Université de Montréal
Faculté des arts et des sciences
Département de sciences économiques

CAHIER 9551

NONPARAMETRIC TWO-STEP ESTIMATION OF UNKNOWN REGRESSION FUNCTIONS WHEN THE REGRESSORS AND THE REGRESSION ERROR ARE NOT INDEPENDENT

Serena NG[1] and Joris PINKSE[2]

[1] Département de sciences économiques and Centre de recherche et développement en économique (C.R.D.E.), Université de Montréal

[2] Department of Economics, University of British Columbia

October 1995

Both authors thank the Social Sciences and Humanities Research Council of Canada (SSHRC) for financial support. The first author also acknowledges grants from the Fonds FCAR of Québec. We thank Paul Beaudry, John Cragg, Joel Horowitz and Pierre Perron for stimulating comments. Jean Boivin provided valuable research assistance.

C.P. 6128, succursale Centre-ville
Montréal (Québec) H3C 3J7

Télécopieur (FAX): (514) 343-5831
Courrier électronique (E-Mail): econo@tornade.ERE.Umontreal.CA

Ce cahier a également été publié au Centre de recherche et développement en économique (C.R.D.E.) (publication no 4595).

Dépôt légal - 1995
Bibliothèque nationale du Québec
Bibliothèque nationale du Canada

ISSN 0709-9231

RÉSUMÉ

Cet article considère une régression non paramétrique de la forme $Y = g(X) + \varepsilon$, où $E[\varepsilon|X] \neq 0$. On s'intéresse à une procédure en deux étapes (2SNP) qui utilise un ensemble d'instruments, Z, et une régression non paramétrique auxiliaire de la forme $X = h(Z) + \eta$. Une régression polynomiale de Y sur X et $\hat{\eta}$ est ensuite utilisée pour obtenir un estimé *consistant* de g. Il est démontré que l'estimateur donne des estimés de g dont l'erreur quadratique moyenne échantillonnale converge en probabilité vers zéro. L'estimateur peut être utilisé dans un contexte de régressions non paramétriques lorsque l'hypothèse d'indépendance entre les régresseurs et les résidus n'est pas respectée.

Mots clés : régressions non paramétriques, variables instrumentales, doubles moindres carrés, biais de simultanéité

ABSTRACT

This paper considers nonparametric regressions of the form $Y = g(X) + \varepsilon$, when $E[\varepsilon|X] \neq 0$. We consider a two-step procedure (2SNP) which makes use of a set of instruments, Z, and an auxiliary nonparametric regression of the form $X = h(Z) + \eta$. A polynomial of Y on X and $\hat{\eta}$ is then used to obtain a consistent estimate of g, which is shown to have sample mean square error that converges in probability to zero. The estimate can be used in nonparametric regressions when independence between the regressors and the regression error fails.

Key words : nonparametric regressions, instrumental variables, two-stage least squares, simultaneity bias

1 Introduction

Consider the linear regression model (e.g. a consumption function):

$$Y_t = X_t^T \beta_0 + \varepsilon_t \qquad t = 1, \ldots, n \qquad \varepsilon_t \sim \text{IID}(0, \sigma^2 I), \tag{1}$$

where X_t is a k–vector of explanatory variables and n is the sample size. Hereafter we denote the $n \times k$ matrix of regressors by X, and Y and ε are n–vectors. Assuming the model is correctly specified, least squares estimation yields $(\hat{\beta} - \beta_0) = (X^T X)^{-1} X^T \varepsilon$. Hence, $\hat{\beta}$ is consistent for β_0 if $n^{-1} X^T \varepsilon = o_p(1)$, and is unbiased if $E(\varepsilon | X) = 0$. However, there are at least two cases when the independence assumption between X and ε is violated[1]. The first is errors–in–variables. When the X_t's are measured with error, the observed X_t's will no longer be independent of the disturbances. The second arises as a result of endogeneity of the regressors. An example is the simultaneous determination of price and quantity in a classical supply–demand model.

In classical linear regression analysis, Instrumental Variables (IV) and Two Stage Least Squares (2SLS) are often used in place of ordinary least squares when the regressors are not independent of the regression error. In the IV approach, one searches for a set of instruments Z that is correlated with X but is orthogonal to the error term. The columns of X are projected onto the space spanned by the columns of Z. The IV criterion function is $(Y - X\beta)^T P_Z (Y - X\beta)$, where P_Z is the projection matrix. Identification requires that there are at least as many instruments as regressors. If this condition is satisfied, it can be shown that the IV estimate is $\sqrt{n}$–consistent and asymptotically normal, although not unbiased under regularity conditions.

The 2SLS approach resolves the problem in two steps. In the first stage, all the variables susceptible to violating the independence condition are regressed on a set of exogenous and predetermined variables. In the second stage, the fitted values from the first stage are used in place of the endogenous variables. Two stage least squares is more suitable for estimating the parameters of a structural model. However, it also applies in cases of "generated regressors" [cf. Pagan (1984)].

A different solution to the problem of endogeneity was used in Stock and Watson (1993), albeit in a time series context with non–stationary variables. There, the DGP is

[1] We use the term independence very freely here, sometimes in the sense of uncorrelatedness, sometimes as the conditional mean condition.

assumed to be $Y_t = X_t\beta + \varepsilon_{1t}$, with $X_t = X_{t-1} + \varepsilon_{2t}$ where ε_{1t} and ε_{2t} can be mutually and serially correlated. Evidently, least squares regression of Y_t on X_t will suffer from bias arising from correlation in the two error terms. Stock and Watson suggest augmenting the regression of Y on X by leads and lags of ε_{2t}. This has the effect of orthogonalizing ε_{1t}, so that the error term of the augmented regression is essentially $\varepsilon_{1t} - E[\varepsilon_{1t}|\varepsilon_{2t}]$, which is by construction orthogonal to X_t.

The independence assumption can also be violated in non-linear models. Consider the parametric non-linear regression model

$$Y_t = g(X_t, \beta) + \varepsilon_t, \tag{2}$$

where $E[\varepsilon|X] \neq 0$. Suppose that in addition to the conditional mean of Y, one is also interested in the parameter β. When $g(X, \beta)$ has a known parametric form, a possible solution is Amemiya's (1974) non-linear two-stage least squares estimate (N2SLS). The NL2SLS is really a non-linear IV estimate since it minimizes the objective function $(Y - g(X,\beta))^T P_Z (Y - g(X,\beta))$, in analogy to the linear IV case. The estimate is $\sqrt{n}$-consistent and asymptotically normal under standard conditions. It is noteworthy that a two-step procedure which first forms a projection of X onto Z and then uses the fitted values in place of X in $g(X,\beta)$, will in general not be consistent. This is because the projection only ensures that $P_Z X$ is orthogonal to ε, but not that $g(P_Z X, \beta)$ is orthogonal to ε [See e.g. Davidson and Mackinnon (1993), Section 7.6].

In this paper, we are concerned with cases in which $E[\varepsilon|X] \neq 0$ and, in which in addition, $g(X)$ takes an unknown form and is to be estimated non-parametrically. Non-parametric regressions are becoming increasingly popular in economic analysis as they are very powerful at identifying interesting economic relationships without imposing distributional and functional assumptions *a priori*. However, problems of simultaneity bias and errors-in-variables frequently arise in economic applications. There is a need to deal with these problems in non-parametric regressions just like in parametric models. For example, if one is interested in a possibly non-linear Engle curve, the consumption function cannot be estimated non-parametrically without imposing the seemingly false assumption that innovations in consumption and income are independent. A consistent, non-parametric estimate of the function g is important because from it, we can derive quantities of economic interest such as the marginal propensity to consume. As well,

measurement errors pose an insurmountable problem for practitioners who work with panel data. The methodology provided here is a step towards resolving these problems in a non–parametric framework.

The estimate we propose is in the spirit of but not identical to the 2SLS estimate of the linear regression framework, and it also differs from the IV approach in ways to be made precise. It can also be seen as a non–parametric extension of the DOLS estimate outlined earlier. Section 2 presents the proposed estimate and discusses related work in the literature. The statistical properties of the estimate are given in Section 3, and the empirical properties of the estimate are shown in Section 4. The proofs of the results are provided in a technical appendix.

2 A Two Stage Non–parametric Estimate (2SNP)

Consider non–parametric estimation of the following model.

$$Y_t = g(X_t) + \varepsilon_t, \tag{3}$$

where g is an unknown continuous function and $\{\varepsilon_t\}$ is a sequence of random variables. Under the standard condition that $E[\varepsilon|X] = 0$, a consistent estimate of g can be obtained under fairly general conditions [cf. e.g. Andrews (1991a) or Robinson (1983)]. Now suppose $E[\varepsilon|X] \neq 0$. Our objective is to devise a method to estimate g consistently.

Suppose X can be expressed as the sum of a predetermined and an unpredictable component:

$$X_t = h(Z_t) + \eta_t, \tag{4}$$

where h is an unknown continuous function, Z_t a vector of predetermined or exogenous variables, and $\{\eta_t\}$ a sequence of random variables satisfying $E[\eta_t|Z_t] = 0$. In addition we assume that $E[\varepsilon_t|Z_t, \eta_t] = E[\varepsilon_t|\eta_t]$, which is implied by Z_t and ε_t being independent. It follows from this last assumption that $E[\varepsilon_t|X_t] \neq 0$ arises when $E[\varepsilon_t|\eta_t] \neq 0$. We wish to exploit (4) to obtain a consistent estimate of g. We focus on the simple case in which Y_t is one–dimensional.

The function h relates the regressor X to the variable Z. The latter can be viewed as an "instrument" in a classical sense. Step one of our approach is to obtain a consistent, non–parametric estimate of h, say $\hat{h}$, and a set of residuals $\{\hat{\eta}_t\}$ that converge to their

corresponding elements in $\{\eta_t\}$. This can be accomplished by nonparametric estimation of (4). For kernel estimates, the rate of convergence of $\hat{h}$ is $n^{2/5}$. For series estimates, the rate of convergence depends on additional conditions such as the smoothness of h and the order of truncation of the series [see Robinson (1983) for conditions for kernel regressions and Andrews (1991a) for conditions on series regressions.].

To motivate step two of our estimation procedure, define

$$m(X_t, \eta_t) = E[Y_t|X_t, \eta_t] = g(X_t) + E[\varepsilon_t|\eta_t] = g(X_t) + w(\eta_t), \tag{5}$$

where $w(\eta_t) = E[\varepsilon_t|\eta_t]$. Suppose for the time being that η is known, and consider a series expansion of the functionals $g(x)$ and $w(\eta)$ in terms of their orthonormal basis functions $\{e_{Xl}\}$ and $\{e_{\eta l}\}$. We have

$$\begin{aligned} g(x) &= \sum_{l=0}^{\infty} \alpha_{Xl} e_{Xl}(x), \\ w(\eta) &= \sum_{l=0}^{\infty} \alpha_{\eta l} e_{\eta l}(\eta), \\ m(x, \eta) &= \sum_{l=0}^{\infty} \alpha_{Xl} e_{Xl}(x) + \sum_{l=0}^{\infty} \alpha_{\eta l} e_{\eta l}(\eta). \end{aligned} \tag{6}$$

The series expansion representation of $m(x, \eta)$ given by (6) implies that an estimate of m can be obtained if we can estimate the unknown α's. Step two of our approach is therefore to obtain the $\hat{\alpha}$'s by series regression. For the series expansions to be approximations of the respective functionals, it is important that $\{e_{Xj}\}$ and $\{e_{\eta j}\}$ indeed form a basis for the support of the densities of X_t and η_t. A popular choice is the Fourier expansion, in which $e_{Xj}(x) = (2\pi)^{-\frac{1}{2}} \cos\{(j+1)x/2\}$ for j odd, and $e_{Xj}(x) = (2\pi)^{-\frac{1}{2}} \sin\{jx/2\}$ for j even, with $e_{X0}(x) = (2\pi)^{-1}$. This gives an orthonormal basis for the class of continuous functions on $[-\pi, \pi]$. Alternatively, we could use Legendre polynomials which constitute an orthogonal system of functions on [-1,1]. The j^{th} term of the expansion around x can be constructed from the recursion

$$(j+1)e_{X,j+1}(x) = (2j+1) \cdot x \cdot e_{Xj}(x) - j \cdot e_{X,j-1}(x) \tag{7}$$

with $e_{X0}(x) = 1/\sqrt{2}$ and $e_{X1}(x) = x/\sqrt{2/3}$.

Once the orthogonal polynomials are formed, we have the regressors for the series regression (since we have assumed η is known). It is obviously impossible to estimate

an infinite number of unknown parameters on the basis of a finite sample. Hence, the number of terms to be estimated is limited to L_n, a practitioner–chosen number that should increase with the sample size. We consider predetermined values of L_n here, but the analysis could be extended, at the cost of lenghtier proofs, to data–dependent methods for selecting L_n [cf. e.g. Andrews (1991b)].

Since the series expansions in X and η each contains a constant term, we could only identify and estimate $\alpha_{\eta 0} e_{\eta 0}(\cdot) + \alpha_{X0} e_{X0}(\cdot)$, but not $\alpha_{\eta 0}$ and α_{X0} separately. However, note that $E[w(\eta_t)] = E\{E[\varepsilon_t|\eta_t]\} = E[\varepsilon_t] = 0$ by the law of iterated expectations. Hence $\alpha_{\eta 0} e_{\eta 0} + \sum_{l=1}^{\infty} \alpha_{\eta l} E[e_{\eta l}(\eta_t)] = 0$. Since $e_{\eta 0}(\cdot)$ and $e_{X0}(\cdot)$ are (known and) assumed constant, $\alpha_{\eta 0}$ and α_{X0} can be estimated using the following relationships.

$$\hat{\alpha}_{\eta 0} e_{\eta 0}(\cdot) = -\sum_{l=1}^{L_n} \hat{\alpha}_{\eta l} \frac{1}{n} \sum_{t=1}^{n} e_{\eta l}(\eta_t), \tag{8}$$

$$\hat{\alpha}_{X0} e_{X0}(\cdot) = \hat{\alpha}_0 - \hat{\alpha}_{\eta 0}, \tag{9}$$

where $\hat{\alpha}_0$ is the estimate of $\alpha_0 = \alpha_{\eta 0} e_{\eta 0}(\cdot) + \alpha_{X0} e_{X0}(\cdot)$.

We have thus far discussed how an estimate of $m(x, \eta)$ can be obtained. Of interest is not the estimate $m(x, \eta)$ per sé, but $g(x)$. However, once the α's are estimated, $\hat{g}(x)$ can be determined according to

$$\hat{g}(x) = \sum_{l=0}^{L_n} \hat{\alpha}_{Xl} e_{Xl}(x), \tag{10}$$

at all y. If estimation of w is of economic interest, it could be similarly evaluated as $\hat{w}(\eta) = \sum_{l=0}^{L_n} \hat{\alpha}_{\eta l} e_{\eta l}(\eta)$, at all η.

We have so far pretended that we could observe the η_t's. This is rarely going to be the case in practice and this fact is precisely the motivation for step 1. It allows us to substitute $\hat{\eta}_t$ for η_t. Because the η's are not observed, the estimation of m is non-standard. The $\hat{\eta}$'s in the second step regression are "generated regressors" in the sense of Pagan (1984). In practice, the second step regression takes the form

$$Y_t = \alpha_0 + \sum_{l=1}^{L_n} \alpha_{Xl} e_{Xl}(X_t) + \sum_{l=1}^{L_n} \alpha_{\eta l} e_{\eta l}(\hat{\eta}_t) + \nu_t^*. \tag{11}$$

As we shall state in Theorem 1 , the sample mean square error of $\hat{m}$ converges in probability to zero in spite of using $\hat{\eta}$ instead of η. Accordingly, the sample mean square

error of $\hat{g}$ also converges in probability to zero. Given consistent estimates of the $\hat{\alpha}$'s, derivatives of $\hat{g}(X)$ can also be derived.

Like in linear models, there are situations in which g cannot identified in our model. One example is when h is constant, in which case, (3) reduces to $Y_t = g(\eta_t) + \varepsilon_t$, where we have assumed (without loss of generality) that $h(z) = 0$ for all z. Unless additional instruments, say $\{W_t\}$, can be found such that $E[Y_t|W_t] = E[g(\eta_t)|W_t]$, g cannot be identified. Generally, if h straddles a continuum of function values across the support of the density of Z_t, identification is generally not an issue.

2.1 Comparison with an IV estimate

It is of interest to compare the two-stage approach here with an IV estimate in a non-parametric regression context. The latter was considered by Newey and Powell (1990), who considered estimating the unknown regression function g defined as in (3) given a set of instruments Z, where $E[\varepsilon|Z] = 0$. Their identification problem is to obtain $g(x)$ from an estimate of $\Pi(Z) = E[g(X)|Z] = \int g(x)f(y|Z)dy$ and $f(y|Z)$, where $f(y|Z)$ is the conditional density of X given Z at y, and noting that $f(y|Z)$ and $\Pi(Z)$ are both functions to be estimated. A related deconvolution problem was considered by Horowitz and Markatou (1995) in the context of errors-in-variables.

Existence of the solution in Newey and Powell's problem requires that $f(y|Z)$ is complete, which can be seen as a generalization of the rank condition in parametric analysis. Assuming this condition holds, Newey and Powell propose a two-stage estimate. As in our analysis, they also use a series expansion $g(x) = \sum_{l=0}^{L_n} \alpha_l e_l(x)$, where the e_l are the basis functions. The first stage of their procedure is to estimate the conditional expectations $E[e_l(X)|Z]$. In the second stage, a series regression of Y on $E[e_l(X)|Z]$ gives $\hat{\alpha}$. The function g is then estimated by $\sum_{l=0}^{L_n} \hat{\alpha}_l e_l(X)$. As in our analysis, consistency of the estimate requires that $L_n \to \infty$, as $n \to \infty$.

Comparing the two methods, Newey and Powell's is an IV approach which works directly with the equation $Y = g(X) + \varepsilon$, at the cost of imposing stronger conditions to ensure the existence of $E[e_l(X)|Z]$. By contrast, our's is in the spirit of 2SLS in the sense that we postulate an equation that relates X and Z. Our analysis differs from 2SLS in that X and Z have a non-parametric relationship. More importantly, we do not use the fitted values of X in estimating g. As mentioned earlier, doing so would not have

yielded consistent estimates of β in a parametric context, and a similar problem arises in a non-parametric context. We make use of the auxiliary regression in the first step to eliminate $E[\varepsilon|X] = E[\varepsilon|\eta]$ so that g can be estimated consistently. Thus, although 2SLS is a special case of the IV estimator in a classical linear setting, our 2SNP estimator is not a special case of Newey and Powell's non-parametric IV estimator in this non-linear, non-parametric setting.

3 Properties of the Two-Step Non-Parametric Estimate

The foregoing discussion suggests that we can write (3) as

$$Y_t = g(X_t) + w(\eta_t) + \nu_t, \tag{12}$$

where $\nu_t = \varepsilon_t - E[\varepsilon_t|\eta_t]$. Applying the series expansions yields

$$Y_t = \sum_{l=0}^{L_n} \alpha_{Xl} e_{Xl}(X_t) + \sum_{l=0}^{L_n} \alpha_{\eta l} e_{\eta l}(\eta_t) + \nu_t + R_t, \tag{13}$$

where $R_t = \sum_{l=L_n+1}^{\infty} \{\alpha_{Xl} e_{Xl}(X_t) + \alpha_{\eta l} e_{\eta l}(\eta_t)\}$. Let $U_t = [X_t, \eta_t]^T$ and $\hat{U}_t = [X_t, \hat{\eta}_t]^T$. Then, for implicitly defined A_U and α_U,

$$Y = A_U \alpha_U + \nu + R. \tag{14}$$

When L_n increases, terms are dropped from each of the R_t's. As $L_n \to \infty$, each of the $|R_t|$'s converges to zero. If $\{\eta\}$ were known, the problem is standard and least squares estimation of (14) gives:

$$\hat{\alpha}_U = (A_U^T A_U)^{-1} A_U^T Y. \tag{15}$$

Under the conditions stated in Andrews (1991a), a non-parametric series estimate of m based upon $\hat{\alpha}_U$, denoted by $\hat{m}_U$, can be obtained.

The non-standard nature of our problem is that in general, we do not observe η and have to replace it by $\hat{\eta}$, leading to $\hat{\alpha}_{\hat{U}} = (A_{\hat{U}}^T A_{\hat{U}})^{-1} A_{\hat{U}}^T Y$, and $\hat{m}_{\hat{U}}$. The properties of $\hat{m}_{\hat{U}}$ therefore depend on $\hat{\alpha}_{\hat{U}}$ instead of $\hat{\alpha}_U$.

We now state the assumptions leading to the main result.

Assumption A

1. *$\{X_t\}$, $\{Z_t\}$, $\{\eta_t\}$ and $\{\varepsilon_t\}$ are sequences of continuous random variables; the ε_t's are scalars, the others have densities with bounded support and the supports are identical across t. Furthermore, h, g and w are continuous functions on the supports of Z_t, X_t and η_t, respectively. Finally, the range of h contains a d_X-dimensional interval of positive volume, where d_X is the dimension of X.*

2. *$E[\nu_t|\eta_t, X_t] = 0$ for all t.*

3. *(a) $V[\nu|\eta, X] = \Omega$, where Ω is a diagonal matrix with elements σ_t^2 for which $\sup_t |\sigma_t^2| < \infty$, or*

   (b) $V[\nu|X, \eta] = \Omega$, where Ω is a symmetric matrix whose rows are absolutely summable, uniformly in X, η, even in the limit.

4. *All $\alpha_{Xl}, \alpha_{\eta l}$'s are bounded, and there exists some number K such that for sufficiently large l, $|\alpha_{Xl}| \leq Kl^{-\delta_X}$ and $|\alpha_{\eta l}| \leq Kl^{-\delta_\eta}$, for some $\delta_X > 1$, $\delta_\eta > 2$.*

5. $\sup_{\eta,\tilde{\eta},x,\tilde{x}}\{|e_{\eta l}(\eta) - e_{\eta l}(\tilde{\eta})| + |e_{Xl}(x) - e_{Xl}(\tilde{x})|\} \leq Kl||\eta - \tilde{\eta}|| + Kl||x - \tilde{x}||$.

6. *For some $\omega > 0$, $\sum_{t=1}^n ||\hat{\eta}_t - \eta_t||^2 = O_p(n^{1-2\omega})$, as $n \to \infty$.*

7. *$L_n/n \to 0$ and $L_n \to \infty$ as $n \to \infty$.*

Condition (1) assumes boundedness and continuity. The boundedness assumption is imposed in order to allow the use of series estimation. If boundedness were unachievable, one could apply a suitable transformation to reduce the support of the relevant densities to a compact interval. The second part of Condition (1) is imposed to ensure identification of g and w. It rules out cases when h is constant, since g and w would then both be functions of η ruling out identification. Condition (2) validates Z_t as an instrument. Conditions (1) and (2) imply continuity of Y_t.

Condition (3a) applies when the errors are (conditionally) uncorrelated and with $\sup_t V[\nu_t|X_t, \eta_t] < \infty$. Part (b) applies when the errors are correlated. It essentially states that the covariance between elements in the ν-vector, conditional upon X and η, should decline as these elements are farther apart in time and that this decline should be suitably fast. It is a summability condition on the covariances that is weaker than covariance stationarity. In effect, it is a "mixing" condition [cf. e.g. Rosenblatt (1956) and Ibragimov and Linnik (1971)].

Condition (4) is standard in the literature, albeit often in a slightly different guise. It is essentially a smoothness condition since the functionals with lower indices tend to be

smoother than those with higher indices [Andrews (1991a)]. It was for instance used by Newey (1995).[2] The smoothness conditions on w are stronger than those on g, because we need to solve the nuisance parameter problem in η.

Condition (5) imposes a Lipschitz-condition on the functionals. For Fourier and Legendre series, the condition is implied by differentiability of $e_{\eta l}$ and e_{Xl}, since the lengths of the vector of first derivatives is bounded by $Kl/2$ for sufficiently large K.

Condition (6) determines the rate of convergence of the estimate of h. If nonparametric estimation is used, $\omega = 2/5$ is a common rate for twice differentiable one-dimensional functions.

Finally, Condition (7) is standard and requires that the number of elements in the series expansion increases as the number of observations increases ad infinitum, but at a slower rate. This ensures that increasing the number of terms in the series expansion does not induce excess variability in the estimate.

We now state the first of our main results.

Theorem 1 *Under the assumption,*

$$n^{-1}\sum_{t=1}^{n}\{\hat{m}(X_t,\eta_t) - m(X_t,\eta_t)\}^2 = O_p(n^{-2\omega} + L_n^{2-2\min\{\delta_X,\delta_\eta\}} + n^{-1}L_n) = o_p(1). \quad (16)$$

Our proof is based on a small number of simple projection results and is similar to that of Andrews and Whang (1990) among others and is given in the Appendix. Theorem 1 says that the sample mean square error converges to zero as $n \to \infty$. This is not the same as a uniform convergence result or indeed a pointwise convergence result. There may be points at which $\hat{m}_{\hat{\eta}}$ does not converge to m, but the fraction of values (X_t, η_t) at which convergence does not occur drops to zero as the sample size tends to infinity. This result is similar to that obtained by Andrews and Whang (1990) and Newey (1995).

The optimal *rate* at which to let L_n increase as $n \to \infty$ is $n^{1/(2\min\{\delta_X,\delta_\eta\}-1)}$, which gives a sample mean square error of $n^{-2\omega} + n^{1/(2\min\{\delta_X,\delta_\eta\}-1)-1}$.[3] If the first step were absent, the rate would have been $n^{1/(2\min\{\delta_X,\delta_\eta\}-1)-1}$. Hence the rate of convergence is essentially

[2] His α-parameter is $\min\{\delta_X, \delta_\eta\} - 1$ in our's.

[3] The optimal rate is determined in the following way. Consider the middle expression in (16). The first term does not depend upon the rate at which L_n increases, but the second and third rates do. Indeed, the faster L_n increases, the greater the second rate and the less the third rate. Hence, the optimal rate is the rate at which the second and third rates are equal, giving the rate stated.

the slowest of that of $\hat{h}$ and that of standard series estimation of m. In Eastwood and Gallant (1991), the parameters δ_X, δ_η must exceed 2.5 for twice differentiable functions, resulting in a convergence rate of the (empirical) mean square error that is the same as our's. If both parameters are equal to 3, we would get a convergence rate of $n^{-\omega} + n^{-2/5}$ for the estimate $\hat{m}$ itself.[4] In a standard nonparametric setting where $\omega = 2/5$, the rate of convergence of $\hat{m}$ is then not affected by the first step. The first step reduces the convergence rate only if m is very smooth and h is not.

We now develop a uniform convergence result. In view of the argument above, uniform convergence is likely to take place at a slower rate than the square root of the mean square error and under stronger assumptions. Indeed, Newey (1995) also requires stronger conditions to prove uniform convergence in his analysis. The following assumptions are sufficient for uniform convergence of our two–step estimate.

Assumption B *Let* $\hat{\zeta}_n^{-1} = \lambda_{\max}(A_{\hat{U}}^T A_{\hat{U}})^{-1}$, *where* $\lambda_{\max}$ *denotes "the maximum eigenvalue of" satisfying*

1. $nL_n^{3-2\min\{\delta_\eta,\delta_X\}}\hat{\zeta}_n^{-1} = o_p(1)$,
2. $L_n^2\hat{\zeta}_n^{-1} = o_p(1)$,
3. $n^{1-2\omega}L_n\hat{\zeta}_n^{-1} = o_p(1)$.

The assumption is technical. In a standard linear regression framework where the number of regressors is fixed, the equivalent of $\lambda_{max}(A_U^T A_U)^{-1}$ is $O_p(n^{-1})$. Because the number of regressors increases with the sample size when series regression is used, the maximum eigenvalue may decline at a slower rate. It is therefore common practice to assume an upper bound on $\lambda_{\max}(A_U^T A_U)^{-1}$, or a lower bound on $\lambda_{\min}(A_U^T A_U)$. Since we do not observe η, Assumption B imposes the condition on $\lambda_{max}(A_{\hat{U}}^T A_{\hat{U}})$ instead. This leads to the following result.

Theorem 2 *If Assumptions A and B hold, then*

$$\begin{aligned}&\sup_{x,\eta}|\hat{m}_{\hat{U}}(x,\eta) - m(x,\eta)| \\ &= O_p(\hat{\zeta}_n^{-\frac{1}{2}}\{n^{\frac{1}{2}}L_n^{\frac{3}{2}-\min\{\delta_\eta,\delta_X\}} + L_n + n^{\frac{1}{2}-\omega}L_n^{\frac{1}{2}}\} + L_n^{1-\min\{\delta_\eta,\delta_X\}}) = o_p(1). \qquad (17)\end{aligned}$$

[4]The convergence rate for the estimate itself is the square root of that of the mean square error.

Although imposing a condition on the maximum eigenvalue of $(A_{\hat{U}}^T A_{\hat{U}})^{-1}$ is the natural adaptation to the current setting, a condition on $(A_{\hat{U}}^T A_{\hat{U}})^{-1}$ involves a restriction on the $\hat{\eta}_t$'s. Since the properties of the $\hat{\eta}_t$'s depend in part upon the relationship between X_t and Z_t, and in part on the estimation method used for h, it is thus more desirable to obtain a result that only involves conditions on the η_t's. However, it is not immediately apparent that Theorem 2 also holds when the eigenvalue condition is imposed on $(A_U^T A_U)^{-1}$. Obtaining such a result is problematic because we have not specified the nature of the relationship between $\hat{\eta}$ and η. However, the next theorem shows that the same uniform convergence result obtains if we impose conditions on the maximum eigenvalue of $(A_U^T A_U)^{-1}$ instead.

Theorem 3 *If, in addition to the previous assumptions,* $n^{2-2\omega} L_n^3 \zeta_n^{-2} = o(1)$, *where* $\zeta_n = \lambda_{\max}(A_U^T A_U)^{-1}$, *then Theorem 2 still holds when* $\hat{\zeta}_n$ *is replaced by* ζ_n.

4 Simulations

In this section, we examine the empirical properties of the two-step non-parametric estimate. For the sake of comparison, the sample mean square error of $\hat{g}(X)$ obtained from the 2SNP is compared to i) a non-parametric polynomial regression of Y on X, and ii) a two-step non-parametric polynomial regression of Y on $\hat{X}$, where $\hat{X}$ are the fitted values from a polynomial regression of X on Z. The former estimate (labelled NP1) is inconsistent if the regression error is not independent of the regressors. The latter (labelled NP2) is in the spirit of the two-stage least squares estimate because it uses the fitted values from the first stage in the second step regression.

The DGP we consider is specified as follows.

$$\begin{aligned} X_t &= \exp(Z_t) + \eta_t, \qquad \eta_t = .1w_t + .2v_{Xt} \\ Y_t &= g(X_t) + \varepsilon_t = -3 + 10\log(X_t) + \varepsilon_t; \qquad \varepsilon_t = .2w_t + .3v_{Yt} \\ Z_t &= 1 + .6v_{Zt} \\ & v_{Zt} \sim \text{i.i.d.}U(0,1); \qquad v_{Yt}, v_{Xt}, w_t \sim \text{i.i.d.}N(0,1). \end{aligned} \tag{18}$$

To give some economic interpretation to the DGP, consider consumption, Y, as a function of earnings, X.[5] The latter can be predicted by years of education, Z, but has

[5]The choice of the parameters is arbitrary. Relating the DGP to the consumption function is an

an idiosyncratic component which is correlated with innovations in Y through w. As specified, $g(X)$ is such that consumption is increasing in X, but at a decreasing rate. The sample means of X and $g(X)$ are 3.3 and 2.8 respectively. The sample correlation between X and ε is about .1. For this parametrization, the variations in $\exp(Z)$ account for over 90% of the variation in X. Thus, Z is a good instrument judged in terms of its predictability for X.

The results reported in Tables 1 and 2 are based on 1000 simulations using Gauss 3.21 running on a 66 mhz PC. We consider different sample sizes, and results for $n = 200$ and 500 are reported without loss of generality. The regressions are based on Legendre polynomials, which seem to perform somewhat better than Fourier series. For the NP1 regressions based on X, the order of the polynomial expansion is denoted L_{n2}. The NP2 and 2SNP estimates are both two–step procedures. In theory, there is no constraint on which non–parametric regression method to use in the first step provided it is consistent. Here, we use orthogonal polynomial regressions for convenience. It is possible that under– or over–smoothing is necessary in the first stage to obtain the minimum mean square error in the second stage. We therefore allow the order of the polynomial in the first stage, L_{n1}, to differ from that in the second stage, L_{n2}. For 2SNP, the mean square error is based on the differences $\hat{g}(X_t) - g(X_t)$, with $\hat{g}(X_t)$ evaluated according to (10). For NP1 and NP2, the mean square error for $\hat{g}$ is based on $\hat{Y}_t - g(X_t)$.

Table 1: Sample Mean Square Error for $\hat{g}(X)$ with $n = 200$

	L_{n2}	2	4	6	8
	NP1	0.00799	0.00404	0.00533	0.00658
L_{n1}	NP2				
2		0.07551	0.07602	0.07658	0.07715
4		0.07819	0.07601	0.07663	0.07728
6		0.07747	0.07527	0.07651	0.07722
6		0.07677	0.07378	0.07499	0.07610
L_{n1}	2SNP				
2		0.00717	0.00290	0.00486	0.01151
4		0.00714	0.00267	0.00455	0.01176
6		0.00714	0.00266	0.00451	0.01121
8		0.00713	0.00266	0.00445	0.01096

attempt to help readers relate the proposed estimate to relevant applications.

Table 2: Sample Mean Square Error for $\hat{g}(X)$ with $n = 500$

	L_{n2}	2	4	6	8
	NP1	0.00727	0.00253	0.00302	0.00352
L_{n1}	NP2				
2		0.07569	0.07588	0.07607	0.07627
4		0.07884	0.07588	0.07610	0.07632
6		0.07853	0.07556	0.07605	0.07631
8		0.07822	0.07524	0.07574	0.07621
L_{n1}	2SNP				
2		0.00645	0.00135	0.00201	0.00340
4		0.00645	0.00112	0.00174	0.00353
6		0.00645	0.00112	0.00175	0.00349
8		0.00645	0.00112	0.00174	0.00346

For benchmark purposes, the mean square error from OLS are .0079 and .00727 for n=200 and 500 respectively. The results quite convincingly reveal that the 2SNP estimate yields the smallest mean square error when L_{n1} and L_{n2} are chosen appropriately. The reason for this is intuitive if we note that η contains two terms, w and v_X, but only one of which violates the independence assumption between X and ϵ. We refer to w as the bad information. Accordingly, v_X is the good information in X from a regression standpoint.

Consider first the NP2 estimate. It uses the fitted values from a polynomial regression of X on Z to form the basis functions in the second step. Since Z contains no information on η, this auxiliary regression ignores not just the bad information w, but also abandons the good information, v_X. The omission of information in v_X in the first stage evidently affects the performance of the IV estimate in the second stage. The NP1 estimate includes the bad information, w, in X when fitting Y. It can be seen that the properties of the estimate are affected even when there is little correlation between X and ϵ (0.1 in this example). The basis functions used for the 2SNP estimate is based upon X and therefore contains all available information. However, the estimate filters out bad information through the inclusion of polynomials in $\hat{\eta}$ in the second step regression. It therefore produces the smallest mean square error.

A good deal of attention has recently been paid to the result that the IV estimates in classical linear models are biased towards the OLS estimates when the instruments are weak [see Nelson and Startz (1990)]. Our simulations reveal a different problem with the use of instrumental variables in a non-linear and non-parametric setting. As seen

from the results above, the NP2 (non-parametric IV) estimates have a large mean square error. The problem is that we are estimating $g(h(Z))$, and nothing separates g from h per se. Thus, although the instrument Z has good explanatory power for X in the above simulations, it is not desirable to estimate the unknown regression function by using Z to instrument out X. In the 2SNP estimation procedure, the instruments do not play a direct role in the equation that determines g. This has the advantage that the estimates will be less susceptible to the problem of weak instruments. In addition, identification of g is immediate since it is $g(X)$ itself that appears in the second step regression.

A comparison of the results in Tables 1 and 2 shows that the mean square error of $\hat{g}$ from the 2SNP procedure falls rapidly as the sample size increases. For a given sample size, the estimate also appears to require a higher level of smoothing in the first stage than in the second stage. While the choice of L_n in the two steps inevitably depends on the DGP considered, we have not found a systematic method for choosing L_{n1} and L_{n2} to minimize the mean square error in the second step. This remains an issue for future research.

5 Conclusion

In this paper, we consider the non-parametric estimation of the function g in the model $Y_t = g(X_t) + \varepsilon_t$ when the assumption $E[\varepsilon|X] = 0$ is violated. The method consists of two steps. First, given an appropriate instrument Z, we estimate the relationship between X and Z non-parametrically and save the residuals $\hat{\eta}$. Second, we form an orthonormal basis for the classes of continuous functions on the supports of the densities of X_t and η_t, respectively and perform a series regression of Y_t on X_t and $\hat{\eta}_t$. An estimate of g can then be immediately determined from $\hat{m}_{\hat{Y}}$.

The 2SNP estimate discussed here extends naturally to partially linear models where Y_t is linear in all but one variable.

$$Y_t = g(X_t) + W_t^T\beta + \varepsilon_t, \tag{19}$$

where W is a $n \times k$ matrix, and β is a k-vector of parameters. As in Section 2, we first obtain the residuals from $X_t = h(Z_t) + \eta_t$, and then determine the basis functions for X_t and η_t. If $E[\varepsilon|W] \neq 0$, then one would need to instrument out W also. Once this is accomplished, a series regression can be used to obtain a consistent estimate of g. The

only difference in the proof is that the matrix of regressors is now augmented by a matrix W of dimension $n \times k$ which is independent of the sample size, n.

We have so far concentrated on the simple case where Y is one-dimensional. More demanding is the case when the range of g is greater than one. On this issue, more work needs to be done.

A Technical Appendix

We begin with a series of useful Lemmas.

Lemma 1 $||(A_{\hat{U}} - A_U)\alpha_U||^2 = O_p(n^{1-2\omega})$.

Proof: The left hand side is $||(A_{\hat{\eta}} - A_\eta)\alpha_\eta||^2$, which is. $\sum_{l,m=1}^{L_n}\sum_{t=1}^{n}\alpha_{\eta l}\alpha_{\eta m}\{e_{\eta l}(\hat{\eta}_t) - e_{\eta l}(\eta_t)\}\{e_{\eta m}(\hat{\eta}_t) - e_{\eta m}(\eta_t)\}$. We can bound the last expression in absolute value by $\sum_{l,m=1}^{L_n}\sum_{t=1}^{n}|\alpha_{\eta l}\alpha_{\eta m}|K^2 lm||\hat{\eta}_t - \eta_t||^2$, by assumption. Hence we get $(K\sum_{l=1}^{L_n}|\alpha_{\eta l}l|)^2\sum_t||\hat{\eta}_t - \eta_t||^2$. The first factor is bounded, by assumption, and the second is $O_p(n^{1-2\omega})$, again by assumption. **Q.E.D.**

Lemma 2 $||P_{\hat{U}}\nu||^2 = O_p(L_n)$.

Proof: The expectation of the left hand side condition upon $\hat{U}$ is $E[\nu^T P_{\hat{U}}\nu|\hat{U}] = \mathrm{tr}(P_{\hat{U}}E[\nu\nu^T|\hat{U}]) = \mathrm{tr}(P_{\hat{U}}E[\nu\nu^T|Z,\eta]) = \mathrm{tr}(P_{\hat{U}}\Omega) \leq L_n \sup_i \sigma_i^2 = O_p(L_n)$, by assumption. **Q.E.D.**

To re–prove Lemma 2 under Assumption (3b), it suffices to establish that $\mathrm{tr}(P_{\hat{U}}V[\nu|X,\eta]P_{\hat{U}}) = \mathrm{tr}(P_{\hat{U}}\Omega P_{\hat{U}}) = O_p(L_n)$. **Proof:** Let the elements of the matrix $P_{\hat{U}}$ be denoted by $p_{ij} = p_{ji}$ and those of Ω by $\sigma_{ij} = \sigma_{ji}$. Then $\mathrm{tr}(P_{\hat{U}}\Omega P_{\hat{U}}) = \sum_{t,s,l=1}^{n} p_{sl}p_{ts}\sigma_{lt}$, which is bounded by $\sum_{t,s,l=1}^{n}|\sigma_{lt}|\{p_{sl}^2 + p_{ts}^2\}$. Now, $\sum_{t,s,l=1}^{n}|\sigma_{lt}|p_{sl}^2 = \sum_{s,l=1}^{n}(\sum_{t=1}^{n}|\sigma_{lt}|)p_{sl}^2 \leq \sum_{s,l=1}^{n}p_{sl}^2 = \sum_{s=1}^{n}p_{ss} = \mathrm{tr}(P_{\hat{U}}) = L_n$, where the inequality follows from the summability condition on the covariances and the subsequent equality from the fact that $P_{\hat{U}} = P_{\hat{U}}P_{\hat{U}}^T$.

Lemma 3 $||R||^2 = O_p(nL_n^{2-2\min\{\delta_\eta,\delta_X\}})$.

Proof: The left hand side is $\sum_{t=1}^{n}[\sum_{l=L_n+1}^{\infty}\{\alpha_{\eta l}e_{\eta l}(\eta_t) + \alpha_{Xl}e_{Xl}(X_t)\}]^2$, which is bounded by $nK^2(\sum_{l=L_n+1}^{\infty}|\alpha_{\eta l}| + \sum_{l=L_n+1}^{\infty}|\alpha_{Xl}|)^2 \leq nK^4[\sum_{l=L_n+1}^{\infty}(l^{-\delta_\eta} + l^{-\delta_X})]^2 = O(nL_n^{2-2\min\{\delta_\eta,\delta_X\}})$. **Q.E.D.**

Proof of Theorem 1

Let m be the vector containing the $m(U_t)$'s, with $U_t = [X_t, \eta_t]^T$, and $\hat{m}_{\hat{U}}$ without argument be the (vector) estimate of m. For the purpose of simplicity of notation we eliminate the constant in the expansion of w but retain the constant in the expansion on g, rather than including a separate constant term. Then the model is

$Y = m + \nu = A_U\alpha_U + R + \nu$. Hence, $\hat{m}_{\hat{U}}$ is the projection of Y onto the columns of $A_{\hat{U}}$, i.e. $\hat{m} = P_{\hat{U}}Y$, with $P_{\hat{U}} = A_{\hat{U}}(A_{\hat{U}}^TA_{\hat{U}})^{-1}A_{\hat{U}}^T$, provided the inverse exists. If not, one should remove the offending rows columns until it does [cf. Andrews (1991a)]. Now, $\hat{m}_{\hat{U}} - m = M_{\hat{U}}(A_{\hat{U}} - A_U)\alpha_U - M_{\hat{U}}R + P_{\hat{U}}\nu$, where $M_{\hat{U}} = I_n - P_{\hat{U}}$. Because $M_{\hat{U}}$ is a projection matrix, $||M_{\hat{U}}v||^2 \leq ||v||^2$ for all vectors v. In Lemma 1, we show that $||(A_{\hat{U}} - A_U)\alpha_U||^2 = O_p(n^{1-2\omega})$, and in Lemma 2 that $||P_{\hat{U}}\nu||^2 = O_p(L_n)$. Finally, Lemma 3 shows that $||R||^2 = O_p(nL_n^{2-2\min\{\delta_\eta,\delta_X\}})$. **Q.E.D.**

Proof of Theorem 2

For any fixed x, η, let a and a_R be vectors with functionals evaluated at (x,η): a contains the functionals whose coefficients are estimated, and a_R those that are not. Thus, $m(x,\eta) = a^T\alpha + a_R^T\alpha_R$. Note that $\hat{m}_{\hat{U}}(x,\eta) = a^T(A_{\hat{U}}^TA_{\hat{U}})^{-1}A_{\hat{U}}^TY$ for a suitable vector a whose L_n elements are uniformly bounded. Let $Q_{\hat{U}} = (A_{\hat{U}}^TA_{\hat{U}})^{-1}A_{\hat{U}}^T$ and note that $Y = A_U\alpha_U + R + \nu = A_{\hat{U}}\alpha_U + (A_U - A_{\hat{U}})\alpha_U + R + \nu$. Further, let α_{UR} and a_R be defined such that $m(x,\eta) = a^T\alpha_U + a_R^T\alpha_{UR}$, where the vector α_{UR} contains the coefficients to the functionals beyond the L_n-th. Thus, $\hat{m}_{\hat{U}}(x,\eta) - m(x,\eta) = a^TQ_{\hat{U}}(A_U - A_{\hat{U}})\alpha_U + a^TQ_{\hat{U}}\nu + a^TQ_{\hat{U}}R - a_R^T\alpha_{UR}$. Note that $|a_R^T\alpha_{UR}| = O_p(L_n^{1-\min\{\delta_\eta,\delta_X\}})$, uniformly in a_R (noting that its elements are uniformly bounded) which can be established in a way that is similar to the way in which we proved Lemma 3. Now, since $Q_{\hat{U}}Q_{\hat{U}}^T = (A_{\hat{U}}^TA_{\hat{U}})^{-1}$, $|a^TQ_{\hat{U}}R| \leq \sqrt{||Q_{\hat{U}}^Ta||^2||R||^2} = O_p(L_n^{\frac{3}{2}-\min\{\delta_\eta,\delta_X\}}\hat{\zeta}_n^{-\frac{1}{2}}n^{\frac{1}{2}})$, uniformly in a, where the equality follows from $\hat{\zeta}_n^{-1} = \lambda_{\max}(A_{\hat{U}}^TA_{\hat{U}})^{-1}$, $||R||^2 = O_p(nL_n^{2-2\min\{\delta_\eta,\delta_X\}})$ (Lemma 3), and $||a||^2 \leq KL_n$. In a similar way we can use Lemma 1 to show that $|a^TQ_{\hat{U}}(A_U - A_{\hat{U}})\alpha_U| = O_p(n^{\frac{1}{2}-\omega}\hat{\zeta}_n^{-\frac{1}{2}}L_n^{\frac{1}{2}})$. It thus remains to be shown that $|a^TQ_{\hat{U}}\nu| = O_p(\hat{\zeta}_n^{-\frac{1}{2}}L_n)$. Note that $||a||^2 \leq KL_n$. Consider $\nu^TA_{\hat{U}}(A_{\hat{U}}^TA_{\hat{U}})^{-2}A_{\hat{U}}^T\nu \leq \nu^TP_{\hat{U}}\nu\hat{\zeta}_n^{-1} = O_p(L_n\hat{\zeta}_n^{-1})$, in view of Lemma 2. Hence $|a^TQ_{\hat{U}}\nu| = O_p(L_n\hat{\zeta}_n^{-\frac{1}{2}})$. **Q.E.D.**

Proof of Theorem 3

It suffices to show that $\lambda_{\max}(A_{\hat{U}}^TA_{\hat{U}}/\zeta_n)^{-1} = O_p(1)$. Note that $\lambda_{\max}(A_{\hat{U}}^TA_{\hat{U}}/\zeta_n)^{-1} = \{\lambda_{\min}(A_{\hat{U}}^TA_{\hat{U}}/\zeta_n)\}^{-1}$. Now, $\min_{x\neq 0}x^TA_{\hat{U}}^TA_{\hat{U}}x/(\zeta_nx^Tx) \geq \min_{x\neq 0}x^TA_{\hat{U}}^TA_{\hat{U}}x/(\zeta_nx^Tx) - \max_{x\neq 0}x^T(A_{\hat{U}}^TA_{\hat{U}} - A_{\hat{U}}^TA_{\hat{U}})x/(\zeta_nx^Tx)$. The first term on the right hand side is 1, because

of the definition of ζ_n. Note that $A_{\hat{U}}^T A_{\hat{U}} - A_U^T A_U = (A_{\hat{U}} - A_U)^T(A_{\hat{U}} - A_U) + A_U^T(A_{\hat{U}} - A_U) + (A_{\hat{U}} - A_U)^T A_U$. Because $||A_U x||^2/||x||^2 = O_p(n)$, uniformly in $x \neq 0$ by the uniform boundedness of the elements of A_U, we only need to examine the behavior of $||(A_{\hat{U}} - A_U)x||^2/(\zeta_n^2||x||^2)$, uniformly in $x \neq 0$. The left hand side is $\sum_{l,m=1}^{L_n} x_{\eta l} x_{\eta m} \sum_{t=1}^{n} \{e_{\eta l}(\hat{\eta}_t) - e_{\eta l}(\eta_t)\}\{e_{\eta m}(\hat{\eta}_t) - e_{\eta m}(\eta_t)\}/(\zeta_n^2||x||^2)$, where x_η is the part of x corresponding to the expansion over η. The expression of interest is bounded by $K^2 \sum_{l,m=1}^{L_n} \sum_{t=1}^{n} |x_{\eta l} x_{\eta m}| lm||\hat{\eta}_t - \eta_t||^2/(\zeta_n^2||x||^2) = O_p(n^{1-2\omega}\zeta_n^{-2}) \sum_{l,m=1}^{L_n} |x_{\eta l} x_{\eta m} lm|/||x||^2$, uniformly in $x \neq 0$ by assumption. Now, let $\iota = [1, 2, \ldots, L_n]^T$, such that the afore double sum is $x_\eta^T \iota\iota^T x_\eta/||x||^2$. Since $||x||^2 \geq ||x_\eta||^2$, the expression of interest is thus bounded by $\lambda_{\max}(\iota\iota^T) = ||\iota||^2 = \sum_{l=1}^{L_n} l^2 = O(L_n^3)$. Hence $||(A_{\hat{U}} - A_U)x||^2/(\zeta_n^2||x||^2) = O_p(n^{1-2\omega}\zeta_n^{-2}L_n^3) = o_p(1)$, by assumption. **Q.E.D.**

References

Amemiya, Takeshi. (1974) "The Non-Linear Two-Stage Least Square Estimator", *Journal of Econometrics* 2, 105-110.

Andrews, Donald W.K. (1991a), "Asymptotic Normality of Series Estimators for Nonparametric and Semiparametric Regression Models", *Econometrica* 59, 307–346.

Andrews, Donald W.K. (1991b), "Asymptotic Optimality of Generalized Y_L, Cross–Validation, and Generalized Cross–Validation in Regression with Heteroskedastic Errors", *Journal of Econometrics* 47, 359–377.

Andrews, Donald W.K. and Yoon-Jae Whang (1990), "Additive Interactive Regression Models: Circumvention of the Curse of Dimensionality", *Econometric Theory* 6, 466–479.

Davidson, James (1994), "Stochastic Limit Theory", Oxford University Press.

Davidson, Russell and James G. MacKinnon (1993), "Estimation and Inference in Econometrics", Oxford University Press.

Eastwood, Brian J. and A. Ronald Gallant (1991), "Adaptive Rules for Seminonparametric Estimators that Achieve Asymptotic Normality", *Econometric Theory* 7, 307–340.

A. Ronald Gallant and Geraldo Souza (1991), "On the Asymptotic Normality of Fourier Flexible Form Estimates", *Journal of Econometrics* 50, 329–353.

Horowitz, Joel L. and Marianthi Markatou (1995), "Semiparametric Estimation of Regression Models for Panel Data", mimeo (University of Iowa).

Ibragimov, Ildar A. and Yu V. Linnik (1971), "Independent and Stationary Sequences of Random Variables", Wolters–Noordhoff (Groningen, The Netherlands).

Nelson, Charles R. and Richard Startz (1990), "The Distribution of the Instrumental Variables Estimator and is t-ratio when the Instrument is a Poor one", *Journal of Business* 63, S125-40.

Newey, Whitney K. (1995), "Convergence Rates for Series Estimators", in "Advances in Econometrics and Quantitative Economics", G.S. Maddala, P.C.B. Phillips and T.N. Srinivisan eds., Basil Blackwell (Cambridge MA), 254–275.

Newey, Whitney K. and James L. Powell (1990), "Nonparametric Instrumental Variables Estimation", mimeo (Princeton University).

Pagan, Adrian R. (1984), "Econometric Issues in the Analysis of Regressions with Generated Regressors", *International Economic Review*, 25, 221-247.

Robinson, Peter M. (1983), "Nonparametric Estimators for Time Series", *Journal of Time Series Analysis* 4, 185–207.

Rosenblatt, Murray (1956), "A Central Limit Theorem and a Strong Mixing Condition", *Proceedings of the National Academy of Science* 42, 43–47.

Stock, James H. and Mark W. Watson (1993), "A Simple Estimator of Cointegrating Vectors in Higher Order Integrated Systems", *Econometrica*, 61, 783-820.

Université de Montréal
Département de sciences économiques
Centre de documentation
C.P. 6128, succursale Centre-ville
Montréal (Québec)
H3C 3J7

Cahiers de recherche (Discussion papers)
1994 à aujourd'hui (1994 to date)

Si vous désirez obtenir un exemplaire, vous n'avez qu'à faire parvenir votre demande et votre paiement (5 $ l'unité) à l'adresse ci-haut mentionnée. / To obtain a copy ($ 5 each), please send your request and prepayment to the above-mentioned address.

9401 : Mercenier, Jean et Bernardin Akitoby, "On Intertemporal General-Equilibrium Reallocation Effects of Europe's Move to a Single Market", janvier 1994, 41 pages.

9402 : Gauthier, Céline et Michel Poitevin, "Using Ex Ante Payments in Self-Enforcing Risk-Sharing Contracts", février 1994, 38 pages.

9403 : Ghysels, Eric et Joanna Jasiak, "Stochastic Volatility and Time Deformation : an Application of Trading Volume and Leverage Effects", février 1994, 37 pages.

9404 : Dagenais, Marcel G. et Denyse L. Dagenais, "GMM Estimators for Linear Regression Models with Errors in the Variables", avril 1994, 33 pages.

9405 : Bronsard, C., Fabienne Rosenwald et Lise Salvas-Bronsard, "Evidence on Corporate Private Debt Finance and the Term Structure of Interest Rates", avril 1994, 42 pages.

9406 : Dinardo, John, Nicole M. Fortin et Thomas Lemieux, "Labor Market Institutions and the Distribution of Wages, 1973-1992 : A Semiparametric Approach", avril 1994, 73 pages.

9407 : Campbell, Bryan et Jean-Marie Dufour, "Exact Nonparametric Tests of Orthogonality and Random Walk in the Presence of a Drift Parameter", avril 1994, 32 pages.

9408 : Bollerslev, Tim et Eric Ghysels, "Periodic Autoregressive Conditional Heteroskedasticity", mai 1994, 29 pages.

9409 : Cardia, Emanuela, "The Effects of Government Financial Policies : Can We Assume Ricardian Equivalence?", mai 1994, 42 pages.

9410 : Kollmann, Robert, "Hidden Unemployment : A Search Theoretic Interpretation", mai 1994, 9 pages.

9411 : Kollmann, Robert, "The Correlation of Productivity Growth Across Regions and Industries in the US", juin 1994, 14 pages.

9412 : Gaudry, Marc, Benedikt Mandel et Werner Rothengatter, "Introducing Spatial Competition through an Autoregressive Contiguous Distributed (AR-C-D) Process in Intercity Generation-Distribution Models within a Quasi-Direct Format (QDF)", juin 1994, 64 pages.

9413 : Gaudry, Marc et Alexandre Le Leyzour, "Improving a Fragile Linear Logit Model Specified for High Speed Rail Demand Analysis in the Quebec-Windsor Corridor of Canada", août 1994, 39 pages.

9414 : Lewis, Tracy et Michel Poitevin, "Disclosure of Information in Regulatory Proceedings", juillet 1994, 38 pages.

9415 : Ambler, Steve, Emanuela Cardia et Jeannine Farazli, "Export Promotion and Growth", août 1994, 41 pages.

9416 : Ghysels, Eric et Haldun Sarlan, "On the Analysis of Business Cycles Through the Spectrum of Chronologies", août 1994, 37 pages.

9417 : Martel, Jocelyn et Timothy C.G. Fisher, "The Creditors' Financial Reorganization Decision : New Evidence from Canadian Data", août 1994, 21 pages.

9418 : Cannings, Kathy, Claude Montmarquette et Sophie Mahseredjian, "Entrance Quotas and Admission to Medical Schools : A Sequential Probit Model", septembre 1994, 26 pages.

9419 : Cannings, Kathy, Claude Montmarquette et Sophie Mahseredjian, "Major Choices : Undergraduate Concentrations and the Probability of Graduation", septembre 1994, 26 pages.

9420 : Nabeya, Seiji et Pierre Perron, "Approximations to Some Exact Distributions in the First Order Autoregressive Model with Dependent Errors", septembre 1994, 40 pages.

9421 : Perron, Pierre, "Further Evidence on Breaking Trend Functions in Macroeconomic Variables", octobre 1994, 50 pages.

9422 : Vogelsang, Timothy J. et Pierre Perron, "Additional Tests for a Unit Root Allowing for a Break in the Trend Function at an Unknown Time", novembre 1994, 57 pages.

9423 : Ng, Serena et Pierre Perron, "Unit Root Tests in ARMA Models with Data Dependent Methods for the Selection of the Truncation Lag", décembre 1994, 41 pages.

9424 : Perron, Pierre, "The Adequacy of Asymptotic Approximations in the Near-Integrated Autoregressive Model with Dependent Errors", décembre 1994, 37 pages.

9425 : Ghysels, Eric et Pierre Perron, "The Effect of Linear Filters on Dynamic Time Series with Structural Change", décembre 1994, 35 pages.

9426 : Boyer, Marcel, Jean-Jacques Laffont, Philippe Mahenc et Michel Moreaux, "Sequential Location Equilibria Under Incomplete Information", décembre 1994, 38 pages.

9427 : Perron, Pierre et Serena NG, "Useful Modifications to Some Unit Root Tests with Dependent Errors and their Local Asymptotic Properties", décembre 1994, 41 pages.

9428 : Garcia, René et Pierre Perron, "An Analysis of the Real Interest Rate Under Regime Shifts", décembre 1994, 42 pages.

9501 : Boyer, Marcel et Jean-Jacques Laffont, "Environmental Risks and Bank Liability", janvier 1995, 46 pages.

9502 : Margolis, David. N., "Firm Heterogeneity and Worker Self-Selection Bias Estimated Returns to Seniority", décembre 1994, 29 pages.

9503 : Abowd, John M., Francis Kramarz et David N. Margolis, "High-Wage Workers and High-Wage Firms", janvier 1995, 73 pages

9504 : Cardia, Emanuela et Steve Ambler, "Indexation Lags and Heterodox Stabilization Programs", janvier 1995, 29 pages.

9605 : Garcia, René et Huntley Schaller, "Are the Effects of Monetary Policy Asymmetric?", février 1995, 42 pages.

9506 : Parent, Daniel, "Survol des contributions théoriques et empiriques liées au capital humain", février 1995, 70 pages.

9507 : Parent, Daniel, "Wages and Mobility : The Impact of Employer-Provided Training", février 1995, 34 pages.

9508 : Parent, Daniel, "Industry-Specific Capital and the Wage Profile : Evidence from the NLSY and the PSID", février 1995, 21 pages.

9509 : Parent, Daniel, "Matching, Human Capital, and the Covariance Structure of Earnings", février 1995, 54 pages.

9510 : Garcia, René, "Asymptotic Null Distribution of the Likelihood Ratio Test in Markov Switching Models", mars 1995, 50 pages.

9511 : Garcia, René, Annamaria Lusardi et Serena Ng, "Excess Sensitivity and Asymmetries in Consumption : An Empirical Investigation", mars 1995, 26 pages.

9512 : Sprumont, Yves, "An Axiomatization of the Pazner-Schmeidler Rules in Large Fair Division Problems", mars 1995, 26 pages.

9513 : Ghysels, Eric, Lynda Khalaf et Cosmé Vodounou, "Simulation Based Inference in Moving Average Models", mars 1995, 10 pages.

9514 : Ng, Serena, "Looking for Evidence of Speculative Stockholding in Commodity Markets", mars 1995, 25 pages.

9515 : Ng, Serena et Huntley Schaller, "The Risky Spread, Investment, and Monetary Policy Transmission : Evidence on the Role of Asymmetric Information", mars 1995, 26 pages.

9516 : Ng, Serena, "Testing for Homogeneity in Demand Systems when the Regressors are Non-Stationary", mars 1995, 26 pages.

9517 : Ghysels, Eric, Clive W.J. Granger et Pierre L. Siklos, "Is Seasonal Adjustment a Linear or Nonlinear Data Filtering Process?", mars 1995, 34 pages.

9518 : Ghysels, Eric, Alastair Hall et Hahn S. Lee, "On Periodic Structures and Testing for Seasonal Unit Roots", mars 1995, 45 pages.

9519 : Sprumont, Yves, "On the Game-Theoretic Structure of Public-Good Economies", mars 1995, 21 pages.

9520 : Charles, Sandra, François Vaillancourt et Nicolas Marceau, "The Impact of Decentralization on Growth and Democracy : A Note", mars 1995, 13 pages.

9521 : Sprumont, Yves, "Balanced Egalitarian Redistribution of Income", mars 1995, 17 pages.

9522 : Bronsard, Camille, Lise Salvas-Bronsard et Alain Trognon, "On the Residual Dynamics Implied by the Rational Expectations Hypothesis", mars 1995, 18 pages.

9523 : Campbell, Bryan et Eric Ghysels, "An Empirical Analysis of the Canadian Budget Process", mars 1995, 30 pages.

9524 : Ghysels, Eric, Alain Guay et Alastair Hall, "Predictive Tests for Structural Change with Unknown Breakpoint", avril 1995, 29 pages.

9525 : Ghysels, Eric, "On Stable Factor Structures in the Pricing of Risk", avril 1995, 37 pages.

9526 : Kollmann, Robert, "Mark Up Fluctuations in U.S. Manufacturing and Trade : Empirical Evidence Based on A Model of Optimal Storage", avril 1995, 18 pages.

9527 : Laberge-Nadeau, Claire, Georges Dionne, Urs Maag, Denise Desjardins, Charles Vanasse et J.-M. Ékoé, "Medical Conditions and the Severity of Commercial Motor Vehicle (CMV) Drivers' Road Accidents", mai 1995, 26 pages.

9528 : Dionne, Georges, Manuel Artis et Montserrat Guillén, "On the Repayment of Personal Loans Under Asymmetrical Information : A Count Data Model Approach", mai 1995, 32 pages.

9529 : Ruge-Murcia, Francisco J., "Government Expenditure and the Dynamics of High Inflation", juin 1995, 45 pages.

9530 : Persaran, M. Hashem et Francisco J. Ruge-Murcia, "A Discrete-Time Version of Target Zone Models with Jumps", juin 1995, 50 pages.

9531 : Dionne, Georges et N. Fombaron, "Non-Convexities and Efficiency of Equilibria in Insurance Markets with Asymmetric Information", juin 1995, 22 pages.

9532 : Bronsard, C., P. Michel et L. Salvas-Bronsard, "Singular Demand Systems as an Instrument of Scientific Discovery", juillet 1995, 17 pages.

9533 : Dionne, G. et C. Fluet, "Incentives in Multi-Period Regulation and Procurement : A Graphical Analysis", juillet 1995, 46 pages.

9534 : Ng, Serena et Pierre Perron, "Estimation and Inference in Nearly Unbalanced, Nearly Cointegrated Systems", août 1995, 38 pages.

9535 : Ng, Serena et Pierre Perron, "The Exact Error in Estimating the Spectral Density at the Origin", août 1995, 33 pages.

9536 : Ghysels, Eric, Christian Gouriéroux et Joanna Jasiak, "Market Time and Asset Price Movements : Theory and Estimation", septembre 1995, 61 pages.

9537 : Dionne, Georges, Robert Gagné et Charles Vanasse, "Inferring Technological Parameters from Incomplete Panel Data", septembre 1995, 40 pages.

9538 : Dufour, Jean-Marie et Eric Renault, "Short-Run and Long-Run Causality in Time Series : Theory", septembre 1995, 40 pages.

9539 : Dufour, Jean-Marie, "Some Impossibility Theorems in Econometrics with Applications to Instrumental Variables, Dynamic Models and Cointegration, septembre 1995, 42 pages.

9540 : Mercenier, Jean et Philippe Michel, "Temporal Aggregation in a Multi-Sector Economy with Endogenous Growth", septembre 1995, 17 pages.

9541 : Loranger, Jean-Guy, "Neoliberalism and the Overwhelming Influence of Financial Markets : a Comparative Analysis between NAFTA Countries and Other G-7 Countries", septembre 1995, 34 pages.

9542 : Proulx, Pierre-Paul, "La décentralisation : facteur de développement ou d'éclatement du Québec", août 1995, 26 pages.

9543 : Arcand, Jean-Louis, Ghazi Boulila et Christian Tritten, "Intergenerational Contracts, Remittances, and Growth", octobre 1995, 57 pages.

9544 : Kichian, Maral, René Garcia et Eric Ghysels, "On the Dynamic Specification of International Asset Pricing Models", octobre 1995, 29 pages.

9545 : Arcand, Jean-Louis, "Land Ownership, Working Capital, and Agricultural Output : Egypt", 1913-1958, octobre 1995, 93 pages.

9546 : Arcand, Jean-Louis et Marcel G. Dagenais, "The Empirics of Economic Growth in a Cross Section of Countries : Do Errors in Variables Really Not Matter?", octobre 1995, 65 pages.

9547 : Dufour, Jean-Marie et Jan F. Kiviet, "Exact Inference Methods for First-Order Autoregressive Distributed Lag Models", octobre 1995, 52 pages.

9548 : Dufour, Jean-Marie et Jan F. Kiviet, "Exact Tests for Structural Change in First-Order Dynamic Models", octobre 1995, 39 pages.

9549 : Kiviet, Jan F. et Jean-Marie Dufour, "Exact Tests in Single Equation Autoregressive Distributed Lag Models", octobre 1995, 38 pages.

9550 : Briys, Eric et Pascale Viala, "Optimal Insurance Design under Background Risk", octobre 1995, 14 pages.

9551 : Ng, Serena et Joris Pinkse, "Nonparametric Two-Step Estimation of Unknown Regression Functions when the Regressors and the Regression Error Are not Independent", octobre 1995, 22 pages.