

TI 2012-133/III
Tinbergen Institute Discussion Paper



Transformed Polynomials for Nonlinear Autoregressive Models of the Conditional Mean

Francisco Blasques

*Faculty of Economics and Business Administration, VU University Amsterdam, and
Tinbergen Institute.*

Tinbergen Institute is the graduate school and research institute in economics of Erasmus University Rotterdam, the University of Amsterdam and VU University Amsterdam.

More TI discussion papers can be downloaded at <http://www.tinbergen.nl>

Tinbergen Institute has two locations:

Tinbergen Institute Amsterdam
Gustav Mahlerplein 117
1082 MS Amsterdam
The Netherlands
Tel.: +31(0)20 525 1600

Tinbergen Institute Rotterdam
Burg. Oudlaan 50
3062 PA Rotterdam
The Netherlands
Tel.: +31(0)10 408 8900
Fax: +31(0)10 408 9031

Duisenberg school of finance is a collaboration of the Dutch financial sector and universities, with the ambition to support innovative research and offer top quality academic education in core areas of finance.

DSF research papers can be downloaded at: <http://www.dsf.nl/>

Duisenberg school of finance
Gustav Mahlerplein 117
1082 MS Amsterdam
The Netherlands
Tel.: +31(0)20 525 8579

Transformed Polynomials for Nonlinear Autoregressive Models of the Conditional Mean

Francisco Blasques*

Department of Finance, VU University Amsterdam
Department of Econometrics, VU University Amsterdam
Tinbergen Institute

December 3, 2012

Abstract

This paper proposes a new set of transformed polynomial functions that provide a flexible setting for nonlinear autoregressive modeling of the conditional mean while at the same time ensuring the strict stationarity, ergodicity, fading memory and existence of moments of the implied stochastic sequence. The great flexibility of the transformed polynomial functions makes them interesting for both parametric and semi-nonparametric autoregressive modeling. This flexibility is established by showing that transformed polynomial sieves are sup-norm-dense on the space of continuous functions and offer appropriate convergence speeds on Hölder function spaces.

*The author is thankful to Andre Lucas and Siem Jan Koopman for helpful comments and suggestions, and to the Dutch Science Foundation (NWO) for financial support. Corresponding address: VU University Amsterdam, FEWEB/FIN, de Boelelaan 1105, 1081 HV Amsterdam, Netherlands, email: f.blasques@vu.nl.

1 INTRODUCTION

Econometricians are often interested in the estimation of autoregressive models for the conditional expectation of a time-series as means of forecasting or analyzing its dynamic behavior. In the last few decades several nonlinear autoregressive models have been proposed for the study of univariate Markov stochastic sequences $\{x_t\}_{t \in \mathbb{Z}}$ generated by,

$$x_t = f_0(x_{t-1}) + \epsilon_t \quad , \quad t \in \mathbb{Z} \quad , \quad (1)$$

where $\{\epsilon_t\}_{t \in \mathbb{Z}}$ is a white noise sequence with variance $\text{Var}(\epsilon_t) = \sigma^2$. Several examples of such models can be found e.g. in Granger and Teräsvirta (1993) and Teräsvirta et al. (2010). For instance, a conditional expectation given by,

$$f_0(x_{t-1}) = \mathbb{E}(x_t | x_{t-1}) = \theta_1 + \theta_2 \exp(-\theta_3 x_{t-1}^2) x_{t-1}$$

gives rise to the exponential autoregressive model introduced in Ozaki (1980). Unfortunately, economic theory rarely provides guidance about the form of f_0 . Hence, in practice, econometricians are often interested in finding a good approximation to the unknown conditional expectation f_0 in the hope that the approximate process has similar stochastic properties.

Linear approximations to the conditional expectation are convenient to work with due to their simplicity and the fact that the dynamic properties of linear stochastic sequences are well understood. In certain settings however, the assumption of linearity might be inappropriate and the researcher is forced to search for alternative nonlinear specifications; see again Granger and Teräsvirta (1993) or Teräsvirta et al. (2010) for several examples.

The current paper introduces a new class of transformed polynomials that can be used to approximate the unknown conditional mean f_0 and renders an approximate process of the form,

$$x_t = \theta_0 + \theta_1(x_{t-1} - x_0) + \sum_{i=2}^k \theta_i(x_{t-1} - x_0)^i \exp\left(\beta(x_{t-1} - x_0)^2\right) + \epsilon_t \quad , \quad t \in \mathbb{Z} \quad , \quad (2)$$

where β is a strictly negative parameter. This class includes the linear autoregressive model as a special case. Yet, it allows also for very general nonlinear autoregressive dynamics. In effect, from a semi-nonparametric perspective, this class is shown to be capable of approximating arbitrarily well any continuous f_0 , and to do so at certain convergence rates when f_0 lies in smooth Hölder spaces. Most importantly, this paper establishes also the conditions under which the approximate autoregressive process is strictly stationary, ergodic, has fading memory and bounded moments. The consistency, convergence rate and asymptotic normality of several extremum estimators is easily obtained as a result.

In what follows, Section 2 introduces the transformed polynomials. Section 3 studies their approximation properties. Section 4 studies the stochastic properties of transformed polynomial autoregressions and Section 5 addresses briefly the asymptotic and small sample behavior of a least squares estimator of the conditional mean.

2 TRANSFORMED POLYNOMIALS

Let $\mathbb{P}_k(\mathcal{X})$ denote the space of k -order *polynomial functions* defined on a compact subset of the real line $\mathcal{X} = [a, b] \subset \mathbb{R}$. Then, for every $p_k \in \mathbb{P}_k(\mathcal{X})$, and every point $x_0 \in \mathcal{X}$, there exists $\boldsymbol{\theta} := (\theta_0, \theta_1, \dots, \theta_k) \in \mathbb{R}^{k+1}$ such that the function p_k admits the

algebraic representation,

$$p_k(x; \boldsymbol{\theta}) = \sum_{i=0}^k \theta_i (x - x_0)^i \quad \forall x \in \mathcal{X} \quad , \quad k \in \mathbb{N},$$

where the normalization by some $x_0 \in \mathcal{X}$ is unnecessary yet convenient in later sections. In what follows, we denote by $\tilde{\mathbb{P}}_k(\mathcal{X})$ the space of k -order T -transformed polynomial functions defined on \mathcal{X} . Elements of $\tilde{\mathbb{P}}_k(\mathcal{X})$ are given by,

$$\tilde{p}_{k,\beta}(x; \boldsymbol{\theta}) = \theta_0 + \theta_1(x - x_0) + \left(\sum_{i=2}^k \theta_i (x - x_0)^i \right) \cdot \phi_\beta(x - x_0),$$

where the ‘*transformation function*’ ϕ_β is defined according to,

$$\phi_\beta(x - x_0) = \exp\left(\beta(x - x_0)^2\right) \quad \forall x \in \mathcal{X} \quad ,$$

where β is a scalar satisfying $\beta < 0$. Note that while $\mathbb{P}_k(\mathcal{X})$ is spanned by power monomials of up to k -order,

$$\mathbb{P}_k(\mathcal{X}) \subseteq \text{Lin}\left\{1, (x - x_0), (x - x_0)^2, \dots, (x - x_0)^k\right\},$$

the space $\tilde{\mathbb{P}}_{k,\beta}(\mathcal{X})$, for fixed $\beta < 0$, is spanned by the alternative basis functions,

$$\tilde{\mathbb{P}}_{k,\beta}(\mathcal{X}) \subseteq \text{Lin}\left\{1, (x - x_0), (x - x_0)^2 \phi_\beta(x - x_0), \dots, (x - x_0)^k \phi_\beta(x - x_0)\right\}$$

and $\tilde{\mathbb{P}}_k(\mathcal{X})$ is then defined as $\tilde{\mathbb{P}}_k(\mathcal{X}) := \bigcup_{\beta < 0} \tilde{\mathbb{P}}_{k,\beta}(\mathcal{X})$. Some basis functions of up to fourth order that span $\mathbb{P}_k(\mathcal{X})$ and $\tilde{\mathbb{P}}_{k,\beta}(\mathcal{X})$ are plotted in Figure 1.

As we shall see, while transformed polynomials inherit the approximation properties of polynomials, they do not inherit their diverging behavior as $|x| \rightarrow \infty$. This

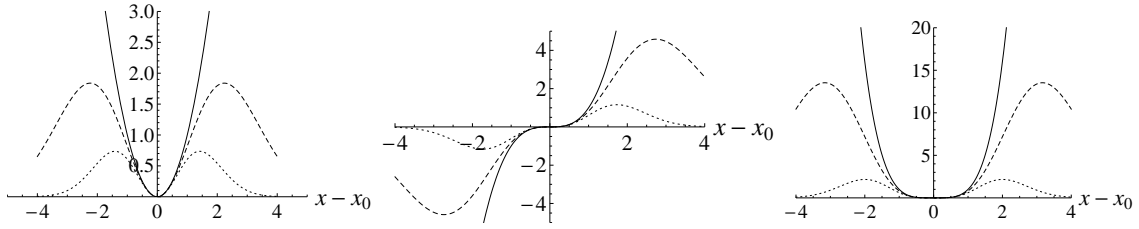


Figure 1: Plots of power monomial x^k (line), transformed monomial $x^k \exp(\beta x^2)$ with $\beta = -1/5$ (dashed line) and $\beta = -1/2$ (dotted line) for $k = 2$ (left), $k = 3$ (center) and $k = 4$ (right).

difference plays a critical role in the dynamic properties of autoregressive processes such as those given by the transformed polynomial autoregression in (2).

3 APPROXIMATION PROPERTIES

Let $\mathbb{C}(\mathcal{X})$ denote the space of continuous functions on \mathcal{X} and $\mathbb{P}(\mathcal{X}) = \bigcup_{k \in \mathbb{N}} \mathbb{P}_k(\mathcal{X})$ denote the space of all polynomial functions on \mathcal{X} . Weierstrass's Theorem establishes that $\mathbb{P}(\mathcal{X})$ is sup-norm dense in $\mathbb{C}(\mathcal{X})$ and hence that for every $f \in \mathbb{C}(\mathcal{X})$, there exists a sequence in $\mathbb{P}(\mathcal{X})$ that converges to f uniformly on \mathcal{X} . The following proposition reveals that this approximation property also holds true for the class of transformed polynomials $\tilde{\mathbb{P}}(\mathcal{X}) = \bigcup_{k \in \mathbb{N}} \tilde{\mathbb{P}}_k(\mathcal{X})$.

PROPOSITION 1. *The space of transformed polynomials $\tilde{\mathbb{P}}(\mathcal{X})$ is dense on $\mathbb{C}(\mathcal{X})$ in sup-norm for every compact $\mathcal{X} \subset \mathbb{R}$.*

Proposition 1 was obtained by making use of Weierstrass's Theorem and by noting that $\tilde{\mathbb{P}}(\mathcal{X})$ is dense on $\mathbb{P}(\mathcal{X})$. It should be clear however that the same result can be achieved if we restrict ourselves to subsets $\tilde{\mathbb{P}}_\beta(\mathcal{X}) \subset \tilde{\mathbb{P}}(\mathcal{X})$ with β near zero. In particular, it is important to note that the uniform convergence of sequences in $\tilde{\mathbb{P}}(\mathcal{X})$ to continuous functions $f \in \mathbb{C}(\mathcal{X})$ can always be obtained on ‘paths of vanishing

transformation', i.e. for those paths satisfying $\beta \rightarrow 0$.¹

In the proposition that follows we make use of a family of operators,

$$\{\Phi_\beta : \mathbb{P}(\mathcal{X}) \rightarrow \tilde{\mathbb{P}}_\beta(\mathcal{X}), \beta < 0\}$$

conveniently defined as $\Phi_\beta(p_k) = \tilde{p}_{k,\beta}$, such that $\forall (x, k, \beta) \in \mathcal{X} \times \mathbb{N} \times \mathbb{R}_-$,

$$\Phi_\beta\left(\sum_{i=0}^k \theta_i(x - x_0)^i\right) = \theta_0 + \theta_1(x - x_0) + \left(\sum_{i=2}^k \theta_i(x - x_0)^i\right) \cdot \phi_\beta(x - x_0).$$

PROPOSITION 2. *Let $p \in \mathbb{P}(\mathcal{X})$. Then $\lim_{\beta \rightarrow 0} \sup_{x \in \mathcal{X}} |p(x) - \Phi_\beta(p)(x)| = 0$. Furthermore, for every $f \in \mathbb{C}(\mathcal{X})$ and every sequence $\{p^j\}_{j \in \mathbb{N}} \subset \mathbb{P}(\mathcal{X})$ satisfying $\sup_{x \in \mathcal{X}} |f - p^j| \rightarrow 0$ as $j \rightarrow \infty$, it holds true that $\lim_{\substack{j \rightarrow \infty \\ \beta \rightarrow 0}} \sup |f(x) - \Phi_\beta(p^j)(x)| = 0$ jointly along any diagonal path $(j, \beta) \rightarrow (\infty, 0)$.*

Proposition 1 established that every point in $\mathbb{C}(\mathcal{X})$ is a limit point of a sequence in $\tilde{\mathbb{P}}(\mathcal{X})$ and Proposition 2 explained how this convergence takes place on *paths of vanishing transformation*. In what follows, we characterize further this convergence by analyzing the rate at which such sequences approximate any $f \in \mathbb{C}(\mathcal{X})$.

DEFINITION 1. *For every $f \in \mathbb{C}(\mathcal{X})$, and any $k \in \mathbb{N}$, the polynomial function $p_k^* \in \mathbb{P}_k(\mathcal{X})$ satisfying $p_k^* = \inf_{p_k \in \mathbb{P}_k(\mathcal{X})} \sup_{x \in \mathcal{X}} |f(x) - p_k(x)|$ is called the best uniform approximation to f from $\mathbb{P}_k(\mathcal{X})$ and $E_k(f) = \sup_{x \in \mathcal{X}} |f(x) - p_k^*(x)|$ the best uniform approximation error.*

Existence is ensured by the following lemma in Timan (1963; Section 2.2).

¹Sequences in $\mathbb{P}(\mathcal{X})$ might be composed of polynomials of different orders. When explicit reference to the order k is desirable, then sequences $\{p^i\}_{i \in \mathbb{N}}$ in $\mathbb{P}(\mathcal{X})$ are denoted $\{p_{k_j}^i\}_{i \in \mathbb{N}}$ where k_j is the order of the j^{th} polynomial in the sequence. Similarly, sequences in $\tilde{\mathbb{P}}(\mathcal{X})$ can be denoted $\{\tilde{p}_{k_j, \beta_j}^i\}_{i \in \mathbb{N}}$ when explicit reference to k and β is desired.

LEMMA 1. For every $f \in \mathbb{C}(\mathcal{X})$, there exist a unique element $p_k^* \in \mathbb{P}_k(\mathcal{X})$ which is the best uniform approximation to f from $\mathbb{P}_k(\mathcal{X})$.

Note that Weierstrass's Theorem implies that $E_k(f) \rightarrow 0$ as $k \rightarrow \infty$. The much celebrated Jackson's Theorem complements this result by relating the speed at which $E_k(f) \rightarrow 0$ to the differentiability of f and the modulus of continuity $\omega(f, \delta)$ of its derivatives, with $\omega(f^{(r)}, \delta)$ defined as,

$$\omega(f^{(r)}, \delta) = \sup_{(x, x') \in \mathcal{X} \times \mathcal{X} : |x - x'| < \delta} |f^{(r)}(x) - f^{(r)}(x')|$$

for any $r \geq 0$. For completeness, Jackson's Theorem is now stated as a lemma; see e.g. Timan (1963, p.261).

LEMMA 2. Let $f \in \mathbb{C}^r(\mathcal{X}) \equiv \mathbb{C}^r([a, b])$ then $\exists M_r > 0$ such that,²

$$E_k(f) = M_r \cdot \left(\frac{d}{k}\right)^r \cdot \omega(f^{(r)}, d/k) = O(k^{-r}) \text{ as } k \rightarrow \infty,$$

where $d := b - a$ and M_r depends only on r for every $r \geq 0$ and $k \geq 1$.

Note that Jackson's Theorem implies that $E_k(f) = M_0 \cdot \omega(f, d/k)$ for continuous functions $f \in \mathbb{C}^r(\mathcal{X})$, it implies $E_k(f) = M'_r(d/k)^{r+\alpha} = O(k^{-(r+\alpha)})$ for Hölder functions $f \in \mathbb{C}^{k, \alpha}(\mathcal{X})$ and $E_k(f) = M'_r(d/k)^{r+1} = O(k^{-(r+1)})$ for functions with Lipschitz r^{th} derivative $f \in \mathbb{C}^{k, 1}(\mathcal{X})$, and also, $M'_r = M_r$ for functions with quasi-smooth r^{th} derivative $\omega(f^{(r)}, \delta) \leq \delta$. In the context of semi-nonparametric autoregressive modeling, Jackson's Theorem allows us also to obtain as a corollary the lower bound on the speed at which the polynomial order k must diverge, as a function of sample size T , in order to obtain a bound for the approximation rate of the sequence of sieves.

²The added limit result uses the fact that $\omega(f, \delta) \rightarrow 0$ as $\delta \rightarrow 0$ for every $f \in \mathbb{C}(\mathcal{X})$.

COROLLARY 1. *Let the polynomial order k be indexed by T satisfying $k_T = O(T^q)$, $q > 0$, as $T \rightarrow \infty$. Then, $E_{k_T}(f) = O(T^{-qr})$ for $f \in \mathbb{C}^r(\mathcal{X})$, $E_k(f) = O(T^{-q(r+\alpha)})$ for $f \in \mathbb{C}^{k,\alpha}(\mathcal{X})$ and $E_k(f) = O(T^{-q(r+1)})$ for f with quasi-smooth r^{th} derivative.*

Proposition 3 builds on Lemma 2 to deliver asymptotic approximation rates for sequences of transformed polynomials. The result is obtained by considering transformations of best polynomial approximations $\tilde{p}_{k,\beta}^* = \Phi_\beta(p_k^*) \in \tilde{\mathbb{P}}_{k,\beta}(\mathcal{X})$. The corresponding uniform approximation error to a function $f \in \mathbb{C}(\mathcal{X})$ from $\tilde{\mathbb{P}}_{k,\beta}(\mathcal{X})$ is then defined as $\tilde{E}_{k,\beta}(f) = \sup_{x \in \mathcal{X}} |f(x) - \tilde{p}_{k,\beta}^*(x)|$. Note that for any given function $f \in \mathbb{C}(\mathcal{X})$, the transformed polynomial $\tilde{p}_{k,\beta}^* \in \tilde{\mathbb{P}}_{k,\beta}(\mathcal{X})$ is not necessarily the best approximation to f from $\tilde{\mathbb{P}}_{k,\beta}(\mathcal{X})$. Instead, it is the transformation of the best polynomial approximation p_k^* to f from $\mathbb{P}_k(\mathcal{X})$. This leaves room to possible improvement in the convergence rate bounds.

PROPOSITION 3. *Let $f \in \mathbb{C}^r(\mathcal{X}) \equiv \mathbb{C}^r([a, b])$ and β be indexed by the order k such that $\beta_k = O(\log(k^{-sr}))$, $s > 0$, as $k \rightarrow \infty$. Then $\tilde{E}_{k,\beta_k}(f) = O(k^{-r \min\{s,1\}})$ as $k \rightarrow \infty$. Furthermore, if $f \in \mathbb{C}^{k,\alpha}(\mathcal{X})$ and $\beta_k = O(\log(k^{-s(r+\alpha)}))$, $s > 0$, as $k \rightarrow \infty$. Then $\tilde{E}_{k,\beta_k}(f) = O(k^{-(r+\alpha) \min\{s,1\}})$ as $k \rightarrow \infty$.*

Once again, it is possible to establish as a corollary the speed at which the polynomial order k must diverge and β must vanish, as a function of sample size T , in order to obtain a given approximation rate bound.

COROLLARY 2. *Let $k_T = O(T^q)$, $q > 0$, as $T \rightarrow \infty$. Then, $E_{k_T, \beta_{k_T}}(f) = O(T^{-qr \min\{s,1\}})$ holds if $f \in \mathbb{C}^r(\mathcal{X})$ and $\beta_k = O(\log(k^{-sr}))$, $s > 0$, as $k \rightarrow \infty$ and $E_{k_T, \beta_{k_T}}(f) = O(T^{-q(r+\alpha) \min\{s,1\}})$ if $f \in \mathbb{C}^{k,\alpha}(\mathcal{X})$ and $\beta_k = O(\log(k^{-s(r+\alpha)}))$, $s > 0$.*

4 STABILITY, STATIONARITY AND ERGODICITY

In what follows we analyze the strict stationarity, ergodicity, fading memory and existence of moments of stochastic sequences generated by the transformed polynomial autoregressive model in (2) with iid innovations $\{\epsilon_t\}_{t \in \mathbb{Z}}$. These are analyzed in a general setting allowing k to take any value in \mathbb{N} . This generality ensures model flexibility in parametric settings and plays a crucial role in semi-nonparametric settings where k is set to diverge to infinity with sample size.

Proposition 4 derives a condition on the parameter vector $\boldsymbol{\theta} := (\theta_1, \dots, \theta_k) \in \Theta \subset \mathbb{R}^k$ that ensures the strict stationarity and ergodicity of the stochastic sequence $\{x_t\}_{t \in \mathbb{Z}}$ generated by (2). This is done by appealing to a special case of the result in Bougerol (1991). In particular, Proposition 4 finds a norm $\|\cdot\|$ on \mathbb{R}^k such that $\{x_t\}_{t \in \mathbb{Z}}$ is strictly stationary and ergodic for every $\boldsymbol{\theta}$ satisfying $\|\boldsymbol{\theta}\| < 1$.

PROPOSITION 4. *Let $\{x_t\}_{t \in \mathbb{Z}}$ be generated according to (2) where $\{\epsilon_t\}_{t \in \mathbb{Z}}$ is an iid sequence with $\mathbb{E}(\epsilon_t) = 0$ and $\mathbb{E}|\epsilon_t|^2 < \infty$. Then, $\{x_t\}_{t \in \mathbb{Z}}$ is strictly stationary and ergodic for every $\theta_0 \in \mathbb{R}$ and $\boldsymbol{\theta} := (\theta_1, \dots, \theta_k)$ satisfying $\ell_1(\boldsymbol{\theta}, \mathbf{w}(\beta)) < 1$ where $\ell_1(\boldsymbol{\theta}, \mathbf{w}(\beta)) := \sum_{i=1}^k w_i(\beta)|\theta_i|$ is a weighted l_1 -norm on $\boldsymbol{\theta}$ with weights given by*

$$w_1(\beta) = 1 \quad \text{and} \quad w_i(\beta) = i(1-2\beta) \left(-\frac{i+\iota}{2\beta} \right)^{\frac{i+\iota}{2}} \exp\left(-\frac{i+\iota}{2} \right), \quad i = 2, \dots, k, \quad (3)$$

$$\text{with } \iota = \begin{cases} -1 & \forall(\beta, i) : \beta < -(i+1)^{\frac{i+1}{2}} / 2e(i-1)^{\frac{i-1}{2}} \\ 1 & \text{otherwise} \end{cases}. \quad (4)$$

Proposition 4, allows us to verify, for any given $\boldsymbol{\theta}$, if the process $\{x_t\}_{t \in \mathbb{Z}}$ is strictly stationary and ergodic. In the context of extremum estimation one can then design algorithms that optimize the criterion function over the parameter space and verify, at

every step, if $\ell_1(\theta, \mathbf{w}(\beta)) < 1$. However, in certain occasions, it is sometimes preferable to restrict the parameter space Θ in some simple way, rather than having to verify if the condition holds for every θ . In particular, it is natural in applications with large k to impose decreasing bounds to the sequence of parameters $\{\theta_i\}$ as $i \rightarrow \infty$ that vanish at some predefined speed. Proposition 5 establishes bounds on the parameter space $\Theta \subseteq \mathbb{R}^{k+1}$ that ensure the strict stationarity and ergodicity of the transformed polynomial autoregressive process for every $\theta := (\theta_0, \theta_1, \dots, \theta_k) \in \Theta = \prod_{i=0}^k \Theta_i$ with $\Theta_i \subseteq \mathbb{R} \forall i$.

For convenience we define $\Theta_1 = (-\delta, \delta)$ and let the bounds on each Θ_i , $i \geq 2$ be written in terms of a multiple c_i of a scalar Δ , hence taking the form $\Theta_i = (-c_i\Delta, c_i\Delta) \subset \mathbb{R} \forall i \geq 2$. Centering at the origin is imposed for simplicity since, from the outset, there is no reason to restrict intercepts, slopes and curvatures in any particular direction. Proposition 5 then provides the maximal allowed Δ for every choice of δ , coefficients $\mathbf{c} := (c_2, \dots, c_k)$ and transformation parameter β .

PROPOSITION 5. *Let $\{x_t\}_{t \in \mathbb{Z}}$ be generated according to (2) where $\{\epsilon_t\}_{t \in \mathbb{Z}}$ is an iid sequence with $\mathbb{E}(\epsilon_t) = 0$ and $\mathbb{E}|\epsilon_t|^2 < \infty$. Let $\Theta \subset \mathbb{R}^{k+1}$ be given by,*

$$\Theta = \mathbb{R} \times (-\delta, \delta) \times (-c_2\Delta, c_2\Delta) \times \dots \times (-c_k\Delta, c_k\Delta). \quad (5)$$

$$\text{with } \Delta < \frac{1 - \delta}{\ell_1(\mathbf{c}, \mathbf{w}(\beta))} \quad \text{where } \ell_1(\mathbf{c}, \mathbf{w}(\beta)) := \sum_{i=2}^k |c_i| w_i(\beta) \quad (6)$$

so that $\ell_1(\mathbf{c}, \mathbf{w}(\beta, \iota))$ is a weighted l_1 -norm on the vector $\mathbf{c} := (c_2, \dots, c_k)$ with weights given in (3) and (4). Then, $\{x_t\}_{t \in \mathbb{Z}}$ is strictly stationary and ergodic for every $\theta \in \Theta$.

Note that the bound on Δ depends crucially on the bound δ that restricts the slope parameter θ_1 and on the transformation parameter β through the vector of weights $\mathbf{w}(\beta)$. Figure 2 plots the bounds for Δ as a function of $\beta \in [-2, 0)$ for different choices

of δ and order k , as well as alternative \mathbf{c} . The choice of vector \mathbf{c} takes into account the idea mentioned above that in applications with large k , it is natural to impose vanishing bounds on the sequence of parameters $\{\theta_i\}$ so that $c_i \rightarrow 0$ as $i \rightarrow \infty$. These cases are covered by the harmonically and geometrically vanishing bounds in Figure 2. In applications where k is small, the vector \mathbf{c} might however be chosen differently. The uniform bounds in Figure 2 illustrate the effects of one such alternative.

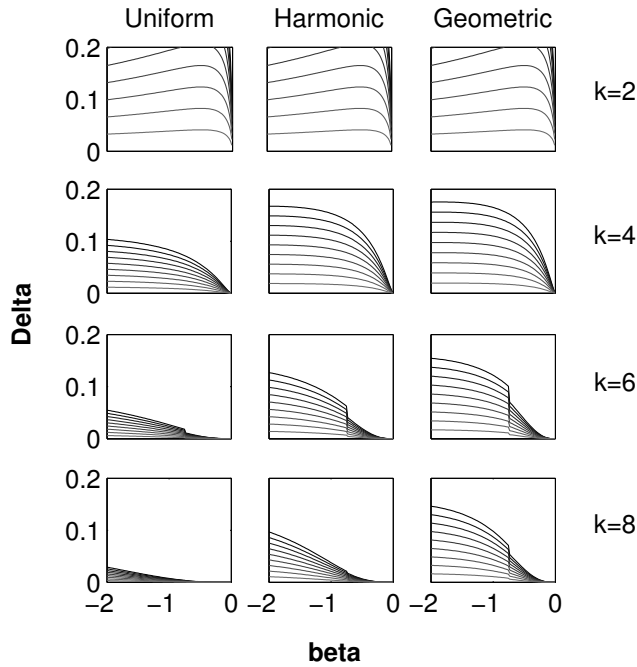


Figure 2: Plots of Δ as a function of β for different $\delta = 0.1, 0.2, \dots, 0.9$ and $k = 2, 4, 6, 8$ increasing from top row to bottom row, and uniform bounds $c_2 = \dots = c_k = 1$ (left column), arithmetic bounds $c_i = 1/(i - 1)$, $i = 2, \dots, k$ (center column) and geometric bounds $c_i = \gamma^{i-2}$, $\gamma = 0.5$, $i = 2, \dots, k$.

In each subplot of Figure 2, the graphs allowing for higher Δ curves over β are those that correspond to the smaller $\delta = 0.1$. These curves then decrease vertically as δ rises to $\delta = 0.9$. Note also that, in each subplot of Figure 2, the lack of smoothness of the graphs over certain β points is caused by the switch of bound implied by the changing ι in (4) as a function of β . Naturally, Figure 2 shows that for $k = 2$, the nature of the parameter space restriction (uniform, arithmetic or geometric) is

irrelevant. For large k however, the geometric bounds allow for larger Δ since they impose stricter bounds as $k \rightarrow \infty$. In contrast, the uniform bounds render the smallest Δ for increasing k .

Finally, Proposition 6 establishes the fading memory of the stochastic process $\{x_t\}_{t \in \mathbb{Z}}$ in (2) by appealing to a special case of Theorem 6.10 in Pötscher and Prucha (1997). In particular, it establishes conditions under which $\{x_t\}_{t \in \mathbb{Z}}$ is L_p -approximable by a mixing sequence and $\sup_t \|\xi\|_p = (\int |\xi|^p dP)^{1/p} < \infty$. The concept of L_p -approximability generalizes the well known notion of near epoch dependence on an α -mixing sequence (which corresponds essentially to L_2 -approximability). Similarly to strict stationarity and ergodicity, these notions of limited dependence can be used to obtain laws of large numbers and central limit theorems under appropriate conditions; see Theorems 6.2, 6.3, 10.1 and 10.2 in Pötscher and Prucha (1997).

PROPOSITION 6. *Let $\{x_t\}_{t \in \mathbb{Z}}$ be generated according to (2) where $\{\epsilon_t\}_{t \in \mathbb{Z}}$ is an iid sequence with $\|\epsilon_t\|_p < \infty$ and $\|x_0\|_p < \infty$ for some $p \geq 1$. Then, $\{x_t\}_{t \in \mathbb{Z}}$ is L_p -approximable by an α -mixing sequence and $\sup_t \|x_t\|_p < \infty \forall \theta_0 \in \mathbb{R}$ and $\boldsymbol{\theta} := (\theta_1, \dots, \theta_k)$ satisfying and $\ell_1(\boldsymbol{\theta}, \mathbf{w}(\beta)) < 1$ with weights defined in (3) and (4). The same holds for every $\boldsymbol{\theta} := (\theta_0, \theta_1, \dots, \theta_k) \in \Theta \subset \mathbb{R}^{k+1}$ with Θ defined in (5) and (6).*

5 SOME REMARKS ON ESTIMATION:

ASYMPTOTIC PROPERTIES AND FINITE SAMPLE BEHAVIOR

Under the results established above it is easy to obtain the asymptotic properties of various extremum estimators for both parametric or semi-nonparametric settings. For example, Tjostheim (1986) establishes, under appropriate regularity conditions, the consistency and asymptotic normality of least squares estimator for the conditional

mean of strictly stationary and ergodic time-series. Gallant (1987) and Gallant and White (1988) obtain similar results for extremum estimators under near epoch dependence limiting memory assumptions that are further generalized by Pötscher and Prucha (1997) to allow for processes that are L_p -Approximable by a mixing sequence. These properties, together with the approximation results of Section 3 form also the basis for the asymptotic results of sieve extremum estimators of semi-nonparametric models; see e.g. Chen (2007).

The details involved in obtaining the asymptotic properties of extremum estimators lie outside the scope of this paper. In any case, it is easy to show that e.g. Propositions 4 or 5 can be used in conjunction with Theorems 3.1 and 3.2 in Tjøstheim (1986) to obtain the consistency and asymptotic normality of the least squares estimator in transformed polynomial autoregression.

The following figures document the finite sample properties of the least squares estimator of the conditional mean under two alternative data generating processes. In both cases, the transformed polynomial autoregression is used to ‘approximate’ an unknown conditional mean, and hence, the autoregression is misspecified. In particular, the Monte Carlo exercise obtains artificial ‘observed’ sequences $\{x_t\}_{t \in \mathbb{N}}$ generated according to (1) with innovations $\{\epsilon_t\}_{t \in \mathbb{N}}$ are that are iid Gaussian $N(0, \sigma^2)$. In the first case, the conditional expectation function f_0 is smooth in x_{t-1} and given by,

$$f_0(x_{t-1}) = \alpha_0 + \alpha_1 \tanh(\alpha_2 x_{t-1}) \quad \forall t \in \mathbb{N}. \quad (7)$$

In the second case, f_0 is continuous in x_{t-1} and can be seen as a time-varying parameter autoregressive model with threshold dynamics,

$$f_0 = \alpha_0 + \alpha_t x_{t-1} \quad \text{with} \quad \alpha_t = \alpha_1 \mathbb{I}_{x_{t-1} \leq 0} + \alpha_2 \mathbb{I}_{x_{t-1} > 0} \quad \forall t \in \mathbb{N}. \quad (8)$$

A scatter of 100 simulated data points from both DGPs at the adopted parameter values for this Monte Carlo study is plotted in Figure 3.

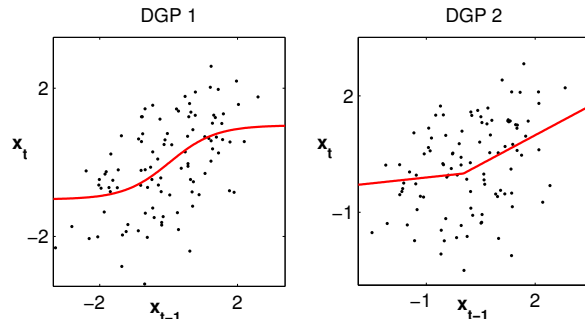


Figure 3: Plots of the conditional means (in red) of DGP 1 in (7) [left] and DGP 2 in (8) [right], and scatter plot of 100 pairs (x_t, x_{t-1}) generated using (7) [left] for $(\alpha_0, \alpha_1, \alpha_2, \sigma^2) = (0, 0.1, 0.5, 1.2)$ and (8) [right] for $(\alpha_0, \alpha_1, \alpha_2, \sigma^2) = (0, 1, 0.75, 1.2)$.

The two DGPs in (7) and (8) are quite different in several respects. In the first, x_t is a smooth bounded function of x_{t-1} with infinitely many derivatives. In the second case, the conditional expectation is only continuous on x_{t-1} as it is not differentiable at $x_{t-1} = 0$. This DGP generates two distinct regimes for $x_{t-1} \leq 0$ and $x_{t-1} > 0$, in virtue of the discontinuous threshold behavior of the time-varying coefficient α_t .

Figure 4 plots the density of the least squares estimator of the conditional mean function for $k = 1$ and $T = 100$ (obtained from 1000 Monte Carlo draws). For a small sample of $T = 100$ data points, a low order $k = 1$ seems appropriate. Clearly, this amounts to a simple linear autoregressive approximation of the nonlinear AR process and hence the choice of β is irrelevant. As expected, since the conditional mean is forced to take values in a space of linear functions $\tilde{\mathbb{P}}_1(\mathbb{R}) = \mathbb{P}_1(\mathbb{R})$, the approximation is naturally incapable of generating the nonlinear features that characterize the conditional expectation f_0 . In particular, the linear autoregressive model is incapable of describing the changes in slope that characterize the DGP's conditional expectation.

Despite model misspecification, the limit behavior of the estimator as $T \rightarrow \infty$

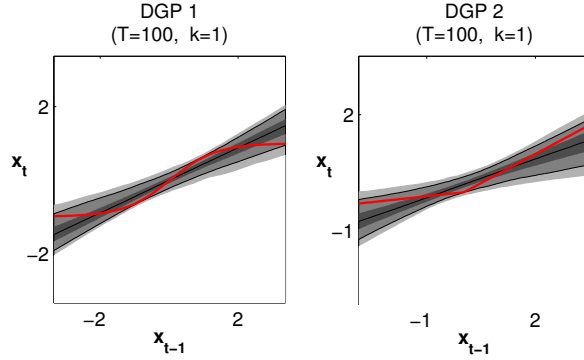


Figure 4: Estimates of conditional expectation obtained as $\hat{f}(x) = \hat{\theta}_0 + \hat{\theta}_1 x$ under the linear autoregressive model defined in (2) when $\theta_i = 0 \forall i \geq 2$, from $T = 100$ points of data generated by (7) [left: DGP 1] and (8) [right: DGP 2]. Light gray, gray and dark gray areas contain contain 99%, 95% and 75% of the mass respectively. Red line is the true conditional expectation f_0 . Dark line within dark gray area is the average estimated conditional expectation.

is generally well understood. In particular, under typical extremum estimation conditions, for fixed k and β , the least squares estimator converges as $T \rightarrow \infty$ to the limit point that provides a best linear approximation (in the least squares sense) to the underlying DGP; see e.g. Pötscher and Prucha (1997). Figure 5 documents the convergence of the linear regression estimates of the conditional expectation to a limit linear approximation.

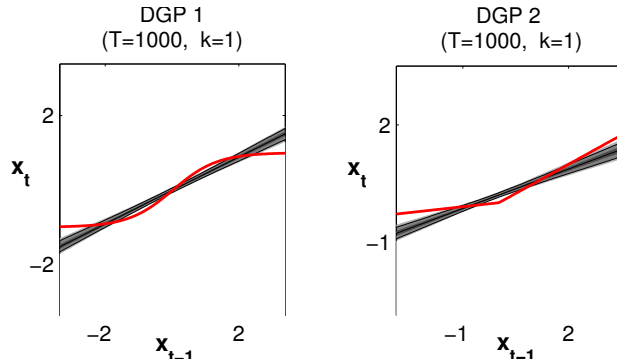


Figure 5: Estimates of conditional expectation obtained under the linear autoregressive model defined in (2) when $\theta_i = 0 \forall i \geq 2$, from $T = 1000$ points of data generated by (7) [left: DGP 1] and (8) [right: DGP 2]. Light gray, gray and dark gray areas contain contain 99%, 95% and 75% of the mass respectively. Red line is the true conditional expectation f_0 . Dark line within dark gray area is the average estimated conditional expectation.

For larger sample sizes, it can be however interesting to choose larger orders k .

Figure 6 plots the density of the estimated conditional expectation for certain pairs

(T, k) using the transformed polynomial approximation proposed in this paper. Here the estimates of conditional expectation lie in the space of transformed polynomials $\mathbb{P}_k(\mathbb{R})$ generated by a parameter space Θ with geometric decaying bounds with decay parameter $\gamma = 0.6$ and δ satisfying the bounds derived in Proposition 5 and plotted in Figure 2.

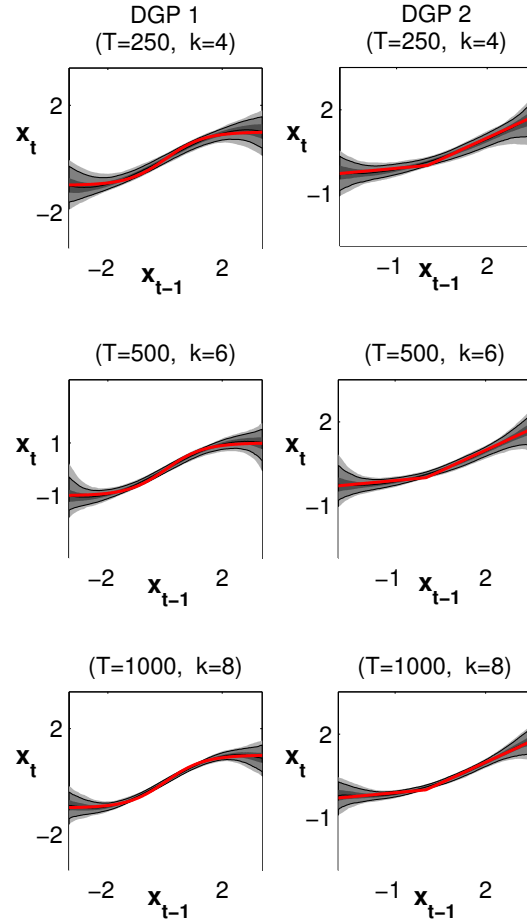


Figure 6: Estimates of conditional expectation for alternative pairs (T, k) using the transformed polynomial autoregression in (2) on data generated from (7) [left column: DGP 1] and data generated from (8) [right column: DGP 2]. Light gray, gray and dark gray areas contain contain 99%, 95% and 75% of the mass respectively. Red line is the true conditional expectation f_0 . Dark line within dark gray area is the average estimated conditional expectation.

Figure 6 reveals that the flexibility of the transformed polynomial is capable of approximating important features of both DGPs even at reasonably small sample

sizes T and low order k . These plots reveal the practical interest of transformed polynomials as a basis for both parametric and semi-nonparametric estimation of the conditional expectation. Most importantly, estimated conditional expectations define approximate autoregressive processes with properties that are well understood both in terms of approximation quality, through the results in Section 3, and stochastic behavior, through the results established in Section 4.

A PROOFS

PROOF OF PROPOSITION 1

We show first that $\tilde{\mathbb{P}}(\mathcal{X})$ is dense on $\mathbb{P}(\mathcal{X})$ in sup-norm, i.e. that $\forall(\epsilon, p) \in \mathbb{R}_0^+ \times \mathbb{P}(\mathcal{X}), \exists \tilde{p} \in \tilde{\mathbb{P}}(\mathcal{X}) : \sup_{x \in \mathcal{X}} |p(x) - \tilde{p}(x)| < \epsilon$. This is achieved by selecting, for every p , the element $\tilde{p} \in \tilde{\mathbb{P}}(\mathcal{X})$ with k and $\{\theta_i\}_{i=0}^k$ identical to those of p to obtain,

$$\sup_{x \in \mathcal{X}} |p(x) - \tilde{p}(x)| = \sup_{x \in \mathcal{X}} \left| (1 - \phi_\beta(x - x_0)) \cdot \sum_{i=2}^k \theta_i (x - x_0)^i \right|.$$

As a result,

$$\sup_{x \in \mathcal{X}} |p(x) - \tilde{p}(x)| \leq \left| 1 - \exp(\beta z^2) \right| \sum_{i=2}^k |\theta_i| z^i \quad \text{for } z = \left| \sup_{x \in \mathcal{X}} |x - x_0| \right|.$$

Hence, $\sup_{x \in \mathcal{X}} |p(x) - \tilde{p}(x)| < \epsilon$ is implied by,

$$\left| 1 - \exp(\beta z^2) \right| < \epsilon/A \quad \text{for any } A \geq \sum_{i=2}^k |\theta_i| z^i,$$

which holds for every ϵ by selecting a small enough β since $\left| 1 - \exp(\beta z^2) \right|$ can be made arbitrarily small by letting β vanish. Together with Weierstrass's Theorem this implies denseness of $\tilde{\mathbb{P}}(\mathcal{X})$ on $\mathbb{C}(\mathcal{X})$ since $\forall(\epsilon, f) \in \mathbb{R}_0^+ \times \mathbb{C}(\mathcal{X}), \exists(\tilde{p}, p) \in \tilde{\mathbb{P}}(\mathcal{X}) \times \mathbb{P}(\mathcal{X})$

such that,

$$\sup_{x \in \mathcal{X}} |f(x) - \tilde{p}(x)| \leq \sup_{x \in \mathcal{X}} |f(x) - p(x)| + \sup_{x \in \mathcal{X}} |p(x) - \tilde{p}(x)| < \epsilon/2 + \epsilon/2 = \epsilon. \quad \square$$

PROOF OF PROPOSITION 2

The first claim follows immediately by inspection of the proof of Proposition 1 where it was shown that,

$$\sup_{x \in \mathcal{X}} |p(x) - \Phi_\beta(p)(x)| \leq \left| 1 - \exp(\beta z^2) \right| \cdot A$$

for any $A > \sum_{i=2}^k |\theta_i| z^i$ where $z = \|\sup_{x \in \mathcal{X}} |x - x_0|\|$, and hence,

$$\limsup_{\beta \rightarrow 0} \sup_{x \in \mathcal{X}} |p(x) - \Phi_\beta(p)(x)| \leq A \cdot \limsup_{\beta \rightarrow 0} \left| 1 - \exp(\beta z^2) \right| = A \cdot 0 = 0.$$

The second claim follows first by norm sub-additivity,

$$\limsup_{\substack{j \rightarrow \infty \\ \beta \rightarrow 0}} \sup_{x \in \mathcal{X}} |f(x) - \Phi_\beta(p^j)(x)| \leq \limsup_{\substack{j \rightarrow \infty \\ \beta \rightarrow 0}} \sup_{x \in \mathcal{X}} |f(x) - p^j(x)| + \limsup_{\substack{j \rightarrow \infty \\ \beta \rightarrow 0}} \sup_{x \in \mathcal{X}} |p^j(x) - \Phi_\beta(p^j)(x)|$$

Then by noting that

$$\limsup_{\substack{j \rightarrow \infty \\ \beta \rightarrow 0}} \sup_{x \in \mathcal{X}} |f(x) - p^j(x)| = \limsup_{j \rightarrow \infty} \sup_{x \in \mathcal{X}} |f(x) - p^j(x)| = 0$$

holds by construction and that,³

$$\limsup_{\substack{j \rightarrow \infty \\ \beta \rightarrow 0}} \sup_{x \in \mathcal{X}} |p^j(x) - \Phi_\beta(p^j)(x)| = \limsup_{\substack{j \rightarrow \infty \\ \beta \rightarrow 0}} \sup_{x \in \mathcal{X}} \left| 1 - \phi_\beta(x - x_0) \right| \cdot \sup_{x \in \mathcal{X}} \left| \sum_{i=2}^{k_j} \theta_{i,j} (x - x_0)^i \right|$$

³Note that here we cannot use the previous construction and have $\lim_{j \rightarrow \infty, \beta \rightarrow 0} \sup_{x \in \mathcal{X}} |p^j(x) - \Phi_\beta(p^j)(x)| \leq \lim_{j \rightarrow \infty} A_j \cdot \lim_{\beta \rightarrow 0} \sup_{x \in \mathcal{X}} |1 - \exp(\beta z^2)|$ because $A_j = \sum_{i=2}^{k_j} |\theta_{i,j}| z^i$ may diverge as $j \rightarrow \infty$ since the coefficients $\theta_{i,j}$ need not be absolutely summable.

which implies,

$$\lim_{\substack{j \rightarrow \infty \\ \beta \rightarrow 0}} \sup_{x \in \mathcal{X}} |p^j(x) - \Phi_\beta(p^j)(x)| = \lim_{\beta \rightarrow 0} |1 - \phi_\beta(z)| \cdot \lim_{j \rightarrow \infty} \left| \sum_{i=2}^{k_j} \theta_{i,j} z^i \right| \leq 0 \cdot B = 0$$

where $z = |\sup_{x \in \mathcal{X}} |x - x_0||$ and $\lim_{j \rightarrow \infty} \left| \sum_{i=2}^{k_j} \theta_{i,j} z^i \right| \leq B$ for some $B < \infty$ because $f \in \mathbb{C}(\mathcal{X})$ implies that f is bounded on the compact \mathcal{X} , and uniform convergence of $p^j \rightarrow f$ implies that p^j is bounded on \mathcal{X} uniformly on $j \geq j^*$ for larger enough $j^* \in \mathbb{N}$.

PROOF OF PROPOSITION 3

Note first that for any $f \in \mathbb{C}(\mathcal{X})$, and every $k \in \mathbb{N}$,

$$\sup_{x \in \mathcal{X}} |f(x) - \tilde{p}_{k,\beta}^*(x)| \leq \sup_{x \in \mathcal{X}} |f(x) - p_k^*(x)| + \sup_{x \in \mathcal{X}} |p_k^*(x) - \tilde{p}_{k,\beta}^*(x)|.$$

Note also that,

$$\begin{aligned} \sup_{x \in \mathcal{X}} |p_k^*(x) - \tilde{p}_{k,\beta}^*(x)| &= \sup_{x \in \mathcal{X}} \left| (1 - \phi_\beta(x - x_0)) \sum_{i=2}^k \theta_i (x - x_0)^i \right| \\ &\leq |1 - \exp(\beta z^2)| \cdot A \end{aligned}$$

where A denotes the upper bound $\sup_k \left| \sum_{i=2}^k \theta_i z^i \right| \leq A$ with $z = |\sup_{x \in \mathcal{X}} |x - x_0||$ as in the proof of Proposition 1, and A does not depend on k and satisfies $A < \infty$ because by Weierstrass's theorem $p_k^* \rightarrow f$ uniformly as $k \rightarrow \infty$ and hence $\{p_k^*\}$ is uniformly bounded by the same argument as in Proposition 2.

Finally, we obtain for $f \in \mathbb{C}^r(\mathcal{X})$, $r \geq 1$, and $\beta = O(\log(k^{-sr}))$,

$$\begin{aligned} \sup_{x \in \mathcal{X}} \left| f(x) - \tilde{p}_{k,\beta}^*(x) \right| &\leq \sup_{x \in \mathcal{X}} \left| f(x) - p_k^*(x) \right| + A \cdot \left| 1 - \exp(\beta z^2) \right| \\ &= O(k^{-r}) + O(1) \cdot \left| 1 - \exp(O(\log(k^{-sr}))) \right| \\ &= O(k^{-r}) + O(1) \cdot O(k^{-sr}) = O(k^{-r \min\{s,1\}}) \text{ as } k \rightarrow \infty \end{aligned}$$

since $\sup_{x \in \mathcal{X}} \left| f(x) - p_k^*(x) \right| = O(k^{-r})$ as $k \rightarrow \infty$ by Lemma 2. By the same argument, we obtain $\sup_{x \in \mathcal{X}} \left| f(x) - \tilde{p}_{k,\beta}^*(x) \right| = O(k^{-(r+\alpha) \min\{s,1\}})$ for $f \in \mathbb{C}^{k,\alpha}(\mathcal{X})$ and $\beta = O(\log(k^{-s(r+\alpha)}))$, and $\sup_{x \in \mathcal{X}} \left| f(x) - \tilde{p}_{k,\beta}^*(x) \right| = O(k^{-(r+1) \min\{s,1\}})$ for $f \in \mathbb{C}^{k,1}(\mathcal{X})$ and $\beta = O(\log(k^{-s(r+1)}))$.

PROOF OF PROPOSITION 4

It follows immediately by Theorem 3.1 in Bougerol (1991), that a sequence $\{x_t\}_{t \in \mathbb{N}}$ converges to a unique strictly stationary and ergodic solution if $|x_0| < \infty$, $\{\epsilon_t\}_{t \in \mathbb{N}}$ is iid with $\mathbb{E}|\epsilon_t|^2 < \infty$ for some t and the uniform contraction condition $\sup_{x \in \mathbb{R}} \left| \partial \tilde{p}_{k,\beta}(x) / \partial x \right| < 1$ holds. We conclude that the time-series $\{x_t\}_{t \in \mathbb{Z}}$ is strictly stationary and ergodic, under the maintained assumptions, if the uniform contraction condition holds. Using the definition of $\tilde{p}_{k,\beta}(x)$ and norm sub-additivity,

$$\begin{aligned} \sup_{x \in \mathbb{R}} \left| \frac{\partial \tilde{p}_{k,\beta}(x)}{\partial x} \right| &= \sup_{x \in \mathbb{R}} \left| \theta_1 + \left(\sum_{i=2}^k i \theta_i (x - \theta_0)^{i-1} \right) \cdot \phi_\beta(x - \theta_0) \right. \\ &\quad \left. + \left(\sum_{i=2}^k \theta_i (x - \theta_0)^i \right) \cdot 2\beta(x - \theta_0) \cdot \phi_\beta(x - \theta_0) \right| \\ &\leq |\theta_1| + \sum_{i=2}^k i |\theta_i| \sup_{x \in \mathbb{R}} \left| (x - \theta_0)^{i-1} \cdot \phi_\beta(x - \theta_0) \right| \\ &\quad + \sum_{i=2}^k |\theta_i| \sup_{x \in \mathbb{R}} \left| (x - \theta_0)^i \cdot 2\beta(x - \theta_0) \cdot \phi_\beta(x - \theta_0) \right| \end{aligned}$$

and hence,

$$\begin{aligned} \sup_{x \in \mathbb{R}} \left| \frac{\partial \tilde{p}_{k,\beta}(x)}{\partial x} \right| &\leq |\theta_1| + \sum_{i=2}^k |\theta_i| \left(i \sup_{x \in \mathbb{R}} \left| (x - \theta_0)^{i-1} \cdot \phi_\beta(x - \theta_0) \right| \right. \\ &\quad \left. - 2\beta \sup_{x \in \mathbb{R}} \left| (x - \theta_0)^{i+1} \cdot \phi_\beta(x - \theta_0) \right| \right). \end{aligned} \quad (9)$$

Now, since we can split the inequality as,

$$\sup_{x \in \mathbb{R}} \left| \frac{\partial \tilde{p}_{k,\beta}(x)}{\partial x} \right| = \max \left\{ \sup_{x \in \mathbb{R} : |x - \theta_0| \leq 1} \left| \frac{\partial \tilde{p}_{k,\beta}(x)}{\partial x} \right|, \sup_{x \in \mathbb{R} : |x - \theta_0| > 1} \left| \frac{\partial \tilde{p}_{k,\beta}(x)}{\partial x} \right| \right\}$$

and use (9) to obtain $\forall x \in \mathbb{R} : |x - \theta_0| \leq 1$ the inequality,

$$\sup_{x \in \mathbb{R} : |x - \theta_0| \leq 1} \left| \frac{\partial \tilde{p}_{k,\beta}(x)}{\partial x} \right| \leq |\theta_1| + \sum_{i=2}^k |\theta_i| (i - 2\beta) \sup_{x \in \mathbb{R}} \left| (x - \theta_0)^{i-1} \cdot \phi_\beta(x - \theta_0) \right|$$

and $\forall x \in \mathbb{R} : |x - \theta_0| > 1$ the inequality,

$$\sup_{x \in \mathbb{R} : |x - \theta_0| > 1} \left| \frac{\partial \tilde{p}_{k,\beta}(x)}{\partial x} \right| \leq |\theta_1| + \sum_{i=2}^k |\theta_i| (i - 2\beta) \sup_{x \in \mathbb{R}} \left| (x - \theta_0)^{i+1} \cdot \phi_\beta(x - \theta_0) \right|$$

it follows that the unrestricted supremum satisfies,

$$\sup_{x \in \mathbb{R}} \left| \frac{\partial \tilde{p}_{k,\beta}(x)}{\partial x} \right| \leq \ell_1(\boldsymbol{\theta}, \mathbf{w}(\beta))$$

where $\ell_1(\boldsymbol{\theta}, \mathbf{w}(\beta)) := \sum_{i=1}^k w_i |\theta_i|$ is a weighted l_1 -norm on the parameter vector $\boldsymbol{\theta} := (\theta_1, \dots, \theta_k)$ with weights given by,

$$w_1(\beta) = 1 \quad \text{and} \quad w_i(\beta) = (i - 2\beta) \cdot \max \left\{ \xi_i(\beta, -1), \xi_i(\beta, 1) \right\} \quad \text{for } i = 2, \dots, k$$

$$\text{with } \xi_i(\beta, \iota) = \sup_{x \in \mathbb{R}} \left| (x - \theta_0)^{i+\iota} \cdot \phi_\beta(x - \theta_0) \right|.$$

Furthermore, by standard calculus optimization it follows that the weights $w_i(\beta, \iota)$

can be given the exact expression,

$$w_i(\beta, \iota) = (i - 2\beta) \left(-\frac{i + \iota}{2\beta} \right)^{\frac{i+\iota}{2}} \exp\left(-\frac{i + \iota}{2}\right), \quad i = 2, \dots, k$$

$$\text{with } \iota = \begin{cases} -1 & \forall(\beta, i) : \beta < -\frac{(i+1)^{\frac{i+1}{2}}}{2e(i-1)^{\frac{i-1}{2}}}, \\ 1 & \text{otherwise} \end{cases},$$

where e is Euler's number, by noting that,

$$\arg \max_{x \in \mathbb{R}} \left| (x - \theta_0)^{i+\iota} \exp(\beta(x - \theta_0)^2) \right| = \theta_0 \pm \sqrt{-\frac{i + \iota}{2\beta}} \quad \forall \iota \in \mathbb{N},$$

and hence that,

$$\xi_i(\beta, \iota) = \sup_{x \in \mathbb{R}} \left| (x - \theta_0)^{i+\iota} \exp(\beta(x - \theta_0)^2) \right| = \left(-\frac{i + \iota}{2\beta} \right)^{\frac{i+\iota}{2}} \exp\left(-\frac{i + \iota}{2}\right) \quad \forall \iota \in \mathbb{N}.$$

and this implies that,

$$\max \left\{ \xi_i(\beta, -1), \xi_i(\beta, 1) \right\} = \begin{cases} \xi_i(\beta, -1) & \forall(\beta, i) : \beta < -\frac{(i+1)^{\frac{i+1}{2}}}{2e(i-1)^{\frac{i-1}{2}}}, \\ \xi_i(\beta, 1) & \text{otherwise} \end{cases},$$

because,

$$\begin{aligned} \xi_i(\beta, -1) > \xi_i(\beta, 1) &\Leftrightarrow \left(-\frac{i-1}{2\beta} \right)^{\frac{i-1}{2}} \exp\left(-\frac{i-1}{2}\right) > \left(-\frac{i+1}{2\beta} \right)^{\frac{i+1}{2}} \exp\left(-\frac{i+1}{2}\right) \\ \Leftrightarrow -2\beta \frac{(i-1)^{\frac{i-1}{2}}}{(i+1)^{\frac{i+1}{2}}} > \exp(-1) &\Leftrightarrow -2\beta > \frac{(i+1)^{\frac{i+1}{2}}}{e(i-1)^{\frac{i-1}{2}}} \Leftrightarrow \beta < -\frac{(i+1)^{\frac{i+1}{2}}}{2e(i-1)^{\frac{i-1}{2}}}. \end{aligned}$$

PROOF OF PROPOSITION 5

The desired result obtained by noting that (9), together with $\sup_{\theta_1 \in \Theta_1} |\theta_1| = \delta$ and $\sup_{\theta_i \in \Theta} |\theta_i| = c_i \Delta$, $i \geq 2$, implies,

$$\begin{aligned} \sup_{\theta \in \Theta} \sup_{x \in \mathbb{R}} \left| \frac{\partial \tilde{p}_{k,\beta}(x)}{\partial x} \right| &\leq \sup_{\theta \in \Theta} |\theta_1| + \sum_{i=2}^k i \sup_{\theta_i \in \Theta} |\theta_i| \sup_{x \in \mathbb{R}} \left| (x - \theta_0)^{i-1} \cdot \phi_\beta(x - \theta_0) \right| \\ &\quad - 2\beta \sum_{i=2}^k \sup_{\theta_i \in \Theta} |\theta_i| \sup_{x \in \mathbb{R}} \left| (x - \theta_0)^{i+1} \phi_\beta(x - \theta_0) \right| \\ &\leq \delta + \Delta \sum_{i=2}^k i c_i \sup_{x \in \mathbb{R}} \left| (x - \theta_0)^{i-1} \cdot \phi_\beta(x - \theta_0) \right| \\ &\quad - 2\beta \delta \sum_{i=2}^k c_i \sup_{x \in \mathbb{R}} \left| (x - \theta_0)^{i+1} \phi_\beta(x - \theta_0) \right|. \end{aligned}$$

This implies, by the same argument as encountered in the proof of 4, that,

$$\sup_{\theta \in \Theta} \sup_{x \in \mathbb{R}} \left| \frac{\partial \tilde{p}_{k,\beta}(x)}{\partial x} \right| < 1 \Leftrightarrow \delta + \Delta \ell_1(\mathbf{c}, \mathbf{w}(\beta, \iota)) < 1 \Leftrightarrow \delta < \frac{1 - \delta}{\ell_1(\mathbf{c}, \mathbf{w}(\beta, \iota))}$$

where $\ell_1(\mathbf{c}, \mathbf{w}(\beta, \iota)) := \sum_{i=2}^k w_i(\beta, \iota) |c_i|$ is a weighted l_1 -norm on the vector $\mathbf{c} := (c_2, \dots, c_k)$ with weights, as derived in the proof of Proposition 4, given by,

$$w_i(\beta, \iota) = (i - 2\beta) \left(-\frac{i + \iota}{2\beta} \right)^{\frac{i+\iota}{2}} \exp\left(-\frac{i + \iota}{2}\right), \quad i = 2, \dots, k$$

$$\text{with } \iota = \begin{cases} -1 & \forall (\beta, i) : \beta < -\frac{(i+1)^{\frac{i+1}{2}}}{2e(i-1)^{\frac{i-1}{2}}} \\ 1 & \text{otherwise} \end{cases}.$$

PROOF OF PROPOSITION 6

Follows immediately by Theorem 6.10 of Pötscher and Prucha (1997) that the uniform bound $\sup_{x \in \mathbb{R}} \left| \partial \tilde{p}_{k_T}(x) / \partial x \right| < 1$ together with $\{\epsilon_t\}_{t \in \mathbb{N}}$ being iid with $\|\epsilon_t\|_p < \infty$ and $\|x_0\|_p < \infty$ imply that $\{x_t\}_{t \in \mathbb{N}}$ is L_p -approximable by the mixing sequence $\{\epsilon_t\}_{t \in \mathbb{N}}$ with $\sup_t \|x_t\| < \infty$. The proofs of Propositions 4 and 5 reveal that $\sup_{x \in \mathbb{R}} \left| \partial \tilde{p}_{k_T}(x) / \partial x \right| < 1$ under the stated conditions.

REFERENCES

- Bougerol, P. (1991). *Kalman Filtering with Random Coefficients and Contractions*. Prépublications de l'Institut Elie Cartan. Univ. de Nancy.
- Chen, X. (2007). Large sample sieve estimation of semi-nonparametric models. In Heckman, J. and Leamer, E., editors, *Handbook of Econometrics*, volume 6 of *Handbook of Econometrics*, chapter 76. Elsevier.
- Gallant, A. R. (1987). Identification and consistency in seminonparametric regression. *Bewley, Truman F., ed. (1987), Advances in Econometrics Fifth World Congress*, 1:145–170.
- Gallant, R. and White, H. (1988). *A Unified Theory of Estimation and Inference for Nonlinear Dynamic Models*. Cambridge University Press.
- Granger, C. W. J. and Teräsvirta, T. (1993). *Modelling Non-Linear Economic Relationships*. Oxford University Press.
- Ozaki, T. (1980). Non-linear time series models for non-linear random vibrations. *Journal of Applied Probability*, 17:84–93.
- Pötscher, B. M. and Prucha, I. R. (1997). *Dynamic Nonlinear Econometric Models: Asymptotic Theory*. Springer-Verlag.
- Teräsvirta, T., Tjøstheim, D., and Granger, C. W. J. (2010). *Modelling Nonlinear Economic Time Series*. Oxford University Press.
- Timan, A. (1963). *Theory of approximation of functions of a real variable*. International series of monographs in pure and applied mathematics. Macmillan.
- Tjøstheim, D. (1986). Estimation in nonlinear time series models. *Stochastic Process Appl.*, 21:251–273.