Journal of Econometrics 176 (2013) 59–79

Contents lists available at SciVerse ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

Semiparametric estimation in triangular system equations with nonstationarity

Jiti Gao[a,b,*], Peter C.B. Phillips[c,d,e,f,1]

[a] University of Adelaide, Australia
[b] Monash University, Australia
[c] Yale University, United States
[d] University of Auckland, New Zealand
[e] University of Southampton, United Kingdom
[f] Singapore Management University, Singapore

ARTICLE INFO

Article history:
Received 6 June 2010
Received in revised form
9 April 2013
Accepted 15 April 2013
Available online 24 April 2013

JEL classification:
C23
C25

Keywords:
Endogeneity
Integrated process
Nonstationarity
Partial linear model
Simultaneity
Vector semiparametric regression

ABSTRACT

A system of multivariate semiparametric nonlinear time series models is studied with possible dependence structures and nonstationarities in the parametric and nonparametric components. The parametric regressors may be endogenous while the nonparametric regressors are assumed to be strictly exogenous. The parametric regressors may be stationary or nonstationary and the nonparametric regressors are nonstationary integrated time series. Semiparametric least squares (SLS) estimation is considered and its asymptotic properties are derived. Due to endogeneity in the parametric regressors, SLS is not consistent for the parametric component and a semiparametric instrumental variable (SIV) method is proposed instead. Under certain regularity conditions, the SIV estimator of the parametric component is shown to have a limiting normal distribution. The rate of convergence in the parametric component depends on the properties of the regressors. The conventional $\sqrt{n}$ rate may apply even when nonstationarity is involved in both sets of regressors.

1. Introduction

Existing studies show that both nonstationarity and nonlinearity are common features of much economic data. Modeling such data in a way that allows for possible nonstationarity helps to avoid dependence on stationarity assumptions and mixing conditions for all of the variables in the system. At present there is a large literature on parametric linear modeling of nonstationary time series and interest has primarily focused on time series with a unit root or near unit root structure (for an overview, see, for example Phillips and Xiao, 1998, and the references therein). In practical work, much attention is given to multivariate systems and cointegration models. Inferential methods for these linear systems include both parametric and semiparametric (e.g., Phillips, 1995, 1998, forthcoming) approaches.

In comparison with work on linear parametric models, there have been only a few studies of parametric nonlinear models with integrated variables. Park and Phillips (1988, 1989, 1999, 2001) introduced techniques for developing asymptotics for certain classes of nonlinear nonstationary parametric systems and aspects of this work have been extended by Pötscher (2004), Jeganathan (2004, 2008), and Berkes and Horváth (2006). Interest has also developed in nonparametric modeling methods to deal with nonlinearity of unknown form involving nonstationary variables. Existing studies in the field of nonparametric autoregression and cointegration estimation include Phillips and Park (1998), Karlsen and Tjøstheim (2001), Wang and Phillips (2009a,b), Karlsen et al. (2007), Kasparis and Phillips (2009), Cai et al. (2009), Schienle (2009), and Phillips (2009). The last paper examines in a nonparametric setting spurious time series models of the type for which the asymptotic theory was given in Phillips (1986, 1998).

* Correspondence to: Department of Econometrics and Business Statistics, Monash University, Caulfield East, VIC 3145, Australia. Tel.: +61 3 99031675; fax: +61 3 99032007.
E-mail addresses: jiti.gao@monash.edu (J. Gao), peter.phillips@yale.edu (P.C.B. Phillips).
1 Present address: Cowles Foundation for Research in Economics, Yale University, New Haven, CT 06520, USA.

0304-4076/$ – see front matter
http://dx.doi.org/10.1016/j.jeconom.2013.04.018

Among nonparametric studies of nonstationarity, two different mathematical approaches have been developed. In one approach, a so-called "Markov splitting technique" has been used in Karlsen and Tjøstheim (2001), and Karlsen et al. (2007) to model univariate time series with a null-recurrent structure; and Chen et al. (2012) consider univariate semiparametric regression modeling of null-recurrent time series, in which there is neither endogeneity nor heteroskedasticity. In the other approach, Phillips and Park (1998), Phillips (2009), and Wang and Phillips (2009a,b) have developed 'local-time' methods to derive an asymptotic theory for nonparametric estimation of univariate models involving integrated time series.

As we explain in detail in the paragraph between Eqs. (2.12) and (2.13) below, completely nonparametric regression estimation is limited only to the univariate integrated time series case. In other words, existing studies given in the literature, such as Karlsen and Tjøstheim (2001), Karlsen et al. (2007), and Wang and Phillips (2009a,b), for the univariate integrated time series case, may not be extendable to the multivariate nonstationary regressor case. This motivates the discussion in the literature for using varying-coefficient regression models, such as in Cai et al. (2009), and semiparametric regression models, as proposed in model (1.1) below, to deal with nonparametric and semiparametric estimation of multivariate nonstationary regressors. In applied work, such nonparametric and semiparametric methods have been shown to be particularly useful in modeling economic data in a way that retains generality where it is most needed while reducing dimensionality problems.

The present paper seeks to pursue these advantages in a wider context that allows for nonstationarities and endogeneities within a vector semiparametric regression model. The null recurrent structure of integrated time series typically reduces the amount of time that such time series spend in the vicinity of any one point, thereby exacerbating the sparse data problem or "curse of dimensionality" in nonparametric and semiparametric modeling of multivariate integrated time series. On the other hand, recurrence means that nonlinear shape characteristics of unknown form may be captured over unbounded domains and endogeneity may be often accommodated without specialized methods (Wang and Phillips, 2009b).

A common motivation for the use of semiparametric formulations such as (1.1) below is that they reduce nonparametric dimensionality through the presence of a linear parametric component. In our setting, the time series $\{(Y_t, X_t, V_t) : 1 \leq t \leq n\}$ are assumed to be modeled in a system of multivariate nonstationary time series models the form

$$\begin{aligned} &Y_t = A\,X_t + g(V_t) + e_t, \\ &X_t = H(V_t) + U_t, \quad t = 1, 2, \ldots, n, \\ &E[e_t|V_t] = E[e_t] = 0 \quad \text{and} \quad E[U_t|V_t] = 0, \end{aligned} \tag{1.1}$$

where n is the sample size, A is a $p \times d$-matrix of unknown parameters, $Y_t = (y_{t1}, \ldots, y_{tp})'$, $X_t = (x_{t1}, \cdots, x_{td})'$, and V_t is a sequence of univariate integrated time series regressors, $g(\cdot) = (g_1(\cdot), \ldots, g_p(\cdot))'$ and $H(\cdot) = (h_1(\cdot), \ldots, h_d(\cdot))'$[2] are all unknown functions, and both e_t and U_t are vectors of stationary time series. Note that $\{X_t\}$ can be stationary when $\{X_t\}$ and $\{V_t\}$ are independent. An extended version of model (1.1) is given in (2.21) in Section 2.2 below to deal with cases where X_t itself is an integrated time series regressor.

Model (1.1) corresponds to similar structures that have been used in the independent case (see Newey et al., 1999; Su and Ullah, 2008). The condition $E[e_t|V_t] = E[e_t]$ is generally needed to ensure that the model is identified. For, if there were an unknown function $\lambda(\cdot)$ such that $e_t = \lambda(V_t) + \varepsilon_t$ with $E[\varepsilon_t|V_t] = 0$, then only $g(\cdot)+\lambda(\cdot)$ would normally be estimable. However, recent research has revealed that some cases where e_t is correlated with V_t may be included. In particular, in studying nonparametric regressions of the form $Y_t = g(V_t) + e_t$, Wang and Phillips (2009b) consider a nonstationary endogenous regressor case where V_t is correlated with ϵ_t and show that conventional nonparametric regression is applicable in spite of the endogeneity. Phillips and Su (2011) show that the same phenomena holds in cross section cases where there are continuous location shifts in the regressor, which play the role of an instrumental variable in tracing out the nonparametric regression function.

The identification condition $E[e_t|V_t] = E[e_t] = 0$ eliminates endogeneity between ϵ_t and V_t while retaining endogeneity between e_t and X_t and potential nonstationarity in both X_t and V_t. The condition $E[e_t|V_t] = E[e_t] = 0$ in our setting corresponds to the condition $E[e_t|V_t, U_t] = E[e_t|U_t]$ that is assumed in Newey et al. (1999) and Su and Ullah (2008), the former being implied by $E[e_t|V_t] = E\left(E\left[e_t|U_t, V_t\right]|V_t\right) = E\left(E\left[e_t|U_t\right]|V_t\right) = E\left(E\left[e_t|U_t\right]\right) = E\left[e_t\right]$ when U_t is independent of V_t and $E[e_t] = 0$. The identification conditions in (1.1) allow for both conditional heteroskedasticity and endogeneity in e_t, permitting e_t to depend on U_t[3]. These conditions are also less restrictive than the exogeneity condition between e_t and (X_t, V_t) that is common in the literature for the stationary case (see, for example Härdle et al., 2000; Gao, 2007).

The present paper treats model (1.1) as a vector semiparametric structural model and considers the case where X_t is a vector of nonstationary regressors and may be endogenous, V_t may be a univariate integrated regressors and uncorrelated with e_t. In the case where endogeneity is involved in semiparametric regression modeling of independent data, some related developments include Robinson (1988), Newey et al. (1999), Ai and Chen (2003), Newey and Powell (2003), Li and Racine (2007), Su and Ullah (2008), and Florens et al. (2012). While estimation of partially linear models with endogeneity is discussed in each of these papers, neither the proposed structures nor the estimation methods may be used to deal with our case.

The contributions of the paper are as follows. We first consider a semiparametric least squares (SLS) estimator of A. When there is endogeneity in X_t, the SLS estimator of A is inconsistent. This may be seen from model (2.9) below when $E\left[e_t'|U_t\right] \neq 0$. Accordingly, the paper proposes a semiparametric instrumental variable least squares (SIV) estimate of A to deal with endogeneity in X_t and a nonparametric estimator for the function $g(\cdot)$. The SIV estimator of A is shown to be consistent with a conventional $\sqrt{n}$ rate of convergence in some cases even when X_t is stochastically nonstationary. This rate arises because nonstationarity in the regression may be eliminated by means of stochastic detrending.

The semiparametric procedure given here may be used on a system of nonlinear simultaneous equations with the following features: (i) nonstationarity and endogeneity in the parametric regressors; (ii) nonlinearity and nonstationarity in the nonparametric regressors; and (iii) stationary residuals. As such, the paper complements existing results on parametric modeling with endogeneity, nonparametric and semiparametric estimation of nonlinear time series (such as Fan and Yao, 2003; Gao, 2007),

[2] $F'(\cdot)$ denotes transpose of the vector function $F(\cdot)$, and $F^{(i)}(\cdot)$ denotes the i-th derivative of $F(\cdot)$.

[3] The additive case where $e_t = \lambda(U_t) + \mu_t$ with $E[\mu_t|V_t] = 0$ is covered in the first part of (1.1) because $E\left[e_t|V_t\right] = E\left[\lambda(U_t)|V_t\right] + E[\mu_t|V_t] = E\left[\lambda(U_t)\right] = E\left[e_t\right]$ when U_t is independent of V_t. The multiplicative case where $e_t = \sigma(U_t)\nu_t$ is also covered in the first part of (1.1) because $E\left[e_t|V_t\right] = E\left[\sigma(U_t)\nu_t|V_t\right] = E\left[e_t\right]$ when (U_t, ν_t) is assumed to be independent of V_t.