Econometric Theory, **18**, 2002, 1408–1448. Printed in the United States of America.
DOI: 10.1017/S0266466602186075

NON- AND SEMIPARAMETRIC IDENTIFICATION OF SEASONAL NONLINEAR AUTOREGRESSION MODELS

LIJIAN YANG
Michigan State University

ROLF TSCHERNIG
University of Maastricht

Non- or semiparametric estimation and lag selection methods are proposed for three seasonal nonlinear autoregressive models of varying seasonal flexibility. All procedures are based on either local constant or local linear estimation. For the semiparametric models, after preliminary estimation of the seasonal parameters, the function estimation and lag selection are the same as nonparametric estimation and lag selection for standard models. A Monte Carlo study demonstrates good performance of all three methods. The semiparametric methods are applied to German real gross national product and UK public investment data. For these series our procedures provide evidence of nonlinear dynamics.

1. INTRODUCTION

In nonlinear time series analysis, nonparametric estimators provide great flexibility because no parametric function class must be chosen a priori. On the other hand, most existing results on nonparametric estimators require the data generating process to be stationary, a condition often violated by economic time series. Although the most common source of nonstationarity is trends, seasonal patterns also play an important role.

The flexibility of nonparametric techniques has not been available for seasonal time series, because of a lack of nonparametric autoregression models that incorporate seasonal nonstationarity. A popular approach for removing sea-

This research was supported by the Deutsche Forschungsgemeinschaft via Sonderforschungsbereich 373 "Quantifikation und Simulation Ökonomischer Prozesse" at Humboldt-Universität zu Berlin. In part, the first author's research was also supported by NSF grant DMS 9971186. The authors thank Richard Blundell and Herman van Dijk for suggesting investigation of the seasonal shift model and Jörg Breitung and Helmut Lütkepohl for valuable comments on an earlier draft. The helpful comments of two anonymous referees, the associate editor, and the editor are also very much appreciated. Address correspondence to: Lijian Yang, Department of Statistics and Probability, Michigan State University, East Lansing, MI 48824, USA; e-mail: yang@stt.msu.edu; and Rolf Tschernig, Department of Quantitative Economics, University of Maastricht, Postbus 616, 6200 MD Maastricht, The Netherlands; e-mail: r.tschernig@ke.unimass.nl.

  0266-4666/02 $9.50

sonal nonstationarity is to use seasonally adjusted data. This, however, is not justified for nonlinear modeling for at least four reasons. First, the effect of such seasonal filters on data exhibiting nonlinearities is unclear as virtually all seasonal adjustment procedures have been designed for linear processes. Second, Ghysels, Granger, and Siklos (1996) show that some of these procedures such as X-11 involve nonlinear transformations that may change the properties of the original data. Third, data adjusted with most model-based seasonal adjustment procedures and procedures with model-based interpretation are noninvertible. Such procedures include those used by official agencies. See, e.g., Maravall (1995) for details.[1] Thus, an additional approximation error is introduced if finite-order AR models, either parametric or nonparametric, are used. Finally, using seasonally adjusted data is misleading if a useful orthogonal decomposition of the original data into a trend, a seasonal, and an irregular component does not exist. See the examples in the work by Franses (1996, Ch. 6), who advocates periodic linear autoregressive models with autoregression parameters that vary with the seasons. Therefore, standard nonparametric models are not appropriate for seasonally adjusted data.

In this paper we consider three seasonal nonlinear autoregressive models of varying seasonal flexibility. The most general model allows for changing conditional means across seasons and generalizes periodic autoregressive models. For this model, which provides a very flexible way to model seasonality without imposing much structure, we suggest nonparametric estimation and lag selection methods and state some asymptotic properties.

This generality has its price. The effective sample size of the nonparametric procedures is given by the size of a single season. This model may therefore be less useful for small data sets typical in macroeconomics. For such a task we propose a less flexible seasonal nonlinear model whose seasonal regression functions are equal to additive constants. We suggest three semiparametric estimators that are shown to have the standard effective sample size.

The same is shown for the semiparametric estimator of an alternative seasonal process that can be decomposed into a nonseasonal nonlinear autoregressive component and additive deterministic seasonal shifts. These seasonal shifts can be used for seasonally adjusting data. An alternative model is analyzed by Burman and Shumway (1998), who allow the seasonal shifts to be multiplied by a nonlinear function of time, however, at the cost of assuming the nonseasonal component to be linear. Orbe, Ferreira, and Rodriguez-Poo (2000) develop nonparametric estimators of time-varying coefficients under seasonal constraints.

In practice, the relevant lags of the autoregression are unknown, and a lag selection procedure is needed. We therefore extend the final prediction error methods of Tschernig and Yang (2000) to the seasonal models and show that the probability of selecting the correct lags approaches one asymptotically. Our simulation study shows that the proposed seasonal lag selection methods work in small samples. Moreover, we find that the non- and semiparametric procedures for nonlinear processes outperform linear methods in terms of the prediction power if the processes are nonlinear. This holds for prediction based on both the correct or the selected set of lags.

To illustrate the semiparametric procedures, we model two macroeconomic time series, German real gross national product (GNP) and UK public investment, for which series we find evidence of nonlinear dynamics. For the former, which exhibits stronger nonlinearity, the semiparametric procedure forecasts substantially better.

It is well known that multivariate function estimation suffers from inaccuracy, commonly referred to as the "curse of dimensionality." This nuisance can only be reduced by imposing special restrictions on a general multivariate function to obtain a less flexible structure. Such examples include the generalized additive structure and varying-coefficient structure proposed by Hastie and Tibshirani (1990, 1993) and in time series analysis, the additive structure of the conditional mean in Chen and Tsay (1993), Tjøstheim and Auestad (1994a), Masry and Tjøstheim (1996), and Yang (2000) or the multiplicative structure of the conditional volatility in Yang, Härdle, and Nielsen (1999). Further research could provide guidance on how such functional restrictions should be imposed on the seasonal models we propose here.

In contrast, lag selection that has to precede any function estimation suffers less from the curse of dimensionality. This robustness of nonparametric lag selection can be attributed to its discrete nature and has already been observed for nonseasonal AR processes in Auestad and Tjøstheim (1990) and Tjøstheim and Auestad (1994b) and more recently in Tschernig and Yang (2000). These authors suggest the idea of "de-linking" lag selection from function estimation, so that the relevant lags may be first selected using a very general nonparametric procedure, after which appropriate structures may be imposed on the selected variables for improved function estimation. For example, the multiplicative modeling of the conditional volatility function in Yang et al. (1999) was carried out not on arbitrary lagged variables but on those selected according to the nonparametric lag selection method of Tschernig and Yang (2000). In our Monte Carlo study the robustness of lag selection is also corroborated for seasonal processes. The identification rates of non- or semiparametric methods for the correct lags are about 50%, whereas those of linear methods may be close to zero (see the correct identification rates represented by dark rectangles in Figures 5 and 6). Therefore we regard the lag selection methods proposed in this paper as preliminary steps for imposing additive or other structures on seasonal autoregression. As such, they are quite satisfactory.

As a final remark we note that the three models only cover various kinds of deterministic seasonality. Nonstationarity due to stochastic seasonality has to be removed prior to the non- or semiparametric modeling (just like trends). To avoid overdifferencing and thus a noninvertible series one may use the HEGY test (Hylleberg, Engle, Granger, and Yoo, 1990).

The paper is organized as follows. In the next section we discuss three seasonal nonlinear autoregressive models with different kinds of seasonal flexibility. Section 3 presents nonparametric estimation and lag selection for the general seasonal nonlinear autoregressive process. In Section 4 we present semiparametric estimators for the two restricted seasonal models. Section 5 describes details of implementing the various procedures. The results of the Monte Carlo study are presented in Section 6. The empirical applications are contained in Section 7, and Section 8 concludes. All assumptions, lemmas, and proofs are in the Appendix. JMULTI, which is a menu-driven software based on GAUSS, contains almost all procedures that are presented in this paper. It is available from http://ise.wiwi.hu-berlin.de/oekonometrie/.

2. SEASONAL NONLINEAR AUTOREGRESSIONS

Assume now that the process $\{Y_t\}_{t\geq 0}$ has a stationary distribution for each of the S seasons. It will often be convenient to write the time index t as $t = s + S\tau$ where $s = 0,1,\ldots,S-1$ denotes the season and $\tau = 0,1,\ldots$ represents a new time index. Throughout this paper, we consider a realization $\{Y_t\}_{t=0}^{n}$ of sample size $n+1$.

The most general seasonal process that we consider is the seasonal nonlinear autoregressive (SNAR) model given by

$$Y_{s+\tau S} = f_s(X_{s+\tau S}) + \sigma_s(X_{s+\tau S})\xi_{s+\tau S}, \tag{1}$$

where $X_t = (Y_{t-i_1}, Y_{t-i_2}, \ldots, Y_{t-i_m})^T$ is the vector of all the correct lagged values, $i_1 < \cdots < i_m$, and the ξ_t's are independent and identically distributed (i.i.d.) with $E(\xi_t) = 0$, $E(\xi_t^2) = 1$, $t = s + \tau S = i_m, i_m + 1, \ldots$, and they are independent of the start-up condition X_{i_m}. Note that the conditional volatility functions $\sigma_s(\cdot)$ may depend on a subvector of $X_{s+\tau S}$ or even be constant. The case in which $\sigma_s(\cdot)$ depends on lags not in $f_s(\cdot)$ is beyond the scope of this paper. In contrast to the standard nonlinear autoregression model the regression functions $\{f_s\}_{s=0}^{S-1}$ here are allowed to vary with the S seasons. This is a nonlinear generalization of the periodic AR (PAR) model

$$Y_{s+\tau S} = b_s + \sum_{i=1}^{p} \alpha_{is} Y_{s+\tau S-i} + \epsilon_{s+\tau S} \tag{2}$$

(see, e.g., Franses, 1996, p. 93; Lütkepohl, 1991, p. 391). For this reason, one can also view the SNAR model as a periodic nonlinear autoregression. The nonparametric estimation of the S regression functions $f_s(\cdot)$ and selection of the lags $i_1 < \cdots < i_m$ will be discussed in Section 3. Note that i_m can be much larger than m. For example, the selected lag vector for the German real GDP is given by (1,4,7), so $m = 3, i_m = 7$ (see Table 4).

We do not allow the set of lags $i_1 < \cdots < i_m$ to vary with the seasons. In the latter case the task of estimation and lag selection has to be carried out separately for each of the S data sets $\{(X_{s+\tau S}, Y_{s+\tau S})\}_{\tau=0}^{[n/S]}, s = 0,1,\ldots,S-1$ using methods for nonseasonal models. One therefore can directly apply, e.g., the procedure of Tschernig and Yang (2000). Pooling the information for all seasons is only useful if some features of the seasonal process, e.g., the lags, are the same across seasons.

As will be seen in Section 3, the effective sample size for estimation and lag selection of model (1) is n/S. For some macroeconomic applications this may be too small. For example, model (1) provides too much flexibility for 30 years of quarterly data. One may, however, restrict the seasonal flexibility in the conditional mean functions to $f_s(\cdot) = f(\cdot) + b_s$, $s = 0,1,2,\dots,S-1$ so that the seasonal variation of the functions between the sth and the 0th season is restricted to the constant shifts b_s. By definition $b_0 = 0$. The resulting process,

$$Y_{s+\tau S} = f(X_{s+\tau S}) + b_s + \sigma_s(X_{s+\tau S})\xi_{s+\tau S}, \tag{3}$$

is a restricted seasonal nonlinear autoregression. We call this second model a seasonal dummy nonlinear autoregressive (SDNAR) model because it is a generalization of the seasonal dummy linear autoregressive (SDAR) model

$$Y_{s+\tau S} = b_s + \sum_{i=1}^{p} \alpha_i Y_{s+\tau S-i} + \epsilon_{s+\tau S}. \tag{4}$$

In Section 4 we show that after estimating the seasonal shifts b_s, the nonparametric function $f(\cdot)$ in the SDNAR model (3) can be estimated with an effective sample size of n. The same also holds for lag selection.

Another way of restricting the seasonal nonlinear autoregression model (1) is to assume that the seasonal process is additively separable into a seasonal mean shift δ_s, $s = 0,1\dots,S-1$, and a nonseasonal nonlinear autoregression $\{U_t\}$, i.e., $Y_{s+\tau S} = \delta_s + U_{s+\tau S}$. One may call

$$\begin{aligned} Y_{s+\tau S} - \delta_s &= f(Y_{s+\tau S-i_1} - \delta_{\{s-i_1\}},\dots,Y_{s+\tau S-i_m} - \delta_{\{s-i_m\}}) \\ &\quad + \sigma(Y_{s+\tau S-i_1} - \delta_{\{s-i_1\}},\dots,Y_{s+\tau S-i_m} - \delta_{\{s-i_m\}})\xi_{s+\tau S} \end{aligned} \tag{5}$$

a seasonal shift nonlinear autoregressive (SHNAR) model. Here we define $\{a\}$ for any integer a as the unique integer between 0 and $S-1$ that is in the same congruence class as a modulo S. For identifiability, one assumes that $\delta_0 = 0$. This SHNAR model is another way of generalizing the SDAR model (4) where the constants $\delta_0,\dots,\delta_{S-1}$ of the linear model are obtained up to an additive constant via the system of linear equations $b_s = \delta_s - \sum_{i=1}^{p} \alpha_i \delta_{\{s-i\}}$, $s = 0,1,\dots,S-1$.

For estimating the seasonal mean shifts $\delta_1,\dots,\delta_{S-1}$ in the SHNAR model (5), a simple parametric method is proposed in Section 4 that allows us to estimate and analyze the nonseasonal process $\{U_{s+\tau S} = Y_{s+\tau S} - \delta_s\}$ by standard nonparametric methods.[2] We remark that none of the proposed models allows the seasonal features to vary with time. Although this also is possible, it is likely to provide too much flexibility for typical sample sizes. Alternatively, one may restrict the nonseasonal part of the conditional mean function to be linear in X_t but allow the seasonal shift to depend on time (see, e.g., Burman and Shumway, 1998).

3. SNAR IDENTIFICATION

Model identification requires two steps: lag selection and estimation. For the lag selection procedure presented in Section 3.2 one a priori has to select M, which is the largest lag considered in the lag search. Therefore, we reserve the first M observations as starting values and for each $M \le t \le n$ define the full lag vector $X_{t,M} = (Y_{t-1}, \dots, Y_{t-M})^T$. Note that the candidate lag vector X_t is a subvector of $X_{t,M}$. In the sequel, x and x_M denote values of X_t and $X_{t,M}$, respectively. If one is only concerned with estimation and if the correct lag vector X_t is known, the largest lag M is set to the largest lag i_m contained in the lag vector X_t. We need some additional notation. Let $i_{M,S}$ be the smallest integer equal to or greater than M/S, $n_S = [(n+1)/S] - 1$, and $n_{M,S} = n_S + 1 - i_{M,S}$ and for each $s = 0, 1, \dots, S-1$, denote $\mathbf{Y}_s = (Y_{s+i_{M,S}S}, Y_{s+(i_{M,S}+1)S}, \dots, Y_{s+n_S S})^T$.

3.1. Estimation

For any $x \in \mathbb{R}^m$, the Nadaraya–Watson estimate $\hat{f}_{1,s}(x)$ and local linear estimate $\hat{f}_{2,s}(x)$ of the seasonal functions $f_s(x)$ in the SNAR model (1) are given by

$$\hat{f}_{a,s}(x) = \hat{f}_{a,s}(x,h) = e_a^T \{Z_{a,s}^T(x) W_s(x) Z_{a,s}(x)\}^{-1} Z_{a,s}^T(x) W_s(x) \mathbf{Y}_s, \qquad a = 1,2, \tag{6}$$

in which

$$Z_{1,s}(x) = (1 \quad \dots \quad 1)^T_{1 \times n_{M,S}}, \qquad Z_{2,s}(x) = \begin{pmatrix} 1 & \dots & 1 \\ \dfrac{X_{s+Si_{M,S}} - x}{h} & \dots & \dfrac{X_{s+Sn_S} - x}{h} \end{pmatrix}^T,$$

$$e_1 = 1, \qquad e_2 = (1, 0_{1\times m})^T, \qquad W_s(x) = \operatorname{diag}\left\{ \frac{1}{n_{M,S}} K_h(X_{s+\tau S} - x) \right\}_{\tau = i_{M,S}}^{n_S},$$

where $K : \mathbb{R}^1 \to \mathbb{R}^1$ is a symmetric probability density with compact support and

$$K_h(u) = \frac{1}{h^m} \prod_{j=1}^m K\left(\frac{u_j}{h}\right)$$

for $u \in \mathbb{R}^m$; $h = h_n$ is a positive number (bandwidth), $h \to 0$, $nh^m \to \infty$ as $n \to \infty$. Further we denote $\|K\|_2^2 = \int K^2(u)du$, $\sigma_K^2 = \int K(u)u^2 du$.

THEOREM 1. *Under Assumptions (A1) and (A2) in the Appendix, for $a = 1, 2$, as $n \to \infty$, the estimation bias of the nonparametric estimates $\hat{f}_{a,s}(x,h)$ is $r_{a,s}(x)\sigma_K^2 h^2/2$ where*

$$r_{1,s}(x) = \operatorname{Tr}\{\nabla^2 f_s(x)\} + \frac{2\nabla^T \mu_s(x) \nabla f_s(x)}{\mu_s(x)}, \qquad r_{2,s}(x) = \operatorname{Tr}\{\nabla^2 f_s(x)\}, \tag{7}$$

whereas the estimation variance is

$$\|K\|_2^{2m} \frac{\sigma_s^2(x)}{\mu_s(x) n_{M,S} h^m}. \tag{8}$$

Derivation of these terms for the standard case ($S = 1$) can be found in Härdle, Tsybakov, and Yang (1998). The $\mu_s(x)$ in equations (7) and (8) represents the stationary density of season s, which exists according to assumption (A1). Among many others, Ango Nze (1992) and Cline and Pu (1999) provide simple conditions to ensure strict stationarity and β-mixing. These conditions are extended to seasonal processes by Theorem 2, which follows, and can be checked for many given processes. See the example in Section 6.1.

(E1) The error ξ_t has a density function that is positive everywhere, and so are the seasonal volatility functions $\sigma_s^2(\cdot), s = 0,1,\dots,S-1$.

(E2) There exist an integer $k \geq \max(M/S,1)$, a constant $R > 0$, and a matrix of coefficients $(a_{sj})_{0\leq s\leq S-1, 1\leq j\leq kS}$ with all $a_{sj} \geq 0$ and $\max_{0\leq s\leq S-1} \sum_{j=1}^{kS} a_{sj} = a < 1$ such that for $0 \leq s \leq S-1, \tau \geq k$

$$E(|Y_{s+\tau S}| \,|\, Y_{s+\tau S-1} = y_1, \dots, Y_{s+\tau S-kS} = y_{kS}) \leq \sum_{j=1}^{kS} a_{sj} |y_j| \tag{9}$$

when $\min_{1\leq j\leq kS} |y_j| \geq R$.

THEOREM 2. *Under conditions (E1) and (E2), the process* $V_\tau = (Y_{\tau S}, Y_{\tau S-1}, \dots, Y_{\tau S-kS+1})^T$, $\tau = k, k+1, \dots$ *is geometrically ergodic. If the initial* V_k *has a stationary distribution, then the process is both strictly stationary and geometrically* β*-mixing.*

3.2. Lag Selection

We now adapt lag selection procedures for standard nonlinear autoregressive time series to the SNAR model (1). For lag selection in the standard case with $S = 1$ Auestad and Tjøstheim (1990) and Tjøstheim and Auestad (1994b) introduce nonparametric versions of the final prediction error (FPE), which were analyzed theoretically and significantly improved upon by Tschernig and Yang (2000). For seasonal time series we define the FPE of the estimates $\{\hat{f}_s\}_{s=0}^{S-1}$ of $\{f_s\}_{s=0}^{S-1}$ as the following functional:

$$FPE(\{\hat{f}_s\}_{s=0}^{S-1}) = \frac{1}{S} \sum_{s=0}^{S-1} E[\{\breve{Y}_{s+\tau S} - \hat{f}_s(\breve{X}_{s+\tau S})\}^2 w(\breve{X}_{s+\tau S,M})],$$

where w denotes a weight function and $\{\breve{Y}_t\}$ is another series with exactly the same distribution as $\{Y_t\}$ but independent of $\{Y_t\}$. Because $\hat{f}_{a,s}(x)$ and therefore the FPE depends primarily on h we denote:

$$FPE_a(h) = FPE(\{\hat{f}_{a,s}\}_{s=0}^{S-1}).$$

The next theorem extends Theorem 2.1 in Tschernig and Yang (2000) to SNAR processes. Note that in this theorem and throughout this paper, we drop the lag reference and denote all seasonal densities by $\mu_s(x_M)$ or $\mu_s(x)$.

THEOREM 3. *Under Assumptions (A1)–(A6), for* $a = 1, 2$, *as* $n \to \infty$ *the* FPE *based on the correct set of lags is*

$$FPE_a(h) = AFPE_a(h) + o\{h^4 + n_{M,S}^{-1} h^{-m}\},$$

in which the asymptotic FPEs *(*AFPEs*) are*

$$AFPE_a(h) = A + b(h)B + c(h)C_a, \tag{10}$$

where

$$A = \frac{1}{S}\sum_{s=0}^{S-1}\int \mu_s(x_M)\sigma_s^2(x)w(x_M)\,dx_M, \tag{11}$$

$$B = \frac{1}{S}\sum_{s=0}^{S-1}\int \mu_s(x_M)\sigma_s^2(x)/\mu_s(x)w(x_M)\,dx_M, \tag{12}$$

$$C_a = \frac{1}{S}\sum_{s=0}^{S-1}\int r_{a,s}^2(x)\mu_s(x_M)w(x_M)\,dx_M, \tag{13}$$

and where

$$b(h) = \|K\|_2^{2m} n_{M,S}^{-1} h^{-m}, \qquad c(h) = \sigma_K^4 h^4/4.$$

Solving the variance-bias trade-off in (10) allows one to derive an asymptotically optimal bandwidth

$$h_{a,opt} = \{m\|K\|_2^{2m} B n_{M,S}^{-1} C_a^{-1} \sigma_K^{-4}\}^{1/(m+4)}, \tag{14}$$

which can be estimated by plug-in methods. See Section 5 for details.

To estimate the asymptotic FPEs the following estimates of A and B are needed:

$$\hat{A}(h) = \frac{1}{S}\sum_{s=0}^{S-1}\frac{1}{n_{M,S}}\sum_{\tau=i_{M,S}}^{n_S}\{Y_{s+\tau S} - \hat{f}_{a,s}(X_{s+\tau S}, h)\}^2 w(X_{s+\tau S,M}),$$

$$\hat{B}(\mathbf{h}_B) = \frac{1}{S}\sum_{s=0}^{S-1}\frac{1}{n_{M,S}}\sum_{\tau=i_{M,S}}^{n_S}\frac{\{Y_{s+\tau S} - \hat{f}_{a,s}(X_{s+\tau S}, h_{B,s})\}^2 w(X_{s+\tau S,M})}{\hat{\mu}_s(X_{s+\tau S}, h_{B,s})},$$

in which the bandwidths h and $\mathbf{h}_B = (h_{B,0}, \ldots, h_{B,S-1})^T$ are all of the same order $n_{M,S}^{-1/(m+4)}$ as the optimal bandwidth $h_{a,opt}$ and $\hat{\mu}_s$ is a kernel estimator of the density μ_s using bandwidth $h_{B,s}$.

Because A is the dominant term in the AFPE expressions (10), one takes into account the bias of its estimator $\hat{A}(h)$ and inserts the bias corrected estimate into (10). This delivers the following estimator for $AFPE_a(h)$:

$$AFPE_a = \hat{A}(h_{a,opt}) + 2K(0)^m n_{M,S}^{-1} h_{a,opt}^{-m} \hat{B}(\mathbf{h}_B). \tag{15}$$

From both asymptotic consideration and simulation results, Tschernig and Yang (2000) conclude that when using $AFPE_a$, the probability of including extra lags in addition to the correct ones is larger than that of missing some of the correct ones. In other words, overfitting is more likely than underfitting. Based on this, Tschernig and Yang propose a corrected AFPE (CAFPE) by multiplying the $AFPE_a$ with a penalizing factor for overfitting and find in simulations that this corrected AFPE selects lags correctly much more often than the uncorrected AFPE. Similar to equation (4.1), p. 466, in Tschernig and Yang (2000), the corrected $AFPE_a$s for the seasonal case are given by

$$CAFPE_a = AFPE_a\{1 + mn_{M,S}^{-4/(m+4)}\}. \tag{16}$$

Whatever FPE criterion one wants to use, one selects the subset $\{\hat{i}_1, \ldots, \hat{i}_{\hat{m}}\}$ with the smallest $(C)AFPE_a'$ where $(C)AFPE_a'$ denotes the quantities according to (15) or (16) for every subset $\{i_1', \ldots, i_{m'}'\}$ of $\{1, \ldots, M\}$.

As in Tschernig and Yang (2000), one can show that the following theorem holds.

THEOREM 4. *Under Assumptions (A1)–(A6) the lag selection procedure based on either (15) or (16) consistently selects the correct lags, i.e., if $\hat{i}_1, \ldots, \hat{i}_{\hat{m}}$ are the selected lags, then as $n \to \infty$*

$$P[\hat{m} = m, \hat{i}_j = i_j, j = 1,2,\ldots,m] \to 1.$$

Note that if the true process is linear, i.e., if all functions $f_s, s = 0,\ldots,S-1$ are linear, then all the $r_{2,s}(x) \equiv 0$ by equation (7), which implies by (13) that $C_2 = 0$ for the local linear CAFPE. This causes assumption (A6) to fail, and a variance-bias trade-off is no longer available. As noted in Tschernig and Yang (2000), the local linear CAFPE becomes inconsistent in this case, but the alternative local constant CAFPE remains consistent. Despite this, we suggest using the local linear CAFPE as it is faster to compute and also has performed quite satisfactorily for linear processes in our Monte Carlo study in Section 6.

4. SDNAR AND SHNAR IDENTIFICATION

Although the function estimators for the SNAR model discussed in the previous section provide ample seasonal and nonlinear flexibility, it is important to realize that they use only $n_{M,S}$ observations, a number much smaller than n and even n/S. This may render estimation of the seasonal nonlinear autoregression (1) difficult if the sample size n is already small. In this section we develop semiparametric estimators for the seasonal dummy nonlinear autoregression (3) and the seasonal shift nonlinear autoregression (5). In both models, the parameters can be estimated with a faster rate of convergence and then effectively "removed" so that the regression function becomes the same for all seasons and all $n - M$ observations can be used for estimation and lag selection. Therefore the standard lag selection methods of Tschernig and Yang (2000) can be applied.

4.1. Estimation of the SDNAR Model

Note that for the SDNAR model (3) one has

$$b_s = f_s(X_{s+\tau S}) - f_0(X_{s+\tau S}). \tag{17}$$

Based on (17) we will present the following three estimation methods for b_s that, when the correct lags are used, all exhibit a rate of convergence $O_p(n^{-4/(m+4)} + n^{-1/2})$. This rate is faster than the rate $O_p(n^{-2/(m+4)})$ for function estimation.

The full dummy method. For seasons $s' = 1, \ldots, S-1$, define the dummy variable $D_{s+\tau S, s'}$, which equals 1 if $s = s'$ and 0 otherwise. This allows us to rewrite (3) as

$$Y_t = f(X_t) + \sum_{s=1}^{S-1} b_s D_{t,s} + \sigma(X_t)\xi_t, \qquad t = i_{M,S} S, \ldots, n.$$

One can then jointly estimate the seasonal parameters and the function nonparametrically. For the seasonal parameters b_s one obtains the local estimators at x:

$$\hat{b}_s(x,h) = e_s^T\{(Z_D^T W Z_D)^{-1} Z_D^T W\}(x)\mathbf{Y}, \tag{18}$$

where e_s denotes the $m + S$ vector whose sth element is 1 and all other elements 0, $s = 1, \ldots, S-1$,

$$Z_D(x) = \begin{pmatrix} 1 & \cdots & 1 \\ D_{i_{M,S}S,1} & \cdots & D_{n1} \\ \vdots & \vdots & \vdots \\ D_{i_{M,S}S,S-1} & \cdots & D_{n,S-1} \\ \dfrac{X_{i_{M,S}S} - x}{h} & \cdots & \dfrac{X_n - x}{h} \end{pmatrix}^T,$$

$$W(x) = \operatorname{diag}\left\{ \frac{K_h(X_t - x)}{n_{M,S}} \right\}_{t=i_{M,S}S}^{n},$$

and $\mathbf{Y} = (Y_{i_{M,S}}, Y_{i_{M,S}+1}, \ldots, Y_n)^T$. One then defines the weighted average estimator

$$\bar{b}_s = \frac{\sum_{\tau=i_{M,S}}^{n_S} w(X_{s+\tau S,M}) \hat{b}_s(X_{s+\tau S}, h)}{\sum_{\tau=i_{M,S}}^{n_S} w(X_{s+\tau S,M})}, \tag{19}$$

where $w(\cdot)$ is the same weight function as used in the FPE. We next establish the asymptotic behavior of $\bar{b}_s$.

THEOREM 5. *Under Assumptions (A1)–(A5), for any* $1 \le s \le S-1$, *as* $n \to \infty$, $h = \beta n_S^{-1/(m+4)}$, *the estimator* $\bar{b}_s$ *based on the correct set of lags satisfies*

$$\bar{b}_s - b_s = O_p(h^4 + n^{-1/2}).$$

The proof of this theorem rests on two lemmas that can be found in the Appendix. Lemma A.4 shows that the usual bias of order h^2 is canceled by the differencing of seasons 0 and s, whereas Lemma A.3 states that the usual variance of order $(nh)^{-1}$ gets smoothed out to order n^{-1} as a result of the averaging of $\hat{b}_s(X_{s+\tau S}, h)$. Therefore, if one subtracts $\bar{b}_s$ from $\mathbf{Y}_s$ for all $s = 1, \ldots, S-1$ to obtain $\tilde{\mathbf{Y}}_s = \mathbf{Y}_s - \bar{b}_s$, then by the faster convergence of the estimator (19) one has

$$\tilde{Y}_{s+\tau S} = f(X_{s+\tau S}) + \sigma(X_{s+\tau S})\xi_{s+\tau S} + O_p(n^{-4/(m+4)} + n^{-1/2}).$$

Thus, using the adjusted data $\{(X_t, \tilde{Y}_t)\}_{t=M}^n$, one can estimate the function f with an effective sample size n as in the nonseasonal case.

Two alternative procedures for the estimation of b_s. To obtain the first alternative one may include in the "regressors matrix" only the dummy of season s instead of all dummies, i.e.,

$$Z_{s,D}(x) = \begin{pmatrix} 1 & \cdots & 1 \\ D_{i_{M,S}S,s} & \cdots & D_{n,s} \\ \dfrac{X_{i_{M,S}S} - x}{h} & \cdots & \dfrac{X_n - x}{h} \end{pmatrix}^T$$

and then define the local estimator of b_s at x as

$$\hat{b}_s(x,h) = e_2^T\{(Z_{s,D}^T W Z_{s,D})^{-1} Z_{s,D}^T W\}(x)\mathbf{Y}, \qquad s = 1, \ldots, S-1, \tag{20}$$

in which e_2 is the $m+2$ vector whose second element is 1 and all other elements 0. One then defines $\bar{b}_s$ by the averaging formula (19) using (20). We call this approach the "partial dummy method," in contrast to the "full dummy method" based on (18).

The second alternative may be termed the "two estimators method"; namely, define

$$\hat{b}_s(x,h) = \hat{f}_{2,s}(x,h) - \hat{f}_{2,0}(x,h), \tag{21}$$

where $\hat{f}_{2,s}(x,h)$ is the local linear estimator as defined in (6). In this case one inserts (21) into the averaging formula (19) to obtain $\bar{b}_s$.

For both approaches the cancellation-and-smooth-out effect of Theorem 5 remains valid if $\bar{b}_s$ is the average of $\hat{b}_s(X_{s+\tau S},h)$, $\tau = i_{M,S},\dots,n_S$, defined by either (20) or (21). The relative merit of these two methods is simplicity, whereas the full dummy method is more robust.

4.2. Lag Selection in the SDNAR Model

Because one can estimate the function f with an effective sample size n using the adjusted data $\{(X_t,\tilde{Y}_t)\}_{t=M}^{n}$, one can treat the adjusted data as being nonseasonal. This suggests applying the local constant or local linear (C)AFPE lag selection criteria for the standard nonlinear autoregression as in Tschernig and Yang (2000). These estimators are obtained by setting S equal to 1 in equations (15) and (16) and taking into account that the process $\{X_t\}_{t=M}^{n}$ can be treated as if it has the average seasonal density $\bar{\mu}(x) = 1/S\sum_{s=0}^{S-1}\mu_s(x)$.

However, before the nonseasonal (C)AFPE criteria can be used one has to check how the seasonal parameter estimator (19) for obtaining the adjusted data behaves if an incorrect set of lags is used. If one uses a set of lags that overfits, i.e., if the set includes all the lags $i_1,\dots,i_m$ and more, then one still has $\bar{b}_s' - b_s = O_p(h'^4 + n^{-1/2})$. This can be seen by examining the proof of Theorem 5 in the Appendix.

The case of underfitting is slightly more complicated. For simplicity, suppose lags $i_1',\dots,i_{m'}'$ are used where $\{i_1',\dots,i_{m'}'\}$ is a proper subset of $\{i_1,\dots,i_m\}$. Denote by $x' = (x_{i_1'},\dots,x_{i_{m'}'})^T$ the variable vector corresponding to the lags, and $x = (x',x'')$. Further, denote for each season $s = 0,1,\dots,S-1$ the discrepancy between $f(x)$ and its conditional expectation on x' as

$$f_s^{\perp}(x) = f(x) - \mu_s(x')^{-1}\int f(x',u'')\mu_s(x',u'')\,du'' = f(x) - E_s\{f(x)|x'\}.$$

We now assume that every function $f_s^{\perp}(x)$ has at least one nonzero point in the interior of the support of w, and hence for each season, the squared projection error into the submodel is positive. This is satisfied by simply enlarging the support of w so that its interior includes at least one nonzero point from each $f_s^{\perp}(x)$, which is easy as all the $f_s^{\perp}(x)$'s are nonzero functions on the support of μ_s.

The following theorem is a refined version of Theorem 5.

THEOREM 6. *Under Assumptions (A1)–(A5), for any underfitting model and* $1 \le s \le S-1$, *as* $n \to \infty$, $h' = \beta n_S^{-1/(m'+4)}$,

$$\bar{b}_s' - b_s = \int \{ f_s^{\perp}(x) - f_0^{\perp}(x)\} w(x_M)\mu_s(x_M)dx_M + O_p(h'^2).$$

So the estimates $\bar{b}_s$ approximate the parameters b_s to the order of $h^4 + n^{-1/2} = n^{-4/(m+4)} + n^{-1/2}$, which is higher than the order $n^{-2/(m+4)}$ of the function estimates, provided all the correct lags are used in the computation, whereas the bias would be nonvanishing if the lags underfit. In the latter case the estimation of $f(\cdot)$ by the adjusted data $\{(X_t', \tilde{Y}_t')\}_{t=M}^n$ will have a nonvanishing bias, as in the case of the standard autoregression model, plus the nonzero bias introduced by the estimation of the b_s's. One then faces the same situation as in Tschernig and Yang (2000) because the (C)AFPEs then obtained will be larger than the true prediction error by a positive constant up to higher order terms. One therefore can select the lags based on (C)AFPE for each set of lags after the data are adjusted by the estimated seasonal parameters based on that set. Therefore these (C)AFPEs have the same properties as in Tschernig and Yang (2000).

4.3. SHNAR Model

The seasonal shift model (5) is easier to analyze than the seasonal dummy model (3). In this case, $\mathbf{U}_{s+\tau S} = \mathbf{Y}_{s+\tau S} - \delta_s$, $s = 0,\dots,S-1, \tau = 0,1,\dots,$ is a stationary process that satisfies

$$U_{s+\tau S} = f(U_{s+\tau S-i_1},\dots,U_{s+\tau S-i_m}) + \sigma(U_{s+\tau S-i_1},\dots,U_{s+\tau S-i_m})\xi_{s+\tau S}.$$

Hence, one defines $\tilde{\mathbf{U}}_s = \mathbf{Y}_s - \hat{\delta}_s$ as a substitute of $\mathbf{U}_s$, where $\hat{\delta}_s = n_{M,S}^{-1}\sum_{\tau=i_{M,S}}^{n_S}(Y_{s+\tau S} - Y_{\tau S})$ is the estimated sth mean shift, for all $s = 1,\dots,S-1$. This is based on the following theorem.

THEOREM 7. *Under Assumptions (A1′), (A2′), (A3), and (A4′)–(A6′) in the Appendix, and assuming that the* SHNAR *model (5) is true,*

$$\sqrt{n}(\hat{\delta}_s - \delta_s) \to N(0, S\sigma_s^2)$$

for all $s = 1,\dots,S-1$, *where* $\sigma_s^2 = E(Y_s - Y_0 - \delta_s)^2 + 2\sum_{\tau=1}^{\infty} E(Y_s - Y_0 - \delta_s)(Y_{s+\tau S} - Y_{\tau S} - \delta_s)$.

Hence $U_t - \hat{U}_t = O_p(n^{-1/2})$ for all $t = M,\dots,n$. One thus can use the seasonally adjusted data $\{\hat{U}_t\}_{t=M}^n$ as a substitute for $\{U_t\}_{t=M}^n$ for estimating $f(\cdot)$. The same applies to lag selection; hence one applies the AFPE and CAFPE criteria of Tschernig and Yang (2000) to the process $\{\hat{U}_t\}_{t=M}^n$ to determine the lags.

In the presence of nonlinearities the SDNAR and SHNAR model are mutually exclusive. Therefore, one may choose for any given data set the model with the smaller CAFPE because the prediction error of a given model indicates the departure of the data set from the imposed model structure.

5. IMPLEMENTATION

In this section we describe how to estimate the unknown quantities B and C_a, $a = 2$, needed for the plug-in optimal bandwidth (14) and the CAFPEs of the SNAR model and also of its restricted SDNAR version. Only local linear ($a = 2$) procedures are implemented, as we want to avoid the complicated bias terms of local constant procedures ($a = 1$). We use the Gaussian kernel for all nonparametric estimates. For all procedures, the weight function $w(x)$ is the indicator function on the range of the observed data. For robustification, 5% of those observations whose densities are the lowest are screened off, and leave-one-out features are implemented for all estimations.

5.1. SNAR Model

To estimate the seasonal densities μ_s in $\hat{B}(\mathbf{h}_B)$, we use the rule-of-thumb bandwidth of Silverman (1986, equation (4.14), Table 4.1, pp. 86–87) $h_{B,s} = h(m + 2, \hat{\sigma}_s, n_{M,S})$, $s = 0, \ldots, S - 1$ where

$$h(k, \sigma, n) = \sigma\{4/k\}^{1/(k+2)} n^{-1/(k+2)}$$

and $\hat{\sigma}_s = \{\prod_{j=1}^{m} \sqrt{\mathrm{Var}(\mathbf{Y}_{s-i_j})}\}^{1/m}$ denotes the geometric mean of the standard deviation of the regressors in each season.

The seasonal functions f_s are estimated by local linear estimators $\hat{f}_{2,s}$ defined in (6) with the same bandwidth $h_{B,s}$. This simple bandwidth has the appropriate rate and performs in our small sample experiments nearly as well as the rule-of-thumb bandwidth of Yang and Tschernig (1999).

For the estimation of the second derivatives in C_2, we use a local quadratic estimator that excludes all cross derivatives, with a simple bandwidth rule $h_{C,s} = h(m + 4, 3\hat{\sigma}_s, n_{M,S})$. As a simplification of the partial local cubic estimator of Yang and Tschernig (1999), this is sufficient for lag selection, which requires less precision than function estimation.

5.2. SDNAR Model

To obtain the adjusted $\widetilde{Y}_t$, $t = M, \ldots, n$, the first step is the estimation of seasonal dummies by the full dummy method given by (18) and (19). By Theorem 5, there does not exist the usual bias-variance trade-off that leads to an optimal bandwidth because the usual bias of order h^2 cancels out. We take $h_B = h(m + 2, \hat{\sigma}, Sn_{M,S})$ with $\hat{\sigma} = \{\prod_{j=1}^{m} \sqrt{\mathrm{Var}(\mathbf{Y}_{-i_j})}\}^{1/m}$, which has the optimal rate for function estimation. For estimating the unknown quantities in the CAFPE for the adjusted data $\{(X_t, \widetilde{Y}_t)\}_{t=M}^{n}$ we use all specifications of Section 5.1 setting S to 1.

6. MONTE CARLO STUDY

In this section we document the practical performance of the estimation and lag selection methods derived for the three seasonal nonlinear autoregressive models on data of moderate samples.

6.1. Setup

All processes are homoskedastic and fulfill the relevant assumptions in the Appendix. We used conditions (E1) and (E2) of Section 3.1 to obtain geometrically ergodic processes. To start, in the stationary distribution we generated $n + 400$ observations and discarded the first 400. Geometric ergodicity ensures convergence to the stationary distribution. Explicit burn-in times can be computed using Theorem 12 of Rosenthal (1995) or Theorem 5.1 in the more recent Roberts and Tweedie (1999), which both provide bounds on the total variation norm. Both theorems involve lengthy calculations of many constants; hence they are not carried out for the examples. Instead we used simulations based on the total variation norm and found that $n = 400$ is more than sufficient. In total, we consider 8 different processes and always allow all lag combinations up to lag $M = 6$. For every experiment we conduct $R = 100$ replications.[3]

SNAR processes. To investigate the general identification devices discussed in Section 3 that can handle SNAR models (1), we consider one periodic autoregression and one seasonal nonlinear autoregressive process, each with two seasons, 200 observations, and standard normal errors, $\xi_t \sim N(0,1)$, and one seasonal nonlinear autoregressive process with four seasons, 400 observations, and standard normal errors:

PAR2. Periodic autoregressive process (2) of order $p = 3$ with two seasons and parameters $\alpha_{10} = 0.55$, $\alpha_{11} = -0.3$, $\alpha_{20} = \alpha_{21} = 0$, $\alpha_{30} = -0.4$, and $\alpha_{31} = 0.3$.

SNAR2. Seasonal nonlinear autoregressive process of order $p = 3$ with two seasons:

$$Y_t = \sum_{i=1}^{p} \alpha_{is} Y_{t-i} + \left(\sum_{i=1}^{p} \beta_{is} Y_{t-i} \right) \frac{1}{1 + \exp\{-\gamma_s (Y_{t-l_s} - c_s)\}} + \xi_t \qquad \textbf{(22)}$$

with $\alpha_{10} = 0.55$, $\alpha_{11} = 0.3$, $\alpha_{20} = \alpha_{21} = 0$, $\alpha_{30} = 0.4$, $\alpha_{31} = 0.55$ and $\beta_{10} = -1.1$, $\beta_{11} = -0.6$, $\beta_{20} = \beta_{21} = 0$, $\beta_{30} = -0.8$, $\beta_{31} = -1.1$, $\gamma_0 = \gamma_1 = 3$, $c_0 = c_1 = 0$, $l_0 = 3$, and $l_1 = 1$.

SNAR4. Seasonal nonlinear autoregressive process (22) of order $p = 2$ with four seasons and parameters $\alpha_{10} = 0.55$, $\alpha_{11} = 0.3$, $\alpha_{12} = -0.3$, $\alpha_{13} = 0.55$, $\alpha_{20} = 0.4$, $\alpha_{21} = 0.55$, $\alpha_{22} = -0.55$, $\alpha_{23} = -0.4$ and $\beta_{10} = -1.1$, $\beta_{11} = -0.6$, $\beta_{12} = 0.6$, $\beta_{13} = 1.1$, $\beta_{20} = -0.8$, $\beta_{21} = -1.1$, $\beta_{22} = 1.1$, $\beta_{23} = 0.8$, $\gamma_0 = \gamma_1 = \gamma_2 = \gamma_3 = 3$, $c_0 = c_1 = c_2 = c_3 = 0$, and $l_0 = 2$, $l_1 = 1$, $l_2 = 1$, $l_3 = 2$.

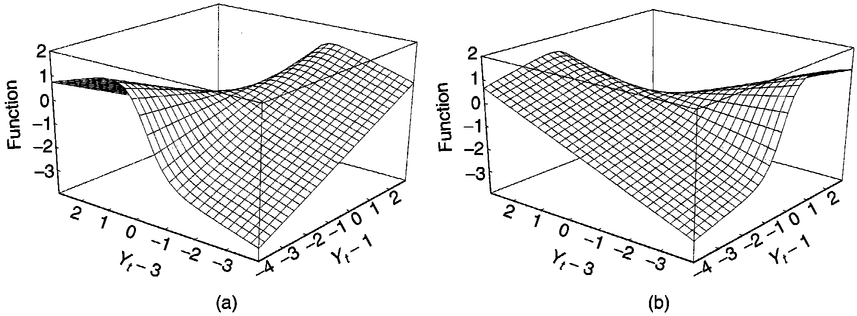

FIGURE 1. Functions of the SNAR2 process: (a) for season 1; (b) for season 2.

To make lag selection difficult the PAR2 and SNAR2 processes contain the nonconsecutive lags 1 and 3. The two functions in the SNAR2 process were chosen for their contrasting shape as can be seen from Figures 1a and b, which display them on the relevant range. Figure 2 displays the four seasonal functions of the SNAR4 process.

Using elementary techniques one can check that conditions (E1) and (E2) are met by all processes in this section. This is demonstrated for the process

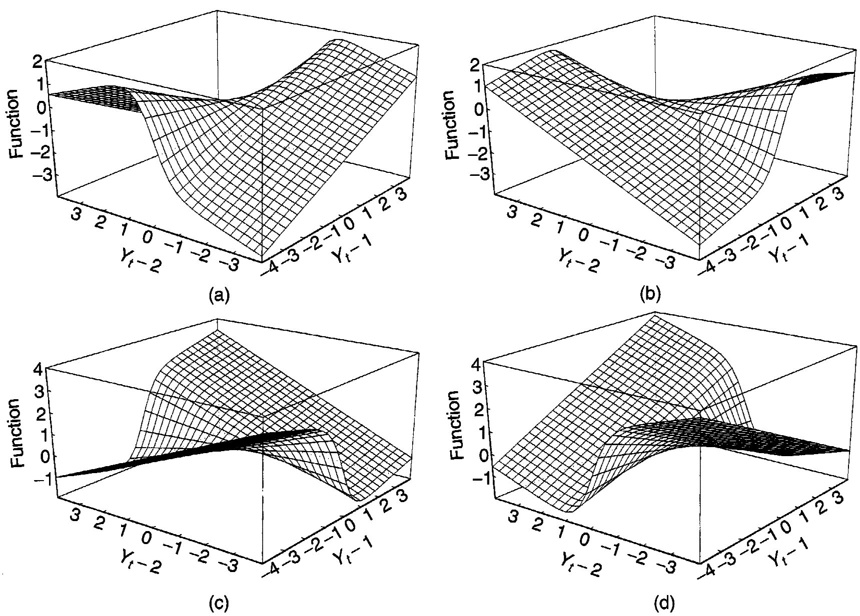

FIGURE 2. Functions of the SNAR4 process: (a) for season 1; (b) for season 2; (c) for season 3; (d) for season 4.

SNAR2. First, we set $M = 4$ and $k = 2$. Because $\xi_t \sim N(0,1)$, (E1) is clearly met. To check (E2), notice that for $a_s = \{1 + \exp[-\gamma_s(y_{s+2\tau-l_s} - c_s)]\}^{-1}$ one has

$$E(|Y_{s+2\tau}||Y_{s+2\tau-1} = y_1, \dots, Y_{s+2\tau-4} = y_4) \le \sum_{j=1}^{3} |\alpha_{js} + a_s\beta_{js}||y_j| + E|\xi_t|$$

$$= \sum_{j=1}^{3} |\alpha_{js} + a_s\beta_{js}||y_j| + C,$$

where $C = (1/\sqrt{2\pi})\int|x|e^{-x^2/2}dx$. Note that as a function of a_s, $|\alpha_{js} + a_s\beta_{js}|$ is convex; hence

$$\begin{aligned}
|\alpha_{js} + a_s\beta_{js}| &\le (1-a_s)|\alpha_{js} + 0 \times \beta_{js}| + a_s|\alpha_{js} + 1 \times \beta_{js}| \\
&= (1-a_s)|\alpha_{js}| + a_s|\alpha_{js} + \beta_{js}| \\
&= \begin{cases} (1-a_0)|0.55| + a_0|0.55 - 1.1| = 0.55 & s = 0, j = 1 \\ (1-a_0)|0| + a_0|0-0| = 0 & s = 0, j = 2 \\ (1-a_0)|0.4| + a_0|0.4 - 0.8| = 0.4 & s = 0, j = 3 \\ (1-a_1)|0.3| + a_1|0.3 - 0.6| = 0.3 & s = 1, j = 1 \\ (1-a_1)|0| + a_1|0-0| = 0 & s = 1, j = 2 \\ (1-a_1)|0.55| + a_1|0.55 - 1.1| = 0.55 & s = 1, j = 3. \end{cases}
\end{aligned}$$

Therefore one has

$$E(|Y_{s+2\tau}||Y_{s+2\tau-1} = y_1, \dots, Y_{s+2\tau-4} = y_4) \le C + \begin{cases} 0.55|y_1| + 0.4|y_3| & s = 0 \\ 0.3|y_1| + 0.55|y_3| & s = 1. \end{cases}$$

Hence if one takes the matrix

$$(a_{sj})_{0\le s\le 1, 1\le j\le 4} = \begin{pmatrix} 0.55 + \delta/2 & 0 & 0.4 + \delta/2 & 0 \\ 0.3 + \delta/2 & 0 & 0.55 + \delta/2 & 0 \end{pmatrix}$$

for some $\delta \in (0,0.05)$, set $R = 1 + \delta^{-1}C$, and $a = \max_{0\le s\le 1}\sum_{j=1}^{4} a_{sj} = 0.95 + \delta$, then $a < 1$ and (9) holds.

SDNAR and SHNAR processes. For the SDAR model (4), the SDNAR model (3), and the SHNAR model (5) we consider the following specifications, all with 100 observations, four seasons, and standard normal errors ξ_t.

SDAR. Seasonal dummy linear autoregression (4) of order $p = 3$ with seasonal parameters $b_1 = 1$, $b_2 = 0.3$, $b_3 = -0.6$ and autoregressive parameters $\alpha_1 = 0.3$, $\alpha_2 = 0$, and $\alpha_3 = 0.4$.

SDNAR1. Seasonal dummy nonlinear autoregression

$$Y_t = f(Y_{t-1}, Y_{t-2}) + 0.5D_{t,1} + 1.5D_{t,2} + 0.8D_{t,3} + \xi_t$$

with

$$f(Y_{t-1}, Y_{t-2}) = -0.6Y_{t-1} + 0.2Y_{t-2} + (Y_{t-1} - 0.5Y_{t-2})\,\frac{1}{1+\exp(-5Y_{t-2})}. \quad \textbf{(23)}$$

SDNAR2. Seasonal dummy nonlinear autoregression

$$Y_t = -0.6Y_{t-1} + 0.2Y_{t-2} + (Y_{t-1} - 0.5Y_{t-2})\,\frac{1}{1+\exp(-2Y_{t-2})} + D_{t,1} + 2D_{t,2} + D_{t,3} + \xi_t.$$

SDNAR3. Seasonal dummy nonlinear autoregression

$$Y_t = \frac{1}{1+\exp\{-4(-0.05 + Y_{t-1} - 3Y_{t-2})\}} + 0.3Y_{t-1} + 0.3Y_{t-2} - 0.3D_{t,1} - 0.5D_{t,2} - 0.1D_{t,3} + \sqrt{0.1}\xi_t.$$

SHNAR. Seasonal shift nonlinear autoregression

$$Y_t - \delta_s = f(Y_{t-1} - \delta_{\{s-1\}}, Y_{t-2} - \delta_{\{s-2\}}) + \xi_t$$

with $f(\cdot)$ given by (23) and $\delta_1 = 1$, $\delta_2 = 1.5$, $\delta_3 = 0.5$.

All nonlinear functions were chosen to represent various degrees of smoothness and complexity. Their plots are displayed in Figures 3a–c. The seasonal parameters were selected such that seasonality is substantial but not dominating.

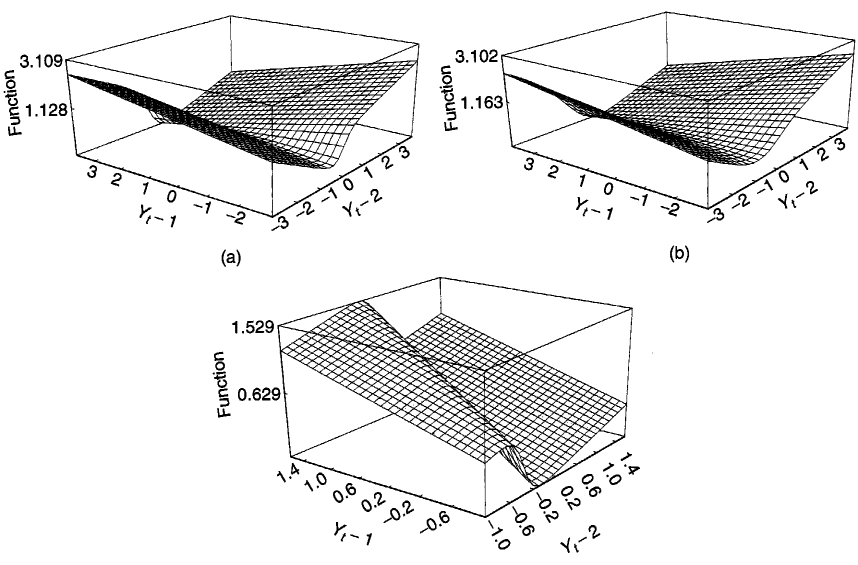

FIGURE 3. Functions of the various seasonal dummy processes: (a) SDNAR1; (b) SDNAR2; (c) SDNAR3.

For lag selection we consider four criteria: first of all, we compute the local linear versions of the CAFPEs developed in the previous sections. In addition, we compute the linear FPE, AIC, and Schwarz criterion. Furthermore, we investigate the performance of the suggested non- and semiparametric estimators in terms of their mean integrated squared error and their prediction power relative to linear procedures.

For each replication, let x' denote a subvector of x_M, whereas x represents the correct lag vector. Define the integrated squared error

$$ISE(\hat{f}_s) = \frac{1}{10{,}000} \sum_i \{\hat{f}_s(X_i') - f_s(X_i)\}^2, \tag{24}$$

where the sum is over 10,000 observations. The mean integrated squared error $MISE(\hat{f}_s)$ is defined as the average of $ISE(\hat{f}_s)$ over the R replications. The vector x' can consist of the correct set of lags or can be selected from the respective criteria.

We also define the one-step prediction error by averaging

$$\{Y_{n_S+1} - \hat{f}_0(X'_{n_S+1})\}^2 \tag{25}$$

over the R replications.

6.2. Results on Lag Selection

All results of the Monte Carlo simulations are shown in Figures 4–6. Each bar graph corresponds to one of the processes described previously and displays the empirical frequencies of the four criteria to correctly fit (black bar), overfit (bar with vertical lines), and underfit. Underfitting is further split into two cases: no wrong lags included (bar with horizontal lines) and wrong lags included (white bar). The last case can be considered the worst outcome of all four as it is neither parsimonious nor includes the correct model.

Figure 4 displays the results for the linear periodic PAR2 process and the seasonal dummy linear SDAR process. As expected the linear Schwarz criterion (SC) performs best in terms of correct selections. The nonparametric CAFPE performs better than AIC and linear FPE for the PAR2 process whereas all three are comparable for the SDAR process. Recall from the comment after Theorem 4 that the local linear CAFPE is inconsistent for linear processes, and it is well known that both the linear FPE and AIC are inconsistent as well. Therefore, it seems preferable to use the much more general CAFPE than the linear AIC or linear FPE.

If one is interested in minimizing underfitting, one should use the linear AIC or FPE. These criteria exhibit the largest empirical frequencies to include the correct lags.

Figures 5 and 6 show the results for the seasonal nonlinear processes. The main conclusion to be drawn here is that for all processes considered the non-

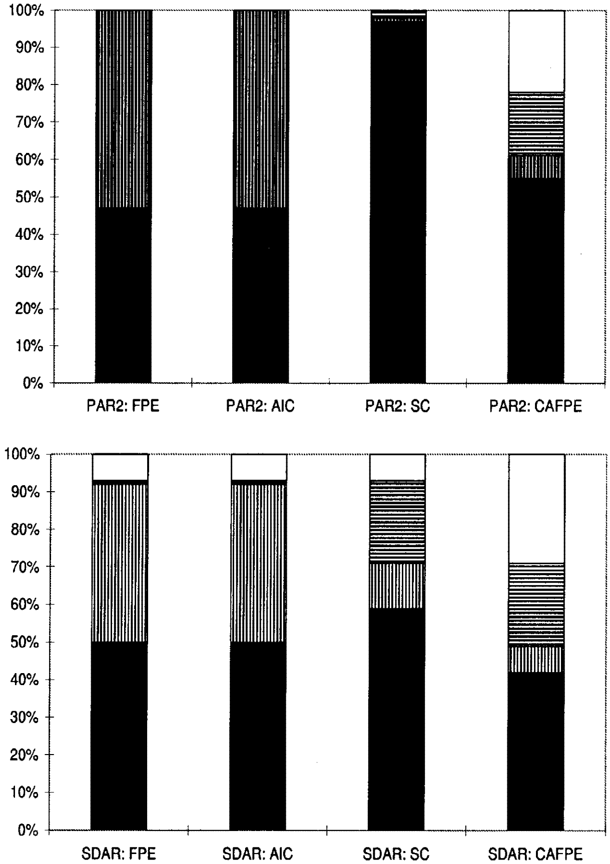

FIGURE 4. Empirical frequencies of the lag selections for the PAR2 and SDAR processes: ■: correct fitting; ▥: overfitting; ▤: underfitting without wrong lags; □: underfitting with wrong lags.

parametric CAFPE scores highest in terms of correct selections and in most cases it is the only useful procedure because the linear criteria can fail completely. This failure is drastic for the seasonal nonlinear SNAR4 process with four seasons, the seasonal dummy nonlinear SDNAR1 and SDNAR2 processes, and the seasonal shift nonlinear SHNAR process. In contrast, the CAFPE's performance loss for linear processes is much less severe.

The results for the seasonal nonlinear processes SNAR2 and SNAR4 in Figure 5 show that the number of seasons is not important as long as there are enough observations for each season, 100 in both cases. Certainly the observed selection rates would drop if the total number of observations decreased. To maintain similar success rates, one has to reduce the seasonal flexibility, which was the main motivation for the SDNAR and SHNAR models. For the latter

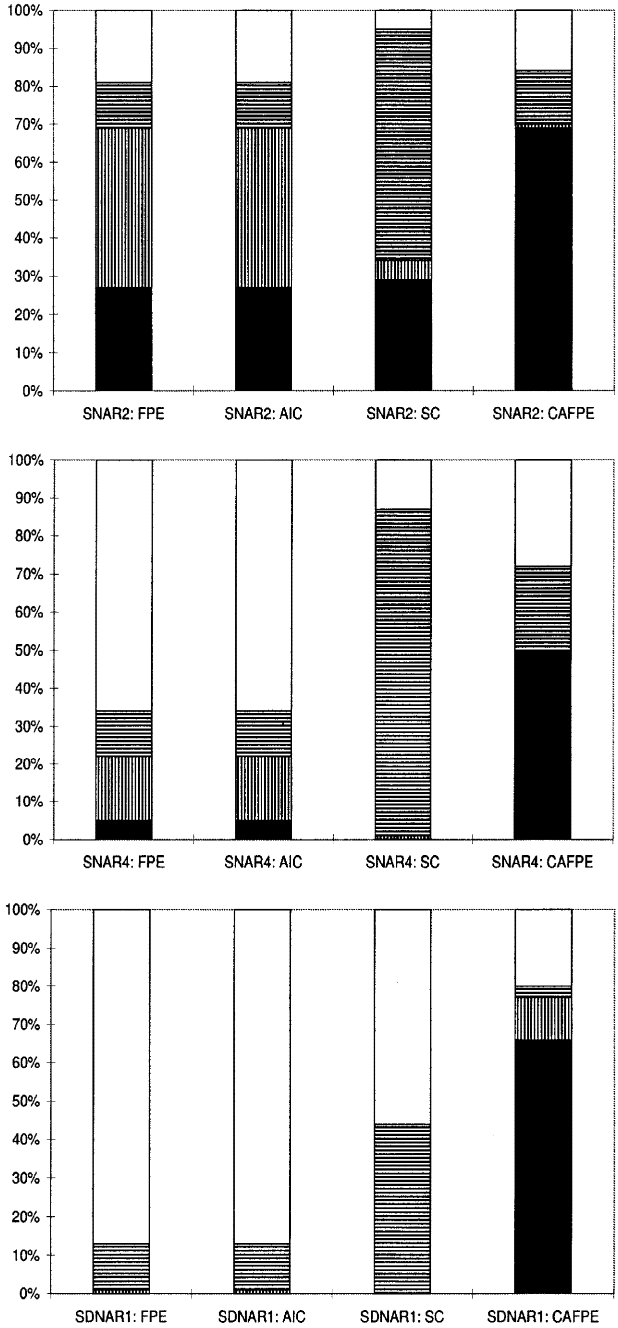

FIGURE 5. Empirical frequencies of the lag selections for the SNAR2, SNAR4, and SDNAR1 processes: ■: correct fitting; ▥: overfitting; ▤: underfitting without wrong lags; □: underfitting with wrong lags.

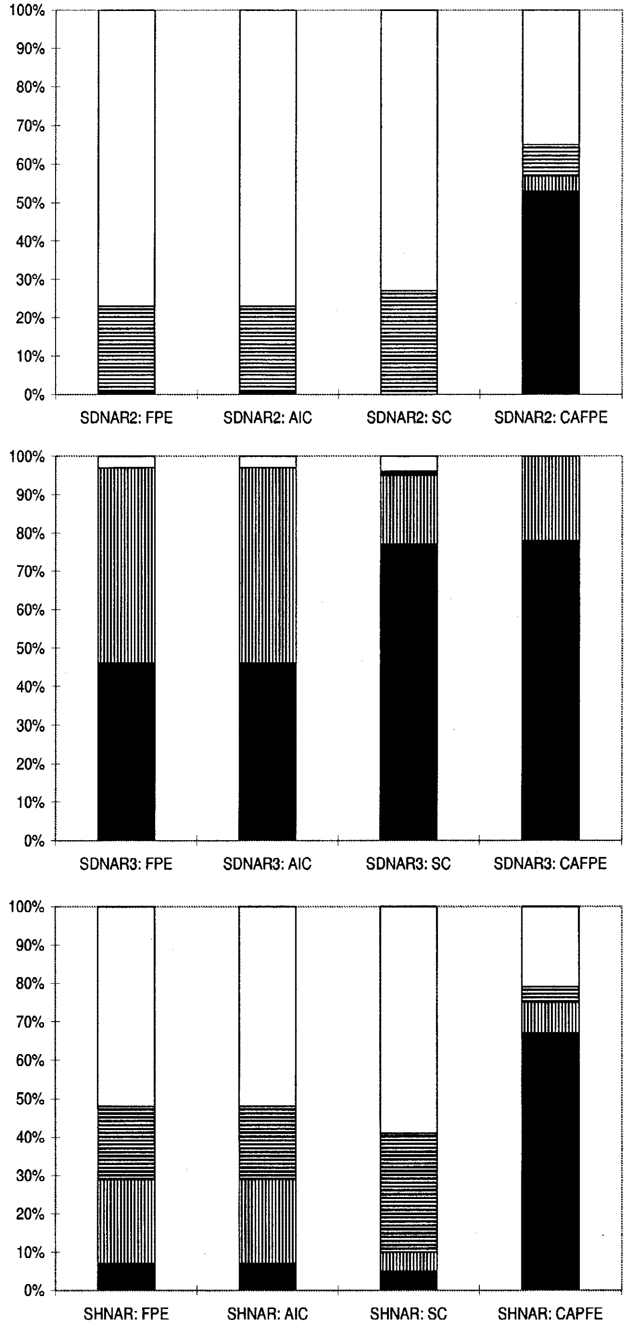

FIGURE 6. Empirical frequencies of the lag selections for the SDNAR2, SDNAR3, and SHNAR processes: ■: correct fitting; ▥: overfitting; ▤: underfitting without wrong lags; □: underfitting with wrong lags.

models even 100 observations are sufficient to guarantee reasonable lag selection results as can be seen from the lower plot of Figure 5 and Figure 6. The plot for the SDNAR3 process also shows that the success rate can reach almost 80%.

If one only requires the selection procedures to include the correct lags, i.e., minimize underfitting, then the CAFPE criterion also performs best. This holds for underfitting without wrong lags and also with wrong lags.

In sum, the local linear CAFPE criterion shows good performance for all seasonal processes. It may therefore provide a useful compromise between reliable lag selection in the presence of nonlinearities and a tolerable decrease in the correct lag selection probability for linear DGPs that is comparable to that between the linear Schwarz criterion and the linear AIC.

6.3. Results on Estimation and Prediction

Besides identifying the relevant lags, it is also important to be able to accurately estimate the conditional mean functions and to predict. Therefore, we compare the mean integrated squared errors (MISE) and prediction errors of the proposed estimators and of simple linear ones. The computing of these errors is described in Section 6.1.

The upper part of Table 1 displays the results for the linear periodic autoregressive PAR2 process. As expected, fitting a nonparametric SNAR model to a linear process increases the MISE, both for correct lags and selected lags. The same phenomenon is true for the prediction errors. Similar conclusions can be drawn from the lower part of Table 1 for the seasonal dummy linear

TABLE 1. Mean integrated squared error and prediction error: Periodic autoregressive processes

	Selected lags				Correct lags	
	CAFPE	FPE	AIC	SC	Nonp.	Linear
			PAR2 *process*			
Under-/Correct/Overfitting	39/55/6	0/47/53	0/47/53	2/97/1		
MISE($\hat{f}_0$)	0.2022	0.0580	0.0580	0.0398	0.0811	0.0346
MISE($\hat{f}_1$)	0.1476	0.0531	0.0531	0.0349	0.0927	0.0329
One-step prediction error	1.1743	1.0384	1.0384	0.9976	1.0392	0.9974
			SDAR *process*			
Under-/Correct/Overfitting	51/42/7	8/50/42	8/50/42	29/59/12		
MISE($\hat{f}$)	0.3877	0.1617	0.1617	0.1843	0.2284	0.0992
MSE($\hat{b}_1$)	0.1246	0.1125	0.1125	0.1027	0.1135	0.0782
MSE($\hat{b}_2$)	0.3109	0.1901	0.1901	0.2321	0.1791	0.1410
MSE($\hat{b}_3$)	0.1588	0.1067	0.1067	0.1141	0.1177	0.0867
MISE($\hat{f} + \sum_{s=1}^{3} \hat{b}_s D_{\cdot,s}$)	0.3118	0.1144	0.1144	0.1290	0.1859	0.0769
One-step prediction error	0.9736	1.0080	1.0080	0.9982	1.0075	0.9855

TABLE 2. Mean integrated squared error and prediction error: Seasonal nonlinear autoregressive processes

	Selected lags				Correct lags	
	CAFPE	FPE	AIC	SC	Nonp.	Linear
			SNAR2 *process*			
Under-/Correct/Overfitting	30/69/1	31/27/42	31/27/42	66/29/5		
MISE($\hat{f}_0$)	0.3928	0.4904	0.4904	0.5638	0.2570	0.4503
MISE($\hat{f}_1$)	0.4096	0.4263	0.4263	0.5306	0.3163	0.3982
One-step prediction error	1.2838	1.4404	1.4404	1.7376	1.1263	1.4070
			SNAR4 *process*			
Under-/Correct/Overfitting	50/50/0	78/5/17	78/5/17	99/0/1		
MISE($\hat{f}_0$)	0.2626	0.3638	0.3638	0.3565	0.1563	0.3524
MISE($\hat{f}_1$)	0.2456	0.3402	0.3402	0.3056	0.1467	0.3334
MISE($\hat{f}_2$)	0.2523	0.3265	0.3265	0.3442	0.1363	0.3122
MISE($\hat{f}_3$)	0.2330	0.3222	0.3222	0.3274	0.1312	0.3440
One-step prediction error	1.1014	1.0861	1.0861	1.2925	0.9380	1.1525

autoregressive SDAR process. We note for the SDAR process that although the seasonal parameters b_s are estimated poorly with the semiparametric procedure, the semiparametric lag selection is roughly as good as the linear procedures, as seen in Figure 4.

Are there substantial gains of the non- and semiparametric procedures in the case of nonlinear seasonal processes? Table 2 contains the results for the seasonal nonlinear autoregressive processes. The largest improvement in prediction error of the nonparametric procedure over Schwarz based lag selection and linear estimation is about 26% and occurs for the SNAR2 process ($1.2838/1.7376 - 1 = -26\%$). For the MISE of function estimation the largest reduction is 30% as in the case of f_0 of the SNAR2 process.

Table 3 contains the results for the seasonal dummy and seasonal shift nonlinear processes SDNAR1, SDNAR3, and SHNAR. There is a substantial reduction in MISE and prediction error for the SDNAR1 process if the semiparametric method is employed. For the SDNAR3 the gain in prediction error becomes less pronounced if lags have to be selected first and vanishes otherwise. For the SHNAR process the semiparametric reduction in MISE and prediction error for f are larger than for the SDNAR1 process that is generated by the same f function. One possible explanation is that the shift process is only one simple transformation from a standard process whereas the dummy process can never be transformed into a standard process.

We conclude that the non- and semiparametric identification procedures suggested in Sections 3 and 4 are overall superior to linear procedures if the generated process is nonlinear, whereas the loss in MISE and prediction power remains tolerable in case of linear processes. Regardless of prediction, lag selection is still quite robust. This is traditionally attributed to the discrete nature

Table 3. Mean integrated squared error and prediction error: Seasonal dummy and seasonal shift nonlinear autoregressive processes

	Selected lags				Correct lags	
	CAFPE	FPE	AIC	SC	Nonp.	Linear
			SDNAR1 *process*			
Under-/Correct/Overfitting	23/66/11	99/0/1	99/0/1	100/0/0		
MISE($\hat{f}$)	0.3907	0.5005	0.5005	0.4739	0.2816	0.4868
MSE($\hat{b}_1$)	0.1180	0.1280	0.1280	0.1045	0.1169	0.0918
MSE($\hat{b}_2$)	0.1672	0.1503	0.1503	0.1600	0.1470	0.1359
MSE($\hat{b}_3$)	0.1462	0.1190	0.1190	0.1199	0.1250	0.0967
MISE($\hat{f} + \sum_{s=1}^{3} \hat{b}_s D_{\cdot,s}$)	0.3574	0.4707	0.4707	0.4735	0.2558	0.4685
One-step prediction error	1.1697	1.4848	1.4848	1.3950	1.1031	1.4183
			SDNAR3 *process*			
Under-/Correct/Overfitting	0/78/22	3/46/51	3/46/51	5/77/18		
MISE($\hat{f}$)	0.0720	0.0672	0.0672	0.0632	0.0653	0.0550
MSE($\hat{b}_1$)	0.0147	0.0186	0.0186	0.0118	0.0135	0.0142
MSE($\hat{b}_2$)	0.0158	0.0462	0.0462	0.0397	0.0154	0.0244
MSE($\hat{b}_3$)	0.0191	0.0264	0.0264	0.0228	0.0180	0.0150
MISE($\hat{f} + \sum_{s=1}^{3} \hat{b}_s D_{\cdot,s}$)	0.0678	0.0551	0.0551	0.0535	0.0619	0.0507
One-step prediction error	0.1457	0.1515	0.1515	0.1489	0.1549	0.1485
			SHNAR *process*			
Under-/Correct/Overfitting	25/67/8	71/7/22	71/7/22	90/5/5		
MISE($\hat{f}$)	0.3490	0.4327	0.4327	0.4273	0.2669	0.4508
One-step prediction error	1.4031	1.7306	1.7306	1.7074	1.2145	1.5276

of the latter. In any case, the proposed procedures turn out to be applicable in practice. The next section will illustrate this.

7. EMPIRICAL APPLICATION

We apply the semiparametric methods introduced in Section 4 to two seasonal macroeconomic time series: quarterly real West German GNP from 1960:1 to 1990:4 compiled by Wolters (1992, p. 424, note 4) and quarterly UK public investment in 1985 prices from 1962:1 to 1988:4 taken from Osborn (1990). These series were chosen because there exists a detailed analysis using linear periodic models and seasonal unit root testing by Franses (1996) and because they are available via the World Wide Web.[4] All data are analyzed in logs.

In the introduction it was noted that the nonparametric analysis requires removing (non)seasonal unit roots from the data first. To save space we refer to the detailed analysis in Franses (1996, pp. 66–72). He uses the HEGY procedure and extensions. Allowing for a time trend and seasonal dummies and choosing the lag order by means of F-tests, he finds for the German real GNP roots at 1 and $-i, i$ and for the UK data roots at all seasonal frequencies and the zero frequency. However, for the German real GNP the evidence for the seasonal roots at $-i, i$ vanishes if one allows for shifting means in the middle of the sample. We therefore investigate the two series after taking first and fourth differences, respectively. Moreover, for the lag selection procedure we divide each resulting series by its standard deviation.

For lag selection the SDNAR model (3) and the SHNAR model (5) are applied, whereas the SNAR model (1) is not fitted because of the small number of years. In addition, the three linear criteria FPE, AIC, and SC are used. In all cases the procedures search over all possible lag combinations up to lag $M = 8$. Table 4 summarizes the selection results for both data sets. It contains the selected lags, the values of the selection criteria, and the estimated optimal bandwidths. The lags are ordered with respect to their contribution to reducing the selection criterion. The first column indicates the model. The model $f = 0$ contains only seasonal dummies but no lagged regressors. The variance of the resulting noise is equivalent to that of $\hat{U}_s = Y_s - \hat{\delta}_s$, the adjusted data after removing the mean shifts. Because the original data have been divided by the standard deviation, the variance of the adjusted data is at most one.

We compare the forecasting performance of all procedures using recursive prediction errors computed as follows. At each time t, observations up to time $t-1$ are used to select the lags and predict $\hat{Y}_t$ for approximately the last 20% of observations of each time series. This amounts to the last six and five years of the German and UK data, respectively. Table 5 shows for each model and both data sets the average of the squared deviations of the predicted values from their corresponding true ones.

TABLE 4. Semiparametric lag selection

Model	Criterion	Selected lags	Est. criterion	$\hat{h}_{opt}$
		First differences of German real GNP data		
$f = 0$	$\mathrm{Var}(\hat{Y}_t)$	0	0.23	—
Dummy	CAFPE	4,1,7	0.075	0.53
Shift	CAFPE	4,2,8	0.084	0.41
Linear	FPE	4,2,8	0.091	—
Linear	AIC	4,2,8	−2.393	—
Linear	SC	4,2	−2.233	—
		Fourth differences of UK public investment		
$f = 0$	$\mathrm{Var}(\hat{Y}_t)$	0	0.991	—
Dummy	CAFPE	1,6,5	0.407	1.24
Shift	CAFPE	1,6,5	0.392	1.07
Linear	FPE	1,2,4,6,8	0.438	—
Linear	AIC	1,2,4,6,8	−0.826	—
Linear	SC	1,2,4	−0.610	—

Note: The selected lags are listed with respect to their contribution to reducing the CAFPE. The maximal lag considered is 8, and all possible lag combinations are considered. The linear models are SDAR models.

Table 5. Recursive prediction errors

	Dummy				Shift			
Data	CAFPE	FPE	AIC	SC	CAFPE	FPE	AIC	SC
German	0.0413	0.0534	0.0504	0.0528	0.0524	0.0578	0.0582	0.0598
UK	0.9623	0.8808	0.8562	0.9458	0.8291	0.7972	0.7933	0.9183

Table 4 shows that regressing the first differences of the German real GNP data only on seasonal dummies reduces the variance to 0.23. Using the SDNAR model lags 4,1,7 are selected. The corresponding CAFPE of 0.075 implies a further substantial reduction in the error variance. Taking the SHNAR model instead leads to lags 4,2,8 and a CAFPE value of 0.084. The FPE estimate of the linear model selects the same lags and its value is 0.091.

We inspected the estimated residuals of both nonlinear models using Godfrey misspecification tests (e.g., Godfrey, 1979), which test an $AR(p)$ model against an $AR(p+r)$ model or an $ARMA(p,r)$ model. Here we use $r = 1$ and $r = 4$. In addition, the Jarque–Bera (JB) test statistic is computed. Because the Godfrey tests are designed for linear models, the results have to be interpreted with care. To save space we left out all test statistics. However, they are all far below the linear critical values, and so the results may be quite robust. In particular, there is no seasonality left in the residuals, and so using the first differencing filter is justified.

Following the discussion at the end of Section 4 on how to select between the SDNAR and SHNAR model, we prefer the SDNAR model for the first differences of the German real GNP as the corresponding CAFPE is about 11% smaller than that of the SHNAR model.

To represent the SDNAR model graphically, we keep the least important lag, lag 7, fixed and vary lags 1 and 4 on a grid. Figure 7 displays three surfaces of the estimated function on the domain of the observations. To look only on the domain of the observations reduces irrelevant boundary effects. This domain is plotted in the plane of the regressors. For visualization, all values have been multiplied by the standard deviation of the first differences, which is 0.052. Lag 7 is fixed at -0.05, 0, 0.05. The estimates of the seasonal dummy parameters are 0.023, 0.018, -0.050. The dummy model shows pronounced nonlinearities. For example, comparing the upper plot with the other two shows that if the quarterly growth was down all quarters, then today's growth is overproportionally low also. The influence of the growth performance in the last quarter on today's growth is ambiguous. It depends very much on the direction and magnitude of last quarter's growth and also on the situation one year ago. If one fits a linear model such effects may easily average out and turn out insignificant. Indeed, the linear criteria do not select lag 1.

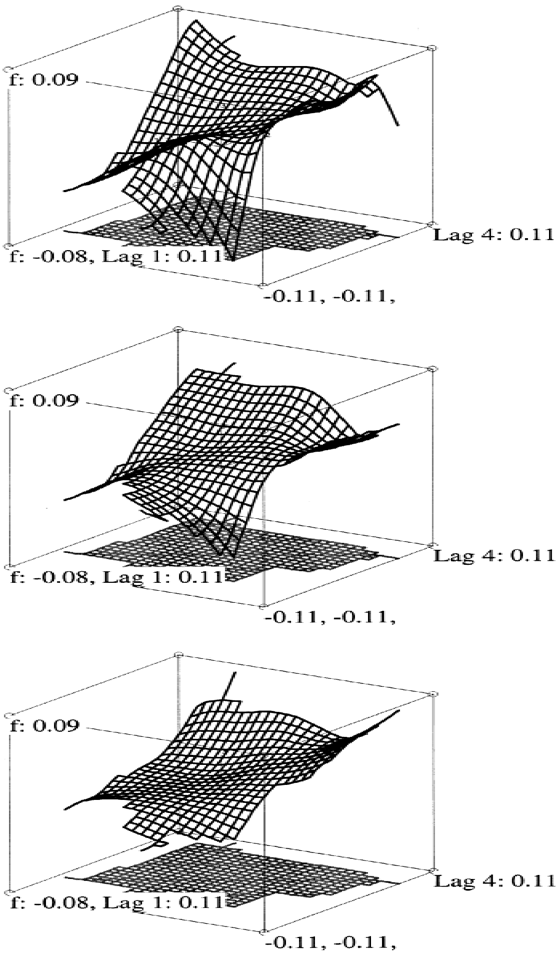

FIGURE 7. German real GNP: dummy model with lags 1 and 4 given lag 7 fixed at −0.05, 0, 0.05.

The recursive prediction errors in Table 5 support the superiority of the SDNAR model. The closest competitor is the linear dummy model using AIC lag selection, which is still 22% larger (0.504/0.413 = 122%). Comparing to the linear models and the nonlinear shift model, this example illustrates the empirical benefits of the more sophisticated nonlinear dummy model.

For the fourth differences of the UK public investment data the selection results are presented in Table 4. After having applied the fourth-order differencing filter, there is no relevant seasonality left in the data because removing seasonal shifts from the series does not change its variance. One may therefore expect the SDNAR and SHNAR models to perform similarly. Indeed, the chosen lag vector 1,6,5 is the same for both models. In contrast, all linear criteria contain the different vector 1,2,4.

As in the previous case, the residual diagnostics of the nonlinear models indicate no sign of misspecification and are therefore not reported here. The surfaces of the estimated regression function of the SHNAR model are shown in Figure 8, where the value of lag 5 is fixed at -0.10, 0, 0.10. All surfaces look quite smooth. The deviation from linear hyperplanes is less pronounced than

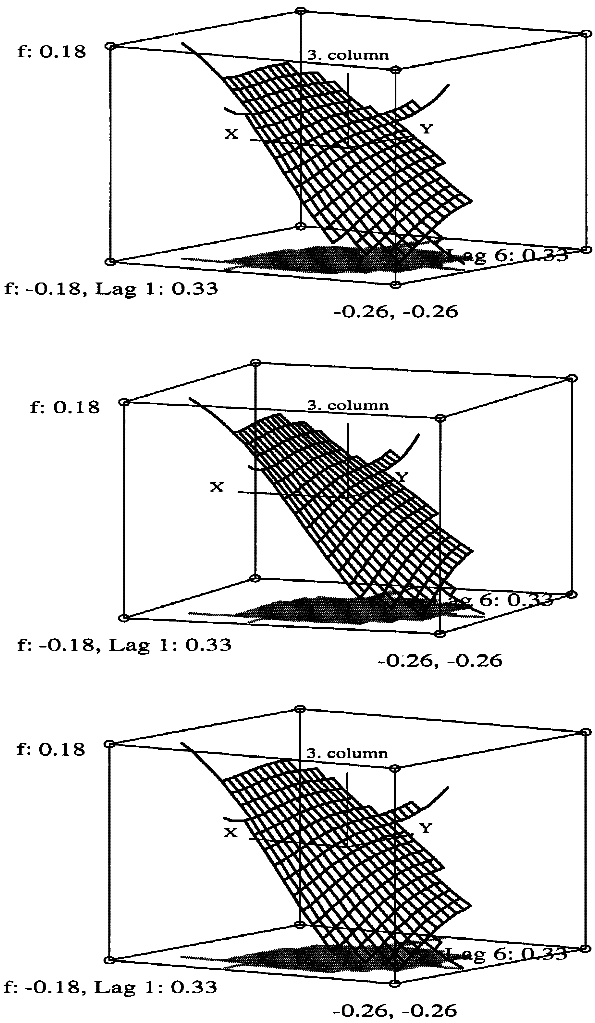

FIGURE 8. UK public investment: shift model with lags 1 and 6 given lag 5 fixed at -0.10, 0, 0.1.

for the German real GNP data although still evident. Therefore, a linear model could be superior for forecasting. This is confirmed by the prediction errors reported in Table 5, where the recursive errors of the nonlinear shift model are 4.5% larger than those of the linear shift model using AIC lag selection (0.8291/0.7933 = 104.5%). The linearity hypothesis can be tested, e.g., with the nonparametric linearity tests of Hjellvik and Tjøstheim (1995, 1996) and Hjellvik, Yao, and Tjøstheim (1998), which can be applied to both the adjusted data $\{X_t, \tilde{Y}_t\}$ of the SDNAR model and the deseasonalized data $\{\hat{U}_t\}$ of the SHNAR model. The prediction results also show that for the UK data set the shift model is the appropriate model.

8. CONCLUSION

In this paper, we have proposed nonparametric estimation and lag selection methods for seasonal nonlinear autoregressive models and derived semiparametric estimators for two restricted versions, one allowing the seasonal function to shift across seasons only by a constant parameter, the other generalizing linear unobserved components models. All methods allow for either local constant or local linear estimation. For the semiparametric models, after preliminary estimation of the seasonal parameters, the function estimation and lag selection are the same as nonparametric estimation and lag selection for standard models. A Monte Carlo study demonstrates good performance of all three methods. The methods are applied to the German real GNP data and UK public investment. It was found that the semiparametric lag selection and estimation procedures work even with moderate sample sizes. They help to identify more complicated dynamics in economic time series and can improve forecasting.

NOTES

1. Most prominent examples for model-based seasonal adjustment procedures are the unobserved components approach that includes the structural time series approach and the ARIMA model-based approach. The Census X-11 and the Hodrick–Prescott filter are representatives of procedures with model-based interpretation.

2. The formulation of models (3) and (5) might lead one to think that it is also possible to have a model such as

$$Y_t = f(Y_{t-i_1} - \delta_{\{t-i_1\}}, \ldots, Y_{t-i_m} - \delta_{\{t-i_m\}}) + \epsilon_t.$$

This is in fact not useful if one still wants to have some kind of strict stationarity. If the process $\{Y_t\}_{t\geq 0}$ itself is stationary, then unless all the parameters $\delta_{\{t-i_1\}}, \ldots, \delta_{\{t-i_m\}}$ are all equal, in which case they can be all set to 0 and one gets back to a standard process, the Y_t defined by this expression will not be stationary, even not for each season. On the other hand, if one wants to have stationarity of $\{Y_t - \delta_{\{t\}}\}_{t\geq 0}$, then this equation makes Y_t stationary, and therefore all parameter $\delta_{\{t-i_1\}}, \ldots, \delta_{\{t-i_m\}}$ equal. We have therefore restricted our semiparametric study to models (3) and (5).

3. All procedures were programmed in GAUSS using DLLs written in C++ and run on Sun workstations.

4. http://www.few.eur.nl/few/people/franses/research/book1.htm.

REFERENCES

Ango Nze, P. (1992) Critères d'ergodicité de quelques modèles à représentation Markovienne. *Compte Rendu à l'Académie des Sciences de Paris, séries 1* 315, 1301–1304.

Auestad, B. & D. Tjøstheim (1990) Identification of nonlinear time series: First order characterization and order determination. *Biometrika* 77, 669–687.

Burman, P. & R.H. Shumway (1998) Semiparametric modeling of seasonal time series. *Journal of Time Series Analysis* 19, 127–145.

Chen, R. & R.S. Tsay (1993) Nonlinear additive ARX models. *Journal of the American Statistical Association* 88, 955–967.

Cline, D.B.H. & H.H. Pu (1999) Geometric ergodicity of nonlinear time series. *Statistica Sinica* 9, 1103–1118.

Doukhan, P. (1994) *Mixing: Properties and Examples*. New York: Springer-Verlag.

Franses, H.F. (1996) *Periodicity and Stochastic Trends in Economic Time Series*. Oxford: Oxford University Press.

Ghysels, E., C.W. Granger, & P. Siklos (1996) Is seasonal adjustment a linear or nonlinear data filtering process? *Journal of Business and Economics Statistics* 14, 374–386.

Godfrey, L.G. (1979) Testing the adequacy of a time series model. *Biometrika* 66, 67–72.

Härdle, W., A.B. Tsybakov, & L. Yang (1998) Nonparametric vector autoregression. *Journal of Statistical Planning and Inference* 68, 221–245.

Hastie, T.J. & R.J. Tibshirani (1990) *Generalized Additive Models*. London: Chapman and Hall.

Hastie, T.J. & R.J. Tibshirani (1993) Varying-coefficient models. *Journal of the Royal Statistical Society, Series B* 55, 757–796.

Hjellvik, V. & D. Tjøstheim (1995) Nonparametric tests of linearity for time series. *Biometrika* 82, 351–368.

Hjellvik, V. & D. Tjøstheim (1996) Nonparametric statistics for testing linearity and series independence. *Journal of Nonparametric Statistics* 6, 223–251.

Hjellvik, V., Q. Yao, & D. Tjøstheim (1998) Linearity testing using polynomial approximation. *Journal of Statistical Planning and Inference* 68, 295–321.

Hylleberg, S., R.F. Engle, C.W.J. Granger, & B.S. Yoo (1990) Seasonal integration and cointegration. *Journal of Econometrics* 44, 215–238.

Liptser, R.S. & A.N. Shirjaev (1980) A functional central limit theorem for martingales. *Theory of Probability and Its Applications* 25, 667–688.

Lütkepohl, H. (1991) *Introduction to Multiple Time Series Analysis*. Heidelberg: Springer-Verlag.

Maravall, A. (1995) Unobserved components in economic time series. In M. Pesaran & M.R. Wickens (eds.), *Handbook of Applied Econometrics*, pp. 12–72. Oxford: Blackwell.

Masry, E. & D. Tjøstheim (1996) Additive nonlinear ARX time series and projection estimates. *Econometric Theory* 13, 214–252.

Orbe, S., E. Ferreira, & J. Rodriguez-Poo (2000) A nonparametric method to estimate time varying coefficients under seasonal constraints. *Journal of Nonparametric Statistics* 12, 779–806.

Osborn, D.R. (1990) A survey of seasonality in UK macroeconomic variables. *International Journal of Forecasting* 6, 327–336.

Roberts, G.O. & R.L. Tweedie (1999) Bounds on regeneration times and convergence rates for Markov chains. *Stochastic Processes and Their Applications* 80, 211–229.

Rosenthal, J.S. (1995) Minorization conditions and convergence rates for Markov chain Monte Carlo. *Journal of the American Statistical Association* 90, 558–566.

Silverman, B. (1986) *Density Estimation for Statistics and Data Analysis*. London: Chapman & Hall.

Tjøstheim, D. & B. Auestad (1994a) Nonparametric identification of nonlinear time series: Projections. *Journal of the American Statistical Association* 89, 1398–1409.

Tjøstheim, D. & B. Auestad (1994b) Nonparametric identification of nonlinear time series: Selecting significant lags. *Journal of the American Statistical Association* 89, 1410–1419.

Tschernig, R. & L. Yang (2000) Nonparametric lag selection for time series. *Journal of Time Series Analysis* 21, 457–487.
Wolters, J. (1992) Persistence and seasonality in output and employment of the Federal Republic of Germany. *Recherceches Economiques de Louvain* 58, 421–439.
Yang, L. (2000) Finite nonparametric GARCH model for foreign exchange volatility. *Communications in Statistics: Theory and Methods* 5 and 6, 1347–1365.
Yang, L., W. Härdle, & J.P. Nielsen (1999) Nonparametric autoregression with multiplicative volatility and additive mean. *Journal of Time Series Analysis* 20, 579–604.
Yang, L. & R. Tschernig (1999) Multivariate bandwidth selection for local linear regression. *Journal of the Royal Statistical Society, Series B* 61, 793–815.

APPENDIX

For the results in this paper, we need the following assumptions.

(A1) For some integer $M \geq i_m$ and for each season, the vector process $X_{t,M}$ is strictly stationary and β-mixing with $\beta(n) \leq c_0 n^{-(2+\delta)/\delta}$ for some $\delta > 0$, $c_0 > 0$. Here

$$\beta(n) = E \sup\{|P(A|\mathcal{F}_M^k) - P(A)| : A \in \mathcal{F}_{n+k}^\infty\},$$

where $\mathcal{F}_t^{t'}$ is the σ-algebra generated by $X_{M,t}, X_{M,t+1}, \ldots, X_{M,t'}$. For each season $s = 0,1,\ldots,S-1$, $\{X_{s+\tau S,M}\}_{\tau=i_{M,S}}^\infty$ has a stationary distribution with density $\mu_{s,M}(x_M)$, $x_M \in \mathbb{R}^M$, which is continuous. Henceforth, we use $\mu_s(\cdot)$ to denote both $\mu_{s,M}(\cdot)$ and all of its marginal densities. If the Nadaraya–Watson estimator is used, $\mu_{s,M}(\cdot)$ has to be continuously differentiable.

(A2) The functions $f_s(\cdot)$, $s = 0,1,\ldots,S-1$ are twice continuously differentiable, whereas each $\sigma_s(\cdot)$ is continuous and positive on the support of $\mu_s(\cdot)$, $s = 0,1,\ldots,S-1$.

(A3) The errors $\{\xi_t\}_{t \geq i_m}$ have a finite fourth moment m_4.

(A4) The support of $w(\cdot)$ is compact with nonempty interior. The function $w(\cdot)$ is continuous and nonnegative and $\mu_s(x) \geq c$ for some constant $c > 0$ if $x_M \in \text{supp}(w)$.

(A5) There exists a constant $c > 0$ such that for any proper subset $\{i_1', \ldots, i_{m'}'\}$ of $\{i_1, \ldots, i_m\}$, and any S functions $\{f_s'\}_{s=0}^{S-1}$ of m' variables,

$$E\left[\sum_{s=0}^{S-1} \{ f_s(Y_{t-i_1}, \ldots, Y_{t-i_m}) - f_s'(Y_{t-i_1'}, \ldots, Y_{t-i_{m'}'})\}^2 w(Y_{t-1}, \ldots, Y_{t-M})\right] \geq c.$$

See Theorem 2 for lower level conditions to guarantee (A1). Note that just as in Assumption (A1) of Tschernig and Yang (2000, p. 459), the strict stationarity condition in the preceding (A1) is not necessary, but we include it here for simplicity of the proof. Assumption (A5) guarantees that all the lags $\{i_1, \ldots, i_m\}$ are needed in all S functions to fit the model correctly. In other words, the functions $f_s(y_{t-i_1}, \ldots, y_{t-i_m})$, $s = 0,1,\ldots,S-1$ do not reduce to functions with fewer variables on the support of the weight function $w(\cdot)$; hence none of the variables $Y_{t-i_1}, \ldots, Y_{t-i_m}$ can be left out to predict Y_t conditional on past observations. The next assumption is needed for the existence of $h_{a,opt}$.

(A6) For $a = 1,2$, the C_a defined in (13) are positive and finite.

For the seasonal shift (SHNAR) model, several of these assumptions have to be modified.

(A1′) For some integer $M \geq i_m$, the vector process $V_{t,M} = (U_{t-1}, \ldots, U_{t-M})$ is strictly stationary with continuous density function $\mu_M(\cdot)$ and β-mixing with $\beta(n) \leq c_0 n^{-(2+\delta)/\delta}$ for some $\delta > 0$, $c_0 > 0$. If the Nadaraya–Watson estimator is used, $\mu_M(\cdot)$ has to be continuously differentiable. The density of $V_t = (U_{t-i_1}, \ldots, U_{t-i_m})$ is denoted as $\mu(\cdot)$.

(A2′) The function $f(\cdot)$ is twice continuously differentiable, whereas $\sigma(\cdot)$ is continuous and positive on the support of $\mu(\cdot)$.

(A4′) The support of $w(\cdot)$ is compact with nonempty interior. The function $w(\cdot)$ is continuous, nonnegative, and bounded below from 0 on the support of $w(\cdot)$.

(A5′) There exists a constant $c > 0$ such that for any proper subset $\{i'_1, \ldots, i'_{m'}\}$ of $\{i_1, \ldots, i_m\}$, and any function f' of m' variables,

$$E[\{f(U_{t-i_1}, \ldots, U_{t-i_m}) - f'(U_{t-i'_1}, \ldots, U_{t-i'_{m'}})\}^2 w(U_{t-1}, \ldots, U_{t-M})] \geq c.$$

(A6′) For $a = 1,2$,

$$0 < \int r_a^2(u)\mu(u_M)w(u_M)\,du_M < +\infty, \tag{A.1}$$

where

$$r_1(u) = \mathrm{Tr}\{\nabla^2 f(u)\} + \frac{2\nabla^T \mu(u)\nabla f(u)}{\mu(u)}, \qquad r_2(u) = \mathrm{Tr}\{\nabla^2 f(u)\}. \tag{A.2}$$

Proof of Theorem 2. Note first that condition (E1) guarantees the recurrence of the process $\{V_\tau\}_{\tau \geq k}$. Observe next that by repeated application of condition (E2) equation (9) implies the existence of a constant $R' > 0$ and a matrix of coefficients $(b_{sj})_{1\leq s\leq S, 1\leq j\leq kS}$ with all $b_{sj} \geq 0$ and $\max_{1\leq s\leq S} \sum_{j=1}^{kS} b_{sj} = b < 1$ such that

$$E(|Y_{\tau S-s+1}|\,|Y_{\tau S-S} = y_1, \ldots, Y_{\tau S-S-kS+1} = y_{kS}) \leq \sum_{j=1}^{kS} b_{sj}|y_j|,$$

$$1 \leq s \leq S, \qquad \tau > k \tag{A.3}$$

when $\min_{1\leq j\leq kS}|y_j| \geq R'$. Now define a function $g(v) = \sum_{i=1}^{kS} g_i|v_i|$ for $v = (v_1, \ldots, v_{kS})$ with positive coefficients g_i, $1 \leq i \leq kS$, to be determined subject to the condition that for some constants $\varepsilon > 0$ and $0 < c < 1$

$$E\{g(V_\tau)|V_{\tau-1} = v\} \leq cg(v) - \varepsilon, \qquad \tau = k+1, k+2, \ldots \tag{A.4}$$

when $\min_{1\le j\le kS}|v_j| \ge R'$. By definition of $g(\cdot)$ and equation (A.3)

$$\begin{aligned}
E\{g(V_\tau)|V_{\tau-1}=v\} &= E\left\{\sum_{i=1}^{kS} g_i|Y_{\tau S-i+1}||V_{\tau-1}=v\right\} \\
&= \sum_{i=1}^{S} g_i E\{|Y_{\tau S-i+1}||V_{\tau-1}=v\} + \sum_{i=S+1}^{kS} g_i|v_{i-S}| \\
&\le \sum_{i=1}^{S} g_i \sum_{j=1}^{kS} b_{ij}|v_j| + \sum_{j=1}^{kS-S} g_{j+S}|v_j| \\
&= \sum_{j=kS-S+1}^{kS} |v_j| \sum_{i=1}^{S} g_i b_{ij} + \sum_{j=1}^{kS-S} |v_j|\left(g_{j+S} + \sum_{i=1}^{S} g_i b_{ij}\right).
\end{aligned}$$

To obtain (A.4), one needs to have a constant $0 < c' < 1$ such that

$$\sum_{j=kS-S+1}^{kS} |v_j| \sum_{i=1}^{S} g_i b_{ij} + \sum_{j=1}^{kS-S} |v_j|\left(g_{j+S} + \sum_{i=1}^{S} g_i b_{ij}\right) < c' \sum_{j=1}^{kS} g_j|v_j|$$

when $\min_{1\le j\le kS}|v_j| \ge R'$, or equivalently

$$\begin{aligned}
g_{j+S} + \sum_{i=1}^{S} g_i b_{ij} &< g_j, \qquad 1 \le j \le kS-S, \\
\sum_{i=1}^{S} g_i b_{ij} &< g_j, \qquad kS-S+1 \le j \le kS,
\end{aligned}$$

which are simply

$$\begin{pmatrix} g_{S+1} \\ \vdots \\ g_{2S} \end{pmatrix} < \begin{pmatrix} g_1 \\ \vdots \\ g_S \end{pmatrix} - (b_{ij})_{1\le i\le S, 1\le j\le S} \begin{pmatrix} g_1 \\ \vdots \\ g_S \end{pmatrix},$$

$$\begin{pmatrix} g_{2S+1} \\ \vdots \\ g_{3S} \end{pmatrix} < \begin{pmatrix} g_{S+1} \\ \vdots \\ g_{2S} \end{pmatrix} - (b_{ij})_{1\le i\le S, S+1\le j\le 2S} \begin{pmatrix} g_1 \\ \vdots \\ g_S \end{pmatrix},$$

$$\vdots$$

$$\begin{pmatrix} g_{kS-S+1} \\ \vdots \\ g_{kS} \end{pmatrix} < \begin{pmatrix} g_{kS-2S+1} \\ \vdots \\ g_{kS-S} \end{pmatrix} - (b_{ij})_{1\le i\le S, kS-2S+1\le j\le kS-S} \begin{pmatrix} g_1 \\ \vdots \\ g_S \end{pmatrix},$$

$$(b_{ij})_{1\le i\le S, kS-S+1\le j\le kS} \begin{pmatrix} g_1 \\ \vdots \\ g_S \end{pmatrix} < \begin{pmatrix} g_{kS-S+1} \\ \vdots \\ g_{kS} \end{pmatrix},$$

where all the inequalities are taken to be elementwise. Apparently, one can solve this system of inequalities if and only if one can find positive g_i, $1 \le i \le S$, such that

$$(b_{ij})_{1\le i\le S,\, kS-S+1\le j\le kS}\begin{pmatrix} g_1\\ \vdots\\ g_S\end{pmatrix} < \begin{pmatrix} g_1\\ \vdots\\ g_S\end{pmatrix} - (b_{ij})_{1\le i\le S, 1\le j\le S}\begin{pmatrix} g_1\\ \vdots\\ g_S\end{pmatrix}$$

$$- \cdots (b_{ij})_{1\le i\le S,\, kS-2S+1\le j\le kS-S}\begin{pmatrix} g_1\\ \vdots\\ g_S\end{pmatrix}$$

or

$$\left(\sum_{l=0}^{k-1} b_{i,j+lS}\right)_{1\le i\le S, 1\le j\le S}\begin{pmatrix} g_1\\ \vdots\\ g_S\end{pmatrix} < \begin{pmatrix} g_1\\ \vdots\\ g_S\end{pmatrix}.$$

Now $\max_{1\le s\le S}\sum_{j=1}^{kS} b_{sj} = b < 1$ implies that $g_1 = \cdots = g_S = 1$ will solve the preceding inequality. Hence one can construct a function $g(\cdot)$ that satisfies (A.4), which entails that the process $\{V_\tau\}_{\tau\ge k}$ is geometrically ergodic by Theorem 3 of Doukhan (1994, p. 91), from which the rest of the lemma follows also. ∎

Proof of Theorem 3 and Theorem 4. For each $s = 0,1,\ldots,S-1$, the following decomposition holds:

$$E[\{\breve{Y}_{s+\tau S} - \hat{f}_s(\breve{X}_{s+\tau S})\}^2 w(\breve{X}_{s+\tau S,M})] = I_s + II_s + III_s,$$

in which

$$I_s = E[\{\breve{Y}_{s+\tau S} - f_s(\breve{X}_{s+\tau S})\}^2 w(\breve{X}_{s+\tau S,M})] = E[\sigma_s^2(\breve{X}_{s+\tau S}) w(\breve{X}_{s+\tau S,M})\breve{\xi}^2_{s+\tau S}],$$

$$II_s = E[\{\hat{f}_s(\breve{X}_{s+\tau S}) - f_s(\breve{X}_{s+\tau S})\}^2 w(\breve{X}_{s+\tau S,M})],$$

$$III_s = 2E[\{\breve{Y}_{s+\tau S} - f_s(\breve{X}_{s+\tau S})\}\{f_s(\breve{X}_{s+\tau S}) - \hat{f}_s(\breve{X}_{s+\tau S})\} w(\breve{X}_{s+\tau S,M})],$$

$$= 2E[\{f_s(\breve{X}_{s+\tau S}) - \hat{f}_s(\breve{X}_{s+\tau S})\} w(\breve{X}_{s+\tau S,M})\sigma_s(\breve{X}_{s+\tau S})\breve{\xi}_{s+\tau S}] = 0.$$

Now the innovations $\breve{\xi}_{s+\tau S}$ are i.i.d. white noise, so

$$\frac{1}{S}\sum_{s=0}^{S-1} I_s = \frac{1}{S}\sum_{s=0}^{S-1} E[\sigma_s^2(\breve{X}_{s+\tau S}) w(\breve{X}_{s+\tau S,M})] = \frac{1}{S}\sum_{s=0}^{S-1}\int \sigma_s^2(x) w(x_M)\mu_s(x_M)\,dx_M = A,$$

where A was defined in (11).

The second term II_s is

$$II_s = \int E\{\hat{f}_s(x) - f_s(x)\}^2 w(x_M)\mu_s(x_M)\,dx_M,$$

in which

$$E\{\hat{f}_s(x) - f_s(x)\}^2 = \{r_{a,s}(x)\sigma_K^2 h^2/2\}^2 + \|K\|_2^{2m} \frac{\sigma_s^2(x)}{\mu_s(x) n_{M,S} h^m} + o\left(h^4 + \frac{1}{nh^m}\right)$$

according to Theorem 1 and especially equation (8). Thus

$$\begin{aligned} II_s &= \frac{\sigma_K^4 h^4}{4} \int r_{a,s}^2(x) w(x_M) \mu_s(x_M) dx_M + \frac{\|K\|_2^{2m}}{n_{M,S} h^m} \int \frac{\sigma_s^2(x)}{\mu_s(x)} w(x_M) \mu_s(x_M) dx_M \\ &\quad + o\left(h^4 + \frac{1}{nh^m}\right) \\ &= c(h) \int r_{a,s}^2(x) w(x_M) \mu_s(x_M) dx_M + b(h) \int \frac{\sigma_s^2(x)}{\mu_s(x)} w(x_M) \mu_s(x_M) dx_M \\ &\quad + o\left(h^4 + \frac{1}{nh^m}\right). \end{aligned}$$

In summary

$$\begin{aligned} &\frac{1}{S} \sum_{s=0}^{S-1} E[\{\breve{Y}_{s+\tau S} - \hat{f}_s(\breve{X}_{s+\tau S})\}^2 w(\breve{X}_{s+\tau S,M})] \\ &\quad = A + c(h) \frac{1}{S} \sum_{s=0}^{S-1} \int r_{a,s}^2(x) w(x_M) \mu_s(x_M) dx_M \\ &\quad\quad + b(h) \frac{1}{S} \sum_{s=0}^{S-1} \int \frac{\sigma_s^2(x)}{\mu_s(x)} w(x_M) \mu_s(x_M) dx_M + o\left(h^4 + \frac{1}{nh^m}\right), \end{aligned}$$

which yields Theorem 3.

For establishing Theorem 4, one applies similar techniques to the cases when X_t does not include all correct lags and when it includes the m correct plus l additional ones. See Tschernig and Yang (2000, Sec. 3), for details. ■

To prove Theorem 5 one first decomposes the estimator into two parts:

$$\bar{b}_s = \frac{\sum_{\tau=i_{M,S}}^{n_S} w(X_{s+\tau S,M}) \hat{b}_s(X_{s+\tau S}, h)}{\sum_{\tau=i_{M,S}}^{n_S} w(X_{s+\tau S,M})} = \frac{P_1 + P_2}{\frac{1}{n_S} \sum_{\tau=i_{M,S}}^{n_S} w(X_{s+\tau S,M})}, \tag{A.5}$$

in which

$$P_1 = \frac{1}{n_S} \sum_{\tau=i_{M,S}}^{n_S} w(X_{s+\tau S,M}) e_s^T \{(Z_D^T W Z_D)^{-1} Z_D^T W\}(X_{s+\tau S}) \boldsymbol{\epsilon},$$

$$P_2 = \frac{1}{n_S} \sum_{\tau=i_{M,S}}^{n_S} w(X_{s+\tau S,M}) e_s^T \{(Z_D^T W Z_D)^{-1} Z_D^T W\}(X_{s+\tau S}) \mathbf{f},$$

and here one denotes by $\boldsymbol{\epsilon} = (\sigma_s(X_{s+\tau S})\xi_{s+\tau S})^T$ the innovation vector and $\mathbf{f} = (f(X_{s+\tau S}) + b_s)^T$ the prediction vector. Using the mixing conditions, similar to Härdle et al. (1998), one has

$$\frac{1}{n_S}\sum_{\tau=i_{M,S}}^{n_S} w(X_{s+\tau S,M}) = \int \mu_s(x_M)w(x_M)dx_M\{1+o_p(1)\},$$

and so Theorem 5 is proved by analyzing P_1 and P_2 using the following lemmas.

The following auxiliary lemma is a standard result.

LEMMA A.1. *Under conditions (A1)–(A4), as* $h \to 0$, $nh^m \to \infty$, *for any compact set* $\mathcal{K}$

$$\sup_{x\in\mathcal{K}} |(Z_D^T W Z_D)^{-1}(x) - \Sigma_D^{-1}(x)| \xrightarrow{a.s} 0,$$

where

$$\Sigma_D^{-1}(x) = \begin{bmatrix} \mu_0^{-1}(x) & -\mu_0^{-1}(x) & \dots & -\mu_0^{-1}(x) & 0_{1\times m} \\ -\mu_0^{-1}(x) & \mu_1^{-1}(x)+\mu_0^{-1}(x) & \dots & \mu_0^{-1}(x) & 0_{1\times m} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ -\mu_0^{-1}(x) & \mu_0^{-1}(x) & \dots & \mu_{S-1}^{-1}(x)+\mu_0^{-1}(x) & 0_{1\times m} \\ 0_{m\times 1} & 0_{m\times 1} & \dots & 0_{m\times 1} & S^{-1}\mu^{-1}(x)\sigma_K^{-2}I_m \end{bmatrix}$$

$$= \begin{bmatrix} \mu_0^{-1}(x) & -\mu_0^{-1}(x)1_{1\times(S-1)} & 0_{1\times m} \\ -\mu_0^{-1}(x)1_{(S-1)\times 1} & \operatorname{diag}\{(\mu_s(x)^{-1}\}_{s=1}^{S-1} + \mu_0^{-1}(x)1_{(S-1)\times(S-1)} & 0_{(S-1)\times m} \\ 0_{m\times 1} & 0_{m\times(S-1)} & S^{-1}\mu^{-1}(x)I_m\sigma_K^{-2} \end{bmatrix}.$$

Proof. Observe that each element of the matrix $(Z_D^T W Z_D)(x)$ can be written as

$$\sum_{s=0}^{S-1}\sum_{\tau=i_{M,S}}^{n_S} Z_{jt}Z_{it}K_h(X_{s+\tau S}-x), \qquad i,j=1,\dots,m+S.$$

Because for each season a stationary density $\mu_s(x)$ exists, one obtains for the S sums using the mixing property

$$(Z_D^T W Z_D)(x) = \begin{bmatrix} S\mu(x) & \mu_1(x) & \dots & \mu_{S-1}(x) & 0_{1\times m} \\ \mu_1(x) & \mu_1(x) & \dots & 0 & 0_{1\times m} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \mu_{S-1}(x) & 0 & \dots & \mu_{S-1}(x) & 0_{1\times m} \\ 0_{m\times 1} & 0_{m\times 1} & \dots & 0_{m\times 1} & S\mu(x)\sigma_K^2 I_m \end{bmatrix}(I_{m+S}+o_p(1))$$

uniformly over the compact set $\mathcal{K}$. It is straightforward to verify the inversion. ■

Note in particular, because the random vector $X_t 1_{\{w(X_{t,M})>0\}}$ has a density with compact support by Assumption (A4), Lemma A.1 applies to its support $\mathcal{K}$. To analyse P_1, one considers the effect of weighted averaging on $\{(Z_D^T W Z_D)^{-1} Z_D^T W\}(X_{s+\tau S})$. Denote x_M by (x, x'_M), where x_M corresponds to the largest lag vector, and define

$$\mu_{s,w}(x) = \int w(x, x'_M)\mu_s(x, x'_M)\,dx'_M \tag{A.6}$$

as the weighted density at x. The next lemma is important.

LEMMA A.2. *As $h \to 0$, $nh^m \to \infty$,*

$$\frac{1}{n_{M,S}} \sum_{\tau=i_{M,S}}^{n_S} w(X_{s+\tau S,M})\{(Z_D^T W Z_D)^{-1} Z_D^T W\}(X_{s+\tau S})$$
$$= \int \{\Sigma_D^{-1} Z_D^T W\}(x)\mu_{s,w}(x)\,dx\{1 + o_p(1)\}.$$

Proof. Using the mixing property, one has

$$\frac{1}{n_{M,S}} \sum_{\tau=i_{M,S}}^{n_S} w(X_{s+\tau S,M})\{(Z_D^T W Z_D)^{-1} Z_D^T W\}(X_{s+\tau S})$$
$$= \int w(x_M)\{(Z_D^T W Z_D)^{-1} Z_D^T W\}(x)\mu_s(x_M)\,dx_M\{1 + o_p(1)\}$$
$$= \int \{(Z_D^T W Z_D)^{-1} Z_D^T W\}(x) \times (w\mu_s)(x, x'_M)\,dx\,dx'_M\{1 + o_p(1)\}.$$

Integrating first in x'_M as in (A.6) and then applying Lemma A.1 gives the formula in the lemma. ∎

LEMMA A.3. *As $h \to 0$, $nh^m \to \infty$, $\sqrt{n_S}P_1 \xrightarrow{D} N(0, \sigma_s^2)$ where*

$$\sigma_s^2 = \int \left\{\frac{1}{\mu_s(x)} + \frac{1}{\mu_0(x)}\right\} \mu_{s,w}^2(x)\sigma_s^2(x)\,dx.$$

Proof. By definition

$$P_1 = \frac{1}{n_S} \sum_{\tau=i_{M,S}}^{n_S} w(X_{s+\tau S,M}) e_s^T\{(Z_D^T W Z_D)^{-1} Z_D^T W\}(X_{s+\tau S})\boldsymbol{\epsilon}.$$

Applying Lemma A.2, one has

$$P_1 = \int e_s^T\{\Sigma_D^{-1} Z_D^T W\}(x)\mu_{s,w}(x)\,dx\boldsymbol{\epsilon}\{1 + o_p(1)\}$$
$$= \frac{1}{n_S} \int \mu_{s,w}(x)\left\{\frac{1}{\mu_s(x)} + \frac{1}{\mu_0(x)}\right\} \sum_{\tau=i_{M,S}}^{n_S} K_h(X_{s+\tau S} - x)\sigma_s(X_{s+\tau S})\xi_{s+\tau S}\,dx\{1 + o_p(1)\}$$
$$- \frac{1}{n_S} \int \frac{\mu_{s,w}(x)}{\mu_0(x)} \sum_{0 \le s' \le S-1} \sum_{\tau=i_{M,S}}^{n_S} K_h(X_{s'+\tau S} - x)\sigma_s(X_{s'+\tau S})\xi_{s'+\tau S}\,dx\{1 + o_p(1)\}$$
$$+ \sum_{1 \le s' \le S-1,\, s' \ne s} \frac{1}{n_S} \int \frac{\mu_{s,w}(x)}{\mu_0(x)} \sum_{\tau=i_{M,S}}^{n_S} K_h(X_{s'+\tau S} - x)\sigma_s(X_{s'+\tau S})\xi_{s'+\tau S}\,dx\{1 + o_p(1)\},$$

which, by cancellation and changes of variables, becomes

$$\frac{1}{n_S}\sum_{\tau=i_{M,S}}^{n_S}\frac{\mu_{s,w}(X_{s+\tau S})}{\mu_s(X_{s+\tau S})}\sigma_s(X_{s+\tau S})\xi_{s+\tau S}\{1+o_p(1)\}$$
$$-\frac{1}{n_S}\sum_{\tau=i_{M,S}}^{n_S}\frac{\mu_{s,w}(X_{\tau S})}{\mu_0(X_{\tau S})}\sigma_s(X_{\tau S})\xi_{\tau S}\{1+o_p(1)\}.$$

By a martingale central limit theorem (Liptser and Shirjaev, 1980, Corollary 6), $\sqrt{n_S}P_1$ is asymptotically normal with variance

$$\int\frac{1}{\mu_s^2(x)}\mu_s(x)\mu_{s,w}^2(x)\sigma_s^2(x)dx+\int\frac{1}{\mu_0^2(x)}\mu_0(x)\mu_{s,w}^2(x)\sigma_s^2(x)dx$$
$$=\int\left\{\frac{1}{\mu_s(x)}+\frac{1}{\mu_0(x)}\right\}\mu_{s,w}^2(x)\sigma_s^2(x)dx=\sigma_s^2,$$

which completes the proof. ■

LEMMA A.4. *As* $h\to 0$, $nh^m\to\infty$,

$$P_2-\frac{1}{n_S}\sum_{\tau=i_{M,S}}^{n_S}w(X_{s+\tau S,M})b_s=O_p(h^4).$$

Proof. By definition

$$P_2-\frac{1}{n_S}\sum_{\tau=i_{M,S}}^{n_S}w(X_{s+\tau S,M})b_s$$
$$=\frac{1}{n_S}\sum_{\tau=i_{M,S}}^{n_S}w(X_{s+\tau S,M})[e_s^T\{(Z_D^TWZ_D)^{-1}Z_D^TW\}(X_{s+\tau S})\mathbf{f}-b_s].$$

Using the fact that $e_s^T\{(Z_D^TWZ_D)^{-1}Z_D^TW\}Z_De_{s'}=1$ or 0 depending on whether $s=s'$, the preceding expression equals

$$\frac{1}{n_S}\sum_{\tau=i_{M,S}}^{n_S}w(X_{s+\tau S,M})e_s^T\{(Z_D^TWZ_D)^{-1}Z_D^TW\}(X_{s+\tau S})\{\mathbf{f}-b_sZ_D(X_{s+\tau S})e_s-fZ_D(X_{s+\tau S})e_0\}$$
$$-\frac{1}{n_S}\sum_{\tau=i_{M,S}}^{n_S}w(X_{s+\tau S,M})e_s^T\{(Z_D^TWZ_D)^{-1}Z_D^TW\}(X_{s+\tau S})\{\nabla^TfZ_D(X_{s+\tau S})(e_{S+1},\ldots,e_{S+m})\},$$

which is

$$\left[\int e_s^T\{\Sigma_D^{-1}Z_D^TW\}(x)\mu_{s,w}(x)\{\mathbf{f}-b_sZ_D(x)e_s-fZ_D(x)e_0\}dx\right.$$
$$\left.-\int e_s^T\{\Sigma_D^{-1}Z_D^TW\}(x)\mu_{s,w}(x)\{\nabla^TfZ_D(x)(e_{S+1},\ldots,e_{S+m})\}dx\right]\{1+o_p(1)\}$$

or approximately

$$\left[\frac{1}{n_S}\int \mu_{s,w}(x)\left\{\frac{1}{\mu_s(x)}+\frac{1}{\mu_0(x)}\right\}\sum_{\tau=i_{M,S}}^{n_S} K_h(X_{s+\tau S}-x)\right.$$
$$\times\{f(X_{s+\tau S})-f(x)-\nabla^T f(x)(X_{s+\tau S}-x)\}dx$$
$$-\frac{1}{n_S}\int\frac{\mu_{s,w}(x)}{\mu_0(x)}\sum_{0\le s'\le S-1}\sum_{\tau=i_{M,S}}^{n_S} K_h(X_{s'+\tau S}-x)$$
$$\times\{f(X_{s'+\tau S})-f(x)-\nabla^T f(x)(X_{s'+\tau S}-x)\}dx$$
$$+\sum_{1\le s'\le S-1,\, s'\ne s}\frac{1}{n_S}\int\frac{\mu_{s,w}(x)}{\mu_0(x)}\sum_{\tau=i_{M,S}}^{n_S} K_h(X_{s'+\tau S}-x)$$
$$\left.\times\{f(X_{s'+\tau S})-f(x)-\nabla^T f(x)(X_{s'+\tau S}-x)\}dx\right],$$

which, after cancellation, becomes

$$\left[\frac{1}{n_S}\int\frac{\mu_{s,w}(x)}{\mu_s(x)}\sum_{\tau=i_{M,S}}^{n_S} K_h(X_{s+\tau S}-x)\{f(X_{s+\tau S})-f(x)-\nabla^T f(x)(X_{s+\tau S}-x)\}dx\right.$$
$$\left.-\frac{1}{n_S}\int\frac{\mu_{s,w}(x)}{\mu_0(x)}\sum_{\tau=i_{M,S}}^{n_S} K_h(X_{\tau S}-x)\{f(X_{\tau S})-f(x)-\nabla^T f(x)(X_{\tau S}-x)\}dx\right]$$
$$\times\{1+o_p(1)\}.$$

Now for the second term in the preceding expression, i.e., the sum of the 0-season,

$$\frac{1}{n_S}\int\frac{\mu_{s,w}(x)}{\mu_0(x)}\sum_{\tau=i_{M,S}}^{n_S} K_h(X_{\tau S}-x)\{f(X_{\tau S})-f(x)-\nabla^T f(x)(X_{\tau S}-x)\}dx$$
$$=\int\frac{\mu_{s,w}(x)}{\mu_0(x)}\int K_h(y-x)\{f(y)-f(x)-\nabla^T f(x)(y-x)\}dx\mu_0(y)dy\{1+o_p(1)\}$$
$$\overset{y-x=hu}{=}\int\frac{\mu_{s,w}(x)}{\mu_0(x)}\int K(u)\{f(x+hu)-f(x)-h\nabla^T f(x)u\}dx\mu_0(x+hu)$$
$$\times du\{1+o_p(1)\}$$
$$=\frac{h^2\sigma_K^2}{2}\int\mu_{s,w}(x)\mathrm{Tr}\{\nabla^2 f(x)\}dx+O_p(h^4).$$

Doing the same thing for the s-season, and taking the difference, one arrives at the conclusion of the lemma. ■

Proof of Theorem 5. Putting together equation (A.5) and Lemmas A.3 and A.4 proves the theorem. ■

Proof of Theorem 6. This follows by examining the proofs of Lemmas A.3 and A.4. Lemma A.3 would still hold in this situation, whereas Lemma A.4 has to be modified. Specifically, the term P_2 is

$$\int \mu_{s,w}(x')f_s^{\perp}(x)dx - \int \mu_{s,w}(x')f_0^{\perp}(x)dx + O_p(h^2)$$

or

$$\int \mu_{s,w}(x')\{f_s^{\perp}(x) - f_0^{\perp}(x)\}dx + O_p(h^2) = \int \{f_s^{\perp}(x) - f_0^{\perp}(x)\}w(x_M)\mu_s(x_M)dx_M + O_p(h^2)$$

according to definition (A.6). See Tschernig and Yang (2000) for more details about nonvanishing bias in underfitting. ■

Proof of Theorem 7. This theorem is an obvious consequence of the central limit theorem for mixing processes (see Doukhan, 1994, Theorem 1, p. 46). ■