Instrumental Variables Estimation of Stationary and Nonstationary Cointegrating Regressions

P.M. Robinson[*] and M. Gerolimetto

London School of Economics

Abstract

Instrumental variables estimation is classically employed to avoid simultaneous equations bias in a stable environment. Here we use it to improve upon ordinary least squares estimation of cointegrating regressions between nonstationary and/or long memory stationary variables where the integration orders of regressor and disturbance sum to less than 1, as happens always for stationary regressors, and sometimes for mean-reverting nonstationary ones. Unlike in the classical situation, instruments can be correlated with disturbances and/or uncorrelated with regressors. The approach can also be used in traditional non-fractional cointegrating relations. Various choices of instrument are proposed. Finite sample performance is examined.

Keywords: Cointegration; Instrumental variables estimation; *I (d)* processes.
JEL Classification: C32

1 Introduction

A cointegrating relation of rank $P \geq 1$ between elements of a vector W_t of $P+Q$ observables, for $Q \geq 1$, can be written

$$AW_t = U_t, \quad t = 1, 2, ...,$$ (1.1)

where A is a $P \times (P + Q)$ matrix of rank P and U_t is a $P \times 1$ unobservable sequence. In frequent econometric practice, U_t is assumed to be a vector of $I(0)$ variables, and W_t a vector of $I(1)$ variables. In that case, under minor additional conditions, Stock (1987) showed that ordinary least squares (OLS) estimates of a reduced form of (1.1) are n-consistent, where n is sample size. This outcome does not require orthogonality between the right-hand side variables and U_t, though the limit distribution is non-standard. Estimates with optimality properties which make fuller use of system information have been devised by Johansen (1991), Phillips (1991), for example; these have mixed normal asymptotics and generate Wald statistics with null χ^2 asymptotics. Extensions to allow W_t to include deterministic, $I(0)$ or $I(2)$ variables have been developed (see e.g. Johansen, 1991, Phillips and Hansen, 1990).

However, $I(1)$ and $I(2)$ are particular notions of stochastic trending behaviour, while $I(0)$ is a particular notion of stationarity, and the concept of cointegration is in no way tied to them. Cointegration involving $I(d)$ processes, for real-valued d, has also been of interest, and has been investigated in a number of empirical studies, starting with Cheung and Lai (1993). To discuss this it is convenient to first give a technical definition of $I(0)$ processes: a vector covariance stationary process v_t, $-\infty < t < \infty$, is said to be $I(0)$ if it has zero mean and spectral density matrix that is positive definite and continuous at zero frequency. Now define formally

$$\Delta^{-d} = \sum_{j=0}^{\infty} a_j(d) L^j, \quad a_j(d) = \frac{\Gamma(j + d)}{\Gamma(d)\Gamma(j + 1)},$$ (1.2)

L denoting the lag operator, $\Delta = 1 - L$ and Γ denoting the gamma function, which satisfies $\Gamma(d) = \infty$ for $d = 0, -1, -2, ...$, and $\Gamma(0)/\Gamma(0) = 1$. For any sequence r_t, $-\infty < t < \infty$, define $r_t^{\#} = r_t 1(t \geq 1)$, where $1(.)$ is the indicator function.

We introduce a fairly general concept of an $L \times 1$ vector fractional process $\left(u_t^{(1)}, ..., u_t^{(\ell)} \right)'$, where the $u_t^{(\ell)}$, $\ell = 1, ..., L$, are scalars. If v_t is a $K \times 1$ vector $I(0)$ process, $1 \leq K < \infty$, with k-th element v_{kt}, then for $\ell = 1, ..., L$, $\Delta^{-d_k^{(\ell)}} v_{kt}^{\#}$ is called a basic fractional, $F(d_k^{(\ell)})$ process and

$$u_t^{(\ell)} = \sum_{k=1}^{K} \Delta^{-d_k^{(\ell)}} v_{kt}^{\#}$$ (1.3)

is called an $I(d^{(\ell)})$ process, for $d^{(\ell)} = \max_{1 \leq k \leq K} d_k^{(\ell)}$; d is called the "integration order" of an $I(d)$ process. The $F(d_k^{(\ell)})$ component of (1.3) with the largest $d_k^{(\ell)}$

dominates with respect to asymptotic theory, so reference to other integration orders is suppressed. The truncation in (1.3) is imposed due to lack of convergence of the series (1.2) when $d \geq \frac{1}{2}$; to be consistent it applies for any d, including $d \in [0, \frac{1}{2})$, but for such d we refer to an $I(d)$ as being "stationary", while $d = \frac{1}{2}$ is on the boundary between stationarity and nonstationarity, and $d > \frac{1}{2}$ is the "nonstationary" region. We call $\left(u_t^{(1)}, ..., u_t^{(L)} \right)$ an $I\left(d^{(1)}, ..., d^{(L)} \right)$ process, for $L < \infty$ (where the "degenerate" case $K < L$ is included).

With this terminology we take (W_t', U_t') in (1.1) to be an $I\left(\delta_1, ..., \delta_{P+Q}, \gamma_1, ..., \gamma_P \right)$ process, where cointegration is expressed by

$$0 \leq \gamma_i < \min_{j:a_{ij} \neq 0} \delta_j, \quad i = 1, ..., P, \tag{1.4}$$

where a_{ij} is the (i, j)-th element of A. Note that (1.1) can result from a linear structural model for W_t in terms of unobserved $F(d)$ (or $I(d)$) components (cf. Stock and Watson, 1988).

Assuming either parametric or nonparametric autocorrelation in the underlying $I(0)$ sequence, and allowing the δ_i and γ_i to be unknown, estimates of coefficients of a reduced form of (1.1) with apparently optimal properties have been proposed by Robinson and Hualde (2003), Hualde and Robinson (2004, 2006), in the case $P = Q = 1$, and in a rather different setting by Jeganathan (1999) when $P = 1$, $Q > 1$ but all δ_i are the same; see also Dolado and Marmol (1997), Kim and Phillips (2000). Extending to general P, Q, it seems that when $\delta_j > \frac{1}{2}$ for all j (and incorporating knowledge of "overidentifying" zero elements of A), a non-null a_{ij} can be consistently estimated with rate $n^{\delta_j - \gamma_i}$ when $\delta_j - \gamma_i > \frac{1}{2}$, and with rate $n^{\frac{1}{2}}$ when $\delta_j - \gamma_i < \frac{1}{2}$.

Such estimates are relatively complicated to compute. Also, their dependence on nuisance parameter or nonparametric function estimates might sometimes be associated with poor finite-sample properties, indeed as the nuisance parameters here include integration orders estimates of coefficients of a reduced form of (1.1) might be expected to be worse in finite samples than in the traditional selection in which W_t, U_t are treated as known to be $I(1), I(0)$, respectively. The optimal estimates of Robinson and Hualde (2003), Hualde and Robinson (2004) also depend respectively on a correct parametric specification of the short memory component of series, and smoothing numbers. Moreover the optimal estimates have to be initiated by a simple initial estimate that does not depend on estimates of the γ_i, at least. There is thus still interest in simpler estimates.The most obvious simple estimate is OLS. However, extending results of Robinson and Marinucci (2001) (where $P = Q = 1$), we observe below that in some circumstances OLS converges only slowly, or not at all.

The main purpose of the present paper is to show that simple instrumental variable (IV) estimates, while still having non-standard limit distributions, can have better convergence rates than OLS, and than the narrow band least squares (NBLS) estimates of Robinson (1994), Robinson and Marinucci (2001). In such circumstances, they would thus be expected to provide better initial estimates in optimal procedures. IV estimation has previously been considered in

a cointegration context, mostly in the traditional case of $I(1)$ observables and $I(0)$ cointegrating errors, but typically to provide representations of optimal estimates, that may also have an approximate maximum likelihood interpretation, say (see e.g. Phillips and Hansen, 1990; Kitamura and Phillips, 1995, 1997; Marmol, Escribano and Aparicio, 2002).

Our IV estimates employ only exclusion and normalization restrictions on A, and do not attempt to correct for cross-correlation in U_t, so it suffices to individually estimate the equations of (1.1). The following section compares rates of convergence of OLS and IV estimates of a single equation. Section 3 discusses strategies for selecting instruments. Section 4 examines finite sample performance by means of Monte Carlo simulations. Section 5 offers some concluding remarks.

2 Instrumental Variables Estimates

Consider the first equation of (1.1), with no loss of generality, and write it

$$y_t = \beta' x_t + u_t. \tag{2.1}$$

Here u_t is the first element of U_t, y_t is a scalar and x_t and β are $q \times 1$ vectors, where with no loss of generality we may take $(x_t', y_t)'$ to be the leading $(q+1) \times 1$ sub-vector of W_t, for $q \leq Q$, and no element of β is known to be zero. For brevity write $\gamma = \gamma_1$, and with no loss of generality set

$$\delta_1 \leq \delta_2 \leq ... \leq \delta_q. \tag{2.2}$$

Thus (x_t', y_t, u_t) is an $I(\delta_1, ..., \delta_{q+1}, \gamma)$ process. Note from (1.4) and (2.2) that

$$\gamma < \delta_{q+1} \leq \delta_1, \tag{2.3}$$

δ_{q+1} being the integration order of y_t. Cointegration of (x_t', u_t) is implied by (2.1) and (2.3). We have implicitly adopted the definition used, in a fractional context, by Robinson and Marinucci (2003), rather than, say, that of Johansen (1996) which requires $\gamma < \delta_q$.

Given a $q \times 1$ vector sequence z_t, $t \geq 1$, an IV estimate of β is given by

$$\hat{\beta}_z = M_{zx}^{-1} M_{zy},$$

using the generic notation

$$M_{ab} = \sum_{t=1}^{n} a_t b_t',$$

for any column vectors a_t, b_t, $t \geq 1$, and assuming M_{zx} is non-singular.

We consider first the OLS special case $\hat{\beta}_x$. Denote by G_1 and G_2 the $q \times q$ diagonal matrices whose i-th diagonal elements are respectively

$$n^{\frac{1}{2}} 1(\delta_i < \frac{1}{2}) + (n \log n)^{\frac{1}{2}} 1(\delta_i = \frac{1}{2}) + n^{\delta_i} 1(\delta_i > \frac{1}{2})$$

3

and

$$n1(\gamma+\delta_i < 1)+(n\log n)1(\gamma > 0, \gamma+\delta_i = 1)+n^{\gamma+\delta_i}\{1(\gamma+\delta_i > 1)+1(\gamma = 0, \delta_i = 1)\}.$$

If X_n, $n \geq 1$, and X are matrices of random variables, having the same dimensions, and g_n, $n \geq 1$, are scalars, let $X_n \Rightarrow X$ denote a suitable notion of weak convergence of X_n to X as $n \to \infty$, and $X_n \approx g_n$ denote $X_n/g_n \Rightarrow X$ as $n \to \infty$. Write

$$\delta_* = \min_{i:\delta_i > \frac{1}{2}} \delta_i$$

when the right side exists, and β_i, $\hat{\beta}_{xi}$ for the i-th elements of β, $\hat{\beta}_x$.

Theorem 1 *Let (1.4) and (2.1) hold, and*

$$G_1^{-1}M_{xx}G_1^{-1} \Rightarrow \Omega_{xx}, \quad G_2^{-1}M_{xu} \Rightarrow \Omega_{xu}, \; as \; n \to \infty, \tag{2.4}$$

where Ω_{xx} is a $q \times q$ matrix that is almost surely (a.s.) finite and positive definite, and Ω_{xu} is a $q \times 1$ vector whose elements are a.s. finite. Then as $n \to \infty$, for $i = 1, ..., q$,

$$\hat{\beta}_{xi} - \beta_i \approx 1(\delta_i < \frac{1}{2}) + (\log n)^{-1}1(\delta_i = \frac{1}{2}) + n^{1-\delta_i-\delta_*}1(\delta_i > \frac{1}{2}, \gamma + \delta_* < 1)$$

$$+n^{1-\delta_i-\delta_*}(\log n)1(\gamma > 0, \delta_i > \frac{1}{2}, \gamma + \delta_* = 1) \tag{2.5}$$

$$+n^{\gamma-\delta_i}\{1(\delta_i > \frac{1}{2}, \gamma + \delta_* > 1) + 1(\gamma = 0, \delta_* = 1)\}.$$

Theorem 1 extends the result for $\delta_i = 1$, $\gamma = 0$ (Stock, 1987). As in this case, the limit distribution of $\hat{\beta}_{xi}$ is non-standard and as well as depending on integration orders and other properties varies qualitatively across certain subsets of integration order space, as found by Robinson and Marinucci (2001, 2003). The limit distributions are in no cases normal or mixed-normal, and thus not useful to statistical inference. The proof of Theorem 1 is in the Appendix, where also sufficient conditions for (2.4), and (2.10) below, are discussed.

When $\gamma + \delta_i < 1$, or $\gamma + \delta_i = 1$ with $\gamma > 0$, the i-th element of Ω_{xu} is a constant, and nonzero unless $cov(x_{it}, u_t) = 0$, where x_{it} is the i-th element of x_t. Ω_{xx} has structure

$$\Omega_{xx} = \begin{bmatrix} \Omega_{xx1} & 0 & 0 \\ 0 & \Omega_{xx2} & 0 \\ 0 & 0 & \Omega_{xx3} \end{bmatrix}, \tag{2.6}$$

where Ω_{xxi} is $q_i \times 1$, $i = 1, 2, 3$, such that $q_1 = \sum_{i=1}^q 1(\delta_i < \frac{1}{2})$, $q_2 = \sum_{i=1}^q 1(\delta_i = \frac{1}{2})$, $q_3 = \sum_{i=1}^q 1(\delta_i > \frac{1}{2})$. The matrices Ω_{xx1} and Ω_{xx2} are constant, and positive definite. For i such that $\gamma+\delta_i = 1$ and $\delta_i = \frac{1}{2}$ the constant and nonzero/positive definite properties result, despite the nonstationarity, from domination of the

4

standard deviation by the mean, see Theorems 4.2 and 5.1 of Robinson and Marinucci (2001). The block-diagonal structure in (2.6) does not require any assumptions of orthogonality between elements of x_t, but rather is due to the differing normalisations. The block-diagonality ensures consistency of $\hat{\beta}_{xi}$ for $i > q_1$, despite the inconsistency of $\hat{\beta}_{xi}$ for $i \leq q_1$ (which is due to simultaneous equations bias). For x_{it} on the stationary-nonstationary boundary $\hat{\beta}_{xi}$ is $\log n$-consistent. For nonstationary x_{it} with $\delta_i > \frac{1}{2}$, we obtain the "optimal" $n^{\delta_i - \gamma}$-consistency of $\hat{\beta}_{xi}$ when

$$\gamma + \delta_j > 1 \quad \text{or} \quad \gamma = 0, \delta_j = 1 \tag{2.7}$$

for all j such that $\delta_j > \frac{1}{2}$, but otherwise the rate of $\hat{\beta}_{xi}$ is restricted: not only if $\gamma + \delta_i < 1$ or $\gamma + \delta_i = 1$ with $\gamma > 0$, but also if $\gamma + \delta_i > 1$ and there is a single $\delta_j > \frac{1}{2}$ such that $\gamma + \delta_j < 1$ or $\gamma + \delta_j = 1$ with $\gamma > 0$. The rates in (2.5) seem sharp, except in the event of constant elements of Ω_{xx} or Ω_{xu} taking particular values.

No rate improvement in (2.5) seems possible if $\delta_i > \frac{1}{2}$ and $\gamma + \delta_i > 1$ or $\gamma = 0$ with $\delta_i = 1$ for all i, which includes the traditional case of $I(1)$ regressors and $I(0)$ disturbances. However, while it is common practice to test the $I(1)$ assumption, and it is frequently not rejected, the tests commonly used, such as Dickey-Fuller, do not have particularly good power against fractional alternatives. It may not be possible to rule out the presence of mean-reverting, or even stationary, x_{it}, and visual discrimination based on the observed trajectory can be hazardous. Improvements to OLS are possible when for some i, $\gamma + \delta_i = 1$ with $\gamma > 0$ or $\gamma + \delta_i < 1$, and the latter inequality always holds when x_{it} is stationary or when x_{it} is nonstationary and mean-reverting but $\gamma = 0$ (and also for sufficiently small positive γ). To guard against such possibilities we consider the IV estimate $\hat{\beta}_z$ with z_t an $I(\xi_1, ..., \xi_q)$ process, with

$$\xi_1 \leq \xi_2 \leq \cdots \leq \xi_q. \tag{2.8}$$

For simplicity we consider only cases in which

$$\delta_i + \xi_i > 1, \quad i = 1, ..., q. \tag{2.9}$$

Denote by G_3, G_4 and G_5 the $q \times q$ diagonal matrices with i-th diagonal elements n^{ξ_i}, n^{δ_i} and

$$n1(\gamma + \xi_i < 1) + (n \log n)1(\gamma > 0, \gamma + \xi_i = 1) + n^{\gamma + \xi_i}\{1(\gamma + \xi_i > 1) + 1(\gamma = 0, \xi_i = 1)\},$$

and by $\hat{\beta}_{zi}$ the i-th element of $\hat{\beta}_z$.

Theorem 2 *Let (1.4) and (2.1) hold, and*

$$G_3^{-1} M_{zx} G_4^{-1} \Rightarrow \Omega_{zx}, \quad G_5^{-1} M_{zu} \Rightarrow \Omega_{zu}, \tag{2.10}$$

where Ω_{zx} is an a.s. finite and non-singular matrix and Ω_{zu} is an a.s. finite vector. Then as $n \to \infty$, for $i = 1, ..., q$,

$$
\begin{aligned}
\hat{\beta}_{zi} - \beta_i &\approx n^{1 - \delta_i - \xi_1} 1(\xi_1 + \gamma < 1) + n^{1 - \delta_i - \xi_1}(\log n)1(\gamma > 0, \xi_1 + \gamma = 1) \\
&\quad + n^{\gamma - \delta_i}\{1(\xi_1 + \gamma > 1) + 1(\gamma = 0, \xi_1 = 1)\}.
\end{aligned} \tag{2.11}
$$

The proof is again left to the Appendix. Again, limit distributions are in general non-standard and depend on integration orders and other features, and vary qualitatively across regions of integration order space, as well as with other properties of the instruments z_i. Thus again limit distributions will not be useful in inference, and we have not felt it worthwhile to take up space representing them.

The i-th element of Ω_{zu} is constant when $\gamma + \xi_i \leq 1$, or when $\gamma + \xi_i = 1$ with $\gamma > 0$. In view of (2.8) $\hat{\beta}_{zi}$ is consistent even in case of stationary x_{it}, while the optimal rate $n^{\gamma - \delta_i}$ is achieved for all i when

$$\xi_1 + \gamma > 1 \quad \text{or} \quad \gamma = 0, \ \xi_1 = 1, \tag{2.12}$$

which for $\xi_1 > \delta_*$ is a milder condition than $\delta_* + \gamma > 1$ or $\gamma = 0, \delta_* = 1$ (cf. (2.5)). The dependence of the right side on ξ_1, rather than ξ_i, due to the general non-diagonality of Ω_{zx}.

The NBLS estimate proposed by Robinson (1994) to consistently estimate cointegrating relations with stationary $I(d)$ regressors was also shown by Robinson and Marinucci (2001) to improve convergence rates in the sort of nonstationary environments in which our IV approach provides improvements over OLS. NBLS also has an IV interpretation and is intuitively appealing, but its convergence rate (which depends on the rate of a bandwidth number as $n \to \infty$) is not necessarily better or worse than that of our simple IV estimate $\hat{\beta}_z$ in the stationary regressor case, and it is worse in the nonstationary case when $\gamma + \delta_i < 1$ or $\gamma + \delta_i = 1$ with $\gamma > 0$ for some i, but (2.12) holds. Instead of the arbitrariness of bandwidth choice in the NBLS approach, $\hat{\beta}_z$ suffers from arbitrariness of instrument, but it is computationally simpler, and applied economists may feel more comfortable with its familiar form.

3　Choice of Instrument

In classical, stable environments, instruments are ideally chosen to be orthogonal to disturbances, but highly correlated with the variables they replace. In the unit root cointegration literature this prescription does not apply : instruments and can be correlated with disturbances, and independent of regressors (see, e.g. Phillips, 1986, Phillips and Hansen, 1990). The latter situation is possible in our more general setting, when $\delta_i > \frac{1}{2}, \xi > \frac{1}{2}$ for all i. We identify below several strategies for choosing instruments. All the estimates covered in Theorem 2 are sub-optimal, and it does not seem possible to find a statistically best choice of instrument z_t. Thus there is inevitably considerable ambiguity in the choice of z_t, and our discussion illustrates this, while also identifying other advantages, or disadvantages, of particular ones. In principle the number of possible instruments one might use is limitless, especially as one could use linear combinations of ones suggested below, or lagged versions, or IV versions of the narrow-band estimate (4.1) employed in the Monte Carlo study of the following section.

1. Cointegration in (2.1) requires that at least two elements of (x_t', y_t) have the same integration order. In the special case when this is true of all elements, convergence rates are constant across $\hat{\beta}_x$. There may be other elements of W_t with different integration orders. Robinson and Yajima (2002) investigated "simple cointegration", in which (1.1) consists of sub-systems involving non-overlapping elements of W_t, integration orders being constant within, but not across, sub-systems. Then if they satisfy (2.9), variables in one system might be suitable instruments in another sub-system that involves variables of lower integration order. Robinson and Yajima (2002) provided an algorithm for partitioning W_t into subsets with common integration orders, and thence determining the cointegrating rank of each sub-system.

2. Evidently this approach cannot generate instruments for each equation of (1.1), and uncertainty about integration orders of observables presents some ambiguity. A simple approach is to integrate observables, possibly x_t in (2.1). For simplicity take x_t to be scalar, and define, for any sequence r_t, $r_t^{(\alpha)} = \Delta^{-\alpha} r_t^{\#}$, $\alpha > 0$. Then if x_t is $I(\delta)$, $x_t^{(\alpha)}$ is $I(\delta + \alpha)$. We might thus choose $z_t = x_t^{(\alpha)}$ as an instrument, for suitable α. If it is believed that $\delta > \frac{1}{2}$, then $\alpha_i = \frac{1}{2}$ suffices for (2.9), whereas $\alpha = 1$ suffices for any $\delta > 0$. Clearly $\alpha = 1$ is an attractively simple option.

3. A similar outcome is achieved by instead integrating the model (2.1). Taking $q = 1$ again, and forming

$$y_t^{(\alpha)} = \beta x_t^{(\alpha)} + u_t^{(\alpha)},$$

an IV estimate using x_t as instrument for $x_t^{(\alpha)}$ can be written $\tilde{\beta} = \sum_{t=1}^{n} x_t y_t^{(\alpha)} / \sum_{t=1}^{n} x_t x_t^{(\alpha)}$. In view of (1.2) $\tilde{\beta} = x' B^{(\alpha)} y / x' B^{(\alpha)} x$, where $x = (x_1, ..., x_n)'$, $y = (y_1, ..., y_n)'$ and $B^{(\alpha)}$ is the $n \times n$ lower-triangular matrix whose t-th row is $(a_0(\alpha), a_1(\alpha), ..., a_{t-1}(\alpha), 0, ..., 0)$. But this is also an IV estimate for (2.1) when the instrument for x_t is the t-th element of $B^{(\alpha)'} x$. On the other hand the IV estimate for (2.1) that uses $x_t^{(\alpha)}$ as instrument for x_t is $\hat{\beta}_{x^{(\alpha)}} = x' B^{(\alpha)'} y / x' B^{(\alpha)'} x$. Clearly $\tilde{\beta}$ and $\hat{\beta}_{x^{(\alpha)}}$ have the same convergence rates.

4. It would be possible to generalize our definition of W_t to allow for deterministic effects, in particular each element could be a sum of an $I(d)$ variable and a sequence asymptotic to t^c, for some c, where either stochastic or deterministic components may dominate asymptotically. This suggests using increasing functions of t as instruments. But this can work also for purely stochastic nonstationary z_t, due to a "spurious regression" effect (Phillips, 1988). For example with $q = 1$, using $z_t = t^{\xi - \frac{1}{2}}$ for $\xi > \frac{1}{2}$ produces the same convergence rate as an $I(\xi)$ z_t (albeit a different type of limit distribution, possibly normal).

5. The "spurious regression" phenomenon also suggests artificial generation of instruments. In particular, for nonstationary x_{it}, an $I(\xi_i)$ instrument of the form $\Delta^{-\xi_i}e_t^{\#}$ would suffice, where $\xi_i > \frac{1}{2}$ and $e_1, e_2, ...,$ is a sequence from a Monte Carlo random number generator. Many workers would resist such a device, however, in part because calculations would be difficult to replicate and empirical findings liable to be greeted with suspicion.

4   Finite Sample Behaviour

Our IV proposals are based on consideration of asymptotic rates of convergence, so it is important to investigate performance in finite samples. We focus only on moderate series lengths n, to gauge relevance to macroeconomic analysis. For long financial series our convergence rates would be more directly informative, and simulation evidence of less importance. We compare some of the choices of instrument described in the previous section, including also OLS, as well as NBLS. The latter, when $q = 1$ in (2.1), is

$$\tilde{\beta}_m = \left(\sum_{j=1}^{m} |w_x(\lambda_j)|^2 \right)^{-1} \mathrm{Re} \left\{ \sum_{j=1}^{m} w_x(\lambda_j) w_y(-\lambda_j) \right\}, \qquad (4.1)$$

where

$$w_a(\lambda) = (2\pi n)^{-\frac{1}{2}} \sum_{t=1}^{n} a_t e^{it\lambda}$$

and $\lambda_j = 2\pi j/n$, $1 \leq m < n/2$. The integer m is chosen by the practitioner: for given n, bias tends to increase, and variance to decrease, with m. If (x_t, u_t) is an $I(\delta, \gamma)$ process with $\gamma < \delta$ (cf. (1.4)), then under regularity conditions and with $m \to \infty$, $m/n \to 0$ as $n \to \infty$, $\tilde{\beta}_m$ is consistent for β, and moreover

$$\tilde{\beta}_m - \beta \approx (n/m)^{\gamma-\delta} 1(\delta < \frac{1}{2}) + n^{\gamma-\delta} m^{1-\gamma-\delta} 1(\delta > \frac{1}{2}, \gamma + \delta < 1)$$
$$+ n^{-1} 1(\gamma = 0, \delta = 1) + (\log n/n) 1(\gamma > 0, \gamma + \delta = 1) + n^{\gamma-\delta} 1(\gamma + \delta > 1)$$

(see Robinson, 1994; Robinson and Marinucci 2001). These authors also characterized limit distributions when x_t is nonstationary, and Christensen and Nielsen (2004) recently established asymptotic normality in the stationary case, when also $\gamma + \delta < 1/2$ and x_t and u_t are incoherent at frequency zero. Unlike OLS, NBLS is consistent in case of stationary x_t, whereas for nonstationary x_t NBLS converges faster than OLS when $\gamma + \delta < 1$ or $\gamma > 0$, $\gamma + \delta = 1$, but slower than IV given a suitable choice of instrument (see (2.5), (2.11)). NBLS has been employed in a number of empirical investigations (see e.g. Bandi and Perron, 2004; Christensen and Nielsen, 2004; Robinson and Marinucci, 2003), so it seems appropriate to see how it compares with IV in our numerical study.

Except in our final experiment, all $I(d)$ time series generated were of basic fractional type with $NID(0, 1)$ $I(0)$ source sequence. We allowed for various

integration orders and levels of cross-correlation between series, and compared Monte Carlo bias and standard deviation (SD) across 2000 replications of various estimates computed from $n = 64$ and $n = 128$ observations.

Our first version of (2.1) took $q = 1$, with $\beta = 2$, $(u_t, x_t) \sim I(\gamma, \delta)$. As well as computing OLS ($\hat{\beta}_x$) and NB$_4$ and NB$_5$ (where NB$_m$ is $\tilde{\beta}_m$ (4.1)), we computed three versions of IV $\hat{\beta}_z$, denoted IV$_w$ and IV$_{(\alpha)}$ for $\alpha = 0.5, 1$. IV$_w$ uses $z_t = w_t \sim I(\chi)$ under the assumption that W_t in (1.1) includes an additional variable w_t; IV$_{(\alpha)}$ takes $z_t = x_t^{(\alpha)}$. Thus, IV$_w$, IV$_{(.5)}$ and IV$_{(1)}$ entail $\xi = \chi, \delta + 0.5$ and $\delta + 1$, respectively. We considered four choices of (γ, δ, χ). In case (a) $(\gamma, \delta, \chi) = (0.1, 0.3, 0.7)$. Here, x_t is stationary, w_t satisfies $\delta + \xi = 1$ and $\gamma + \xi < 1$ (cf. (2.9), (2.12)), $x_t^{(.5)}$ satisfies (2.9) and $\gamma + \xi < 1$ (cf. (2.12)), and $x_t^{(1)}$ satisfies (2.9) and (2.12). In case (b), $(\gamma, \delta, \chi) = (0.2, 0.6, 0.9)$. Here, x_t is nonstationary but satisfies $\gamma + \delta < 1$ (cf. (2.7)) while all choices of z_t satisfy (2.9) and (2.12). In case (c) $(\gamma, \delta, \chi) = (0.4, 0.7, 0.8)$. Here, (2.9), (2.9) and (2.12) are satisfied. The same is true in case (d), $(\gamma, \delta, \chi) = (0, 1, 1)$, which covers the usual $I(0)/I(1)$ situation. Convergence rates of the various estimates are displayed in Table 1.

(Table 1 about here)

We also varied the cross-correlation structure of the NID$(0, 1)$ innovations of u_t, x_t, w_t, denoted v_{ut}, v_{xt}, v_{wt}, respectively. In case (i) we took v_{ut}, v_{xt}, v_{wt} to be mutually independent, so that OLS is consistent even when x_t is stationary and w_t is "spurious". In cases (ii)-(iv), $cov(v_{ut}, v_{wt}) = cov(v_{xt}, v_{wt}) = 0.5$ throughout, but we varied $\rho = cov(v_{ut}, v_{xt})$ taking $\rho = 0.5, 0.7, 0.9$, respectively. These four cases are indicated in Tables 2 and 3 by the ρ column, where the different treatment of other correlations in case (i) must be borne in mind.

(Table 2 about here)

Table 2 presents Monte Carlo results when $n = 64$. In case (a), where x_t is stationary, OLS unsurprisingly performs increasingly poorly as ρ increases, with respect to bias as well as SD. IV$_w$ registers considerable improvement, though IV$_{(\alpha)}$ does better, especially when $\alpha = 1$, the results reflecting the ordering of rates in Table 1. However, the NB$_m$ are clearly superior, indicating an advantage in this case for those computationally more complex estimates. The pattern as ρ increases across cases (ii)-(iv) is not very clear where the IV and NBLS estimates are concerned: sometimes the best results are found when $\rho = 0.9$. In case (b) the rather narrow inferiority of OLS seen in Table 1 is borne out in the Monte Carlo results. Generally in cases (b)-(d) there is not much to choose between NBLS and IV, with the latter sometimes superior, but in cases (b) and (c) the same ordering of IV estimates as in case (a) is found. Notice that SD can be affected by choice of z_t even when convergence rate is not, and this may explain why IV$_w$ is less variable than OLS even in case (d). Table 3 contains corresponding results for $n = 128$ The fact that the bias of OLS is not reduced is unsurprising. Bias is otherwise on the whole somewhat reduced, as is SD, except for the NBLS estimates; this reflects the fact that under suitable conditions the asymptotic variance of NB$_m$ is of order m^{-1}. We also obtained results with $z_t = x_t^{(2)}$ and some deterministic z_t; these are unreported because

9

in the former case they differed little from those for $z_t = x_t^{(1)}$, and in the latter case because they were systematically worse than those for the stochastic z_t.

(Table 3 about here)

We also investigated effects of differential integration orders in multiple regression models. We first took $q = 3$ in (2.1), $\beta_1 = 2.5$, $\beta_2 = 2$, $\beta_3 = 1.5$, and $\gamma = 0.1$, $\delta_1 = 0.4$, $\delta_2 = 0.7$, $\delta_3 = 1$. Thus we have a stationary regressor, a nonstationary mean-reverting one and a unit root one. We considered four different covariance structures for the NID$(0, 1)$ innovations v_{ut}, v_{it} of u_t and x_{it}, $i = 1, 2, 3$. Using the notation $\rho_i = cov(v_{ut}, v_{it})$, $\rho_{ij} = cov(v_{it}, v_{jt})$, these are as follows:

Case (I): $\rho_i = \rho_{12} = \rho_{13} = 0$, $i = 1, 2, 3$; $\rho_{23} = 0.5$.
Case (II): $\rho_i = \rho_{ij} = 0.5$, $i = 1, 2, 3$, $j \neq i$.
Case (III): $\rho_2 = \rho_3 = \rho_{ij} = 0.5$, $i = 1, 2, 3$, $j \neq i$; $\rho_1 = 0.7$.
Case (IV): $\rho_2 = \rho_3 = \rho_{ij} = 0.5$, $i = 1, 2, 3$, $j \neq i$; $\rho_1 = 0.9$.

In case (I) there is correlation only between the nonstationary regressors x_{2t} and x_{3t}, whereas in the other three cases all innovation correlations are 0.5 except for that between v_{ut} and v_{1t}, which varies.

(Table 4 about here)

Table 4 contains rates of convergence of OLS and IV$_{(\alpha)}$, $\alpha = 0.5, 1$, where IV$_{(\alpha)}$ is $\hat{\beta}_z$ with $z_t = \left(x_{1t}^{(\alpha)}, x_{2t}^{(\alpha)}, x_{3t}^{(\alpha)} \right)'$: IV$_{(.5)}$ overcomes the inconsistency in the OLS estimate of β_1, and improves on the rates of the OLS estimates of β_2 and β_3, while a small further improvement is registered by IV$_{(1)}$. Tables 5-7 compare finite sample performance of these estimates, each table including results for both $n = 64$ and 128, and the four increasing values of ρ_1 indexing cases (I)-(IV). Apart from the poor performance of the OLS estimate of β_1 in case (I), the results seem quite consistent with the asymptotics. In case (I) OLS of β_2 and β_3 does well, but in the other cases IV$_{(.5)}$ is better and IV$_{(1)}$ better still, perhaps more than one might anticipate from Table 4. More surprising is the systematic way in which SD falls from OLS through IV$_{(.5)}$ to IV$_{(1)}$.

(Tables 5-7 about here)

Our next experiment focussed more explicitly on the curse-of-dimensionality dangers of multiple regression, taking $q = 8$ in (2.1). To keep the design simple, the elements of x_t were mutually independent, and $I(\delta)$ for the same δ and the same scale, and they were independent also of u_t, which was $I(\gamma)$. We took $\beta = (2, 2.5, 1.5, 3, 3.5, 1, 4, 8)^T$, and employed three (γ, δ) combinations. As in the previous experiment, we compared OLS, IV$_{(.5)}$ and IV$_{(1)}$. As expected the results are similar across the elements of the β estimates, albeit with some variation in the direction of bias, and we report in Tables 8 and 9 only results for the first two elements, which seem to exhibit the bias variation. OLS is seriously biased for the stationary case $(\gamma, \delta) = (0.1, 0.4)$, despite the independence between x_t and u_t (as found in previous experiments) with IV$_{(.5)}$ and IV$_{(1)}$ making considerable improvements; they also significantly reduce SD. For the other (γ, δ) combinations OLS performs reasonably, but the IV estimates do

somewhat better. In all cases there is improvement with increasing n. On the whole, $IV_{(1)}$ does slightly better than $IV_{(.5)}$.

(Tables 8-9 about here)

Spurious regression was mentioned in the preceding section. This was originally identified as an issue in case of regression between independent $I(1)$ processes. More recently, Cappuccio and Lubian (1997), Tsay and Chung (2000) have studied its effects in the context of stationary and nonstationary fractional processes, focussing on the behaviour of OLS regression estimates (as well as more basic statistics, and OLS-based statistics such as t-ratios). We compare our IV estimates with OLS and NBLS. The model is (2.1) with $q = 1$ and $\beta = 0$, and x_t and $y_t = u_t$ are independent $I(\delta)$ processes. (Thus, this is actually not a cointegrated model.) In Table 10 IV_w uses an instrument w_t that is $I(\chi)$; the (δ, χ) combinations are (a'): $(.3, .7)$; (b'): $(.6, .8)$; (c'): $(.7, .9)$; and (d'): $(1, 1)$. Spurious effects are, as expected, greatest in the "most nonstationary" unit root case (d'), and we find that NBLS and our IV estimates manage to noticeably reduce them, with NBLS doing best, followed by the $IV_{(\alpha)}$ estimates. The same general pattern is repeated for the other parameter values, with both bias and SD decreasing as δ decreases. Throughout there is some improvement with increasing n.

(Table 10 about here)

Our final experiment partially relaxes the Gaussian prescription by allowing the disturbances u_t to have innovations from a t_5 distribution, so they have finite moment of order 4 only. One expects deterioration of OLS estimates in the presence of fat tails, and we wish to examine how the IV and NBLS estimates compare. The experiment is also designed so as to allow comparisons with Tables 2 and 3: we have $q = 1$, $\beta = 2$, and take the same (γ, δ, ξ) combinations as there, though we now only consider the correlation $\rho = 0.5$ between u_t and x_t. There is little difference in the biases from those of Tables 2 and 3, but unsurprisingly the standard deviations increase. Except when $(\gamma, \delta, \chi) = (0, 1, 1)$, OLS clearly performs worst and the two versions of NBLS narrowly outperform the IV estimates.

5 Final Comments

We have shown that IV estimates of cointegrating relations can improve on OLS when the latter falls short of optimal rate. Our fractional integration setting illustrates much of the scope of the approach but is not essential, the methods applying also in the more traditional case of integer orders: to guard against the possibility that a regressor is $I(0)$ rather than $I(1)$ (or $I(1)$ rather than $I(2)$) an $I(1)$ (or $I(2)$) instrument might be employed. Several choices of instrument are proposed, including a variable that appears elsewhere in the system, and fractional or integer integration of the regressor. In a Monte Carlo study, NBLS estimates, employed in a number of recent empirical studies, sometimes perform better, but perhaps not sufficiently to offset the advantages of simplicity and

familiarity of IV. Of the instruments employed, the simple and always-available one of partially summing the regressor seems generallly to perform best.

6 Appendix

<u>Proof of Theorem 1</u>: Write

$$\hat{\beta}_x - \beta = G_1^{-1} \left(G_1^{-1} M_{xx} G_1^{-1} \right)^{-1} G_1^{-1} G_2 \left(G_2^{-1} M_{xu} \right).$$

From (2.4) this is dominated as $n \to \infty$ by $G_1^{-1} \Omega_{xx}^{-1} G_1^{-1} G_2 \Omega_{xu}$. Now $G_1^{-1} G_2$ has i-th diagonal element

$$n^{\frac{1}{2}} 1(\delta_i \;<\; \frac{1}{2}) + (n/\log n)^{\frac{1}{2}} 1(\delta_i = \frac{1}{2}) + n^{1-\delta_i} 1(\delta_i > \frac{1}{2}, \gamma + \delta_i < 1)$$

$$+ n^{1-\delta_i} (\log n) 1(\gamma \;>\; 0, \delta_i = 1) + n^{\gamma} \{ 1(\gamma + \delta_i > 1) + 1(\gamma = 0, \delta_i = 1) \}.$$

By the block-diagonality of Ω_{xx}, $\hat{\beta}_{xi} - \beta_i = O_p(1)$ for $i = 1, ..., q_1$, and $\hat{\beta}_{xi} - \beta_i = O_p((\log n)^{-1})$ for $i = q_1 + 1, ..., q_1 + q_2$. In general Ω_{xx3} is not a diagonal matrix. Denoting by H the matrix consisting of the last q_3 rows and columns of $G_1^{-1} G_2$, a typical element c_{ij} of $\Omega_{xx3}^{-1} H$ satisfies

$$c_{ij} \approx n^{1-\delta_*} 1(\gamma + \delta_* < 1) + n^{1-\delta_*} (\log n) 1(\gamma > 0, \gamma + \delta_* = 1)$$

$$+ n^{\gamma} \{ 1(\gamma + \delta_* > 1) + 1(\gamma = 0, \delta_* = 1) \}), \; i, j > q_1 + q_2.$$

The proof is now readily completed. □

<u>Proof of Theorem 2</u>: Write

$$\hat{\beta}_z - \beta = G_4^{-1} \left(G_3^{-1} M_{zx} G_4^{-1} \right)^{-1} G_3^{-1} G_5 \left(G_5^{-1} M_{zu} \right).$$

By (2.10) this is dominated by $G_4^{-1} \Omega_{zx}^{-1} G_3^{-1} G_5 \Omega_{zu}$ as $n \to \infty$. The i-th diagonal element of $G_3^{-1} G_5$ is

$$n^{1-\xi_i} 1(\gamma + \xi_i \;<\; 1) + n^{1-\xi_i} (\log n) 1(\gamma > 0, \gamma + \xi_i = 1)$$
$$+ n^{\gamma} \{ 1(\gamma + \xi_i \;>\; 1) + 1(\gamma = 0, \xi_i = 1) \}.$$

Bearing in mind the general non-diagonality of Ω_{zx}, and (2.8), the proof is readily completed. □

<u>Sufficient conditions for (2.4) and (2.10)</u>: We need to discuss the underlying $I(0)$ vector generating x_t, y_t and, when it is involved, z_t (bearing in mind (1.3) and that z_t may have a common or overlapping $I(0)$ source with x_t and y_t). Denote by v_t this $I(0)$ vector, and assume that

$$v_t = \mu + \sum_{j=-\infty}^{\infty} A_j \varepsilon_{t-j}, \quad \sum_{j=0}^{\infty} \sum_{k=j+1}^{\infty} \left\{ \|A_k\|^2 + \|A_{-k}\|^2 \right\} < \infty,$$

where the A_j are square matrices, the ε_t are iid vectors with zero mean, covariance matrix Ω and finite moments of all orders, $\sum_{j=-\infty}^{\infty} A_j$ and Ω are of full rank, and μ is a mean vector. These conditions are sufficient for (2.4) and (2.10). They are designed principally to utilize results of Marinucci and Robinson (2000) (see also Akonom and Gourieroux, 1987) to the extent that convergence of fractional Brownian motion of x_t, u_t or z_t underlies (2.4) and (2.10), but they also imply conditions for Theorems 4.1-4.5 and Theorem 5.1 of Robinson and Marinucci (2001), which (2.4) and (2.10) also use. They easily cover Gaussian ARMA v_t.

7   References

Akonom, J. & C. Gourieroux (1987) A functional central limit theorem for fractional processes. Preprint.

Bandi, F. & B. Perron (2004) Long memory and the relation between implied and realized volatility. Preprint.

Cappuccio, N. & D. Lubian (1997) Spurious regressions between $I(1)$ processes with long memory errors. *Journal of Time Series Analysis* **18**, 341-354.

Cheung, Y.W. & K.S. Lai (1993) A fractional cointegration analysis of purchasing power parity. *Journal of Business and Economic Statistics* **11**, 103-112.

Christensen, B.J. & M.O. Nielsen (2004) Asymptotic normality of narrowband least squares in the stationary fractional cointegration model and volatility forecasting. *Journal of Econometrics*, forthcoming.

Dolado, J. & F. Marmol (1997) Efficient estimation of cointegrating relationships among higher order and fractionally integrated processes. Preprint.

Hualde, J. & P.M. Robinson (2004) Semiparametric estimation of fractional cointegration. Preprint.

Hualde, J. & P.M. Robinson (2006) Root-n-consistent estimation of weak fractional cointegration. *Journal of Econometrics*, forthcoming.

Jeganathan, P. (1999) On asymptotic inference in cointegrated time series with fractionally integrated errors. *Econometric Theory* **15**, 583-621.

Johansen, S. (1991) Estimation and hypothesis testing of cointegrating vectors in Gaussian vector autoregressive models. *Econometrica* **59**, 1551-1580.

Johansen, S. (1996) *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models.* Oxford: Oxford University Press.

Kim, C.S. & P.C.B. Phillips (2000) Fully modified estimation of fractional cointegration models. Preprint.

Kitamura, Y. & P.C.B. Phillips (1995) Efficient IV estimation in nonstationary regression: an overview and simulation study. *Econometric Theory* **11**, 1095-1130.

Kitamura, Y. & P.C.B. Phillips (1997) Fully modified IV, GIVE and GMM estimation with possibly non-stationary regressors and instruments. *Journal of Econometrics* **80**, 85-123.

Marinucci, D. & P.M. Robinson (2000) Weak convergence of multivariate fractional processes. *Stochastic Processes and their Applications* **86**, 103-120.

Marmol, F., A. Escribano & F.M. Aparicio (2002) Instrumental variable interpretation of cointegration with inference results for fractional cointegration. *Econometric Theory* 18, 646-672.

Phillips, P.C.B. (1986) Understanding spurious regressions in econometrics. *Journal of Econometrics* **33**, 311-340.

Phillips, P.C.B. (1991) Optimal inference in cointegrated systems. *Econometrics* **59**, 283-306.

Phillips, P.C.B. & B.E. Hansen (1990) Statistical inference in instrumental variables regression with I(1) processes. *Review of Economic Studies* **57**, 99-125.

Robinson, P.M. (1994) Semiparametric analysis of long memory time series. *Annals of Statistics* **22**, 515-539.

Robinson, P.M. & J. Hualde (2003) Cointegration in fractional systems with unknown integration orders. *Econometrica* **71**, 1727-1766.

Robinson, P.M. & D. Marinucci (2001) Narrow band analysis of nonstationary processes. *Annals of Statistics* **29**, 947-986.

Robinson, P.M. & D. Marinucci (2003) Semiparametric frequency domain analysis of fractional cointegration. In P.M. Robinson (ed.) *Time Series with Long Memory.* Oxford: Oxford University Press, pp. 334-373.

Robinson, P.M. & Y. Yajima (2002) Determination of cointegrating rank in fractional systems. *Journal of Econometrics* **106**, 217-241.

Stock, J. (1987) Asymptotic properties of least squares estimators of cointegrating vectors. *Econometrica* **58**, 1035-1056.

Stock, J. & M.W. Watson (1988) Testing for common trends. *Journal of the American Statistical Society* **83**, 1097-1107.

Tsay, W.-J. & C.-F. Chung (2000) The spurious regression of fractionally integrated processes. *Journal of Econometrics* **96**, 155-182.

Table 1: Convergence rates of deviations of estimates from β in model (2.1)
with $q = 1$ in cases (I)-(IV). $X = $ "inconsistent".

γ, δ, χ	OLS	NB_m	IV_w	$\text{IV}_{(.5)}$	$\text{IV}_{(1)}$
$.1, .3, .7$	X	$(m/n)^{.2}$	$(\log n)^{-1}$	$n^{-.1}$	$n^{-.2}$
$.2, .6, .9$	$n^{-.2}$	$m^{.2} n^{-.4}$	$n^{-.4}$	$n^{-.4}$	$n^{-.4}$
$.4, .7, .8$	$n^{-.3}$	$n^{-.3}$	$n^{-.3}$	$n^{-.3}$	$n^{-.3}$
$0, 1, 1$	n^{-1}	n^{-1}	n^{-1}	n^{-1}	n^{-1}

Table 2: Monte Carlo bias and SD of estimates of β in model (2.1) with $q = 1$
in cases (a)-(d) and (i)-(iv), with $n = 64$, and 2000 replications

		Bias						SD					
γ, δ, χ	ρ	OLS	NB_4	NB_5	IV_w	$\text{IV}_{(.5)}$	$\text{IV}_{(1)}$	OLS	NB_4	NB_5	IV_w	$\text{IV}_{(.5)}$	$\text{IV}_{(1)}$
	0	-1.98	.002	-.002	.021	-.003	.002	3.13	.101	.114	.646	.194	.187
	.5	4.60	.048	.058	.098	.075	-.064	10.27	.521	.511	1.00	.846	.724
$.1, .3, .7$	.7	7.14	-.054	051	.094	.073	.068	15.97	.577	.491	.954	.832	.702
	.9	7.96	.058	.059	.092	-.071	.056	17.98	.618	.635	.994	.782	.697
	0	-.007	.008	-.006	-.052	-.010	.008	.272	.197	.142	.523	.298	.355
	.5	.070	-.045	.049	.051	-.056	.042	.787	.311	.530	.610	.688	.546
$.2, .6, .9$	.7	.077	-.028	-.032	.041	.038	-.031	.694	.286	.332	.485	.666	.518
	.9	-.080	.039	.035	.041	.037	.031	.731	.240	.421	.515	.507	.577
	0	-.007	.008	-.006	-.060	-.009	-.008	.360	.182	.173	.510	.356	.321
	.5	-.074	-.057	.065	.068	-.062	.058	.953	.585	.691	.818	.712	.745
$.4, .7, .8$	.7	.059	.043	-.037	-.046	-.044	.041	.821	.501	.606	.589	.576	.511
	.9	.062	.042	.040	-.040	.041	.037	.725	.496	519	.554	.529	.495
	0	-.004	.002	.002	-.003	.006	.005	.087	.016	.014	.021	.058	.053
	.5	-.009	.005	.002	-.005	-.009	.008	.082	.041	.046	.071	.091	.078
$0, 1, 1$	.7	.005	.007	.009	.006	.006	.006	.072	.048	.034	.038	.095	.083
	.9	-.007	.006	.006	-.006	.005	.005	.081	.033	049	.026	.053	.062

Table 3: Monte Carlo bias and SD of estimates of β in model (2.1) with $q = 1$
in cases (a)-(d) and (i)-(iv), with $n = 128$, and 2000 replications

γ, δ, χ	ρ	Bias						SD					
		OLS	NB$_4$	NB$_5$	IV$_w$	IV$_{(.5)}$	IV$_{(1)}$	OLS	NB$_4$	NB$_5$	IV$_w$	IV$_{(.5)}$	IV$_{(1)}$
	0	1.26	.001	.002	.014	.002	.002	2.17	.092	.085	.575	.139	.128
	.5	4.33	.040	.040	.097	.074	.063	9.69	.504	.463	.985	.779	.713
.1, .3, .7	.7	6.91	-.051	.048	.092	.071	.069	15.95	.509	.463	938	.725	.633
	.9	7.86	.053	.056	.090	.060	.048	17.58	.536	.587	.929	728	.602
	0	.004	-.005	.005	.040	.008	.008	.118	.112	.201	.412	.301	.265
	.5	.060	.038	.040	.043	.050	.034	.711	.145	.349	.542	.425	.502
.2, .6, .9	.7	.062	.025	.039	.034	.040	.029	.618	.267	.344	.339	.626	.433
	.9	.068	.037	.031	.031	.033	.030	.606	.303	.401	.478	460	.493
	0	.003	-.007	.005	.050	.007	.006	.180	.157	.187	.611	.281	.242
	.5	.045	.054	.061	-.051	.058	-.051	.675	.622	.674	.601	.690	.653
.4, .7, .8	.7	.052	.032	.029	.038	.029	.024	.794	.435	.514	.523	.490	.481
	.9	.059	.040	.035	.034	.038	.036	.638	.481	.492	.521	.505	.421
	0	.001	.002	.002	.001	-.005	.004	.010	.010	.012	.019	.051	.047
	.5	.003	.002	.001	.004	.005	-.004	.055	.012	.029	.091	.071	.095
0, 1, 1	.7	.001	.006	-.009	.006	.006	-.004	.045	.029	.020	026	.076	.067
	.9	.002	.003	.003	-.005	.005	.004	.052	.029	.044	.021	.059	.051

Table 4: Convergence rates of deviations of estimates from $\beta_1, \beta_2, \beta_3$ in model (2.1) with $q = 3$. $X =$ "inconsistent".

Parameter	OLS	$\text{IV}_{(.5)}$	$\text{IV}_{(1)}$
β_1	X	$n^{-.3}/\log n$	$n^{-.3}$
β_2	$n^{-.4}$	$n^{-.6}/\log n$	$n^{-.6}$
β_3	$n^{-.7}$	$n^{-.9}/\log n$	$n^{-.9}$

Table 5: Monte Carlo bias and SD of estimates of β_1 in model (2.1) with $q = 3$ in cases (I)-(IV), with $n = 64$ and 128, and 2000 replications

n	ρ_3	Bias			SD		
		OLS	$\text{IV}_{(.5)}$	$\text{IV}_{(1)}$	OLS	$\text{IV}_{(.5)}$	$\text{IV}_{(1)}$
	0	4.93	.021	-.010	14.6	.224	.139
	.5	-5.88	.085	.070	19.7	.869	.748
64	.7	4.01	.073	.064	17.7	.777	.638
	.9	5.27	.076	.065	18.8	.764	.665
	0	4.37	-.014	.009	7.36	.108	.101
	.5	4.45	-.080	.067	8.69	.817	.720
128	.6	-4.94	-.069	.062	16.5	.700	.635
	.9	5.34	.072	.064	18.1	.719	.628

Table 6: Monte Carlo bias and SD of estimates of β_2 in model (2.1) with $q = 3$
in cases (I)-(IV), with $n = 64$ and 128, and 2000 replications

		Bias			SD		
n	ρ_3	OLS	$IV_{(.5)}$	$IV_{(1)}$	OLS	$IV_{(.5)}$	$IV_{(1)}$
	0	.016	.016	.009	.298	.217	.110
	.5	-.093	.078	-.062	1.02	.692	.731
64	.7	.089	.064	.054	.996	.731	.616
	.9	.098	.063	.051	1.21	.704	.631
	0	.015	.011	.009	.253	.103	.089
	.5	.088	.074	.062	.962	.652	.697
128	.6	.087	.060	.047	.901	.627	.568
	.9	.096	.060	-.046	.928	.616	.593

Table 7: Monte Carlo bias and SD of estimates of β_3 in model (2.1) with $q = 3$
in cases (I)-(IV), with $n = 64$ and 128, and 2000 replications

		Bias			SD		
n	ρ_3	OLS	$IV_{(.5)}$	$IV_{(1)}$	OLS	$IV_{(.5)}$	$IV_{(1)}$
	0	.009	.014	.008	.199	.183	.108
	.5	.078	-.057	-.046	.825	.673	.581
64	.7	.082	.056	.045	.889	.693	.652
	.9	.083	.057	.042	.890	.711	.534
	0	.008	.009	-.005	.130	.137	.091
	.5	.072	.053	.042	.797	.648	.548
128	.6	.074	-.051	-.041	.828	.674	.639
	.9	.079	.054	.040	.838	.733	.522

Table 8: Monte Carlo bias and SD of estimates of β_1 in model (2.1) with $q = 8$, $n = 64, 128$, with $\rho_i = \rho_{ij} \equiv 0$, and 2000 replications

		Bias			SD		
γ, δ	n	OLS	$IV_{(.5)}$	$IV_{(1)}$	OLS	$IV_{(.5)}$	$IV_{(1)}$
.1,.4	64	-.955	.011	.010	1.13	.156	.149
	128	.750	.010	.009	.880	.142	.138
.4,.7	64	.016	.008	-.008	.155	.131	.133
	128	.012	.005	.005	.129	.121	.116
0,1	64	-.007	.004	.004	.088	.086	.081
	128	.003	.003	.002	.042	.036	.038

Table 9: Monte Carlo bias and SD of estimates of β_2 in model (2.1) with $q = 8$, $n = 64, 128$, with $\rho_i = \rho_{ij} \equiv 0$, and 2000 replications

		Bias			SD		
γ, δ	n	OLS	$IV_{(.5)}$	$IV_{(1)}$	OLS	$IV_{(.5)}$	$IV_{(1)}$
.1,.4	64	.883	-.011	.011	1.13	.181	.168
	128	.697	.010	-.010	.805	.134	.122
.4,.7	64	.019	.006	-.006	.170	.126	.130
	128	.011	.003	.002	.148	.118	.115
0,1	64	-.009	.004	.003	.093	.073	.072
	128	.005	-.002	.002	.046	.038	.031

Table 10: Monte Carlo bias and SD of estimates of β in model (2.1) with
$q = 1$, in cases (a')-(d'), with $n = 64, 128$, and 2000 replications

		Bias						SD					
δ, χ	n	OLS	NB$_4$	NB$_5$	IV$_w$	IV$_{(.5)}$	IV$_{(1)}$	OLS	NB$_4$	NB$_5$	IV$_w$	IV$_{(.5)}$	IV$_{(1)}$
.3,.7	64	-.016	.014	.013	-.014	.014	-.014	.220	.191	.185	.217	.216	.211
	128	.013	.011	.011	.011	.010	.010	.171	.163	.168	.170	.164	.159
.6,.8	64	.047	.038	-.040	.046	.038	.040	.561	.491	.506	.546	.495	.501
	128	-.041	.030	.031	-.040	.033	.031	.519	.427	.437	.521	.504	.501
.7,.9	64	-.086	-.072	.076	.083	.079	.074	.683	.574	.580	.670	.649	.624
	128	.064	.053	-.052	.058	-.060	.059	.678	.673	.598	.628	.610	.608
1,1	64	.098	.080	.080	.084	.082	.081	1.01	.808	.855	.970	.932	.878
	128	-.065	-.046	.047	.053	.0.50	-.047	.940	.786	.792	.937	.812	.815

Table 11: Monte Carlo bias and SD of estimates in model (2.1) with $q = 1$ in
cases (a)-(d), with t_5 errors, with $n = 64, 128$ and 2000 replications

		Bias						SD					
γ, δ, χ	n	OLS	NB$_4$	NB$_5$	IV$_w$	IV$_{(.5)}$	IV$_{(1)}$	OLS	NB$_4$	NB$_5$	IV$_w$	IV$_{(.5)}$	IV$_{(1)}$
.1,.3,.7	64	5.20	.062	.066	.078	.075	.074	12.38	1.08	1.24	1.53	1.32	1.27
	128	4.90	.048	.047	.099	.079	.044	10.85	.963	.912	1.38	1.12	1.01
.2,.6,.9	64	.081	.059	.058	.075	.069	.068	1.96	.627	.651	.850	.872	.864
	128	.078	.042	.044	.047	.052	.054	1.53	.539	.494	.800	.761	.797
.4,.7,.8	64	.085	.063	.064	.070	.071	.069	1.94	1.11	1.12	1.15	1.13	1.13
	128	.049	.054	.064	.052	.060	.053	1.34	1.02	.985	1.18	1.08	1.09
0,1,1	64	.007	.006	.005	.008	.007	.006	.212	.191	.207	.234	.219	.211
	128	.003	.003	.004	.006	.005	.005	.154	.155	.168	.197	.174	.181