DEPARTMENT OF ECONOMICS
UNIVERSITY OF CYPRUS

NON-LINEARITY INDUCED WEAK INSTRUMENTATION

Ioannis Kasparis, Peter C.B. Phillips & Tassos Magdalinos

Discussion Paper 2-2012

P.O. Box 20537, 1678 Nicosia, CYPRUS Tel.: +357-22893700, Fax: +357-22895028
Web site: http://www.econ.ucy.ac.cy

Non-linearity Induced Weak Instrumentation

Ioannis Kasparis*
Department of Economics
University of Cyprus

Peter C. B. Phillips†
Yale University, University of Auckland,
University of Southampton & Singapore Management University

and

Tassos Magdalinos
Department of Economics
University of Southampton

November 2011

Abstract

In regressions involving integrable functions we examine the limit properties of IV estimators that utilise integrable transformations of lagged regressors as instruments. The regressors can be either $I(0)$ or $I(1)$ processes. We show that this kind of nonlinearity in the regression function can significantly affect the relevance of the instruments. In particular, such instruments become weak when the signal of the regressor is strong, as it is in the $I(1)$ case. Instruments based on integrable functions of lagged $I(1)$ regressors display long range dependence and so remain relevant even at long lags, continuing to contribute to variance reduction in IV estimation. However, simulations show that OLS is generally superior to IV estimation in terms of MSE, even in the presence of endogeneity. Estimation precision is also reduced when the regressor is nonstationary.

*We thank Timos Papadopoulos and Dimitris Mavridis for their assistance with the simulations.

†Partial support is acknowledged from the NSF under Grant No. SES 09-56687.

Keywords: Instrumental variables, Integrable function, Integrated process, Invariance principle, Local time, Mixed normality, Stationarity, Nonlinear cointegration, Unit roots, Weak Instruments.

JEL classification: C22, C32

1 Introduction

Models involving nonlinear functions of serially correlated processes arise in various contexts, especially where economic variables and policy reaction functions are formulated to depend on underlying fundamentals. In economic theory and financial models, fundamentals are often represented in continuous time by stochastic processes such as Brownian motion or diffusions. Examples include the way in which fundamentals are thought to drive real macroeconomic variables such as output and productivity or financial variables such as stock prices and exchange rates returns.

The econometric formulation of such models may involve dependencies of the type

$$y_t = f(x_t, \beta) + u_t, \tag{1}$$

where the regressor x_t is a stationary or non-stationary autoregessive process, u_t is a stationary error process, and the regressor function f is some possibly nonlinear function of x_t and the parameter vector β. In general, x_t and u_t will be contemporaneously correlated, so that the equation may be interpreted as a structural equation within a larger system. When x_t is an $I(1)$ process, the equation is a nonlinear nonstationary relation and is sometimes called a nonlinear cointegrating relation between y_t and x_t. Such systems often prompt the use of instrumental variable (IV) techniques involving lagged variables, on which Les Godfrey has written extensively, particularly in the context of specification testing (Godfrey, 1988).

When the regression function (1) is linear the asymptotic variance of various estimators of β is well known to be inversely related to the strength of the regressor signal. This phenomenon is partly dependent on linearity and can be reversed when the regression function is nonlinear. In particular, if the regression function f is integrable, the asymptotic variance of OLS rises as the signal in x_t becomes stronger. This is true when x_t is $I(0)$ or $I(1)$. IV estimation is also susceptible to a weak instruments effect in which instruments become weaker as the signal increases. Simulation results confirm that the mean squared error (MSE) in IV estimation is significantly larger than that of OLS and bias gains from IV estimation are small relative to increases in variance. Estimation precision is also weaker when the regressor is non-stationary.

The main focus of this work is on the properties of IV estimators of the β parameter given in (1) when the regression function f is integrable. The present paper concentrates on the case where $f(x_t, \beta) = \beta f(x_t)$ where it is convenient to take a class of instrument functions for the regressor $f(x_t)$. IV methods are usually introduced to address issues of endogeneity that can occur in systems where the regressor x_t is contemporaneously correlated with the error u_t. The class of instruments considered are nonlinear and are formed by taking nonlinear functions of instrumental variables that satisfy relevance conditions with regard to x_t and orthogonality conditions with respect to u_t. Within this framework, a limit theory for IV estimation of structural models involving nonlinearities is obtained.

Limit theory for the case where $x_t \sim I(0)$ is standard, and utilises well known results for stationary ergodic sequences. For $x_t \sim I(1)$ the limit distribution of the IV estimator is dependent on the distribution of the innovations, so a full invariance principle does not apply. However, invariance principles do apply in the limit to conventional test statistics and so inference may be conducted in the usual fashion. This outcome is related to recent results of Jeganathan (2006, 2008), whose findings provide a major advance in studying sample functions of nonstationary processes involving endogeneities and general linear process time series innovations. In particular, Jeganathan's results enable a limit theory for least squares regression involving endogenous nonstationary covariates, which are further discussed in work by Jeganathan and Phillips (2009) and Chang and Park (2009). But those results are confined to cases of integrated regressors and they do not cover IV regression which has some clear advantages in stationary models.

The organization of the paper is as follows. Section 2 provides limit theory for IV estimators in the context of integrable regression functions. Section 3 discusses the consequences of nonlinearity on the limit variance of both OLS and IV estimators. Some simulation results are also provided. Section 4 considers IV estimators that utilize many instruments and provides results for the case of infinitely many instruments. Section 5 concludes. Proofs and technical material are given in the Appendix.

Before proceeding to the next section, we introduce some notation. For a vector $x = (x_i)$ or a matrix $A = (a_{ij})$, $|x|$ and $|A|$ denote the vector and matrix respectively of the moduli of their elements. The maximum of the moduli is denoted by $\|.\|$. For a matrix A, $A > \mathbf{0}$ denotes positive definiteness. As usual, $\overset{d}{=}$ denotes distributional equivalence. For a complex number x, $\bar{x}$ is its complex conjugate, and the Fourier transform of an integrable function f is denoted by $\tilde{f}$ (so that $\tilde{f}(\lambda) = \int_{\mathbb{R}} e^{-i\lambda s} f(s) ds$ and upon inversion $f(s) = (2\pi)^{-1} \int_{\mathbb{R}} e^{i\lambda s} \tilde{f}(\lambda) d\lambda$). For a (possibly matrix valued) function f, $\|.\|_B$ denotes its supremum over the subset B of its domain and we write $L_m = L_m(-\infty, \infty)$ for the function space $\left\{ f \mid \int_{-\infty}^{\infty} |f(x)|^m dx < \infty \right\}$. The L_1 family of functions will be also written as I. The real part of the complex number x is denoted by $\text{Re}(x)$. Finally, for a random variable X, we write $\|X\|_p = \{\mathbf{E}|X|^p\}^{1/p}$ and $\mathbf{E}_t(X) = \mathbf{E}(X \mid \mathcal{F}_t)$, where $\{\mathcal{F}_t\}_{t \in \{0\} \cup \mathbb{N}}$ is an appropriate filtration.

2 IV estimation of integrable models

2.1 Limit Theory

To illustrate the main ideas, we start by reviewing the special case of the structural equation (1)

$$y_t = \beta f(x_t) + u_t, \tag{2}$$

where the regressor x_t is either a stationary autoregressive process or an integrated process. The error term u_t is a martingale difference. Further, x_t is correlated with u_t so there is endogeneity in (2). When x_t is a stationary autoregressive process and $f(.)$ satisfies certain regularity conditions, the OLS estimator $\hat{\beta}$ of β is well known to be inconsistent with limit

$$\hat{\beta} \xrightarrow{p} \beta + \frac{\mathbf{E}[f(x_t)u_t]}{\mathbf{E}f(x_t)^2}.$$

When x_t is a unit root process we have the following limit theory. First, suppose that f is locally integrable and asymptotically homogeneous[1] i.e.

$$f(\lambda x) \approx \kappa(\lambda) H_f(x),$$

for λ large and some function $H_f(x)$ with continuous derivative $\dot{H}_f(x) := dH_f(x)/dx$. Then the OLS estimator has limit distribution

$$\sqrt{n}\kappa(\sqrt{n})\left(\hat{\beta} - \beta\right) \xrightarrow{d} \left[\int_0^1 H_f(B_x(r))^2 dr\right]^{-1} \times \left[\int_0^1 H_f(B_x(r))\, dB_u(r) + \sigma_{xu} \int_0^1 \dot{H}_f(B_x(r))\, dr\right] \tag{3}$$

(see de Jong, 2002, Ibragimov and Phillips, 2008, and Kasparis, 2008, for further details). Moreover, for f integrable, it follows from Jeganathan (2008) that

$$\sqrt[4]{n}\left(\hat{\beta} - \beta\right) \xrightarrow{d} MN\left(0, \frac{\sigma^2}{L_{B_x}(1,0)\int_{-\infty}^{\infty} f(r)^2 dr}\right) \tag{4}$$

[1] See Park and Phillips (1999, 2001) for more details about this family of models.

We therefore have the following collection of different asymptotic results depending on the nature of the regressor and function: (i) for $x_t \sim I(0)$ OLS is inconsistent; (ii) for $x_t \sim I(1)$ with locally integrable f, there is second order asymptotic bias in the limit theory given by the term $\sigma_{xu} \int_0^1 \dot{H}_f(B_x(r))\, dr$ in (3); and (iii) for $x_t \sim I(1)$ with integrable f the OLS is consistent and well centred. In view of (4), IV estimation is unecessary, when $x_t \sim I(1)$ and $f \in L_1$. Nevertheless, there is a case for pursuing IV estimation if one is unsure about the time series properties of the regressor such as its degree of integration. Some robustness to the integration properties of the regressor then seems desirable in estimation.

In this paper we study the case where regressor function $f \in L_1$. Such formulations arise naturally in many econometric contexts, such as discrete choice estimation, where we may want to allow for nonstationary data (Park and Phillips, 2000; Hu and Phillips, 2005; Phillips, Jin and Hu, 2009). We consider estimating (2) by instrumental variables using an integrable function g of an instrument z_t that is valid in the sense that it satisfies the usual orthogonality condition with respect to u_t and the relevance condition for x_t. In particular, the instrument z_t is determined by lagged values of the covariate. Let u_t be a martingale difference with respect to a filtration for which z_t is measurable. Also, x_t and u_t are correlated so that (2) is a structural equation. We plan to estimate β using the nonlinear instrument $g(z_t)$, giving

$$\hat{\beta} = \frac{\sum_{t=1}^{n} g(z_t)\, y_t}{\sum_{t=1}^{n} g(z_t)\, f(x_t)} = \beta + \frac{\sum_{t=1}^{n} g(z_t)\, u_t}{\sum_{t=1}^{n} g(z_t)\, f(x_t)}. \tag{5}$$

For stationary, weakly dependent x_t it is well known that limit theory is Gaussian. If x_t is $I(1)$ of the form $x_t = \sum_{s=1}^{t} v_s$ and v_s is a martingale difference sequence with $\{v_s\}_{s=1}^{t-1}$ independent of u_t, $z_t = x_{t-1}$ is a valid instrument. We will consider this case below. By the martingale CLT and standard nonlinear $I(1)$ asymptotics (Park and Phillips, 1999, 2001), the numerator of (5) has the following limit

$$\frac{1}{n^{1/4}} \sum_{t=1}^{n} g(z_t)\, u_t \xrightarrow{d} MN\left(0, \sigma^2 \int_{-\infty}^{\infty} g(s)^2 ds L_x(1,0)\right), \tag{6}$$

where L_x is the local time of the Brownian Motion $B_x(r)$ to which the standardized partial sums $n^{-1/2} \sum_{s=1}^{\lfloor nr \rfloor} v_s$ converge weakly.

The next part involves a Fourier integral approach and follows some earlier work by Borodin and Ibragimov (1995) and more recent research by Jeganathan (2006, 2008). We briefly outline the heuristics here. A formal derivation is given in the proof of Theorem 1. Using the fact that v_t is a martingale difference,

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{n} g(z_t)\, f(x_t) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} g(x_{t-1})\, f(x_{t-1} + v_t)$$

$$= \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{1}{\sqrt{n}} \sum_{t=1}^{n} g\left(x_{t-1}\right) e^{i\lambda x_{t-1} + i\lambda v_t} \tilde{f}(\lambda) d\lambda$$

$$= \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{1}{\sqrt{n}} \sum_{t=1}^{n} g\left(x_{t-1}\right) e^{i\lambda x_{t-1}} \mathbf{E}\left(e^{+i\lambda v_t}\right) \tilde{f}(\lambda) d\lambda + o_p(1)$$

$$\xrightarrow{d} \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(s) e^{i\lambda s} ds \mathbf{E}\left(e^{i\lambda v_t}\right) \tilde{f}(\lambda) d\lambda L_x(1,0)$$

$$= \frac{1}{2\pi} \int_{-\infty}^{\infty} \tilde{g}(-\lambda) \tilde{f}(\lambda) \mathbf{E}\left(e^{i\lambda v_t}\right) d\lambda L_x(1,0). \tag{7}$$

Note that this limit depends on the characteristic function of v_t and hence the result is not an invariance principle (IP). However, this distributional dependence does not prevent statistical testing, where an IP will hold as is shown below.

To proceed we simplify (7) using the convolution inversion

$$\frac{1}{2\pi} \int_{-\infty}^{\infty} \tilde{g}(-\lambda) \tilde{f}(\lambda) e^{i\lambda v} d\lambda = \int_{-\infty}^{\infty} g(s) f(s+v) ds,$$

so that

$$\begin{aligned} \int_{-\infty}^{\infty} \tilde{g}(-\lambda) \tilde{f}(\lambda) \mathbf{E}\left(e^{i\lambda v_t}\right) d\lambda &= \mathbf{E} \int_{-\infty}^{\infty} \tilde{g}(-\lambda) \tilde{f}(\lambda) e^{i\lambda v_t} d\lambda = \mathbf{E}\left(\int_{-\infty}^{\infty} g(s) f(s+v_t) ds\right) \\ &= \int_{-\infty}^{\infty} g(s) \mathbf{E} f(s+v_t) ds. \end{aligned}$$

Then

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{n} g\left(z_t\right) f\left(x_t\right) \xrightarrow{d} \int_{-\infty}^{\infty} g(s) \mathbf{E} f(s+v_t) ds L_x(1,0). \tag{8}$$

Observe that if $z_t = x_t$ then (8) reduces to the familiar result for integrable processes. Thus, (8) is an extension of usual nonlinear nonstationary theory (Park and Phillips, 1999) and the formula shows that the limit function is real even when v_t has a nonsymmetric distribution (in which case the characteristic function $\mathbf{E}\left(e^{i\lambda v_t}\right)$ is complex). Combining (6) and (8) gives the following mixed normal (MN) limit theory

$$
\begin{aligned}
n^{1/4}\left(\hat{\beta}-\beta\right) &= \frac{\frac{1}{n^{1/4}}\sum_{t=1}^{n} g\left(z_t\right) u_t}{\frac{1}{\sqrt{n}}\sum_{t=1}^{n} g\left(z_t\right) f(x_t)} \\
&\xrightarrow{d} MN\left(0, \frac{\sigma^2 \int_{-\infty}^{\infty} g(s)^2 ds}{L_x(1,0)\left[\int_{-\infty}^{\infty} \tilde{g}(-\lambda)\tilde{f}\left(\lambda\right)\mathbf{E}\left(e^{i\lambda v_t}\right) d\lambda\right]^2}\right) \\
&= MN\left(0, \frac{\sigma^2 \int_{-\infty}^{\infty} g(s)^2 ds}{L_x(1,0)\left[\int_{-\infty}^{\infty} g(s)\mathbf{E} f(s+v_t) ds\right]^2}\right). \qquad (9)
\end{aligned}
$$

We now proceed with a formal development of the theory. We consider two cases involving an autoregressive covariate x_t generated by

$$
x_t = \rho_t x_{t-1} + v_t. \qquad (10)
$$

In the first case, x_t has a unit root and in the second x_t is stable autoregressive. We apply the following conditions.

Assumption 1:

The autoregressive coefficient ρ_t is defined as

$$
\rho_t = \begin{cases} 1, \text{ for } t > 0 \\ \rho, \text{ for } t \leq 0 \end{cases},
$$

where $|\rho| < 1$ is constant.

Assumption 1*:

The autoregressive coefficient is $\rho_t = \rho$ with $|\rho| < 1$.

Under (10) and Assumption 1 x_t is a unit root process which for $t > 0$ has the form $x_t = \sum_{i=1}^{t} v_i + x_0$, with $x_0 = O_p(1)$, which is uncorrelated with $\sum_{i=1}^{t} v_i$. Under Assumption 1* x_t is a stationary autoregression. Next, we specify the properties of the variables u_t and v_t that appear in (2) and (10) respectively. Let $\mathcal{F}_t$ be the sigma algebra generated by (u_t, v_t).

Assumption 2:

(i) *$\{v_t, \mathcal{F}_t\}$ is a martingale difference sequence with $\mathbf{E}\left[e^{i\lambda v_t} \mid \mathcal{F}_{t-1}\right] = \mathbf{E}\left[e^{i\lambda v_t}\right] = \varphi(\lambda)$ a.s. In addition, the characteristic function of v_t satisfies $\varphi(\lambda) = o(|\lambda|^{-\delta})$ as $\lambda \to \infty$, for some $\delta > 1$.*

(ii) *$\{u_t, \mathcal{F}_t\}$ is a martingale difference sequence with $\mathbf{E}\left[u_t^2 \mid \mathcal{F}_{t-1}\right] = \sigma^2 < \infty$ a.s.*

(iii) $\sup_t \mathbf{E}(u_t^4 \mid \mathcal{F}_{t-1}) < \infty$ *a.s.*

While v_t is assumed to be a martingale difference sequence in (i), the results of the paper may be extended to the case where v_t is a stationary linear process under some additional conditions using the approach developed in recent work by Jeganathan (2008). The martingale difference condition (ii) is also restrictive, although it has been used in other recent work (Wang and Phillips, 2009, Chang and Park, 2008), and may also be extended. However, relaxation of these conditions introduces major new difficulties that substantially complicate the arguments, as mentioned in Remark (c) below. Such extensions are therefore left for future work.

The limit theory for the IV estimator in the nonstationary case is given in the following result.

Theorem 1 *Let Assumptions 1 and 2 hold and suppose that:*
(i) $g, g^2 \in L_1$ *and are Lipschitz continuous,*
(ii) $f, \tilde{f}, \tilde{g}\tilde{f} \in L_1$,
(iii) $h(s) \equiv (2\pi)^{-1} g(s) \widetilde{\left(\tilde{f}\varphi\right)}(-s)$ *is Lipschitz continuous.*

(a) *Then for* $z_t = x_{t-1}$, *as* $n \to \infty$

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{n} g(z_t) f(x_t) \xrightarrow{d} \int_{-\infty}^{\infty} \tilde{g}(-\lambda) \tilde{f}(\lambda) \mathbf{E}\left(e^{i\lambda v_t}\right) d\lambda L_x(1,0).$$

(b) *Further, if* $g^4 \in L_1$, *as* $n \to \infty$ *we have*

$$n^{1/4}\left(\hat{\beta} - \beta\right) \xrightarrow{d} MN\left(0, \frac{\sigma^2 \int_{-\infty}^{\infty} g(s)^2 ds}{L_x(1,0)\left[\int_{-\infty}^{\infty} \tilde{g}(-\lambda)\tilde{f}(\lambda)\mathbf{E}\left(e^{i\lambda v_t}\right) d\lambda\right]^2}\right). \tag{11}$$

Remarks.

(a) The smoothness condition on g includes a range of possible instrument functions. It can be further relaxed if the methods in Wang and Phillips (2009) are used.

(b) Although (7) and (9) are not distributionally invariant because the limits depend on the characteristic function and distribution of v_t, hypothesis testing on β may be conducted in the usual way with the $t-$ statistic constructed as

$$\hat{t} = \frac{\hat{\beta} - \beta}{s_{\hat{\beta}}}, \tag{12}$$

where $s_{\hat{\beta}}^2 = (n^{-1}\sum_{t=1}^n \hat{u}_t^2)\left(\sum_{t=1}^n g(z_t)^2\right)/\left(\sum_{t=1}^n g(z_t) f(x_t)\right)^2$ and $\hat{u}_t = y_t - \hat{\beta} f(x_t)$. Noting that

$$\frac{1}{\sqrt{n}}\sum_{t=1}^n g(z_t)^2 \xrightarrow{d} \int_{-\infty}^{\infty} g(s)^2 ds L_x(1,0), \quad n^{-1}\sum_{t=1}^n \hat{u}_t^2 \rightarrow_p \sigma^2, \tag{13}$$

we have from (8) and (13)

$$\begin{aligned} n^{1/2}s_{\hat{\beta}}^2 &= n^{-1}\sum_{t=1}^n \hat{u}_t^2 \frac{\frac{1}{\sqrt{n}}\sum_{t=1}^n g(z_t)^2}{\left(\frac{1}{\sqrt{n}}\sum_{t=1}^n g(z_t) f(x_t)\right)^2} \\ &\xrightarrow{d} \frac{\sigma^2 \int_{-\infty}^{\infty} g(s)^2 ds}{L_x(1,0)\left(\int_{-\infty}^{\infty} g(s)\mathbf{E}f(s+v_t)ds\right)^2}. \end{aligned} \tag{14}$$

It follows by (9) and (14) that

$$\begin{aligned} t &= \frac{n^{1/4}\left(\hat{\beta}-\beta\right)}{\left(n^{1/2}s_{\hat{\beta}}^2\right)^{1/2}} \\ &= \frac{\frac{1}{n^{1/4}}\sum_{t=1}^n g(z_t) u_t}{\left\{\frac{1}{\sqrt{n}}\sum_{t=1}^n g(z_t) f(x_t)\right\}\left\{n^{1/2}s_{\hat{\beta}}^2\right\}^{1/2}} \xrightarrow{d} N(0,1), \end{aligned}$$

so that the distributional effect in the limit theory (11) scales out asymptotically in the t-statistic. Hence, conventional methods of inference are possible.

(c) We remark that the least squares estimator $\hat{\beta}_{LS}$ has the following limit

$$n^{1/4}\left(\hat{\beta}_{LS}-\beta\right) = \frac{\frac{1}{n^{1/4}}\sum_{t=1}^n f(x_t) u_t}{\frac{1}{\sqrt{n}}\sum_{t=1}^n f(x_t)^2} \xrightarrow{d} MN\left(0, \frac{\sigma^2}{L_x(1,0)\int_{-\infty}^{\infty} f(s)^2 ds}\right),$$

which applies even in the case of endogenous x_t when (u_t, v_t) is an iid sequence (Jeganathan 2006, 2008; Chang and Park, 2011). The limit distribution has a more complex form and is only a weak invariance principle when (u_t, v_t) is serially dependent. In that case, the variance depends on the distribution of (u_t, v_t), as shown in Jeganathan (2008) and Jeganathan and Phillips (2009).

(d) Note that under Assumption 1*, $f(x_t)$ is a strictly stationary (e.g. Ibragimov and Linnik, 1971), L_2 near epoch dependent sequence of size $-\infty$ of the innovation sequence $\{v_t\}$ (c.f. Theorem 17.12 in Davidson, 1994). Therefore, under Assumption 1* and some additional regularity conditions, we get the well-known limit theory for IV estimation involving mixing time series (e.g. Pötscher and Prucha, 1997; Bierens and Gallant, 1997)

$$\sqrt{n}\left(\hat{\beta}-\beta\right) \xrightarrow{d} N\left(0, \frac{\sigma^2 \mathbf{E} g(x_t)^2}{\left[\mathbf{E} f(x_{t-1}) g(x_t)\right]^2}\right).$$

It follows that when x_t is a stable autoregressive process the t-statistic of (12) satisfies $\hat{t} \xrightarrow{d} N(0,1)$, just as in the unit root case.

2.2 Choice of the Instrument Function

We next consider how limit variance is affected by choice of the instrument function. Denote by $\Omega_{IV}(g)$ the limit variance in (9) i.e.

$$\Omega_{IV}(g) = \frac{\sigma^2 \int_{-\infty}^{\infty} g(s)^2 ds}{L_x(1,0)\left(\int_{-\infty}^{\infty} g(s) \mathbf{E} f(s+v_t) ds\right)^2}. \tag{15}$$

As before, f is the regression function and g is the instrument function. Suppose that the characteristic function of v_t is real valued and positive i.e. $\mathbf{E}\left(e^{isv_t}\right) = \operatorname{Re} \mathbf{E}\left(e^{isv_t}\right) \geq 0$. Define the measure $\mu(ds) = \mathbf{E}\left(e^{isv_t}\right) ds$ and the *energy* of a function $\beta(s) \in L_1 \cap L_2$ by

$$\int_{-\infty}^{\infty} \left|\tilde{\beta}(s)\right|^2 ds.$$

Further, define the *relative energy* of $\beta(s)$ as $\int_{-\infty}^{\infty} \left|\tilde{\beta}(s)\right|^2 ds / \int_{-\infty}^{\infty} \left|\tilde{\beta}(s)\right|^2 \mu(ds)$.[2] It can be shown that $\Omega_{IV}(f) \leq \Omega_{IV}(g)$ *a.s.*, for any instrument function g of larger relative energy than that of the regression function f. We make use of the following result.

[2] Note that Parseval's identity gives $\int_{-\infty}^{\infty} \left|\tilde{\beta}(s)\right|^2 ds = \int_{-\infty}^{\infty} \left|\beta(s)\right|^2 ds$ whilst convolution inversion gives $\int_{-\infty}^{\infty} \left|\tilde{\beta}(s)\right|^2 \mu(ds) = \int_{-\infty}^{\infty} \beta(s) \mathbf{E}\beta(s+v_t) ds$ (e.g. Lang (1993), pp.242-243). Further, simple calculations show that the relative energy satisfies

$$\int_{-\infty}^{\infty} \left|\tilde{\beta}(s)\right|^2 ds / \int_{-\infty}^{\infty} \left|\tilde{\beta}(s)\right|^2 \mu(ds) \geq 1.$$

Proposition 1: (i) *Suppose that* $f, g \geq 0$ *and the characteristic function* $\mathbf{E}\left(e^{isv_t}\right) = \operatorname{Re}\mathbf{E}\left(e^{isv_t}\right) \geq 0$. *Then*

$$\left\{\int_{-\infty}^{\infty} g(s)\mathbf{E}f(s+v_t)ds\right\}^2 \leq \left\{\int_{-\infty}^{\infty} g(s)\mathbf{E}g(s+v_t)ds\right\}\left\{\int_{-\infty}^{\infty} f(s)\mathbf{E}f(s+v_t)ds\right\}$$

(ii) *Further, suppose that*

$$\frac{\int_{-\infty}^{\infty}\left|\tilde{f}\left(s\right)\right|^2 ds}{\int_{-\infty}^{\infty}\left|\tilde{f}\left(s\right)\right|^2 \mu(ds)} \leq \frac{\int_{-\infty}^{\infty}\left|\tilde{g}\left(s\right)\right|^2 ds}{\int_{-\infty}^{\infty}\left|\tilde{g}\left(s\right)\right|^2 \mu(ds)}. \tag{16}$$

Then,

$$\frac{\int_{-\infty}^{\infty} f(s)^2 ds}{\left[\int_{-\infty}^{\infty} f(s)\mathbf{E}f(s+v_t)ds\right]^2} \leq \frac{\int_{-\infty}^{\infty} g(s)^2 ds}{\left[\int_{-\infty}^{\infty} f(s)\mathbf{E}g(s+v_t)ds\right]^2}.$$

Equation (16) postulates that f has smaller relative energy than g. The stated result is a direct consequence of Proposition 1.

Corollary 1. *Suppose that the conditions of Theorem 1 and Proposition 1 hold. Then* $\Omega_{IV}(f) \leq \Omega_{IV}(g)$ *a.s.*

Corollary 1 holds with strict inequality whenever part (i) of Proposition 1 holds with strict inequality. The following conditions postulate that f and g are of the same energy (with respect to measures ds and $\mu(ds)$), and are sufficient for equality in (16):

$$\int_{-\infty}^{\infty}\left|\tilde{f}\left(s\right)\right|^2 ds = \int_{-\infty}^{\infty}\left|\tilde{g}\left(s\right)\right|^2 ds \text{ and } \int_{-\infty}^{\infty}\left|\tilde{f}\left(s\right)\right|^2 \mu(ds) = \int_{-\infty}^{\infty}\left|\tilde{g}\left(s\right)\right|^2 \mu(ds).$$

Therefore, by Corollary 1 for any instrument function g with the same energy as the regression function f we have $\Omega_{IV}(f) \leq \Omega_{IV}(g)$ *a.s.*

Example: Suppose that $f(x) = 1/\pi(1+x^2)$ and $g(x) = e^{-x^2}/\sqrt{\pi}$ with $x_t - x_{t-1} = v_t \sim i.i.d.N(0, \sigma_v^2)$. Then we have

$$\int_{-\infty}^{\infty}\left|\tilde{f}\left(s\right)\right|^2 ds = \frac{1}{2\pi},\ \int_{-\infty}^{\infty}\left|\tilde{f}\left(s\right)\right|^2 \mu(ds) = \frac{1}{\sqrt{2\pi\sigma_v^2}} e^{\frac{2}{\sigma_v^2}}\left(1 - \operatorname{erf}(\sqrt{2/\sigma_v^2})\right),$$

$$\int_{-\infty}^{\infty}\left|\tilde{g}\left(s\right)\right|^2 ds = \frac{1}{\sqrt{2\pi}},\ \int_{-\infty}^{\infty}\left|\tilde{g}\left(s\right)\right|^2 \mu(ds) = \sqrt{\frac{1}{2\pi(1+\sigma_v^2)}},$$

where $\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-z^2} dz$ is the error function. Numerical calculations show that in this case condition (16) is satisfied for all $\sigma_v^2 > 0$. Therefore, we have $\Omega_{IV}(f) \leq \Omega_{IV}(g)$ $a.s.$

Remark. The information loss arising from the use of the instrumental variable $f(x_{t-1})$ in place of the least squares instrument $f(x_t)$ is measured by the difference

$$\begin{aligned}
\frac{1}{\sigma^2}(\Omega_{IV} - \Omega_{LS}) &= \frac{\int_{-\infty}^{\infty} f(s)^2 ds}{L_x(1,0)\left[\int_{-\infty}^{\infty} f(s)\mathbf{E}f(s+v_t)ds\right]^2} - \frac{1}{L_x(1,0)\int_{-\infty}^{\infty} f(s)^2 ds} \\
&= \frac{\left(\int_{-\infty}^{\infty} f(s)^2 ds\right)^2 - \left[\int_{-\infty}^{\infty} f(s)\mathbf{E}f(s+v_t)ds\right]^2}{L_x(1,0)\left[\int_{-\infty}^{\infty} f(s)\mathbf{E}f(s+v_t)ds\right]^2 \int_{-\infty}^{\infty} f(s)^2 ds}.
\end{aligned}$$

Observe the following non-random bound

$$\left|\int_{-\infty}^{\infty} f(s)f(s+v_t)ds\right| \leq \left\{\int_{-\infty}^{\infty} f(s)^2 ds\right\}^{1/2} \left\{\int_{-\infty}^{\infty} f(s+v_t)^2 ds\right\}^{1/2} = \int_{-\infty}^{\infty} f(s)^2 ds,$$

so that

$$\int_{-\infty}^{\infty} f(s)\mathbf{E}f(s+v_t)ds = \mathbf{E}\int_{-\infty}^{\infty} f(s)f(s+v_t)ds \leq \int_{-\infty}^{\infty} f(s)^2 ds,$$

leading to the inequality $\Omega_{IV} \geq \Omega_{LS}$. In general, the information loss is greater, the greater the dispersion of the distribution of v_t. This is demonstrated explicitly in the next section.

3 Effects of Nonlinearity on Limit Variance

This section examines the effects of non-linearity on the limit variance of OLS and IV estimators. As before, we consider the case where the regression function $f \in I$, and $x_t \sim I(0)$ or $I(1)$. It is well known that, for linear f, the asymptotic variance of various estimators for β is inversely related to the regressor signal. This phenomenon depends on functional form. When the regression function is an integrable one, the asymptotic variance-regressor signal relationship is reversed. In particular, if the regression function f is integrable, the asymptotic variance of OLS increases when the signal of x_t increases. This is true whether x_t is $I(0)$ or $I(1)$. The phenonomen may be accentuated when IV techniques are employed. When the regression function has thin tails, there is an additional weak instruments effect. In particular, instruments become weaker as the regressor signal increases. Simulation results show that the MSE of the IV estimator is significantly larger than the MSE of the OLS estimator in this case. Therefore, bias gains from IV estimation are small relative to the increase in variance. Furthermore, estimation precision is reduced when the regressor is non-stationary.

3.1 OLS estimation

For stationary x_t and under exogeneity the limit variance in OLS estimation is

$$\Omega_{LS} = \frac{\sigma^2}{\mathbf{E}f(x_t)^2}.$$

For various integrable functions f and for various distributions of v_t $(= x_t - \rho x_{t-1})$, Ω_{LS} is positively related to the variance of v_t (and x_t). Consider the following example.

Example 1. Let $f(x) = \exp(-x^2)$ and $|\rho| < 1$ with $v_t \sim N(0, \sigma_v^2)$. Then

$$\mathbf{E}f(x_t)^2 = \sqrt{\frac{1}{1 + 2\sigma_v^2 \left(1 - \rho^2\right)^{-1}}} = O\left(\frac{1}{\sigma_v}\right), \tag{17}$$

as $\sigma_v^2 \to \infty$. Hence, $\Omega_{LS} = O\left(\sigma_v\right) \to \infty$ as $\sigma_v^2 \to \infty$. In addition, estimation precision deteriorates when the autoregressive coefficient approaches unity i.e. $|\rho| \to 1$. The latter is expected given the fact that the convergence rate under stationary x_t (viz., $\sqrt{n}$) exceeds that for integrated x_t (viz., $\sqrt[4]{n}$).

For $x_t \sim I(1)$, the limit variance of the OLS estimator is (Park and Phillips, 1999))

$$\Omega_{LS} = \frac{\sigma^2}{L_x(1,0)\int_{-\infty}^{\infty} f(s)^2 ds} = O\left(\sigma_v\right), \tag{18}$$

as the following argument shows. Let $W(r)$ be standard Brownian motion. The "chronological" local time $L_x(1,0)$ of B_x at the origin over $[0,1]$ is

$$L_x(1,0) = \lim_{\varepsilon \downarrow 0} \frac{1}{2\varepsilon} \int_0^1 1\left\{|B_x(r)| < \varepsilon\right\} dr = \lim_{\varepsilon \downarrow 0} \frac{1}{2\varepsilon} \int_0^1 1\left\{\left|\sqrt{\mathbf{E}v_t^2} W(r)\right| < \varepsilon\right\} dr$$

$$= \frac{1}{\sqrt{\mathbf{E}v_t^2}} \lim_{\varepsilon \downarrow 0} \frac{\sqrt{\mathbf{E}v_t^2}}{2\varepsilon} \int_0^1 1\left\{|W(r)| < \varepsilon / \sqrt{\mathbf{E}v_t^2}\right\} dr = \frac{1}{\sqrt{\mathbf{E}v_t^2}} \lim_{\eta \downarrow 0} \frac{1}{2\eta} \int_0^1 1\left\{|W(r)| < \eta\right\} dr$$

$$= \frac{1}{\sqrt{\mathbf{E}v_t^2}} L_W(1,0) = \sigma_v^{-1} L_W(1,0).$$

The limit variance therefore has the form

$$\Omega_{LS} = \frac{\sigma^2}{L_x(1,0)\int_{-\infty}^{\infty} f(s)^2 ds} = \frac{\sigma^2\sqrt{\mathbf{E}v_t^2}}{L_W(1,0)\int_{-\infty}^{\infty} f(s)^2 ds} = O\left(\sigma_v\right). \tag{19}$$

It follows that for both stationary and nonstationary cases $\Omega_{LS} = O\left(\sigma_v\right)$. Nonetheless, as remarked earlier, estimation is less precise under nonstationarity due to the slower convergence rate $n^{1/4}$. This reduction in precision is manifest in simulations.

3.2 IV estimation

We consider the limit variance of the IV estimator. In the following analysis the instrument is $z_t = x_{t-1}$. Suppose $x_t \sim I(0)$. IV can lead to significant deterioration in estimation, as the following extension of Example 1 above shows.

Example 2. Suppose that $f(x) = g(x) = \exp(-x^2)$ and $|\rho| < 1$ with $v_t \sim N(0, \sigma_v^2)$. From (17) $\mathbf{E}f(x_t)^2 = O\left(\frac{1}{\sigma_v}\right)$, and

$$\mathbf{E}f(x_t)f(x_{t-1}) = O\left(\sigma_v^{-1}\right). \tag{20}$$

Correspondingly, instrument relevance goes to zero as the signal of the regressor $(\sigma_v^2/(1-\rho^2))$ approaches infinity. From (17) and (20), the limit variance is

$$\Omega_{IV} = \frac{\sigma^2\mathbf{E}f(x_t)^2}{\left\{\mathbf{E}\left[f(x_{t-1})f(x_t)\right]\right\}^2} = O\left(\sigma_v\right) \text{ as } \sigma_v \to \infty.$$

This expression reveals that the contrary impact of regressor signal on estimation efficiency is of the same order for IV as OLS estimation. In particular, Examples 1 and 2 show

$$\frac{\Omega_{IV}}{\Omega_{LS}} = O\left(1\right) \quad \text{as } \sigma_v \to \infty.$$

For $x_t \sim I(1)$, the effects of nonlinearity on IV estimation are quite different to these results for stationary models, as the following example demonstrates.

Example 3. Suppose that $f(x) = g(x) = \exp(-x^2)$ and $\rho = 1$ with $v_t \sim N(0, \sigma_v^2)$. The following term captures the relevance of the instruments in the limit:

$$\int_{-\infty}^{\infty} f(s)\mathbf{E}f(s+v_t)ds = \sqrt{\frac{1}{2\pi\left(1+\sigma_v^2\right)}}. \tag{21}$$

Therefore, in view of the above, (15) and the fact that $L_x(1,0) = \sigma_v^{-1} L_W(1,0)$ the limit variance of the IV estimator is

$$\begin{aligned}\Omega_{IV} &= \frac{\sigma^2 \int_{-\infty}^{\infty} f(s)^2 ds}{L_x(1,0)\left(\int_{-\infty}^{\infty} f(s)\mathbf{E}f(s+v_t)ds\right)^2} = \frac{\sigma^2\sqrt{\frac{\pi}{2}}}{\sigma_v^{-1}L_W(1,0)\frac{1}{2\pi(1+\sigma_v^2)}} \\ &= \sigma_v\left(1+\sigma_v^2\right)\frac{\sigma^2\sqrt{2}\pi^{3/2}}{L_W(1,0)} = O\left(\sigma_v^3\right).\end{aligned}$$

since $\int_{-\infty}^{\infty} e^{-2x^2}dx = \sqrt{\frac{\pi}{2}}$. Further, since

$$\Omega_{LS} = \frac{\sigma^2}{L_x(1,0)\int_{-\infty}^{\infty} f(s)^2 ds} = \sigma_v \frac{\sqrt{2}\sigma^2}{\sqrt{\pi}L_W(1,0)} = O\left(\sigma_v\right),$$

the ratio

$$\frac{\Omega_{IV}}{\Omega_{LS}} = \frac{\sigma_v\left(1+\sigma_v^2\right)\frac{\sigma^2\sqrt{2}\pi^{3/2}}{L_W(1,0)}}{\sigma_v\frac{\sqrt{2}\sigma^2}{\sqrt{\pi}L_W(1,0)}} = \left(1+\sigma_v^2\right)\pi^2 = O\left(\sigma_v^2\right), \qquad (22)$$

so that IV is substantially more dispersed as $\sigma_v \to \infty$, which is quite different from the behaviour reported in Example 2 for stationary regression.

Finally, we consider an example where the regression function f is heavy tailed. In this case there is no weak instruments effect, and behavior of the IV estimator is analogous to that of OLS.

Example 4 (heavy tailed regression function) Suppose that $f(x) = g(x) = 1/\pi(1+x^2)$ and $\rho = 1$ with $v_t \sim N(0, \sigma_v^2)$. Then the relevance of the instruments is given by:

$$\int_{-\infty}^{\infty} f(s)\mathbf{E}f(s+v_t)ds = \frac{1}{\sqrt{2\pi}\sigma_v}e^{\frac{2}{\sigma_v^2}}\left(1-\mathrm{erf}(\sqrt{2/\sigma_v^2})\right) = \frac{1}{\sqrt{2\pi}\sigma_v}e^{\frac{2}{\sigma_v^2}}\left(\mathrm{erf\,c}(\sqrt{2/\sigma_v^2})\right), \qquad (23)$$

where $\mathrm{erf}\,(x) = \frac{2}{\sqrt{\pi}}\int_0^x e^{-z^2}dz$ is the error function and $\mathrm{erf\,c}(x) = \frac{2}{\sqrt{\pi}}\int_x^{\infty} e^{-z^2}dz = 1-\mathrm{erf}\,(x)$ is the complementary error function. Observe that $e^x\left(1-\mathrm{erf}(x)\right) \to 1$ as $x \to 0$ so that

$$\begin{aligned}\int_{-\infty}^{\infty} f(s)\mathbf{E}f(s+v_t)ds &= \frac{1}{\sqrt{2\pi}\sigma_v}e^{\frac{2}{\sigma_v^2}}\left(1-\mathrm{erf}(\sqrt{2/\sigma_v^2})\right) \\ &= \frac{1}{\sqrt{2\pi}\sigma_v}\{1+o(1)\} = O\left(\sigma_v^{-1}\right),\end{aligned}$$

as $\sigma_v^2 \to \infty$. Thus, just as in (21) of Example 3, the relevance term vanishes as $\sigma_v^2 \to \infty$ in this heavy tailed case. Further,

$$\frac{1}{\pi^2}\int_{-\infty}^{\infty} \frac{1}{(1+x^2)^2} dx = \frac{1}{2\pi}.$$

Then, since $L_x(1,0) = \sigma_v^{-1} L_W(1,0)$ we get

$$\begin{aligned}
\Omega_{IV} &= \frac{\sigma^2 \int_{-\infty}^{\infty} f(s)^2 ds}{L_x(1,0)\left(\int_{-\infty}^{\infty} f(s)\mathbf{E}f(s+v_t)ds\right)^2} \\
&= \frac{\frac{1}{2\pi}\sigma_v \sigma^2}{L_W(1,0)\left(\frac{1}{\sqrt{2\pi}\sigma_v} e^{\frac{2}{\sigma_v^2}}\left(\operatorname{erf c}(\sqrt{2/\sigma_v^2})\right)\right)^2} \\
&= \frac{\sigma_v^3 \sigma^2}{L_W(1,0) e^{\frac{4}{\sigma_v^2}}\left(\operatorname{erf c}(\sqrt{2/\sigma_v^2})\right)^2} \\
&= O\left(\sigma_v^3\right),
\end{aligned}$$

as $\sigma_v^2 \to \infty$. Next, we have

$$\Omega_{LS} = \frac{\sigma^2}{L_x(1,0)\int_{-\infty}^{\infty} f(s)^2 ds} = \sigma_v \frac{2\pi\sigma^2}{L_W(1,0)}.$$

Then

$$\begin{aligned}
\frac{\Omega_{IV}}{\Omega_{LS}} &= \frac{\sigma_v^3 \sigma^2}{L_W(1,0) e^{\frac{4}{\sigma_v^2}}\left(\operatorname{erf c}(\sqrt{2/\sigma_v^2})\right)^2} \times \frac{L_W(1,0)}{2\pi\sigma^2\sigma_v} \\
&\sim \frac{\sigma_v^2}{\pi e^{\frac{4}{\sigma_v^2}}\left(\operatorname{erf c}(\sqrt{2/\sigma_v^2})\right)^2} \sim \frac{\sigma_v^2}{\pi} \to \infty, \quad \text{as } \sigma^2 \to \infty.
\end{aligned}$$

Thus, $\frac{\Omega_{IV}}{\Omega_{LS}} = O\left(\sigma_v^2\right) \to \infty$ as $\sigma_v^2 \to \infty$ and the ratio has the same order as in Example 3. Note, however, that with the heavier tailed function $f(x) = 1/\pi(1+x^2)$,

$$\frac{\Omega_{IV}}{\Omega_{LS}} \sim \frac{\sigma_v^2}{\pi}\left\{1+o(1)\right\} < \left(1+\sigma_v^2\right)\pi^2\left\{1+o(1)\right\},$$

so we may expect that IV will perform better when the regression function is heavier in the tail.

3.3 Simulations

This section provides some brief simulation results for the MSE of the OLS and IV estimators in a simple nonlinear model to illustrate these effects in finite samples. We generated 10,000 replications with sample size $n = 2000$, of the following model:

$$y_t = \beta e^{-0.5x_t^2} + u_t, \ \beta = 1,$$

with $x_t = \rho x_{t-1} + v_t$ and

$$\begin{bmatrix} u_t \\ v_t \end{bmatrix} \sim iid\ N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & R \times \sigma_v \\ R \times \sigma_v & \sigma_v^2 \end{bmatrix}\right), \quad -1 < R < 1.$$

The variance term takes values $\sigma_v^2 = \{1, 2, 3, 4, 5\}$. Further we consider a range of autoregressive parameters $\rho =$0, 0.5 and 1.

Figures 1-6 provide plots of the LS and IV variance against various values of R and σ_v^2. It is apparent that variance increases as the error variance σ_v^2 gets larger. Further, variance increases when the autoregressive coefficient ρ approaches unity. Figures 7-9 provide plots of the ratio MSE_{IV}/MSE_{OLS}, for various values of the autoregressive parameter. The OLS estimator is superior in terms of MSE. Therefore, possible bias reduction gains from IV estimation are relatively small compared to the deterioration in estimation precision. The relative MSE performance of the IV estimator deteriorates as σ_v^2 increases.

4 IV Estimation with Many Instruments

4.1 Stationary regressor case

We start with the stationary regressor and linear model case

$$\begin{aligned} y_t &= \beta x_t + u_t \\ x_t &= \rho x_{t-1} + v_t, \ |\rho| < 1 \end{aligned}$$

and consider IV estimation that utilises K successively lagged values of x_t as instruments i.e. $z_t' = (x_{t-1}, ..., x_{t-K})$. Then $\hat{\beta} = (X'P_Z X)^{-1} X'P_Z Y$ where X, Y, and Z are observation matrices of x_t, y_t and z_t and P_Z is the projection matrix onto the range of Z. In this case, $\hat{\beta}$ has the following limit distribution

$$\sqrt{n}\left(\hat{\beta} - \beta\right) \xrightarrow{d} N\left(0, \sigma^2 \left\{A_K' \Omega_K^{-1} A_K\right\}^{-1}\right), \tag{24}$$

where $A_K' = \frac{\sigma_v^2}{1-\rho^2}\left[\rho, ..., \rho^K\right]$, and Ω_K is Toeplitz with (j,k)'th element $\frac{\sigma_v^2}{1-\rho^2}\rho^{|j-k|}$. Simple calculations then show that $A_K'\Omega_K^{-1}A_K = \frac{\rho^2\sigma_v^2}{1-\rho^2}$ and the variance of the limit distribution (24) is

$$\sigma^2 \left\{A_K' \Omega_K^{-1} A_K\right\}^{-1} = \frac{\sigma^2}{\rho^2 \sigma_v^2} \left(1 - \rho^2\right). \tag{25}$$

Thus $\sigma^2 \left\{A_K' \Omega_K^{-1} A_K\right\}^{-1}$ is independent of the dimension K and exceeds the variance of the limit distribution of the OLS estimator (when x_t is exogenous), which is $\frac{\sigma^2}{\sigma_v^2}(1-\rho^2)$ for all $|\rho| < 1$ and all K. In this linear model case, the Markov property of x_t ensures that additional lagged values of the regressor (beyond x_{t-1}) do not contribute further to reducing the variance of the IV estimator beyond that of the instrument $x_{t-1.}$

4.2 Integrated regressor case

By comparison we now consider the use of lagged instruments in the integrable function model case. In particular, suppose K lagged values of x_t, i.e. $x_{t-1}, ..., x_{t-K}$, are used to construct instruments based on certain specified integrable functions. To fix ideas we consider the IV estimator

$$\hat{\beta} = \arg\min_{\beta \in \mathbb{R}} \hat{Q}_n(\beta), \tag{26}$$

where the objective function is

$$Q_n(\beta) = n^{-1} \left[\sum_{t=1}^{n} Z_t \left(y_t - \beta f(x_t)\right)\right]' W_n^{-1} \left[\sum_{t=1}^{n} Z_t \left(y_t - \beta f(x_t)\right)\right], \tag{27}$$

where $Z_t' = [g_1(x_{t-1}), ..., g_K(x_{t-K})]$, $g_i \in L_1$, and W_n is some weight matrix. We do not consider here the more general case where the nonlinear functions in Z_t may themselves depend on the unknown parameters β. Define the observation matrices

$$X = [f(x_1), ..., f(x_n)]', \; Z = [Z_1, ..., Z_n]' \text{ and } Y = [y_1, ..., y_n]'.$$

The generalised IV (GIV) estimator of β from (26) and (27) is

$$\hat{\beta} = \left[X' Z W_n^{-1} Z' X\right]^{-1} X' Z W_n^{-1} Z' Y.$$

When $W_n = n^{-1/2} \sum_{t=1}^{n} Z_t Z_t'$ the optimal GIV is $\hat{\beta} = [X' P_Z X]^{-1} X' P_Z Y$. The following result gives the limit distribution of $\hat{\beta}$ when $x_t \sim I(1)$.

Theorem 2 *Suppose that Assumptions 1 and 2 hold, and for $k = 1, ..., K$:*

(i) $g_k, g_k^2, g_k^4 \in L_1$ *and are Lipschitz continuous,*
(ii) $f, \tilde{f} \in L_1$,
(iii) $g_k(s)\widetilde{\left(\tilde{f}\varphi\right)}(-s)$ *is Lipschitz continuous.*

Then, as $n \to \infty$

$$n^{1/4}\left(\hat{\beta}-\beta\right) \xrightarrow{d} MN\left(\sigma^2 L_x(1,0)^{-1}\left\{A_K'\Omega_K^{-1}A_K\right\}^{-1}\right), \tag{28}$$

where

$$A_K' = \left[\int_{-\infty}^{\infty} \mathbf{E}f\left(s+\Delta x_t\right)g_1(s)ds, ..., \mathbf{E}\int_{-\infty}^{\infty} f\left(s+\Delta^K x_t\right)g_K(s)ds\right]$$

and

$$\Omega_K = \int_{-\infty}^{\infty}\begin{bmatrix} g_1(s)^2 & \mathbf{E}g_1(s+\Delta x_t)g_2(s) & . & \mathbf{E}g_1(s+\Delta^{K-1}x_t)g_K(s) \\ \mathbf{E}g_2(s+\Delta x_t)g_1(s) & g_2(s)^2 & & . \\ . & & . & . \\ \mathbf{E}g_K(s+\Delta^{K-1}x_t)g_K(s) & . & . & g_K(s)^2 \end{bmatrix} ds$$

In this case, the relevance of each lagged instrument x_{t-k} tends to deteriorate as the lag k increases as we now show. In particular, we have the asymptotic representation

$$\int_{-\infty}^{\infty} \mathbf{E}f\left(s+\Delta^k x_t\right)g_k(s)ds = \frac{1}{\sqrt{2\pi}}\frac{\tilde{f}(0)\,\tilde{g}_k(0)}{\sigma_v\sqrt{k}}\left\{1+O\left(k^{-1}\right)\right\}. \tag{29}$$

To show (29), we proceed first under the assumption that $v_j \sim iidN\left(0,\sigma_v^2\right)$ as follows

$$\begin{aligned}
\int_{-\infty}^{\infty} \mathbf{E}f\left(s+\Delta^k x_t\right)g_k(s)ds &= \frac{1}{2\pi}\int_{-\infty}^{\infty}\mathbf{E}\int_{-\infty}^{\infty} e^{i\lambda\left\{s+\sum_{j=t-k+1}^{t} v_j\right\}}\tilde{f}(\lambda)\,d\lambda g_k(s)ds \\
&= \frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{is\lambda}E\left\{e^{i\lambda\sum_{j=t-k+1}^{t} v_j}\right\}\tilde{f}(\lambda)\,d\lambda g_k(s)ds \\
&= \frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{is\lambda}e^{-\sigma_v^2\lambda^2 k/2}\tilde{f}(\lambda)\,d\lambda g_k(s)ds \\
&= \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-\sigma_v^2\lambda^2 k/2}\tilde{f}(\lambda)\int_{-\infty}^{\infty} e^{is\lambda}g_k(s)dsd\lambda \\
&= \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-\sigma_v^2\lambda^2 k/2}\tilde{f}(\lambda)\,\tilde{g}_k(-\lambda)\,d\lambda \qquad (30) \\
&= \frac{1}{\sqrt{2\pi}}\frac{\tilde{f}(0)\,\tilde{g}_k(0)}{\sigma_v\sqrt{k}} + O\left(k^{-1}\right), \qquad (31)
\end{aligned}$$

by Laplace approximation as $k \to \infty$. This result also applies in the non Gaussian case where $v_t \sim iid\left(0, \sigma^2\right)$ has characteristic function $cf_v\left(\lambda\right)$ and finite moments to the third order. In this case

$$E\left\{e^{i\lambda\sum_{j=t-k+1}^{t} v_j}\right\} = cf_v\left(\lambda\right)^k = e^{k\varphi(\lambda)} = e^{k\left\{-\frac{\lambda^2\sigma^2}{2}+o\left(\lambda^2\right)\right\}} = e^{-k\frac{\lambda^2\sigma^2}{2}}\left\{1+o\left(\lambda^2\right)\right\},$$

since the second characteristic $\varphi\left(\lambda\right)$ of v_t has a valid power series expansion. Then, by formal Laplace approximation (e.g. Bleistein and Handelsman, 1986, chapter 5), we again have

$$\begin{aligned}
\int_{-\infty}^{\infty} \mathbf{E} f\left(s+\Delta^k x_t\right) g_k(s) ds &= \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{k\varphi(\lambda)} \tilde{f}\left(\lambda\right)\tilde{g}_k\left(-\lambda\right) d\lambda \\
&= \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-k\frac{\sigma^2\lambda^2}{2}}\left\{1+o\left(\lambda^2\right)\right\}\tilde{f}\left(\lambda\right)\tilde{g}_k\left(-\lambda\right) d\lambda \\
&= \frac{1}{\sqrt{2\pi}}\frac{\tilde{f}\left(0\right)\tilde{g}_k\left(0\right)}{\sigma\sqrt{k}}\left\{1+O\left(k^{-1}\right)\right\}. \qquad (32)
\end{aligned}$$

For the explicit model with $f(x) = g_k(x) = \exp(-x^2)$ we have

$$\begin{aligned}
\tilde{f}\left(\lambda\right) &= \int_{-\infty}^{\infty} e^{-is\lambda}e^{-s^2} ds = e^{\frac{1}{4}i^2\lambda^2}\int_{-\infty}^{\infty} e^{-\left(s+\frac{1}{2}i\lambda\right)^2} ds \\
&= \sqrt{2\pi}e^{-\frac{1}{4}\lambda^2}\frac{\left(1/2\right)^{1/2}}{\sqrt{2\pi}\left(1/2\right)^{1/2}}\int_{-\infty}^{\infty} e^{-\frac{2}{2}\left(s+\frac{1}{2}i\lambda\right)^2} ds \\
&= \sqrt{\pi}e^{-\frac{1}{4}\lambda^2}. \qquad (33)
\end{aligned}$$

Then by direct calculation

$$\begin{aligned}
\int_{-\infty}^{\infty} \mathbf{E} f\left(s+\Delta^k x_t\right) f(s) ds &= \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-\sigma_v^2\lambda^2 k/2}\tilde{f}\left(\lambda\right)\tilde{f}\left(-\lambda\right) d\lambda \\
&= \frac{1}{2}\int_{-\infty}^{\infty} e^{-\sigma_v^2\lambda^2 k/2}e^{-\frac{1}{2}\lambda^2} d\lambda \\
&= \frac{1}{2}\int_{-\infty}^{\infty} e^{-\frac{1}{2}\lambda^2\left(1+\sigma_v^2 k\right)} d\lambda \\
&= \frac{\sqrt{2\pi}}{2}\frac{\left(1+\sigma_v^2 k\right)^{-1/2}}{\sqrt{2\pi}\left(1+\sigma_v^2 k\right)^{-1/2}}\int_{-\infty}^{\infty} e^{-\frac{1}{2}\lambda^2\left(1+\sigma_v^2 k\right)} d\lambda \\
&= \sqrt{\frac{\pi}{2}}\frac{1}{\left(1+\sigma_v^2 k\right)^{1/2}}
\end{aligned}$$

which accords with the general formula (32) above as $k \to \infty$ since $\tilde{f}(0)\,\tilde{g}_k(0) = \pi$ from (33). Thus, the autocovariances decay according to the power law $1/k^{1/2}$, just like those of a fractionally integrated process with memory parameter $d = 1/4$.

Further, in the stationary case, the limit variance is known to be optimal if all relevant instruments are included. This is also true in the present case when $x_t \sim I(1)$.[3]

4.3 IV Limit Theory when $K \to \infty$

We consider the limit behavior of the variance in (28) as the number of instrument functions $K \to \infty$. We first observe that, unlike the stationary case where the limit variance (25) is independent of K, the variance in (28) in the nonstationary case does depend on K and is decreasing in K so that $\Omega_{IV}^{K=\infty} < \Omega_{IV}^{K=1}$.

It is convenient to consider a special case where explicit formula are available. Accordingly, we consider the model

$$\begin{aligned} y_t &= \beta f(x_t) + u_t, \;\; f(x) = e^{-x^2} \\ x_t &= x_{t-1} + v_t \end{aligned}$$

with instrument functions $g_k(x_{t-k}) = f(x_{t-k})$ for all $k = 1, ..., K$. The limit distribution of the IV estimator $\hat{\beta}$ is given in (28) where in this case the key element in the limit variance is the Toeplitz form $A_K' \Omega_K^{-1} A_K = tr\left\{\Omega_K^{-1} A_K A_K'\right\}$ whose components are the vector

$$A_K' = \sqrt{\frac{\pi}{2}} \left[\frac{1}{(1+\sigma_v^2)^{1/2}}, \frac{1}{(1+2\sigma_v^2)^{1/2}}, ..., \frac{1}{(1+K\sigma_v^2)^{1/2}} \right],$$

[3] Write the limit variance of the IV estimator with m instruments ($m < K$) as

$$\sigma^2 L(1,0)^{-1} \left\{ (A^*)' (\Omega^*)^{-1} A^* \right\}^{-1}$$

where

$$A^* = RA,\; \Omega^* = R\Omega R',\; R = \begin{bmatrix} I_m & \mathbf{0} \end{bmatrix}.$$

Set $C = \Omega^{-1/2} A$, $D = \Omega^{1/2} R'$ and $P_D = I - D(D'D)^{-1} D' \geq 0$. Then

$$\begin{aligned} 0 &\leq C' \left\{ I - D(D'D)^{-1} D \right\} C \\ &= A'\Omega^{-1}A - (A'R')(R\Omega R')^{-1}(RA) \\ &= A'\Omega^{-1}A - (A^*)'(\Omega^*)^{-1}A^*. \end{aligned}$$

and Toeplitz matrix Ω_K whose (i,j)th element is

$$\gamma_h = \sqrt{\frac{\pi}{2}} \frac{1}{\left(1 + |h|\, \sigma_v^2\right)^{1/2}}, \ \ h = i - j.$$

We define the function f_Ω corresponding to Ω_K by the Fourier series constructed from the coefficients γ_h in the Toeplitz matrix, viz.,

$$\begin{aligned} f_\Omega (x) &= \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \gamma_h e^{-ihx} = \frac{\gamma_0}{2\pi} + \frac{1}{\pi}\sqrt{\frac{\pi}{2}} \sum_{h=1}^{\infty} \frac{\cos(hx)}{\left(1 + h\sigma_v^2\right)^{1/2}} \\ &= \frac{1}{2\sqrt{2\pi}} + \frac{1}{\sqrt{2\pi}} \sum_{h=1}^{\infty} \frac{\cos(hx)}{\left(1 + h\sigma_v^2\right)^{1/2}}, \end{aligned}$$

which converges and is continuous for all $x \neq 0$ with the following behavior in the neighborhood of the origin

$$f_\Omega (x) = \frac{1}{2\sqrt{2\pi}} + \frac{1}{\sqrt{2\pi}} \sum_{h=1}^{\infty} \frac{\cos(hx)}{\left(1 + h\sigma_v^2\right)^{1/2}} \tag{34}$$

$$\begin{aligned} &\sim \frac{1}{\sqrt{2\pi}\sigma_v} \frac{\Gamma\left(\frac{1}{2}\right) \sin\left(\frac{\pi}{4}\right)}{x^{1/2}} + O(1), \quad \text{for } x \sim 0 \\ &= \frac{1}{2\sigma_v} \frac{1}{x^{1/2}} + O(1), \quad \text{for } x \sim 0 \end{aligned} \tag{35}$$

in view of the well known formula (e.g., Zygmund, 1959, p.70)

$$\sum_{j=1}^{\infty} \frac{\cos(jx)}{j^\alpha} = \frac{\Gamma(1-\alpha)\sin\left(\frac{\pi\alpha}{2}\right)}{x^{1-\alpha}} + O(1) \quad \text{for } x \in (0, \pi].$$

Similarly, the kth element of the vector A_K is $a_k = \sqrt{\frac{\pi}{2}} \frac{1}{\left(1+k\sigma_v^2\right)^{1/2}}$ and setting $a_0 = 0$ we have the corresponding series

$$a(x) = \sum_{h=0}^{\infty} a_h e^{-ihx} = \sqrt{\frac{\pi}{2}} \sum_{h=1}^{\infty} \frac{\cos(hx) - i\sin(hx)}{\left(1 + h\sigma_v^2\right)^{1/2}},$$

which again converges for $x \neq 0$, noting that

$$\sum_{j=1}^{\infty} \frac{\sin(jx)}{j^\alpha} = \frac{\Gamma(1-\alpha)\cos\left(\frac{\pi\alpha}{2}\right)}{x^{1-\alpha}} + O(1) \quad \text{for } x \in (0, \pi].$$

Thus, for $x \sim 0$ we have

$$\begin{aligned} a(x) &= \sqrt{\frac{\pi}{2}} \sum_{h=1}^{\infty} \frac{\cos(hx) - i\sin(hx)}{(1+h\sigma_v^2)^{1/2}} \\ &\sim \sqrt{\frac{\pi}{2}} \frac{\Gamma\left(\frac{1}{2}\right)\left\{\sin\left(\frac{\pi}{4}\right) - i\cos\left(\frac{\pi}{4}\right)\right\}}{x^{1/2}\sigma_v} + O(1) \\ &= \frac{\pi}{2} \frac{1-i}{x^{1/2}\sigma_v} + O(1) \end{aligned} \tag{36}$$

To evaluate the quadratic form $A_K' \Omega_K^{-1} A_K$ we transform the expression as $A_K' U U^* \Omega_K^{-1} U U^* A_K$ where $U^* = \overline{U}'$ is the complex conjungate transpose of U and U is the unitary matrix with elements $u_{jk} = K^{-1/2} e^{i\omega_j k}$, where $\omega_j = \frac{2\pi j}{K}$, $j = 1, 2, ..., K$. The jth element of $\sqrt{K} U^* A_K$ has the following form for large K

$$\sum_{k=1}^{K} e^{-i\omega_j k} a_k \sim \sum_{k=1}^{\infty} e^{-i\omega_j k} a_{kj} \sim \sqrt{\frac{\pi}{2}} \sum_{j=1}^{\infty} \frac{\cos(k\omega_j) - i\sin(k\omega_j)}{(1+k\sigma_v^2)^{1/2}} = a(\omega_j),$$

and the same transform is known to approximately diagonalize Ω_K^{-1} with jth diagonal element of $U^* \Omega_K^{-1} U$ being $\{2\pi f_\Omega(\omega_j)\}^{-1}$ (see Hannan and Deistler, 1988, page 224; and, in the long memory case for unbounded spectrum at $\omega \sim 0$, Dahlhaus, 1989, and Lieberman and Phillips, 2005). Then

$$\begin{aligned} A_K' \Omega_K^{-1} A_K &= A_K' U U^* \Omega_K^{-1} U U^* A_K \simeq \frac{1}{2\pi} \sum_{k=1}^{K} \frac{|a(\omega_j)|^2}{2\pi f_\Omega(\omega_j)} \frac{2\pi}{K} \\ &\to \frac{1}{(2\pi)^2} \int_{-\pi}^{\pi} \frac{|a(\omega)|^2}{f_\Omega(\omega)} d\omega. \end{aligned}$$

It follows that the limit variance of the IV estimator as $K \to \infty$ is

$$\frac{\sigma^2 (2\pi)^2}{L_x(1,0)} \left\{ \int_{-\pi}^{\pi} \frac{a_1(\omega)^2 + a_2(\omega)^2}{f_\Omega(\omega)} d\omega \right\}^{-1}, \tag{37}$$

where

$$a_1(\omega) = \sqrt{\frac{\pi}{2}} \sum_{j=1}^{\infty} \frac{\cos(k\omega)}{(1+k\sigma_v^2)^{1/2}}, \quad a_2(\omega) = \sqrt{\frac{\pi}{2}} \sum_{j=1}^{\infty} \frac{\sin(k\omega)}{(1+k\sigma_v^2)^{1/2}},$$

and $f_\Omega(\omega)$ is given in (34).

>From the above formulae we see that $\Omega_{IV}^{K=\infty} = O(\sigma_v^2)$**,** whereas from Example 3 we have $\Omega_{IV}^{K=1} = O(\sigma_v^3)$, so that large K instrumentation reduces variance in IV estimation relative to $K = 1$ as σ_v^2 increases.

5 Conclusion

The present paper concentrates on IV estimation of structural relations which involve integrable functions of a nonstationary regressor. The instruments involve lagged values of the regressor and the limit theory reveals how instrument relevance weakens as the regressor signal strengthens leading to a deterioration in the performance of this type of IV regression. The relevance of instruments that are based on integrable functions of lagged nonstationary regressors is shown to decay slowly with the lag according to a power law like that of long range dependence with a memory parameter $d = 1/4$. Hence, persistence in the regressor ensures that instruments remain (weakly) relevant at long lags and that the contribution to variance reduction in IV estimation continues when all such instruments are included in the regression, reaching a well defined limit in the case of infinitely many weak instruments.

6 Appendix

Lemma A

Suppose that the random sequence of functions $T_n : \mathbb{R} \to \mathbb{C}$ satisfies:

(i) $\|T_n(\lambda)\|_1 \to 0$ *as* $n \to \infty$, *a.e. w.r.t. Lebesgue measure;*

(ii) $\|T_n(\lambda)\|_1 \leq T_o(\lambda)$ *a.e. w.r.t. Lebesgue measure, with* $\int T_o(\lambda)d\lambda < \infty$,

(iii) $\int |T_n(\lambda)|\, d\lambda \leq Y_n$, *with* $\mathbf{E}Y_n < \infty$ *and* $|T_n(.)|$ *continuous. Then, as* $n \to \infty$,

$$\left\| \int T_n(\lambda)d\lambda \right\|_1 \to 0.$$

Proof of Lemma A By (ii) $\int \|T_n(\lambda)\|_1\, d\lambda < \infty$. Next Fatou's Lemma gives

$$\begin{aligned} 0 &= \int \liminf_{n\to\infty} \|T_n(\lambda)\|_1\, d\lambda \leq \liminf_{n\to\infty} \int \|T_n(\lambda)\|_1\, d\lambda \\ &\leq \limsup_{n\to\infty} \int \|T_n(\lambda)\|_1\, d\lambda \leq \int \limsup_{n\to\infty} \|T_n(\lambda)\|_1\, d\lambda = 0. \end{aligned}$$

Hence, $\lim_{n\to\infty} \int \|T_n(\lambda)\|_1\, d\lambda = 0$. Next,

$$\left\| \int T_n(\lambda)d\lambda \right\|_1 \leq \mathbf{E} \int |T_n(\lambda)|\, d\lambda = \int \|T_n(\lambda)\|_1\, d\lambda$$

where the last equality is due to (iii) (e.g. Davidson (1994), Theorem 9.32). ■

Proof of Proposition 1: First we shall show (i). Note that because $\mathbf{E}\left(e^{i\lambda v_t}\right) \geq 0$ we can define the following measure $\mu(d\lambda) = \mathbf{E}\left(e^{i\lambda v_t}\right) d\lambda$. Then write

$$\left[\int_{-\infty}^{\infty} g(s)\mathbf{E}f(s+v_t)ds\right]^2 = \left[\int_{-\infty}^{\infty} \tilde{g}(-\lambda)\tilde{f}(\lambda)\,\mathbf{E}\left(e^{i\lambda v_t}\right) d\lambda\right]^2$$

$$= \left[\int_{-\infty}^{\infty} \tilde{g}(-\lambda)\tilde{f}(\lambda)\,\mu(d\lambda)\right]^2 \leq \left[\int_{-\infty}^{\infty} \tilde{g}(-\lambda)\overline{\tilde{g}(-\lambda)}\mu(d\lambda)\right]\left[\int_{-\infty}^{\infty} \tilde{f}(\lambda)\overline{\tilde{f}(\lambda)}\mu(d\lambda)\right]$$

The first equality above follows from Fubini's Theorem (which in turn holds because the integrand is non-negative). Now, note that because $f, g \in \mathbb{R}$, the complex conjugate of the Fourier transforms are (e.g. Lang, 1993)

$$\overline{\tilde{g}(-\lambda)} = \tilde{g}(-(-\lambda)) = \tilde{g}(\lambda) \text{ and } \overline{\tilde{f}(\lambda)} = \tilde{f}(-\lambda).$$

Therefore,

$$\left[\int_{-\infty}^{\infty} \tilde{g}(-\lambda)\tilde{f}(\lambda)\,\mu(d\lambda)\right]^2 \leq \left[\int_{-\infty}^{\infty} \tilde{g}(-\lambda)\tilde{g}(\lambda)\,\mathbf{E}\left(e^{i\lambda v_t}\right) d\lambda\right]\left[\int_{-\infty}^{\infty} \tilde{f}(\lambda)\tilde{f}(-\lambda)\,\mathbf{E}\left(e^{i\lambda v_t}\right) d\lambda\right]$$

(by Cauchy-Swartz)

$$= \left[\int_{-\infty}^{\infty} g(\lambda)\mathbf{E}g(\lambda+v_t)d\lambda\right]\left[\int_{-\infty}^{\infty} f(\lambda)\mathbf{E}f(\lambda+v_t)d\lambda\right]$$

and this shows (i).

Next consider

$$\frac{\int_{-\infty}^{\infty} g(s)^2ds}{\left\{\int_{-\infty}^{\infty} g(s)\mathbf{E}f(s+v_t)ds\right\}^2} \geq \frac{\int_{-\infty}^{\infty} g(s)^2ds}{\left\{\int_{-\infty}^{\infty} g(s)\mathbf{E}g(s+v_t)ds\right\}\left\{\int_{-\infty}^{\infty} f(s)\mathbf{E}f(s+v_t)ds\right\}} \text{ (by (i))}$$

$$= \frac{\int_{-\infty}^{\infty} g(s)^2ds}{\left\{\int_{-\infty}^{\infty} f(s)\mathbf{E}f(s+v_t)ds\right\}^2}\frac{\int_{-\infty}^{\infty} f(s)\mathbf{E}f(s+v_t)ds}{\int_{-\infty}^{\infty} g(s)\mathbf{E}g(s+v_t)ds}$$

$$= \frac{\int_{-\infty}^{\infty} f(s)^2ds}{\left\{\int_{-\infty}^{\infty} f(s)\mathbf{E}f(s+v_t)ds\right\}^2}\frac{\int_{-\infty}^{\infty} g(s)^2ds}{\int_{-\infty}^{\infty} f(s)^2ds}\frac{\int_{-\infty}^{\infty} f(s)\mathbf{E}f(s+v_t)ds}{\int_{-\infty}^{\infty} g(s)\mathbf{E}g(s+v_t)ds}$$

$$\geq \frac{\int_{-\infty}^{\infty} f(s)^2ds}{\left\{\int_{-\infty}^{\infty} f(s)\mathbf{E}f(s+v_t)ds\right\}^2}$$

where the last inquality follows from the assumption

$$\frac{\int_{-\infty}^{\infty} g(s)^2 ds}{\int_{-\infty}^{\infty} f(s)^2 ds} \frac{\int_{-\infty}^{\infty} f(s)\mathbf{E}f(s+v_t)ds}{\int_{-\infty}^{\infty} g(s)\mathbf{E}g(s+v_t)ds} \geq 1$$

i.e.

$$\frac{\int_{-\infty}^{\infty} f(s)^2 ds}{\int_{-\infty}^{\infty} f(s)\mathbf{E}f(s+v_t)ds} \leq \frac{\int_{-\infty}^{\infty} g(s)^2 ds}{\int_{-\infty}^{\infty} g(s)\mathbf{E}g(s+v_t)ds}.$$

Proof of Theorem 1

(a) By Fourier inversion (e.g. Lang (1993) Theorem 5.1) we get

$$\begin{aligned}
\frac{1}{\sqrt{n}}\sum_{t=1}^{n} g(z_t) f(x_t) &= \frac{1}{\sqrt{n}}\sum_{t=1}^{n} g(x_{t-1}) f(x_{t-1}+v_t) \\
&= \frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{1}{\sqrt{n}}\sum_{t=1}^{n} g(x_{t-1}) e^{i\lambda x_{t-1}+i\lambda v_t} \tilde{f}(\lambda)d\lambda \\
&= \frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{1}{\sqrt{n}}\sum_{t=1}^{n} g(x_{t-1}) e^{i\lambda x_{t-1}}\mathbf{E}\left(e^{i\lambda v_t}\right) \tilde{f}(\lambda)d\lambda \\
&\quad + \frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{1}{\sqrt{n}}\sum_{t=1}^{n} g(x_{t-1}) e^{i\lambda x_{t-1}}\left[e^{i\lambda v_t} - \mathbf{E}\left(e^{i\lambda v_t}\right)\right] \tilde{f}(\lambda)d\lambda \\
:&= \frac{1}{\sqrt{n}}\sum_{t=1}^{n} h(x_{t-1}) + \int_{-\infty}^{\infty} T_n(\lambda)d\lambda.
\end{aligned}$$

It can be easily checked that $h \in L_1$. Therefore, by (iii) and Theorem 3.2 of P&P we get.

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n} h(x_{t-1}) \xrightarrow{p} L_x(1,0)\int_{-\infty}^{\infty} h(s)ds.$$

In view of this it suffices to show that

$$\int_{-\infty}^{\infty} T_n(\lambda)d\lambda = o_p(1),$$

which we now do.

Write $z(t,\lambda) \equiv e^{i\lambda v_t} - \mathbf{E}\left(e^{i\lambda v_t}\right)$. Then,

$$\begin{aligned}
\mathbf{E}\left|T_n(\lambda)\right|^2 &\equiv \mathbf{E}\left|\frac{1}{2\pi}\frac{1}{\sqrt{n}}\sum_{t=1}^{n} g(x_{t-1}) e^{i\lambda x_{t-1}} z(t,\lambda)\tilde{f}(\lambda)\right|^2 \\
&= \frac{1}{n}\frac{1}{2\pi}\mathbf{E}\sum_{t=1}^{n}\sum_{j=1}^{n} g(x_{t-1}) g(x_{j-1}) e^{i\lambda x_{t-1}} e^{-i\lambda x_{j-1}} z(t,\lambda)\overline{z(j,\lambda)}\tilde{f}(\lambda)\overline{\tilde{f}(\lambda)}.
\end{aligned}$$

Note that $z(t,\lambda)$, $\overline{z(t,\lambda)}$ are martingale differences w.r.t. $\mathcal{F}_t$, and $\mathbf{E}_{t-1}\left|z(t,\lambda)\right|^2 \leq 1$. Further, by Assumption 1(ii), the density of $t^{-1/2}x_t$, $d_t(x)$, is uniformly bounded (e.g. Pötscher, 2004). Hence,

$$
\begin{aligned}
\mathbf{E}\left|T_n(\lambda)\right|^2 &= \left|\tilde{f}(\lambda)\right|^2 \frac{1}{2\pi}\frac{1}{n}\mathbf{E}\sum_{t=1}^{n} g^2\left(x_{t-1}\right)\mathbf{E}_{t-1}\left|z(t,\lambda)\right|^2 \leq \left|\tilde{f}(\lambda)\right|^2 \frac{1}{n}\mathbf{E}\sum_{t=1}^{n} g^2\left(x_{t-1}\right) \\
&= \left|\tilde{f}(\lambda)\right|^2 \frac{1}{2\pi}\frac{1}{n}\sum_{t=2}^{n}\int_{-\infty}^{\infty} g^2(\sqrt{t-1}x)d_{t-1}(x)dx \\
&\leq \left|\tilde{f}(\lambda)\right|^2 \frac{1}{2\pi}\frac{1}{n}\sum_{t=1}^{n} t^{-1/2}\int_{-\infty}^{\infty} g^2(s)d_t(s/t^{1/2})ds \\
&\leq n^{-1/2}\left|\tilde{f}(\lambda)\right|^2 \frac{1}{2\pi}\left(2\sup_{t\geq 1}\left\|d_t\right\|_{\mathbb{R}}\int_{-\infty}^{\infty} g^2(s)ds + o(1)\right) \to 0. \qquad (38)
\end{aligned}
$$

In addition, it can be easily seen from the above that

$$
\mathbf{E}\left|T_n(\lambda)\right|^2 \leq \sup_{t\geq 1}\left\|d_t\right\|_{\mathbb{R}}\int_{-\infty}^{\infty} g^2(s)ds\left|\tilde{f}(\lambda)\right|^2. \qquad (39)
$$

In view of (38) and (39) we also get

$$
\left\|T_n(\lambda)\right\|_1 \to 0 \text{ and } \int_{-\infty}^{\infty}\left\|T_n(\lambda)\right\|_1 d\lambda \leq \sqrt{\sup_{t\geq 1}\left\|d_t\right\|_{\mathbb{R}}\int_{-\infty}^{\infty} g^2(s)ds}\int_{-\infty}^{\infty}\left|\tilde{f}(\lambda)\right| d\lambda < \infty. \qquad (40)
$$

In addition,

$$
\begin{aligned}
\int_{-\infty}^{\infty}\left|T_n(\lambda)\right| d\lambda &\leq \frac{1}{\sqrt{n}}\frac{1}{2\pi}\int_{-\infty}^{\infty}\sum_{t=1}^{n}\left|g\left(x_{t-1}\right)e^{i\lambda x_{t-1}}z(t,\lambda)\tilde{f}(\lambda)\right| d\lambda \\
&= \frac{1}{\sqrt{n}}\frac{1}{2\pi}\int_{-\infty}^{\infty}\left|\tilde{f}(\lambda)\right|\sum_{t=1}^{n}\left|g\left(x_{t-1}\right)z(t,\lambda)\right| d\lambda \\
&\leq \frac{4}{\sqrt{n}}\frac{1}{2\pi}\int_{-\infty}^{\infty}\left|\tilde{f}(\lambda)\right| d\lambda\sum_{t=1}^{n}\left|g\left(x_{t-1}\right)\right| \quad (\left|z(t,\lambda)\right| < 4).
\end{aligned}
$$

In view of the above, for all n

$$
\mathbf{E}\int_{-\infty}^{\infty}\left|T_n(\lambda)\right| d\lambda < \infty. \qquad (41)
$$

Hence, Lemma 1 together with (40) and (41) implies that

$$
\left\|\int_{-\infty}^{\infty} T_n(\lambda)d\lambda\right\|_1 \to 0.
$$

(b) Consider the martingale $M_n \equiv n^{-1/4} \sum_{t=1}^{n} g(x_{t-1}) u_t$. By Theorem 3.2 of P&P we have

$$M_n \xrightarrow{d} M \equiv \left\{ L_x(1,0) \int_{-\infty}^{\infty} g(s)^2 ds \right\}^{1/2} W, \tag{42}$$

where $W \sim N(0, \sigma^2)$ independent of $L_x(1,0)$. In addition, consider the (discrete) quadratic variation $[M_n] \equiv n^{-1/2} \sum_{t=1}^{n} g^2(x_{t-1}) u_t^2$, of M_n. By Jacod and Shiryaev (1986, VI Corollary 6.7) the following condition

$$\sup_n n^{-1/4} \mathbf{E} \max_{1 \le t \le n} |g(x_{t-1}) u_t| < \infty, \tag{43}$$

ensures that

$$(M_n, [M_n]) \xrightarrow{d} (M, [M]).$$

Notice that for some $\gamma > 2$

$$\begin{aligned}
& n^{-1/4} \mathbf{E} \max_{1 \le t \le n} |g(x_{t-1}) u_t| \\
\le\ & n^{-1/4} \left\{ \mathbf{E} \left(\max_{1 \le t \le n} |g(x_{t-1}) u_t| \right)^{\gamma} \right\}^{1/\gamma} = n^{-1/4} \left\{ \mathbf{E} \left(\max_{1 \le t \le n} |g(x_{t-1}) u_t|^{\gamma} \right) \right\}^{1/\gamma} \\
\le\ & n^{-1/4} \left\{ \mathbf{E} \left(\sum_{t=1}^{n} |g(x_{t-1}) u_t|^{\gamma} \right) \right\}^{1/\gamma} = n^{-1/4} \left\{ \mathbf{E} \left(\sum_{t=1}^{n} |g(x_{t-1})|^{\gamma} \mathbf{E}_{t-1} |u_t|^{\gamma} \right) \right\}^{1/\gamma} \\
\le\ & n^{-1/4} \left\{ C \mathbf{E} \left(\sum_{t=1}^{n} |g(x_{t-1})|^{\gamma} \right) \right\}^{1/\gamma} \le n^{-1/4} \left\{ C \sum_{t=1}^{n} \int_{-\infty}^{\infty} g^{\gamma}(\sqrt{t} x) d_t(x) dx \right\}^{1/\gamma} \\
\le\ & n^{-\frac{\gamma-2}{4\gamma}} \left\{ 2C \sup_t \|d_t\|_{\mathbb{R}} \int_{-\infty}^{\infty} g^{\gamma}(x) dx + o(1) \right\}^{1/\gamma} \to 0,
\end{aligned}$$

which establishes (43). Let $\langle M_n \rangle \equiv n^{-1/2} \sigma^2 \sum_{t=1}^{n} g^2(x_{t-1})$. We have $[M_n] = \langle M_n \rangle + o_p(1)$, because

$$\begin{aligned}
& \mathbf{E} \left| [M_n] - \langle M_n \rangle \right| \\
\le\ & n^{-1/2} \left\{ \mathbf{E} \left(\sum_{t=1}^{n} g^2(x_{t-1}) \left(u_t^2 - \sigma^2 \right) \right)^2 \right\}^{1/2} = n^{-1/2} \left\{ \mathbf{E} \sum_{t=1}^{n} g^4(x_{t-1}) \left(u_t^2 - \sigma^2 \right)^2 \right\}^{1/2} \\
=\ & n^{-1/2} \left\{ \mathbf{E} \left(\sum_{t=1}^{n} g^2(x_{t-1}) \mathbf{E}_{t-1} \left(u_t^2 - \sigma^2 \right) \right)^2 \right\}^{1/2} = n^{-1/2} \left(\mathbf{E} u_t^4 - \sigma^4 \right)^{1/2} \left\{ \mathbf{E} \sum_{t=1}^{n} g^4(x_{t-1}) \right\}^{1/2} \\
\le\ & n^{-1/4} \left(\mathbf{E} u_t^4 - \sigma^4 \right)^{1/2} \left\{ 2 \sup_t \|d_t\|_{\mathbb{R}} \int_{-\infty}^{\infty} g^4(x) dx + o(1) \right\}^{1/2} \to 0.
\end{aligned}$$

Therefore,

$$\left(M_n, \langle M_n \rangle\right) \xrightarrow{d} \left(M, \langle M \rangle\right). \tag{44}$$

Next, the IV estimator

$$\begin{aligned} n^{1/4}\left(\hat{\beta}-\beta\right) &= \frac{n^{-1/4}\sum_{t=1}^{n} g\left(x_{t-1}\right) u_t}{n^{-1/2}\sum_{t=1}^{n} g\left(x_{t-1}\right) f(x_t)} \\ &= \frac{\langle M_n \rangle}{n^{-1/2}\sum_{t=1}^{n} g\left(x_{t-1}\right) f(x_t)} \frac{n^{-1/4}\sum_{t=1}^{n} g\left(x_{t-1}\right) u_t}{\langle M_n \rangle} \equiv A_n B_n. \end{aligned}$$

By part (a) and P&P (Theorem 3.2) we get.

$$A_n \xrightarrow{p} \frac{\sigma^2 \int_{-\infty}^{\infty} g^2(\lambda) d\lambda}{\int_{-\infty}^{\infty} \tilde{g}(-\lambda) \tilde{f}(\lambda) \mathbf{E}\left(e^{i\lambda v_t}\right) d\lambda}.$$

In addition, by (42) and (44)

$$B_n \xrightarrow{d} \sigma^{-2}\left\{L_x(1,0)\int_{-\infty}^{\infty} g(s)^2 ds\right\}^{-1/2} W,$$

which gives the required result.

Proof of Theorem 2. Write

$$\sqrt[4]{n}\left(\hat{\beta}-\beta\right) = \left[n^{-1/2} X'Z\left(ZZ'\right)^{-1} Z'X\right]^{-1} \frac{1}{\sqrt[4]{n}} X'Z\left(ZZ'\right)^{-1} Z'u$$

Then by Theorem 1(i)

$$n^{-1/2} X'Z\left(ZZ'\right)^{-1} Z'X \xrightarrow{d} L(1,0)\left\{A'\Omega^{-1}A\right\}$$

Further, by Theorem 1(i) and the martingale CLT (c.f. the proof of Theorem 1(ii))

$$\begin{aligned} n^{-1/4} X'Z\left(ZZ'\right)^{-1} Z'u &\xrightarrow{d} \sqrt{\sigma^2} L_x(1,0) A'\left\{L_x(1,0)\Omega\right\}^{-1}\left\{L_x(1,0)\Omega\right\}^{1/2} W \\ &= \left(\sigma^2 L_x(1,0)\right)^{1/2} A'\Omega^{-1/2} W \\ &= MN\left(0, \sigma^2 L_x(1,0) A'\Omega^{-1}A\right), \end{aligned}$$

where W is a standard bivariate normal independent of $L(1,0)$. Hence,

$$\begin{aligned} \sqrt[4]{n}\left(\hat{\beta}-\beta\right) &= \left[X'Z\left(ZZ'\right)^{-1} Z'X\right]^{-1} X'Z\left(ZZ'\right)^{-1} Z'u \\ &\xrightarrow{d} MN\left(0, \sigma^2 L_x(1,0)^{-1}\left\{A'\Omega^{-1}A\right\}^{-1} A'\Omega^{-1}A\left\{A'\Omega^{-1}A\right\}^{-1}\right) \\ &= MN\left(0, \sigma^2 L_x(1,0)^{-1}\left\{A'\Omega^{-1}A\right\}^{-1}\right), \end{aligned}$$

as required.■

7 References

Bierens, H. J. and R. A. Gallant (1997) Nonlinear models, Edward Elgar Publishing.

Bleistein, N. and R. A. Handelsman (1976). *Asymptotic Expansions of Integrals.* New York: Holt, Rinehart and Winston.

Borodin A. N. and Ibragimov, I. A. (1995) Limit Theorems for Functionals of Random Walks, *Proceedings of the Steklov Institute of Mathematics,* No. 2 (195).

Chang, Y. and J. Park (2011). Endogeneity in Nonlinear Regressions with Integrated Time Series, *Econometric Reviews* 30, 51-87.

Dahlhaus, R. (1989) Efficient parameter estimation for self-similar processes. *Annals of Statistics*, 28, 1762-1794.

Davidson, J (1994) *Stochastic limit theory*, Oxford

de Jong, R.M. (2002) Nonlinear estimators with integrated regressors but without exogeneity. Unpublished manuscript.

Godfrey, L. G. (1988) *Misspecification tests in econometrics*, Cambridge.

Hannan, E. J. and M. Deistler (1988). *The Statistical Theory of Linear Systems.* New York: John Wiley and Sons.

Hu, L. and P. C. B. Phillips (2004): "Dynamics of the Federal Funds Target Rate: A Nonstationary Discrete Choice Approach," *Journal of Applied Econometric*s, Vol. 19, 2004, pp. 851-867

Ibragimov, R. and P.C.B. Phillips. (2008). "Regression asymptotics using martingale convergence methods." *Econometric Theory* 24 888-947.

Jacod, J. and A.N. Shiryaev (1986) *Limit theorems for stochastic processe*s, Springer-Verlag.

Jeganathan, P. (2006). Limit Theorems for Functionals of Sums That Converge to Fractional Stable Motions, CFDP 1558.

Jeganathan, P. (2008). Limit Theorems for Functionals of Sums That Converge to Fractional Brownian and Stable Motions, CFDP 1649.

Jeganathan, P. and P. C. B. Phillips (2009). A Nonlinear Regression Model with Integrated Time Series (in preparation).

Kasparis, I. (2008). Detection of functional form misspecification in cointegrating relations. *Econometric Theory*, 24 1373-1403.

Lang, S. (1993) *Real and functional analysis*, 3rd ed. Springer-Verlag.

Lieberman, O. and P. C. B. Phillips (2004). Error bounds and asymptotic expansions for Toeplitz product functionals of unbounded spectra. *Journal of Time Series Analysis*, 25, 733-753.

Park, J.Y. and P.C.B. Phillips (2000). "Nonstationary binary choice," *Econometrica*, 68, 1249-1280.

Park, J.Y. and P.C.B. Phillips (2001) Nonlinear regressions with integrated time series. *Econometrica* 69, 117-161.

Phillips, P. C. B., S. Jin and L. Hu (2007). Nonstationary Discrete Choice: A Corrigendum and Addendum", *Journal of Econometrics*, 141, 1115-1130.

Pötscher, B.M. (2004) Nonlinear functions and convergence to Brownian Motion: Beyond the continuous mapping theorem. *Econometric Theory*, 20, 1-22.

Pötscher, B. M. and I.R. Prucha (1997). *Dynamic Nonlinear Econometric Models: Asymptotic Theory.* Springer Verlag

Wang, Q. and P.C.B. Phillips (2009) Asymptotic theory for local time density estimation and nonparametric cointegrating regression. *Econometric Theory*, 25, 710-738.

Zygmund, A. (1958) *Trigonometric Series,* Second Edition. Cambridge.

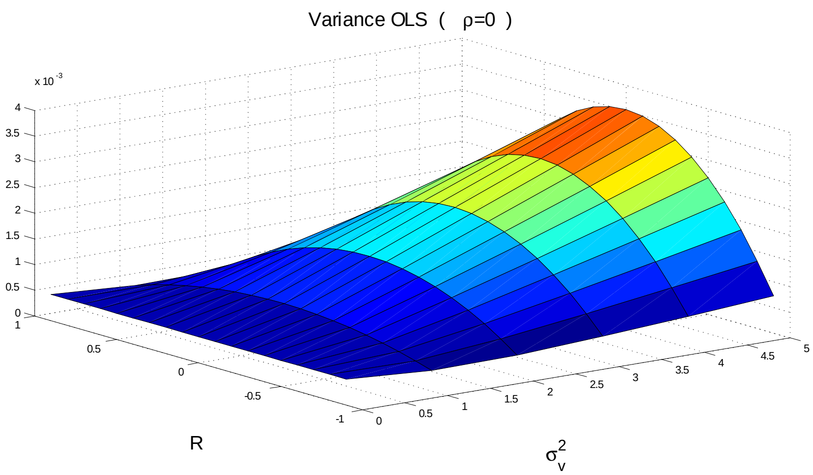

Figure 1: Variance of the OLS estimator ($\rho = 0$)

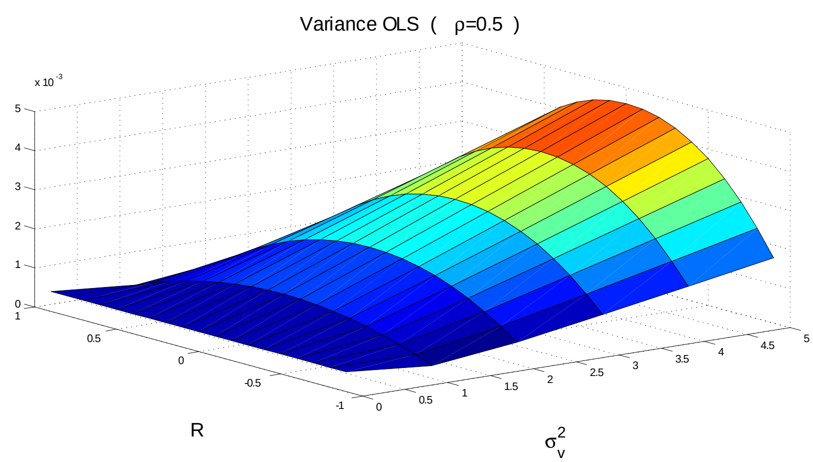

Figure 2: Variance of the OLS estimator ($\rho = 0.5$)

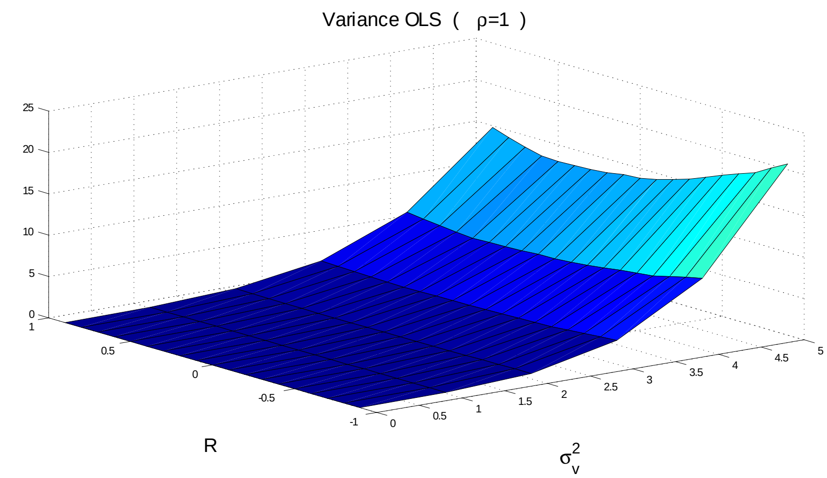

Figure 3: Variance of the OLS estimator ($\rho = 1$)

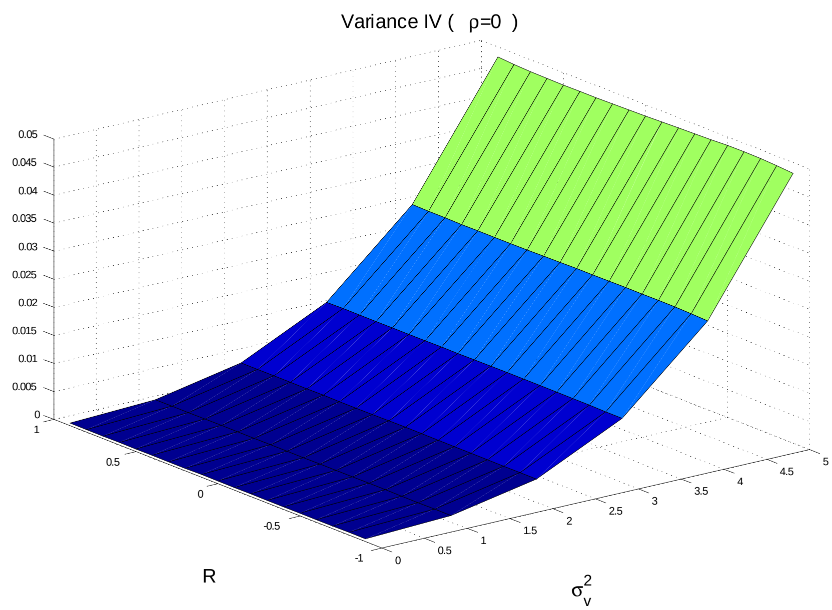

Figure 4: Variance of the IV estimator ($\rho = 0$)

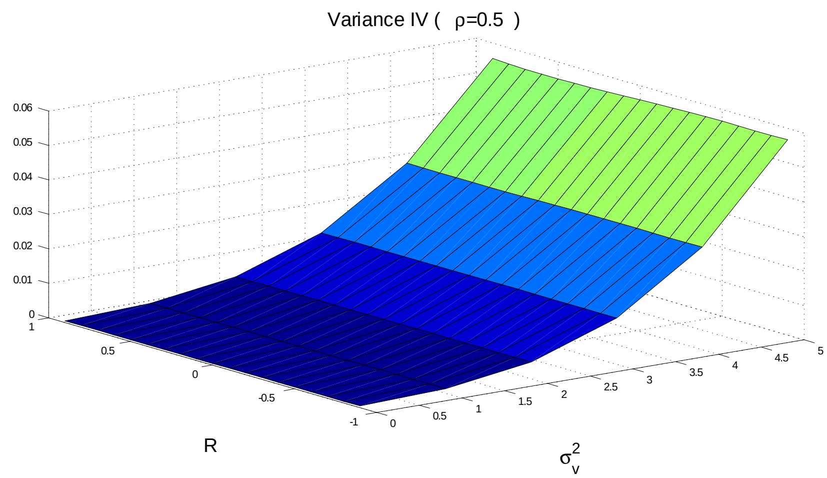

Figure 5: Variance of the IV estimator ($\rho = 0.5$)

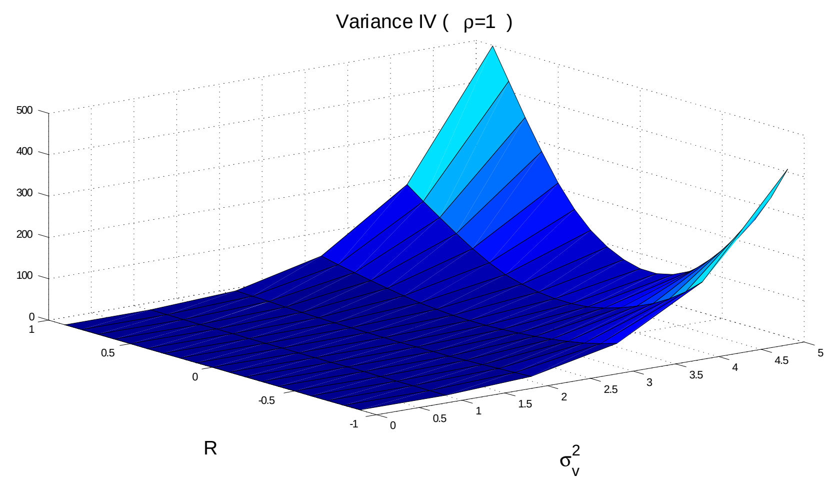

Figure 6: Variance of the IV estimator ($\rho = 1$)

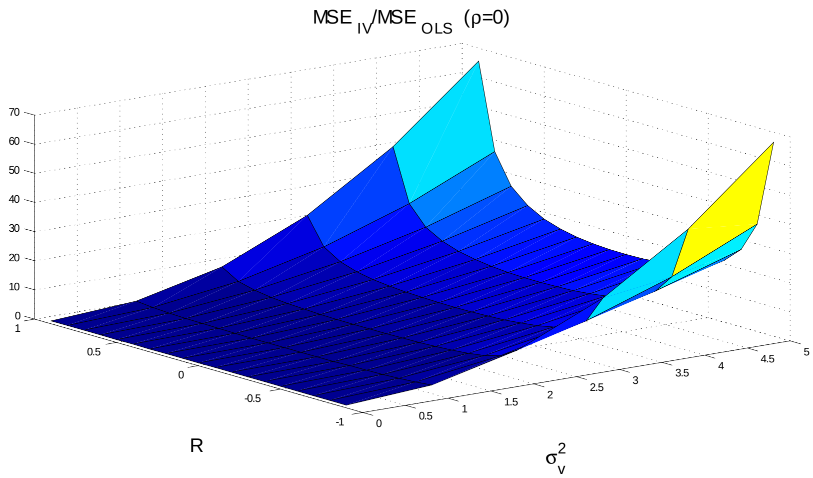

Figure 7: MSE ratio of IV vs OLS ($\rho = 0$)

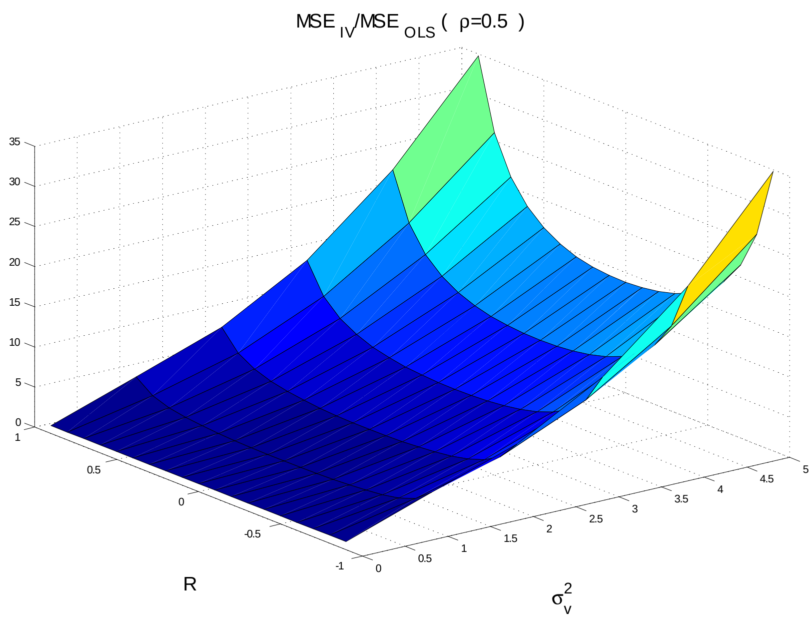

Figure 8: MSE ratio of IV vs OLS ($\rho = 0.5$)

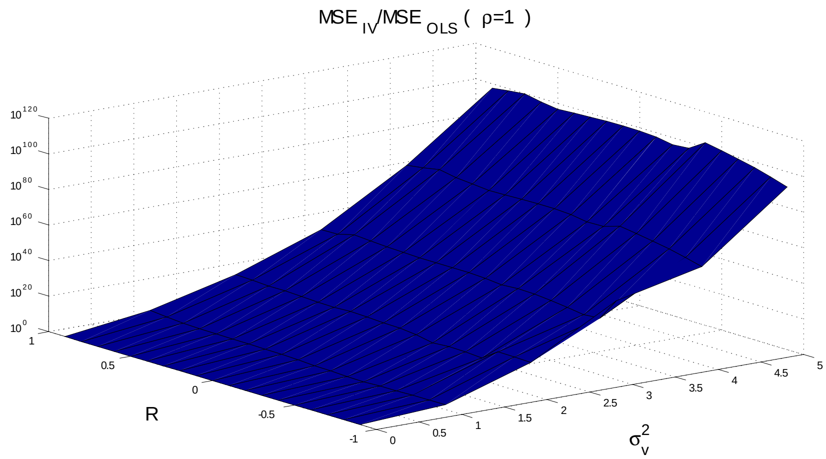

Figure 9: MSE ratio of IV vs OLS ($\rho = 1$)