponding bounds for $Y_{n,t}$ follow immediately since $Y_{n,t}$ is the first coordinate of $\mathbf{Y}_{n,t}$ (and similarly for $Y_t^*(u)$). In what follows, we write the argument for the Markov(1) recursion and apply it to $\mathbf{Y}$.

For $u, v \in [0,1]$, using the Markov(1) recursion and the triangle inequality,

$$\begin{aligned}
\mathbb{E}\left[\|Y_t^*(u) - Y_t^*(v)\|^{\tilde{p}}\right]^{1/\tilde{p}} &= \mathbb{E}\left[\|G\left(Y_{t-1}^*(u), \varepsilon_t, \theta(u)\right) - G\left(Y_{t-1}^*(v), \varepsilon_t, \theta(v)\right)\|^{\tilde{p}}\right]^{1/\tilde{p}} \\
&\leq \mathbb{E}\left[\|G\left(Y_{t-1}^*(u), \varepsilon_t, \theta(u)\right) - G\left(Y_{t-1}^*(u), \varepsilon_t, \theta(v)\right)\|^{\tilde{p}}\right]^{1/\tilde{p}} \\
&\quad + \mathbb{E}\left[\|G\left(Y_{t-1}^*(u), \varepsilon_t, \theta(v)\right) - G\left(Y_{t-1}^*(v), \varepsilon_t, \theta(v)\right)\|^{\tilde{p}}\right]^{1/\tilde{p}}
\end{aligned}$$

https://doi.org/10.1017/S0266466626100449 Published online by Cambridge University Press

$$
\begin{aligned}
&\leq C\left(1+\mathbb{E}\left[\left\|Y_{t-1}^{*}(u)\right\|^{\tilde{p}\delta}\right]^{1/\tilde{p}}\right)|u-v|^{r} \\
&\quad+\rho\mathbb{E}\left[\left\|Y_{t-1}^{*}(u)-Y_{t-1}^{*}(v)\right\|^{\tilde{p}}\right]^{1/\tilde{p}} \\
&\vdots \\
&\leq C_{1}|u-v|^{r},
\end{aligned}
$$

where $C_{1}=C\left(1+\sup_{u\in[0,1]}\mathbb{E}\left[\left\|Y_{t-1}^{*}(u)\right\|^{\delta\tilde{p}}\right]^{1/\tilde{p}}\right)/(1-\rho)$. By setting $v=t/n$, we have that

$$
\mathbb{E}\left[\left\|Y_{t}^{*}(t/n)-Y_{t}^{*}(u)\right\|^{p}\right]^{1/\tilde{p}}\leq C_{1}|t/n-u|^{r}.
$$

In addition,

$$
\begin{aligned}
&\mathbb{E}\left[\left\|Y_{n,t}-Y_{t}^{*}(t/n)\right\|^{\tilde{p}}\right]^{1/\tilde{p}} \\
&\quad=\mathbb{E}\left[\left\|F\left(Y_{n,t-1},\varepsilon_{t},\theta(t/n)\right)-F\left(Y_{t-1}^{*}(t/n),\varepsilon_{t},\theta(t/n)\right)\right\|^{\tilde{p}}\right]^{1/\tilde{p}} \\
&\quad\leq\rho\mathbb{E}\left[\left\|Y_{n,t-1}-Y_{t-1}^{*}(t/n)\right\|^{\tilde{p}}\right]^{1/\tilde{p}} \\
&\quad\leq\rho\left(\mathbb{E}\left[\left\|Y_{n,t-1}-Y_{t-1}^{*}((t-1)/n)\right\|^{\tilde{p}}\right]^{1/\tilde{p}}+\mathbb{E}\left[\left\|Y_{t-1}^{*}(t/n)-Y_{t-1}^{*}((t-1)/n)\right\|^{\tilde{p}}\right]^{1/\tilde{p}}\right) \\
&\quad\leq\rho\mathbb{E}\left[\left\|Y_{n,t-1}-Y_{t-1}^{*}((t-1)/n)\right\|^{\tilde{p}}\right]^{1/\tilde{p}}+\rho C_{1}/n^{r} \\
&\quad\leq\rho^{2}\mathbb{E}\left[\left\|Y_{n,t-2}-Y_{t-2}^{*}((t-1)/n)\right\|^{\tilde{p}}\right]^{1/\tilde{p}}+\rho^{2}C_{1}/n^{r}+\rho C_{1}/n^{r} \\
&\quad\vdots \\
&\quad\leq\rho^{t}\mathbb{E}\left[\left\|Y_{n,0}-Y_{0}^{*}(0)\right\|^{\tilde{p}}\right]^{1/\tilde{p}}+\frac{\rho C_{1}/n^{r}}{1-\rho}.
\end{aligned}
$$

Applying the above Markov(1) argument to the stacked process $\{\mathbf{Y}_{n,t}\}$ gives the same bounds for $\mathbf{Y}_{n,t}-\mathbf{Y}_{t}^{*}(t/n)$ and $\mathbf{Y}_{t}^{*}(u)-\mathbf{Y}_{t}^{*}(v)$. Since $Y_{n,t}$ and $Y_{t}^{*}(\cdot)$ are the first coordinates of $\mathbf{Y}_{n,t}$ and $\mathbf{Y}_{t}^{*}(\cdot)$, the same bounds hold for the original q-Markov process $\{Y_{n,t}\}$. Continuing the above two recursions yields the desired results. □

Proof of Lemma 3. With $\xi_{t}=(\varepsilon_{t},\eta_{t})$ and $F(y,x,\xi_{t},t/n)=[G(y,x,\varepsilon_{t};\theta(t/n)),H(x,\eta_{t};t/n)]$, the Markov process $Z_{n,t}=\left(Y_{n,t},X_{n,t}\right)$ satisfies $Z_{n,t}=F\left(Z_{n,t-1},\xi_{t},t/n\right)$. To show existence of a stationary and τ-weakly dependent solution $Z_{t}^{*}(u)=\left(Y_{t}^{*}(u),X_{t}^{*}(u)\right)$ to $Z_{t}^{*}(u)=F\left(Z_{t-1}^{*}(u),\xi_{t},u\right)$, define $\|z\|_{w}=\|(y,x)\|_{w}=\left(\|y\|^{p}+w\|x\|^{p}\right)^{1/p}$ for some $w>0$ which will be chosen below. For $z=(y,x)$ and $z'=(y',x')$, we have

$$
\begin{aligned}
\left(\mathbb{E}\left[\left\|F(z,\xi_{t},u)-F\left(z',\xi_{t},u\right)\right\|_{w}^{p}\right]\right)^{1/p}&=\left(\mathbb{E}\left[\left\|G(y,x,\varepsilon_{t};\theta(u))-G(y',x',\varepsilon_{t};\theta(u))\right\|^{p}\right]\right. \\
&\quad\left.+w\mathbb{E}\left[\left\|H(x,\eta_{t};u)-H(x',\eta_{t};u)\right\|^{p}\right]\right)^{1/p}.
\end{aligned}
$$

https://doi.org/10.1017/S0266466626100449 Published online by Cambridge University Press

Assumption 11(ii) and (iii) and Minkowski's inequality imply that

$$\begin{aligned}
&\left(\mathbb{E}\left[\|F(z,\xi_t,u)-F(z',\xi_t,u)\|_w^p\right]\right)^{1/p}\\
&\quad\le \left(\left(\rho_y\|y-y'\|+L\|x-x'\|\right)^p+\rho_x w\|x-x'^p\right)^{1/p}\\
&\quad\le \left(\rho_y^p\|y-y'^p+\rho_x\left(w^{1/p}\|x-x'\|\right)^p\right)^{1/p}+L/w^{1/p}\left(w^{1/p}\|x-x'\|\right)\\
&\quad\le \kappa_w\|z-z'\|_w,\qquad \kappa_w:=\max\{\rho_y,\rho_x^{1/p}\}+L/w^{1/p}.
\end{aligned}$$

Since $\rho_y<1$ and $\rho_x^{1/p}<1$, we can choose w large enough so that $L/w^{1/p}<1-\max\{\rho_y,\rho_x^{1/p}\}$. Hence, $F(\cdot,\xi_t,u)$ is a contraction in L^p under $\|\cdot\|_w$. The existence and uniqueness of the stationary solution $Z_t^*(u)$ and geometric τ-weak dependence then follow from Corollary 3.1 of Doukhan and Wintenberger (2008), implying $\tau_{Y,X}(k)\le C\kappa_w^k$.

For the LS$(\tilde{p},r)$ claim, combine the above contraction bound with the parameter-smoothness conditions in Assumption 11 (iv) and (v) and follow the same approximation argument as in the proof of Lemma 5. □

Proof of Corollary 4. We can here apply our theory with $\Theta=\mathbb{R}^{1+2q+d_X}$ since the least-squares criterion used for estimation is concave in θ (cf. the comments following Assumptions 1–3). We first show that $Y_{n,t}$ is locally stationary with $p\ge 2$ moments when $\mathbb{E}\left[\|\varepsilon_t\|^2\right]<\infty$ by verifying the conditions of Lemma 3 for $G(y,x,e,\theta):=\omega+\sum_{i=1}^q\alpha_{1,i}y_i^+ +\sum_{i=1}^q\alpha_{2,i}y_i^-+\gamma' x+e$: First, $\mathbb{E}\left[G(0,0,\varepsilon_t;\theta)^2\right]=\mathbb{E}\left[\varepsilon_t^2\right]<\infty$; second, for all $y,y\in\mathbb{R}^q$ and $x,x'\in\mathbb{R}^{d_X}$,

$$\mathbb{E}\left[\left(G(y,x,\varepsilon_t;\theta)-G\left(y',x',\varepsilon_t;\theta\right)\right)^2\right]^{1/2}\le\sum_{i=1}^q\max\left\{|\alpha_{1,i}|,|\alpha_{2,i}|\right\}\|y-y'\|+\|\gamma\|\,\|x-x'\|;$$

third, for all $\theta,\theta'\in\Theta$,

$$\begin{aligned}
\mathbb{E}\left[\left\|G(y,x,\varepsilon_t;\theta)-G\left(y,x,\varepsilon_t;\theta'\right)\right\|^p\right]^{1/p}&=\left|\omega-\omega'\right|+\left\{\left\|\alpha_1-\alpha_1'\right\|+\left\|\alpha_2-\alpha_2'\right\|\right\}\|y\|\\
&\quad+\left\|\gamma-\gamma'\right\|\|x\|\\
&\le C(1+\|x\|+\|y\|)\left\|\theta-\theta'\right\|.
\end{aligned}$$

Thus, under (22), $\left(Y_{n,t},X_{n,t}\right)$ is locally stationary with $\left(Y_t^*(u),X_t^*(u)\right)$ being τ-weakly dependent, $\mathbb{E}\left[Y_t^*(u)^2\right]<\infty$ and $\mathbb{E}\left[\left\|X_t^*(u)\right\|^2\right]<\infty$. Next, write $\ell_{n,t}(\theta)=-\left(Y_{n,t}-\theta'\tilde{X}_{n,t}\right)^2$, where $\tilde{X}_{n,t}=\left(1,Y_{n,t-1}^+,\ldots,Y_{n,t-q}^+,Y_{n,t-1}^-,\ldots,Y_{n,t-q}^-,X_{n,t-1}'\right)'$, so that $s_{n,t}=\varepsilon_t\tilde{X}_{n,t}$ and $h_{n,t}(\theta)=\tilde{X}_{n,t}\tilde{X}_{n,t}'$. It is easily seen that $\ell_{n,t}(\theta)$, $s_{n,t}s_{n,t}'$, and $h_{n,t}(\theta)$ satisfy the second part of Lemma 1 with $s=1$ and $p_1=2$ so that they are all ULS $(1,1,\Theta)$. It now follows from the first part of Corollary 3 that Theorem 2 applies to the local linear estimator.

Next, we verify the conditions of the second part of Corollary 3. First, note that Lemma 3 implies that if $\mathbb{E}\left[\varepsilon_t^4\right]<\infty$, then $\mathbb{E}[Y_t^*(u)^4]<\infty$ and so $h_t^*(\theta(u)|u)=\tilde{X}_t^*(u)\tilde{X}_t^*(u)'$ is τ-weakly dependent with $\mathbb{E}\left[\left\|h_t^*(\theta(u)|u)\right\|^2\right]<\infty$, and hence Assumption 10 is satisfied. Finally, to verify Assumption 7, we show that $\partial_u h_t^*(\theta(u)|u)=2\tilde{X}_t^*(u)\,\partial_u\tilde{X}_t^*(u)'$ is

https://doi.org/10.1017/S0266466626100449 Published online by Cambridge University Press

well-defined almost surely and has first moments. This will hold if $\partial_u\left(Y_t^*(u), X_t^*(u)\right)$ exists almost surely and has second moments. We show this by applying Theorem 4.8 of Dahlhaus et al. (2019). This theorem requires $G(y,x,\varepsilon_t;\theta)$ to be differentiable w.r.t. (y,x,θ) for all values of (y,x,θ) which fails to hold at $y=0$. However, since ε_t is assumed to have a continuous distribution then $Y_t^*(u)\,|\mathcal{F}_{t-1}^*(u)$ will also have a continuous distribution and so $\Pr\left(Y_t^*(u)=0|\mathcal{F}_{t-1}^*(u)\right)=0$. Thus, $G\left(Y_{t-1}^*(u),\ldots,Y_{t-q}^*(u),X_{t-1}^*(u).\varepsilon_t;\theta\right)$ is differentiable w.r.t. $\left(Y_{t-1}^*(u),\ldots,Y_{t-q}^*(u),X_{t-1}^*(u),\theta\right)$ almost surely. By inspection of the proof of Theorem 4.8, this suffices for the result to hold. □

Proof of Corollary 5. We verify the conditions of Corollary 3. Verification of the conditions for the stationary version, including identification and existence of relevant moments, follows from Kristensen and Rahbek (2005). For the analysis of the local linear estimator, what remains to be shown is uniform local stationarity of $\ell_{n,t}(\theta)$, $s_{n,t}s_{n,t}'$, and $h_{n,t}(\theta)$. First, by Lemma 3 with $G(y,x,\varepsilon,\theta)=\left(\omega+\sum_{i=1}^q\alpha_i y_i+\gamma' x\right)\varepsilon^2$, where $\sum_{i=1}^q\alpha_i<1$, it follows that $\left(Y_{n,t},X_{n,t}\right)$ is locally stationary with $\sup_{n,t}\mathbb{E}\left[Y_{n,t}^2\right]<\infty$ and $\mathbb{E}\left[Y_t^*(u)^2\right]<\infty$. Next, we verify that $\ell_{n,t}(\theta)$ and its first two derivatives satisfy the conditions of Lemma 1.

Recall that $\ell_{n,t}(\theta)=\log\left(\lambda_{n,t}(\theta)\right)+Y_{n,t}/\lambda_{n,t}(\theta)$ with $\tilde{X}_{n,t}=\left(1,Y_{n,t-1},\ldots,Y_{n,t-q},X_{n,t-1}'\right)'$ and $\lambda_{n,t}(\theta)=\theta'\tilde{X}_{n,t}$. It satisfies

$$\left|\frac{\partial\ell_{n,t}(\theta)}{\partial Y_{n,t}}\right|=\frac{1}{\lambda_{n,t}(\theta)}\leq\frac{1}{\omega}\leq\frac{1}{\delta_L},\quad\left\|\frac{\partial\ell_{n,t}(\theta)}{\partial\lambda_{n,t}(\theta)}\right\|\leq\frac{1}{\lambda_{n,t}(\theta)}+\frac{Y_{n,t}}{\lambda_{n,t}^2(\theta)}\leq\frac{1}{\delta_L}+\frac{Y_{n,t}}{\delta_L\lambda_{n,t}(\theta)},$$

where

$$\frac{Y_{n,t}}{\lambda_{n,t}(\theta)}=\frac{\theta(t/n)'\tilde{X}_{n,t}}{\theta'\tilde{X}_{n,t}}\varepsilon_t^2\leq\frac{\sup_u\omega(u)+\sum_{i=1}^q\sup_u\alpha_i(u)+\sum_{i=1}^{d_X}\sup_u\beta_i(u)}{\delta_L}\varepsilon_t^2. \tag{A.3}$$

Since $\varepsilon_t^2=Y_{n,t}/\lambda_{n,t}(\theta)$, $\ell_{n,t}(\theta)$ satisfies the second part of Lemma 1 with $g\left(Z_{n,t}\right)=C\left(1+Y_{n,t}/\lambda_{n,t}(\theta(t/n))\right)=C\left(1+\varepsilon_t^2\right)$, which has second moment by assumption. Thus, Lemma 1 implies that $\ell_{n,t}(\theta)$ is ULS $(1,1,\Theta)$.

Next, we analyze the score function

$$s_{n,t}=\left(1-\frac{Y_{n,t}}{\lambda_{n,t}(\theta(t/n))}\right)\frac{\partial_\theta\lambda_{n,t}(\theta(t/n))}{\lambda_{n,t}(\theta(t/n))}=\left(1-\varepsilon_t^2\right)\frac{\tilde{X}_{n,t}}{\lambda_{n,t}(\theta(t/n))}.$$

Since $\mathbb{E}\left[\varepsilon_t^2|\mathcal{F}_{n,t-1}\right]=\mathbb{E}\left[\varepsilon_t^2\right]=1$, $s_{n,t}(\theta(t/n))$ is an MGD. Moreover,

$$\left|\frac{\partial s_{n,t}}{\partial Y_{n,t}}\right|=\frac{\left\|\tilde{X}_{n,t}\right\|}{\lambda_{n,t}^2(\theta(t/n))}\leq C$$

$$\left\|\frac{\partial s_{n,t}}{\partial\tilde{X}_{n,t}}\right\|\leq\left(1-\varepsilon_t^2\right)\left\{\frac{1}{\lambda_{n,t}(\theta(t/n))}+\frac{\left\|\tilde{X}_{n,t}\right\|^2}{\lambda_{n,t}^2(\theta(t/n))}\right\}\leq C\left|1-\varepsilon_t^2\right|$$

and so $s_{n,t}$ satisfies the conditions of Lemma 1 with $g\left(Z_{n,t}\right)=C\left(1+\varepsilon_t^2\right)$.

https://doi.org/10.1017/S0266466626100449 Published online by Cambridge University Press

The Hessian takes the form

$$h_{n,t}(\theta) = \frac{\partial_\theta \lambda_{n,t}(\theta)\, \partial_\theta \lambda_{n,t}(\theta)'}{\lambda_{n,t}^2(\theta)} \left[\frac{2Y_{n,t}}{\lambda_{n,t}(\theta)} - 1 \right],$$

and recycling the inequalities established above it follows that the Hessian is also ULS(1,1, Θ). This verifies the conditions for Theorem 2.

For the analysis of the local constant estimator, observe that $h^*_{ij,t}(\theta(u)|u)$ is τ-weakly dependent by Lemma 1; moreover, if $\mathbb{E}\left[\|\varepsilon_t\|^{4+\gamma}\right] < \infty$, then $\mathbb{E}\left[\|h^*_{ij,t}(\theta(u)|u)\|^{2+\gamma/2}\right] < \infty$. Thus, Assumption 10 holds. To verify Assumption 9, first note that $\partial_u X^*_t(u)$ exists under the assumptions of the theorem. We then follow the same arguments as in Example 5.5 of Dahlhaus et al. (2019) to show that $\partial_u h^*_{ij,t}(\theta|u)$ exists. □

Proof of Corollary 6. We first show that the $(Y_{n,t}, X_{n,t})$ is locally stationary by verifying the conditions of Lemma 3 with $G(y,x,\varepsilon_t;\theta) := N_t\left(\omega + \sum_{i=1}^q \alpha_i y_i + \gamma' x\right)$, where $\varepsilon_t := N_t(\cdot)$, $t = 1,2,\ldots$, are i.i.d. copies of a Poisson process (see Agosto et al., 2016 for details): First, $\mathbb{E}[|G(0,0,\varepsilon_t;\theta(u))|] \le \mathbb{E}[N_t(\omega(u))] = \omega(u) < \infty$; second, for all x,y,x',y',

$$\begin{aligned}
&\mathbb{E}\left[\left\|G(y,x,\varepsilon_t;\theta(u)) - G\left(y,x',\varepsilon_t;\theta(u)\right)\right\|\right] \\
&\quad \le \mathbb{E}\left[\left\| N_t\left(\sum_{i=1}^q \alpha_i(u)\left|y_i - y_i'\right| + \|\gamma(u)\|\left\|x - x'\right\|\right)\right\|\right] \\
&\quad = \sum_{i=1}^q \alpha_i(u)\left|y_i - y_i'\right| + \|\gamma(u)\|\left\|x - x'\right\|,
\end{aligned}$$

where $\sum_{i=1}^q \alpha_i(u) < 1$. Finally, for any θ, θ',

$$\begin{aligned}
\mathbb{E}\left[\left\|G(y,x,\varepsilon_t;\theta) - G\left(y,x,\varepsilon_t;\theta'\right)\right\|\right] &= \left|\omega - \omega'\right| + \sum_{i=1}^q \left|\alpha_i - \alpha_i'\right| y_i + \left\|\gamma - \gamma'\right\| \|x\| \\
&\le C(1 + \|x\| + \|y\|)\left\|\theta - \theta'\right\|.
\end{aligned}$$

Thus, $\lambda_{n,t}(\theta)$ and $Y_{n,t}$ are ULS $(1,1,\Theta)$. From Agosto et al. (2016), we also know that $\mathbb{E}\left[\lambda^*_t(u)^2\right] < \infty$, and hence $\mathbb{E}[Y^*_t(u)^2] < \infty$ too.

Next, we observe that $\lambda_{n,t}(\theta)$, $\partial_\theta \lambda_{n,t}(\theta)$, and $\partial^2_{\theta\theta} \lambda_{n,t}(\theta)$ are on the same form as in the GARCH model. In particular, it is easily checked that $\lambda_{n,t}(\theta)$, $\partial_\theta \lambda_{n,t}(\theta)$, and $\partial^2_{\theta\theta} \lambda_{n,t}(\theta)$ are ULS $(1,1,\Theta)$ and with their stationary versions having second moments. This in turn implies that the log-likelihood and its first two derivatives w.r.t. θ satisfy the conditions of Lemma 1. First,

$$\left|\frac{\partial \ell_{n,t}(\theta)}{\partial Y_{n,t}}\right| = \left|\log\left\{\lambda_{n,t}(\theta)\right\}\right| \le \max\left\{\left|\log \delta_L\right|, \lambda_{n,t}(\theta)\right\},$$

where $\lambda_{n,t}(\theta)$ has second moment. Second,

$$\left|\frac{\partial \ell_{n,t}(\theta)}{\partial \lambda_{n,t}(\theta)}\right| \le \frac{Y_{n,t}}{\lambda_{n,t}(\theta)} + 1 \le \frac{Y_{n,t}}{\delta_L} + 1,$$

https://doi.org/10.1017/S0266466626100449 Published online by Cambridge University Press

where again the right-hand side has second moment. The score function takes the form $s_{n,t} = \left(Y_{n,t}/\lambda_{n,t}(\theta(t/n)) - 1\right)\partial_\theta \lambda_{n,t}(\theta(t/n))$ which satisfies the MGD condition. It is easily checked that $s_{n,t}$ is LS $(2,1)$. Similarly,

$$h_{n,t}(\theta) = \frac{Y_{n,t}}{\lambda_{n,t}^2(\theta)}\partial_\theta \lambda_{n,t}(\theta)\,\partial_\theta \lambda_{n,t}(\theta)' - \left(\frac{Y_{n,t}}{\lambda_{n,t}(\theta)} - 1\right)\partial_{\theta\theta}^2 \lambda_{n,t}(\theta)$$

is ULS $(1,1,\Theta)$

Finally, for the local constant estimator, we need to show that $\partial_u H(u)$ exists. We do this for the case of $q = 1$ to keep notation simple; the general case follows by the same arguments. We show existence by verifying the conditions SC1–SC3 of Truquet (2020) under which the result will follow (cf. Proposition 2 of the same paper). First observe that the transition kernel of the Markov chain $\left(Y_t^*, X_t^*\right)(u) \mid \left(Y_{t-1}^*, X_{t-1}^*\right)(u)$ takes the form

$$q(y,x|y_0,x_0;u) = \frac{\lambda(y_0,x_0;u)^y e^{-\lambda(y_0,x_0;u)}}{y!} p_X(x|x_0;u),$$

where $\lambda(y_0,x_0;u) = \omega(u) + \alpha(u)\,y_0 + \gamma(u)'x_0$. It satisfies $q(y,x|y_0,x_0;u) > 0$ for all (y,x,y_0,x_0,u) under the assumptions of the corollary. Moreover, by the same arguments as in the proof of Proposition 4 of Truquet (2020),

$$\mathbb{E}\left[1 + Y_t^*(u)^{d_0} \mid \left(Y_{t-1}^*, X_{t-1}^*\right)(u) = (y,x)\right] \leq \rho\left(1 + y^{d_0}\right) + \gamma(u)\|x\|^{d_0} + b(u),$$

for some $b < \infty$ with $d_0 = 2$. This combined with the drift criterion imposed on $X_t^*(u)$ implies that SC1 is satisfied with $\phi(y,x) = 1 + |y| + \|x\|$ and $d_0 = 1$. The transition kernel is differentiable under the assumptions of the corollary and so SC2 is satisfied with

$$\begin{aligned}\partial_u q(y,x|y_0,x_0;u) = &\frac{y\lambda(y_0,x_0;u)^{y-1}\,\partial_u\lambda(y_0,x_0;u)\,e^{-\lambda(y_0,x_0;u)}}{y!} p_X(x|x_0;u)\\ &- \frac{\lambda(y_0,x_0;u)^y\,\partial_u\lambda(y_0,x_0;u)\,e^{-\lambda(y_0,x_0;u)}}{y!} p_X(x|x_0;u)\\ &+ \frac{\lambda(y_0,x_0;u)^y e^{-\lambda(y_0,x_0;u)}}{y!}\partial_u p_X(x|x_0;u)\end{aligned}$$

so that, applying the restrictions on $p_X(x|x_0;u)$,

$$\begin{aligned}\frac{|\partial_u q(y,x|y_0,x_0;u)|}{q(y,x|y_0,x_0;u)} &\leq \frac{y\,|\partial_u\lambda(y_0,x_0;u)|}{\lambda(y_0,x_0;u)} + |\partial_u\lambda(y_0,x_0;u)| + \frac{|\partial_u p_X(x|x_0;u)|}{p_X(x|x_0;u)}\\ &\leq C(1 + y + \|x\|).\end{aligned}$$

Thus, by the same arguments as in the proof of Proposition 4 of Truquet (2020),

$$\begin{aligned}&\int\int \phi(y,x)\,|\partial_u q(y,x|y_0,x_0;u)|\,d\mu(y)\,dx\\ &= \int\int \phi(y,x)\,\frac{|\partial_u q(y,x|y_0,x_0;u)|}{q(y,x|y_0,x_0;u)} q(y,x|y_0,x_0;u)\,d\mu(y)\,dx\\ &\leq C\int\int \phi^2(y,x)\,q(y,x|y_0,x_0;u)\,d\mu(y)\,dx \leq \tilde{C}\phi^2(y_0,x_0),\end{aligned}$$

https://doi.org/10.1017/S0266466626100449 Published online by Cambridge University Press

where $\mu(y)$ denotes the counting measure. Finally, $q(y,x|y_0,x_0;u) \leq \bar{q}(y,x|y_0,x_0)$ where, with $\bar{p}_X^{(0)}(x|x_0)$ defined in the corollary,

$$\bar{q}(y,x|y_0,x_0) = \frac{\bar{\lambda}(y_0,x_0)^y e^{-\underline{\lambda}(y_0,x_0;u)}}{y!} \bar{p}_X^{(0)}(x|x_0);$$

here, $\bar{\lambda}(y_0,x_0) = \bar{\omega} + \bar{\alpha} y_0 + \bar{\gamma}' x_0$, $\underline{\lambda}(y_0,x_0;u) = \underline{\omega} + \underline{\alpha} y_0 + \underline{\gamma}' x_0$ with $\bar{\omega} = \max_{v\in[u-\epsilon,u+\epsilon]} \omega(u)$, $\underline{\omega} = \min_{v\in[u-\epsilon,u+\epsilon]} \omega(u)$ and similarly for the other constants. Since $\int\int \phi^2(y,x)\, \bar{q}(y,x|y_0,x_0)\, d\mu(y)\, dx < \infty$, it follows by dominated convergence that

$$\lim_{h\to 0} \int\int \phi^2(y,x) \left| q(y,x|y_0,x_0;u+h) - q(y,x|y_0,x_0;u) \right| d\mu(y)\, dx = 0.$$

A similar bound can be obtained for $|\partial_u q(y,x|y_0,x_0;u)|$ so that

$$\lim_{h\to 0} \int\int \phi^2(y,x) \left| \partial_u q(y,x|y_0,x_0;u+h) - \partial_u q(y,x|y_0,x_0;u) \right| d\mu(y)\, dx = 0.$$

This completes the proof. □

FUNDING STATEMENT

The authors declare that no specific funding has been received for this article.

COMPETING INTEREST

The authors declare that no competing interests exist.

REFERENCES

Agosto, A., Cavaliere, G., Kristensen, D., & Rahbek, A. (2016). Modeling corporate defaults: Poisson autoregressions with exogenous covariates (PARX). *Journal of Empirical Finance*, 38, 640–663. Part B.

Bardet, J.-M., Doukhan, P., & Wintenberger, O. (2022). Contrast estimation of time-varying infinite memory processes. *Stochastic Processes and their Applications*, 152, 32–85.

Brown, B. (1971). Martingale central limit theorems. *Annals of Mathematical Statistics*, 42, 59–66.

Caldara, D., Fernández-Villaverde, J., Rubio-Ramírez, J. F., & Yao, W. (2012). Computing DSGE models with recursive preferences and stochastic volatility. *Review of Economic Dynamics*, 15(2), 188–206.

Christoffersen, P., Jacobs, K., & Ornthanalai, C. (2012). Dynamic jump intensities and risk premiums: Evidence from S&P 500 returns and options. *Journal of Financial Economics*, 106(3), 447–472.

Dahlhaus, R. (1997). Fitting time series models to nonstationary processes. *Annals of Statistics*, 25(1), 1–37.

Dahlhaus, R., Richter, S., & Wu, W. B. (2019). Towards a general theory for nonlinear locally stationary processes. *Bernoulli*, 25(2), 1013–1044.

Dahlhaus, R., & Subba Rao, S. (2006). Statistical inference for time-varying ARCH processes. *Annals of Statistics*, 34(3), 1075–1114.

Das, S. R., Duffie, D., Kapadia, N., & Saita, L. (2007). Common failings: How corporate defaults are correlated. *The Journal of Finance*, 62(1), 93–117.

Davis, R. A., & Liu, H. (2016). Theory and inference for a class of nonlinear models with application to time series of counts. *Statistica Sinica*, 26(4), 1673–1707.

https://doi.org/10.1017/S0266466626100449 Published online by Cambridge University Press

Doukhan, P., & Wintenberger, O. (2008). Weakly dependent chains with infinite memory. *Stochastic Processes and their Applications*, 118(11), 1997–2013.

Duffie, D., Saita, L., & Wang, K. (2007). Multi-period corporate default prediction with stochastic covariates. *Journal of Financial Economics*, 83(3), 635–665.

Fan, J., & Gijbels, I. (2018). *Local polynomial modelling and its applications*. Routledge.

Fan, J., Heckman, N. E., & Wand, M. P. (1995). Local polynomial kernel regression for generalized linear models and quasi-likelihood functions. *Journal of the American Statistical Association*, 90(429), 141–150.

Fokianos, K., & Tjøstheim, D. (2011). Log-linear Poisson autoregression. *Journal of Multivariate Analysis*, 102(3), 563–578.

Fryzlewicz, P., Sapatinas, T., & Subba Rao, S. (2008). Normalized least-squares estimation in time-varying ARCH models. *Annals of Statistics*, 36(2), 742–786.

Ghysels, E., & Hall, A. (1990). Are consumption-based intertemporal capital asset pricing models structural? *Journal of Econometrics*, 45(1–2), 121–139.

Giacomini, R., & Rossi, B. (2016). Model comparisons in unstable environments. *International Economic Review*, 57(2), 369–392.

Han, H., & Kristensen, D. (2014). Asymptotic theory for the QMLE in GARCH-X models with stationary and nonstationary covariates. *Journal of Business & Economic Statistics*, 32(3), 416–429.

Hillebrand, E. (2005). Neglecting parameter changes in GARCH models. *Journal of Econometrics*, 129(1–2), 121–138.

Inoue, A., Jin, L., & Pelletier, D. (2019). Local-linear estimation of time-varying-parameter GARCH models and associated risk measures. Working Paper.

Inoue, A., & Rossi, B. (2011). Identifying the sources of instabilities in macroeconomic fluctuations. *Review of Economics and Statistics*, 93(4), 1186–1204.

Kristensen, D., & Rahbek, A. (2005). Asymptotics of the QMLE for a class of ARCH(q) models. *Econometric Theory*, 21(5), 946–961.

Lando, D., & Nielsen, M. S. (2010). Correlation in corporate defaults: Contagion or conditional independence? *Journal of Financial Intermediation*, 19(3), 355–372.

Loader, C. (2006). *Local regression and likelihood*. Springer Science & Business Media.

Newey, W. K., & McFadden, D. (1994). Large sample estimation and hypothesis testing. *Handbook of Econometrics*, 4, 2111–2245.

Richter, S., & Dahlhaus, R. (2019). Cross validation for locally stationary processes. *Annals of Statistics*, 47(4), 2145–2173.

Truquet, L. (2019). Local stationarity and time-inhomogeneous Markov chains. *Annals of Statistics*, 47(4), 2023–2049.

Truquet, L. (2020). A perturbation analysis of Markov chains models with time-varying parameters. *Bernoulli*, 26(4), 2876–2906.

Vazquez-Abad, F. J., & Kushner, H. J. (1992). Estimation of the derivative of a stationary measure with respect to a control parameter. *Journal of Applied Probability*, 29(2), 343–352.

https://doi.org/10.1017/S0266466626100449 Published online by Cambridge University Press