Available online at www.sciencedirect.com
ScienceDirect
Journal of Econometrics 142 (2008) 241–264
JOURNAL OF Econometrics
www.elsevier.com/locate/jeconom

Nonparametric transformation to white noise

Oliver B. Linton[a,*], Enno Mammen[b]

[a]Department of Economics, London School of Economics, Houghton Street, London WC2A 2AE, UK
[b]Department of Economics, University of Mannheim, L7, 3-5, 68131 Mannheim, Germany

Received 9 August 2006; received in revised form 2 April 2007; accepted 3 May 2007
Available online 6 June 2007

Abstract

We consider a semiparametric distributed lag model in which the "news impact curve" m is nonparametric but the response is dynamic through some linear filters. A special case of this is a nonparametric regression with serially correlated errors. We propose an estimator of the news impact curve based on a dynamic transformation that produces white noise errors. This yields an estimating equation for m that is a type two linear integral equation. We investigate both the stationary case and the case where the error has a unit root. In the stationary case we establish the pointwise asymptotic normality. In the special case of a nonparametric regression subject to time series errors our estimator achieves efficiency improvements over the usual estimators, see Xiao et al. [2003. More efficient local polynomial estimation in nonparametric regression with autocorrelated errors. Journal of the American Statistical Association 98, 980–992]. In the unit root case our procedure is consistent and asymptotically normal unlike the standard regression smoother. We also present the distribution theory for the parameter estimates, which is nonstandard in the unit root case. We also investigate its finite sample performance through simulation experiments.

JEL classification: C14

Keywords: Efficiency; Inverse problem; Kernel estimation; Nonparametric regression; Time series; Unit roots

1. Introduction

In this paper we discuss the estimation of the unknown quantities in the model

$$B(L)Y_t = A(L)m(X_t) + \varepsilon_t, \tag{1}$$

where ε_t is a martingale difference sequence with respect to the past of Y_t and current and past regressors X_t, while $A(L) = \sum_{j=0}^{\infty} a_j L^j$ and $B(L) = \sum_{j=0}^{\infty} b_j L^j$ are lag polynomial operators with $a_0 = b_0 = 1$ for identification, where $Lx_t = x_{t-1}$. The function $m(.)$ is assumed to be unknown but smooth, and is the object of central interest, although the dynamics of the model represented by $A(L), B(L)$ are also fundamental to the interpretation.

*Corresponding author. Tel.: +44 171 955 7864; fax: +44 171 831 1840.
E-mail addresses: o.linton@lse.ac.uk (O.B. Linton), emammen@rumms.uni-mannheim.de (E. Mammen).

0304-4076/$ - see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.jeconom.2007.05.018

We first discuss a special case of central interest, the nonparametric regression model

$$Y_t = m(X_t) + u_t, \quad t = 1, \ldots, T, \tag{2}$$

where the covariates follow some stationary mixing process, while the residual process u_t satisfies

$$A(L)u_t = \varepsilon_t = \sum_{j=0}^{\infty} a_j u_{t-j}. \tag{3}$$

In this case, $A(L)Y_t = A(L)m(X_t) + \varepsilon_t$, which is a special case of (1) with $A(L) = B(L)$. The parametric version of the regression models (2) and (3) is a standard teaching topic in graduate econometrics, Harvey (1981, Chapter 6). In the semiparametric model there are many standard estimators of m and of the parameters of $A(L)$ that are consistent under summability conditions on A, see, for example, Robinson (1983), Bierens (1983), Masry and Fan (1997), Hidalgo (1997), and Fan and Yao (2003). However, unlike in the parametric case, the standard kernel regression smoothers do not take account of the correlation structure in X_t or u_t and estimate the regression function in the same way as if these processes were independent. Furthermore, the variance of such estimators is proportional to the short run variance of u_t, $\sigma_u^2 = \text{var}(u_t)$ and does not depend on the regressor or error covariance functions $\text{cov}(X_t, X_{t-j})$, $\text{cov}(u_t, u_{t-j})$, $j \neq 0$. This is a bit surprising in comparison with the parametric case. One might think that there is useful information in the autocorrelation structure for estimation of the mean. This point has been addressed recently by Xiao et al. (2003) who proposed a more efficient estimator of m based on a prewhitening transformation

$$Y_t - \sum_{j=1}^{\infty} a_j(Y_{t-j} - m(X_{t-j})) = m(X_t) + \varepsilon_t, \tag{4}$$

where the right-hand side is now a standard nonparametric regression with whitened errors. The transform implicitly takes account of the autocorrelation structure. In practice they replaced the unknown quantities on the left-hand side by preliminary estimates of m and $a_j(\alpha)$. Their procedure improves in terms of variance over the usual kernel smoothers.

Model (1) is more general than nonparametric regression with autocorrelated errors and is perhaps more rightly viewed as a generalization of the distributed lag model. The traditional distributed lag model (with $m(x) = x$) has been very popular in economics, Dhrymes (1971).[1] More recently, Hendry et al. (1984) reviewed the specification of such models and gave a taxonomy of special cases. It can be motivated from some simple economic relationships being distorted by adaptive expectations, partial adjustment, etc., see Harvey (1981, Chapter 7). Suppose there is a latent variable Y^* that has some equilibrium relationship with covariate X, which in general can be nonlinear so that $Y_t^* = m(X_t)$. Then suppose that actual Y only responds to Y^* with some lagging mechanism, for example, $Y_t - Y_{t-1} = \gamma[Y_t^* - Y_{t-1}] + \varepsilon_t$ for some $\gamma \in (0, 1)$, then we obtain a special case of (1).[2] The lags arise because production takes time or because agents take time to respond to a signal or because there are institutional constraints. The traditional applications were in, for example, production studies where Y_t is output and X_t is the capital/labour ratio of a given firm or industry observed over time. More recent applications have been in rational expectations models where the data are at different frequencies, Hansen and Hodrick (1980). The issues concerning formulation and estimation of the lag polynomials A, B are pretty much resolved in the linear case, see Hannan and Deistler (1988) for a more recent discussion in the multivariate case. Linearity of m is just a convenience and was adopted many years ago when computational and technical issues were binding. We allow for nonlinear m because for some problems linear m is not well motivated and at odds with the data. Note that model (1) includes as a special case the so-called NARMAX model introduced in Chen and Billings (1989) and used frequently by systems engineers in which

[1] Sims (1971) and Geweke (1978) consider a continuous time distributed lag model where $Y(t) = \int_{-\infty}^{\infty} a(s)X(t-s)\,\mathrm{d}s + \varepsilon(t)$ and the data are observed at discrete time intervals in which case the (high frequency) discrete time approximation to this is like (1) with $B(L) = 1$ and $A(L) = \sum_{j=-\infty}^{\infty} a_j L^j$ for some a_j related to the function $a(.)$ under some conditions.

[2] The usual properties of linear dynamic regression models can be extended to the nonlinear case. Thus, for example, we can define the average instantaneous impact $\mathrm{E}[\partial Y_t/\partial X_t]$ as equal to the average derivative of the function $m = \mathrm{E}[m'(X_t)]$, a quantity that has been investigated elsewhere. The total dynamic average impact $\sum_{j=0}^{\infty} \mathrm{E}[\partial Y_{t+j}/\partial X_t] = \mathrm{E}[m'(X_t)] \sum_{j=0}^{\infty} (B(L)/A(L))_j$ is proportional to the instantaneous impact.