Estimation with Pairwise Observations

Felix Chan[1], László Mátyás[2], and Kristóf Reizinger[2]

[1]Curtin University, Perth, Australia
[2]Central European University, Budapest, Hungary and Vienna, Austria

May 17, 2023

Acknowledgement: The authors would like to thank Tom Wansbeek for insightful discussion. Contribution by Gyorgy Ruzicska in the early stages of this project is kindly acknowledged.

Abstract: The paper introduces a new estimation method for the standard linear regression model. The procedure is not based on the optimisation of any objective function rather, it is a simple weighted average of slopes from observation pairs. Unlike traditional methods, such as Least Squares and Maximum Likelihood, among others, the residual of this method is not by construction orthogonal to the explanatory variables of the model, therefore can be used for testing endogeneity, i.e., the correlation between the explanatory variables and the disturbance terms. In addition, this approach also enables us to consistently estimate the unknown parameters of a model with endogeneity without relying on any kind of additional external information usually needed in the form of instrumental variables or moment conditions.

Keywords: *Linear regression model, consistent estimation, endogeneity, testing for endogeneity.*
JEL: C01, C13, C20, C26, C51

1 Introduction

To start with, the paper intents to answer a simple question relating to the estimation of a linear regression model. Let us assume that we have a very basic model with only one explanatory variable:

$$y_i = \beta_0 + \beta_1 x_i + u_i, \qquad u_i \sim D(0, \sigma_u^2), \quad i = 1, \ldots, n. \tag{1}$$

Assuming the n observations the data can be divided in clusters in such a way that in each cluster there are only two observations. This clustering may be carried out in two different ways:

- The (x_i, y_i) observations are ordered based on the values of x_i and two adjacent observations are grouped into one cluster. Let us call this the *sorted adjacent* approach. There are $n-1$ such pairs. When we do not apply any sorting and just group each adjacent observations into a cluster, we call such approach *non-sorted adjacent*.
- All feasible combinations

  $$[(x_i, y_i); (x_j, y_j)], \ (i \neq j)$$

  of the observations are considered to form clusters. There are in total $\frac{n(n-1)}{2}$ of such clusters. Let us call this the (non-sorted) *full-pairwise approach*. Here, again we may consider sorting the observations based on the values of x_i, which gives the *sorted full-pairwise approach*.

In each cluster there are only two observations so the "regression line" in a cluster is the line going through these two observations, and there is no estimation involved. In such a way we can get as many β_0 and β_1 parameters as the number of clusters (see Figures 1 and 2 for example).

Our main question of interest is how these "cluster-wise" or let us call them *pairwise* parameters relate to the "true" parameters of the data generation process formulated in (1). Let us denote the pairwise parameters in the ith cluster with $\beta_{i,0}$ and $\beta_{i,1}$ and consider the following two estimators:

- The weighted average of $\beta_{i,1}$

  $$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} w_i \beta_{i,1}}{\sum_{i=1}^{n} w_i} \tag{2}$$

  with some w_i weights.

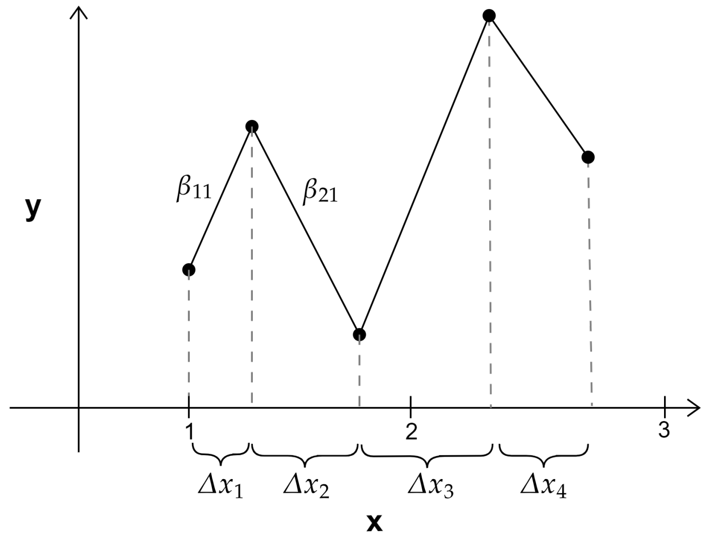

Figure 1: Sorted adjacent

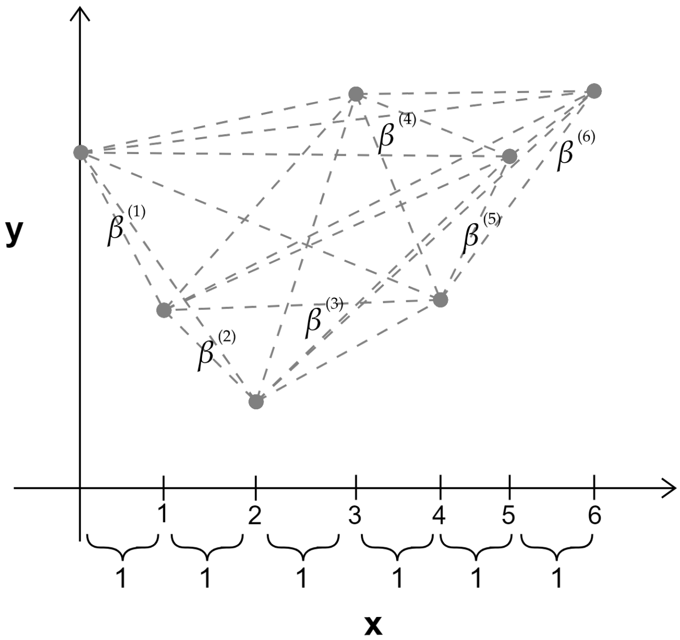

Figure 2: Non-sorted full-pairwise

- Minimise a loss function such as

$$\hat{\beta}_1 = \arg\min_{\beta_1} \sum_{i=1}^{n} (w_i(\beta_{i,1} - \hat{\beta}_1))^2. \quad (3)$$

Estimators for β_0 can also be defined similarly and they will be discussed with more details in Section 3.5.

The question now is how $\hat{\beta}_1$ relates to the true parameter, β_1 formulated in Equation (1). Let us call this estimation procedure *Estimation with Pairwise Observations* (EwPO).

2 Some Monte Carlo Simulation Results

First, we adopted a Petri-dish type approach and carried out a large number of Monte Carlo simulation. The setup and the main results can be found in Appendix A.[1] The results provide some interesting insights in terms of the consistency of the proposed estimators under different choices of w_i. The EwPO turned out to be consistent in the following cases:

1. EwPO as defined in Equation (2) with
   (a) Sorted full-pairwise and $w_{ij} = x_i - x_j$
   (b) Non-sorted adjacent and $w_i = |x_i - x_{i-1}|$
   (c) Sorted full-pairwise and $w_{ij} = |x_i - x_j|$
   (d) Non-sorted full-pairwise and $w_{ij} = |x_i - x_j|$
   (e) Sorted full-pairwise and $w_{ij} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}$
   (f) Non-sorted full-pairwise and $w_{ij} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}$
2. Loss function defined in Equation (3)
   (a) Non-sorted adjacent and $w_i = x_i - x_{i-1}$
   (b) Sorted full-pairwise and $w_{ij} = x_i - x_j$
   (c) Non-sorted full-pairwise and $w_{ij} = x_i - x_j$

   Note: In case of the loss function taking the absolute value of weights does not matter as weights are squared in the functional form!

3 Some Analytical Results

This section presents some analytical explanations on the simulations results. It will also present the notations and definitions used throughout this paper.

[1] Additional simulation results can be found in the Online Supplement Chan et al. (2023).

3.1 Notations and Definitions

In order to develop the theoretical foundation, we introduce a few notations, definitions and concepts. Let Δx_{ij} denotes the difference between x_i and x_j that is,

$$\Delta x_{ij} = x_i - x_j, \qquad i, j = 1, \ldots, n. \tag{4}$$

In the event that the subscript is suppressed, then $j = i - 1$. That is $\Delta x_i = x_i - x_{i-1}$.

The *pairwise* parameter is defined to be

$$\beta_{1,(i,j)} = \frac{\Delta y_{ij}}{\Delta x_{ij}}. \tag{5}$$

In the event that $j = i - 1$ we simplify the notation by writing $\beta_{1,i}$. That is $\beta_{1,i} = \beta_{1,(i,j)}$.

It is useful to also note that

$$\Delta y_{ij} = \beta_1 \Delta x_{ij} + \Delta u_{ij}. \tag{6}$$

3.2 Partial Sum

Some of the expressions for analysing the full-pairwise case contain several partial sums which are asymptotically related to a *Brownian Motion* or *Wiener Process*. So it is useful to introduce some additional notations and state the relevant results here. Define $\lfloor x \rfloor$ as the largest integer that is less than or equal to x and $\lceil x \rceil$ as the smallest integer that is larger than or equal to x. Let $r \in [0, 1]$, $u_i \sim N(0, \sigma_u^2)$ and define

$$W_n(r) = \frac{1}{\sqrt{n}} \sum_{i=1}^{\lfloor rn \rfloor} \frac{u_i}{\sigma_u}, \tag{7}$$

then under the conditional stated in (Billingsley, 1999, Theorem 8.2, P.90), the *Functional Central Limit Theorem* ensures $W_n(r) \xrightarrow{d} W(r)$, where $W(r)$ denotes a Wiener Process.

3.3 Adjacent Case

The estimator considered in this section is

$$\hat{\beta}_1 = \left(\sum_{i=2}^{n} w_i \right)^{-1} \left(\sum_{i=2}^{n} w_i \beta_{1,i} \right) \tag{8}$$

where w_i represents the weight for the i^{th} pairwise parameter.

3.3.1 Case 1: $w_i = \Delta x_i$

In this case, Equation (8) reduces to

$$\begin{aligned}\hat{\beta}_1 =& \left(\sum_{i=2}^{n} w_i\right)^{-1} \left(\sum_{i=2}^{n} w_i \beta_{1,i}\right)\\ =& \left(\sum_{i=2}^{n} \Delta x_i\right)^{-1} \left(\sum_{i=2}^{n} \Delta y_i\right)\\ =& \left(\sum_{i=2}^{n} \Delta x_i\right)^{-1} \left(\sum_{i=2}^{n} \beta_1 \Delta x_i + \Delta u_i\right)\\ =& \beta_1 + \frac{\sum_{i=2}^{n} \Delta u_i}{\sum_{i=2}^{n} \Delta x_i}.\end{aligned}$$

Now, in the case of adjacent pairwise parameter

$$\begin{aligned}\sum_{i=2}^{n} \Delta x_i =& (x_2 - x_1) + (x_3 - x_2) + (x_4 - x_3) + \ldots + (x_n - x_{n-1})\\ =& x_n - x_1.\end{aligned}$$

The last line follows because the first term in every bracket is being cancelled out by the last term in the next bracket. The only two exceptions are x_1 and x_n. So this gives

$$\hat{\beta}_1 = \beta_1 + \frac{u_n - u_1}{x_n - x_1}. \tag{9}$$

Equation (9) holds regardless whether the data was sorted or not. The only difference is that, in the sorted case $x_n = \max\{x_i\}_{i=1}^n$ and $x_1 = \min\{x_i\}_{i=1}^n$. Equation (9) does seem to suggest that $\hat{\beta}_1$ is not going to be consistent in this case, as $= \dfrac{\Delta u_{n1}}{\Delta x_{n1}}$ will remain even when $n \to \infty$.

3.3.2 Case 2: $w_i = |\Delta x_i|$

In this case, it would be useful to first define the sgn (x) function.

$$\operatorname{sgn}(x) = \begin{cases} 1 & \text{if } x \geq 0 \\ -1 & \text{if } x < 0. \end{cases} \tag{10}$$

Note that $\operatorname{sgn}(x)\, x = |x|$, and that unlike the case when $w_i = \Delta x_i$,

$$\sum_{i=1}^{n} |\Delta x_i| = |x_2 - x_1| + |x_3 - x_2| + \ldots + |x_n - x_{n-1}|$$

and no simple cancellation occurs here when the observations are *not sorted*. However, when the observations are *sorted*, it is clear that $\sum_{i=2}^{n} |\Delta x_i| = \sum_{i=2}^{n} \Delta x_i$ since $\Delta x_i = |\Delta x_i|$. So the result for the sorted case is the same as the case when $w_i = \Delta x_i$.
For the non-sorted case,

$$
\begin{aligned}
\hat{\beta}_1 &= \left(\sum_{i=2}^{n} w_i\right)^{-1} \left(\sum_{i=2}^{n} w_i \beta_{1,i}\right) \\
&= \left(\sum_{i=2}^{n} |\Delta x_i|\right)^{-1} \left(\sum_{i=2}^{n} \operatorname{sgn}(\Delta x_i)\, \Delta y_i\right) \\
&= \left(\sum_{i=2}^{n} |\Delta x_i|\right)^{-1} \left(\sum_{i=2}^{n} \beta_1 \operatorname{sgn}(\Delta x_i)\, \Delta x_i + \operatorname{sgn}(\Delta x_i)\, \Delta u_i\right) \\
&= \left(\sum_{i=2}^{n} |\Delta x_i|\right)^{-1} \left(\sum_{i=2}^{n} \beta_1 |\Delta x_i| + \operatorname{sgn}(\Delta x_i)\, \Delta u_i\right) \qquad \because \operatorname{sgn}(\Delta x_i)\, \Delta x_i = |\Delta x_i| \\
&= \beta_1 + \frac{\sum_{i=2}^{n} \operatorname{sgn}(\Delta x_i)\, u_i}{\sum_{i=1}^{n} |\Delta x_i|}.
\end{aligned}
$$

Under the assumption that $\mathbb{E}\left(|\Delta x_i|\right) \leq \infty$, it is clear that $\mathbb{E}\left(|\Delta x_i|\right) > 0$. Note that $\mathbb{E}(|\operatorname{sgn}(\Delta x_i)\, u_i|) = \mathbb{E}(|u_i|)$, so under that assumption that $\mathbb{E}(u_i) = 0$, $\mathbb{E}(|u_i|) < \infty$ and $u_i \perp x_i$ then $\mathbb{E}\left[\operatorname{sgn}(\Delta x_i)\, u_i\right] = \mathbb{E}\left[\operatorname{sgn}(\Delta x_i)\right] \mathbb{E}[u_i] = 0$. Moreover, under Weak Law of Large Number (WLLN),

$$
(n-1)^{-1} \sum_{i=2}^{n} \operatorname{sgn}(\Delta x_i)\, u_i = o_p(1)
$$

and hence

$$
\hat{\beta}_1 - \beta_1 = o_p(1).
$$

3.4 Full-pairwise

An important difference here is that the number of pairwise parameters is much larger. Let N denotes the number of full-pairwise parameters then $N = \dfrac{n(n+1)}{2}$. It is also important to identify the exact pair from each pairwise parameter and redefine the estimator as

$$
\hat{\beta}_1 = \left(\sum_{i=2}^{n} \sum_{j=1}^{i-1} w_{ij}\right)^{-1} \left(\sum_{i=2}^{n} \sum_{j=1}^{i-1} w_{ij} \beta_{1,(i,j)}\right) \tag{11}
$$

where $\beta_{1,(i,j)}$ is defined as in Equation (5).

3.4.1 Case 1: $w_{ij} = \Delta x_{ij}$

We first deal with the *non-sorted* case. In this case, the estimator reduces to

$$\begin{aligned}\hat{\beta}_1 &= \left(\sum_{i=2}^{n}\sum_{j=1}^{i-1} w_{ij}\right)^{-1} \left(\sum_{i=2}^{n}\sum_{j=1}^{i-1} w_{ij}\beta_{1,(i,j)}\right) \\ &= \left(\sum_{i=2}^{n}\sum_{j=1}^{i-1} \Delta x_{ij}\right)^{-1} \left(\sum_{i=2}^{n}\sum_{j=1}^{i-1} \Delta y_{ij}\right) \\ &=\beta_1 + \frac{\sum_{i=2}^{n}\sum_{j=1}^{i-1} \Delta u_{ij}}{\sum_{i=2}^{n}\sum_{j=1}^{i-1} \Delta x_{ij}}. \end{aligned} \tag{12}$$

Let us first focus on the denominator as the argument for the numerator will be similar.

$$\begin{aligned}\sum_{i=2}^{n}\sum_{j=1}^{i-1} \Delta x_{ij} &= \sum_{i=2}^{n}\sum_{j=1}^{i-1} (x_i - x_j) \\ &= \sum_{i=2}^{n}\sum_{j=1}^{i-1} x_i - \sum_{i=2}^{n}\sum_{j=1}^{i-1} x_j \\ &= \sum_{i=2}^{n} (i-1)x_i - \sum_{i=2}^{n}\sum_{j=1}^{i-1} x_j. \end{aligned} \tag{13}$$

For reason that will become clear later, we are going to assume $x_i \sim D(\mu_x, \sigma_x^2)$ with $\sigma^2 < \infty$ and $\mathbb{E}|x_i| < \infty$. We will analyse expression (13) term-by-term. The first term in expression (13) can be rewritten as

$$\begin{aligned}\sum_{i=2}^{n}(i-1)x_i &= N\mu_x + \sum_{i=2}^{n}(i-1)(x_i - \mu_x) \\ &= N\mu_x + \sum_{i=2}^{n} iz_i - \sum_{i=2}^{n} z_i \\ &= N\mu_x - \sum_{i=2}^{n} z_i + \sum_{i=2}^{n} iz_i\,, \end{aligned}$$

where $z_i = x_i - \mu_x$. Note that z_i is the de-meaned version of x_i and therefore $\mathbb{E}(z_i) = 0$. To expand further the expression above,

$$\begin{aligned}
N\mu_x - \sum_{i=2}^{n} z_i + \sum_{i=2}^{n} i z_i =& N\mu_x - \sum_{i=2}^{n} z_i + \sum_{i=1}^{n}\sum_{j=i}^{n} z_j \\
=& N\mu_x - \sum_{i=2}^{n} z_i + \sum_{i=1}^{n}\sum_{j=\lceil rN \rceil}^{n} z_j \text{ where } r = i/N \\
=& N\mu_x - \sum_{i=2}^{n} z_i + \sqrt{n}\sigma_x \sum_{i=1}^{n} (W_n(1) - W_n(r)) \\
=& N\mu_x - \sum_{i=2}^{n} z_i + n\sqrt{n}\sigma_x \sum_{i=1}^{n} \left(\int_{(i-1)/n}^{i/n} W_n(1) - W_n(s) ds \right) \\
=& N\mu_x - \sum_{i=2}^{n} z_i + n\sqrt{n}\sigma_x \left(W_n(1) - \int_0^1 W_n(s) ds \right). \qquad (14)
\end{aligned}$$

Now for the second term in expression (13),

$$\begin{aligned}
\sum_{i=2}^{n}\sum_{j=1}^{i-1} x_j =& N\mu_x + \sum_{i=2}^{n}\sum_{j=1}^{i-1} z_j \\
=& N\mu_x + \sqrt{n}\sigma_x \sum_{i=1}^{n} W_n(r) \quad r = \frac{i-1}{n} \\
=& N\mu_x + n\sqrt{n}\sigma_x \sum_{i=2}^{n} \int_{(i-1)/n}^{i/n} W_n(s) ds \\
=& N\mu_x + n\sqrt{n}\sigma_x \int_0^1 W_n(s) ds.. \qquad (15)
\end{aligned}$$

Substitute expressions (14) and (15) to expression (13) yields

$$\sum_{i=2}^{n}\sum_{j=1}^{i-1} \Delta x_{ij} = n\sqrt{n}\sigma_x \left(W(1) - 2\int_0^1 W(s) ds \right) - \sum_{i=1}^{n} z_i. \qquad (16)$$

Note that $N = O(n^2)$ so

$$N^{-1} \sum_{i=2}^{n}\sum_{j=1}^{i-1} \Delta x_{ij} = o_p(1).$$

The same argument applies to $\sum_{i=1}^{n}\sum_{j=1}^{i-1} \Delta u_{ij}$ and since both approach 0 at the same rate, $\hat{\beta}_1$ is going to be unstable as $n \to \infty$.

The case when the data is sorted means that $\sum_{i=1}^{n}\sum_{j=1}^{i-1}\Delta x_{ij}$ is not going to be 0 asymptotically since $\Delta x_{ij} > 0$ by construction. Equation (12) stills hold when x_i is sorted such that $x_i \geq x_j$ for all $i > j$. Since u_i is independent from x_i this means the numerator in Equation (12) approaches 0 as $n \to \infty$ but the denominator will either approach a finite non-zero positive constant or diverges. In both cases, the $\hat{\beta}_1 - \beta_1 = o_p(1)$ under the Assumptions of Proposition 1, which provides the asymptotic distribution for this case.

3.4.2 Case 2: $w_{ij} = |\Delta x_{ij}|$

Following similar derivation as above, it is straightforward to show that

$$\begin{aligned}
\hat{\beta}_{1,(i,j)} =& \left(\sum_{i=2}^{n}\sum_{j=1}^{i-1}|\Delta x_{ij}|\right)^{-1}\left(\sum_{i=2}^{n}\sum_{j=1}^{i-1}\frac{\Delta y_{ij}}{\Delta x_{ij}}|\Delta x_{ij}|\right)\\
=& \left(\sum_{i=2}^{n}\sum_{j=1}^{i-1}|\Delta x_{ij}|\right)^{-1}\left(\sum_{i=2}^{n}\sum_{j=1}^{i-1}\Delta y_{ij}\text{sgn}\left(\Delta x_{ij}\right)\right)\\
=& \left(\sum_{i=2}^{n}\sum_{j=1}^{i-1}|\Delta x_{ij}|\right)^{-1}\left(\sum_{i=2}^{n}\sum_{j=1}^{i-1}\beta_1|\Delta x_{ij}| + \Delta u_{ij}\text{sgn}\left(\Delta x_{ij}\right)\right)\\
=& \beta_1 + \left(\sum_{i=2}^{n}\sum_{j=1}^{i-1}|\Delta x_{ij}|\right)^{-1}\left(\sum_{i=2}^{n}\sum_{j=1}^{i-1}\Delta u_{ij}\text{sgn}\left(\Delta x_{ij}\right)\right)\\
=& \beta_1 + \left(\sum_{i=2}^{n}\sum_{j=1}^{i-1}|\Delta x_{ij}|\right)^{-1}\left(\sum_{i=2}^{n}\sum_{j=1}^{i-1}u_i\text{sgn}\left(\Delta x_{ij}\right) + \sum_{i=2}^{n}\sum_{j=1}^{i-1}u_j\text{sgn}\left(\Delta x_{ij}\right)\right).
\end{aligned} \tag{17}$$

Given that $x_i \perp u_j$ for all $i, j = 1, \ldots, n$ and $n^{-1}\sum_{i=1}^{n}|\text{sgn}\left(\Delta x_{ij}\right)| \leq 1$, the last two terms in Equation (17) will therefore converge to 0 following a similar argument as in the case $w_i = |\Delta x_{ij}|$. Using the same argument as above, $\sum_{i=1}^{n}\sum_{j=1}^{i-1}|\Delta x_{ij}| = O_p(n^2)$ with a non-zero positive bound. Thus $\hat{\beta}_1 - \beta_1 = o_p(1)$.

3.4.3 Confidence Intervals

The asymptotic distribution of $\hat{\beta}$ in the full-pairwise case with $w_{ij} = \Delta x_{ij}$ can be found in Proposition 2. Analytical results for the asymptotic behaviour of the full-pairwise case with $w_{ij} = |\Delta x_{ij}|$ is more difficult as the presence of absolute value creates some technical challenges. In practice, however, hypothesis testing in this case can still be conducted via the *jackknife* procedure.

Regardless on the choice of w_{ij}, the distribution of $\hat{\beta}$ in the full-pairwise case is likely to be non-standard and involve stochastic integrals. Therefore, the critical values will still required to be simulated even if the analytical distributions can be obtained. From a practical viewpoint, the jackknife procedure does not necessarily induce more computational cost and it is generally much simpler to conduct as the procedure will be the same across different choices of w_{ij}.

An example of the procedure to estimate critical values can be found as follows:

Step 1. Set the level of significant for the test α.

Step 2. Set the number of observations to be removed from the main sample, call it d where $\sqrt{n} < d < n$.

Step 3. Set the number of replication $R < \binom{n}{n-d}$

Step 4. set $i = 1$

Step 5. Create a jackknife sample, J_i, by removing d randomly selected observations from the main sample.

Step 6. Compute $\hat{\beta}$ using the jackknife sample J_i, call it $\hat{\beta}_i$

Step 7. Set $i \leftarrow i + 1$.

Step 8. Repeat Steps 5 to 7 until $i = R$.

Step 9. Sort the set $\{\hat{\beta}_i\}_{i=1}^{R}$ from lowest to highest.

Step 10. The $\lfloor \frac{\alpha}{2} R \rfloor^{th}$ and $\lfloor \frac{(1-\alpha)}{2} R \rfloor^{th}$ elements in the set gives the lower and upper bounds of the confidence interval at the α significant level.

Table 1 provides an example on the performance of the jackknife algorithm. As shown in the table, the algorithm performs well in obtaining the confidence intervals of the full-pairwise estimator.

Full-pairwise MC	Jackknife confidence intervals lower and upper bounds			
	Parameter	Full-pairwise, one single sample MC estimates	Lower bound ($\beta_{0.025}$)	Upper bound ($\beta_{0.975}$)
n=50	Exogen	0.4479	0.3142	0.5618
	$\rho = 0.2$	0.6085	0.4763	0.8134
	$\rho = 0.5$	0.5925	0.4515	0.7357
	$\rho = 0.8$	0.5971	0.4415	0.7929
n=500	Exogen	0.4725	0.4541	0.5352
	$\rho = 0.2$	0.5347	0.4994	0.5809
	$\rho = 0.5$	0.6051	0.5881	0.6683
	$\rho = 0.8$	0.6516	0.6188	0.6971
n=1000	Exogen	0.5027	0.4388	0.5003
	$\rho = 0.2$	0.5558	0.5106	0.5660
	$\rho = 0.5$	0.6101	0.5626	0.6195
	$\rho = 0.8$	0.6527	0.6404	0.6912
n=5000	Exogen	0.5117	0.4899	0.5155
	$\rho = 0.2$	0.5415	0.5291	0.5554
	$\rho = 0.5$	0.6007	0.5904	0.6155
	$\rho = 0.8$	0.6644	0.6561	0.6801

Table 1: Jakckknife simulation of β quantiles for one given sample of the Monte Carlo (MC) simulation results reported in the 2^{nd} column, DGP $\sim$ N(0,5), $|\Delta x|$ weighted full-pairwise estimator, number of repetitions: 10000 and jackknife sample size (d): 2535

3.5 What about the Intercept?

The discussion above focuses on the slope. This section investigates the relation between pairwise parameter and the intercept in Equation (1). The most obvious approach is to consider

$$\hat{\beta}_0 = \bar{y}_n - \hat{\beta}_1 \bar{x}_n \,, \tag{18}$$

where $\hat{\beta}_1$ is any consistent estimator of β_1 with $\bar{y}_n$ and $\bar{x}_n$ denoting the averages of y_i and x_i over a sample of n observations, respectively. That is, $\bar{y}_n = n^{-1} \sum_{i=1}^{n} y_i$ and similarly for $\bar{x}_n$.

It is straightforward to show that $\hat{\beta}_0$ is consistent since

$$\begin{aligned} \hat{\beta}_0 =& \bar{y}_n - \hat{\beta}_1 \bar{x}_n \\ =& \left(\beta_1 - \hat{\beta}_1\right) \bar{x}_n + \bar{u}_n \end{aligned}$$

and as $n \to \infty$, $\beta_1 - \hat{\beta}_1 = o_p(1)$, $\bar{x}_n - \mathbb{E}(x_i) = o_p(1)$ an $\bar{u}_n = o_p(1)$. Thus, by the Continuous Mapping Theorem, $\hat{\beta}_0 - \beta_0 = o_p(1)$.

While this provides a consistent estimator, we would also like to examine if there is a connection between the pairwise parameter of β_0 and β_0. Define the pairwise parameter of the intercept as

$$\beta_{0,(i,j)} = y_i - \beta_{1,(i,j)} x_i. \tag{19}$$

3.5.1 The adjacent case

In the adjacent case we consider the following estimator of β_0

$$\hat{\beta}_0 = \left(\sum_{i=1}^{n} w_i\right)^{-1} \left(\sum_{i=1}^{n} w_i \beta_{0,i}\right). \tag{20}$$

Since with $w_i = |\Delta x_i|$ the non-sorted observations is the only case that produces consistent estimate of β_1, we will focus on this particular case for β_0.

$$\begin{aligned}\hat{\beta}_0 =& \left(\sum_{i=2}^{n} |\Delta x_i|\right)^{-1} \left(\sum_{i=2}^{n} |\Delta x_i| \beta_{0,i}\right) \\ =& \beta_0 + \left(\sum_{i=2}^{n} |\Delta x_i|\right)^{-1} \left[\beta_1 \sum_{i=2}^{n} |\Delta x_i| x_i - \sum_{i=2}^{n} |\Delta x_i| x_i \beta_{1,i} + \sum_{i=2}^{n} |\Delta x_i| u_i\right].\end{aligned} \tag{21}$$

Expanding the following term further gives

$$\begin{aligned}\sum_{i=2}^{n} |\Delta x_i| x_i \beta_{1,i} =& \sum_{i=2}^{n} \operatorname{sgn}(\Delta x_i) x_i \Delta y_i \\ =& \beta_1 \sum_{i=2}^{n} |\Delta x_i| x_i + \sum_{i=2}^{n} \operatorname{sgn}(\Delta x_i) x_i \Delta u_i.\end{aligned}$$

Substitute the last expression above into Equation (21) yields

$$\hat{\beta}_0 = \beta_0 + \left(\sum_{i=2}^{n} |\Delta x_i|\right)^{-1} \left(\sum_{i=2}^{n} |\Delta x_i| u_i - \sum_{i=2}^{n} \operatorname{sgn}(\Delta x_i) x_i \Delta u_i\right). \tag{22}$$

Under the assumption that $x_i \perp u_i$ and the usual regularity assumptions on x_i, then by the WLLN, $\sum_{i=2}^{n} |\Delta x_i| u_i = o_p(1)$ and $\sum_{i=2}^{n} \operatorname{sgn}(\Delta x_i) x_i \Delta u_i$. The former holds, because the sum is converging to $\mathbb{E}(|\Delta x_i| u_i) = \mathbb{E}(|\Delta x_i|)\mathbb{E}(u_i) = 0$ and the second expression holds because the sum is converging to $\mathbb{E}[\operatorname{sgn}(\Delta x_i) x_i \Delta u_i] = \mathbb{E}[\operatorname{sgn}(\Delta x_i) x_i]\mathbb{E}(\Delta u_i) = 0$.

3.6 Loss Function

Next, we analyse the properties of the estimator (3):

$$\hat{\beta} = \arg\min_{\beta} \sum_{i=1}^{n} \left[w_i \left(\beta_i - \beta \right) \right]^2 \ , \tag{23}$$

where

$$\beta_i = \frac{\Delta y_i}{\Delta x_i}.$$

3.6.1 $w_i = |\Delta x_i|$

Given the quadratic loss, this section will show that the case when $w_i = \Delta x_i$ is identical to the case when $w_i = |\Delta x_i|$.

The solution to Equation (23) implies

$$\sum_{i=1}^{n} w_i^2 \left(\beta_i - \hat{\beta} \right) = 0,$$

and therefore

$$\hat{\beta} = \left(\sum_{i=1}^{n} w_i^2 \right)^{-1} \left(\sum_{i=1}^{n} w_i^2 \beta_i \right).$$

Thus when $w_i = |\Delta x_i|$ or when $w_i = \Delta x_i$ we get

$$\hat{\beta} = \left(\sum_{i=1}^{n} |\Delta x_i|^2 \right)^{-1} \left(\sum_{i=1}^{n} \Delta x_i \Delta y_i \right) \ ,$$

which is in fact the OLS estimator for the 'first difference' equation

$$\Delta y_i = \Delta x_i \beta + \Delta u_i \ . \tag{24}$$

3.6.2 Full-pairwise

Reconsider the estimator in the full-pairwise case

$$\hat{\beta} = \arg\min_{\beta} \sum_{i=1}^{n} \sum_{j=1}^{i-1} \left[w_{ij} \left(\beta_{ij} - \beta \right) \right]^2 \ , \tag{25}$$

which gives

$$\hat{\beta} = \left(\sum_{i=1}^{n} \sum_{j=1}^{i-1} w_{ij}^2 \right)^{-1} \left(\sum_{i=1}^{n} \sum_{j=1}^{i-1} w_{ij}^2 \Delta \beta_{ij} \right).$$

Now, if we set $w_{ij} = |\Delta x_{ij}|$ then

$$\hat{\beta} = \beta + \left(\sum_{i=1}^{n} \sum_{j=1}^{i-1} |\Delta x_{ij}|^2 \right)^{-1} \left(\sum_{i=1}^{n} \sum_{j=1}^{i-1} \Delta x_{ij} \Delta u_{ij} \right) , \tag{26}$$

which under the assumption that $x_i \perp u_i$, shows that the estimator is consistent. The same applies to $w_{ij} = \Delta x_{ij}$. The equivalence is mostly due to the quadratic loss.

An interesting case is $w_{ij} = \sqrt{|\Delta x_{ij}|}$, then

$$\hat{\beta} = \left(\sum_{i=1}^{n} \sum_{j=1}^{i-1} |\Delta x_{ij}| \right)^{-1} \left(\sum_{i=1}^{n} \sum_{j=1}^{i-1} |\Delta x_{ij}| \hat{\beta}_{ij} \right) ,$$

which is our full-pairwise estimator.

4 Multivariate Extension

One way to extend the approach to multiple explanatory variables is to utilise the residual matrix. Consider the following Data Generating Process

$$y_i = \beta_0 + \sum_{k=1}^{K} \beta_k x_{ki} + u_i \qquad u_i \sim D(0, \sigma_u^2) \quad i = 1, \ldots, n \tag{27}$$

with the matrix representation

$$\mathbf{y} = \mathbf{X}\beta + u \, , \tag{28}$$

where $\mathbf{y} = (y_1, \ldots, y_n)'$ is a $n \times 1$ vector containing the dependent variable, $\mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_K)$ is a $n \times K$ matrix containing the K explanatory variables such that $\mathbf{X}$ is $\mathbf{x}_k = (x_{k1}, \ldots, x_{kn})'$ for $k = 1, \ldots, K$ and $\mathbf{u} = (u_1, \ldots, u_K)'$ is a $n \times 1$ vector containing the residuals.

It will be convenient to introduce some special matrices here. These matrices allow us to represent the pairwise parameter as a sequence of matrix operations. Define $\mathbf{d}_{i,j}$ as a $1 \times n$ vector such that the i^{th} and j^{th} elements are 1 and -1, respectively while all other elements are 0. Let $\mathbf{D}_F = \left(\mathbf{d}'_{n,n-1}, \mathbf{d}'_{n,n-2}, \ldots, \mathbf{d}'_{n,1}, \mathbf{d}'_{n-1,n-2}, \ldots, \mathbf{d}'_{n-1,1}, \ldots, \mathbf{d}'_{2,1}\right)'$ and $\mathbf{D}_a = \left(\mathbf{d}'_{2,1}, \mathbf{d}'_{3,2}, \mathbf{d}'_{4,3} \ldots, \mathbf{d}'_{n,n-1}\right)'$. Given these matrices, it is straightforward to show that $\mathbf{D}_F\mathbf{x} = \{\Delta x_{ij}\}_{i=2,j=1}^{n,i-1}$ is a $n(n-1)/2 \times 1$ vector and $\mathbf{D}_a\mathbf{x} = \{\Delta x_i\}_{i=2}^{n}$ is a $(n-1) \times 1$ vector. The former gives the vector of Full-pairwise Difference while the latter yields the Adjacent Difference.

Let $\mathbf{S} : \mathbb{R}^n \to \mathbb{R}^n$ be a function that returns a $n \times 1$ zero-one selection matrix such that if $\mathbf{z} = \mathbf{S}(\mathbf{x})\mathbf{x}$ then $\mathbf{z}$ will be a $n \times 1$ vector containing the same elements as $\mathbf{x}$ such that $z_i \geq z_j$ for $j > i$. Thus, $\mathbf{S}(\mathbf{x})\mathbf{x}$ produces the sorted version of $\mathbf{x}$ in descending order. Finally, define the residual maker

$$\mathbf{M}_k = \mathbf{I}_n - \mathbf{X}_{-k}\left(\mathbf{X}'_{-k}\mathbf{X}_{-k}\right)^{-1}\mathbf{X}_{-k}\,, \tag{29}$$

where $\mathbf{X}_{-k}$ denote the $n \times K - 1$ matrix which contains all the columns of $\mathbf{X}$ except the k^{th} column.

The first step is to develop the matrix representation of the univariate case, then the multivariate case can be derived by repeat applications of the residual maker. That is, we will construct the multivariate version by combining the univariate estimator for each of the regressors.

$$\mathbf{y} = \mathbf{1}\beta_0 + \mathbf{x}\boldsymbol{\beta}_1 + \mathbf{u}$$

and note that $\mathbf{D}\mathbf{y} = \mathbf{D}\mathbf{x}\beta_1 + \mathbf{D}\mathbf{u}$ for $\mathbf{D} = \{\mathbf{D}_a, \mathbf{D}_F\}$. The pairwise parameters for the sample can be written as

$$\boldsymbol{\beta}_1 = \mathbf{diag}\left(\mathbf{x}\right)^{-1}\mathbf{D}\mathbf{y}\,, \tag{30}$$

where $\mathbf{diag}\left(\mathbf{x}\right) = \left[\mathbf{I} \otimes (\mathbf{D}\mathbf{x})'\right]\mathbf{S}_n$ is a diagonal matrix with the elements of $\mathbf{x}$ in the diagonal. $\mathbf{S}_n$ is a $n^2 \times n$ zero-one matrix such that all the $((i-1)n+1, i)$ elements are 1 for $i = 1, \ldots, n$ and all the other elements are 0. Note that Equation (30) is reasonably general and covers all the previously discussed case by defining the transformation matrix $\mathbf{D}$ differently. A summary can be found in Table 2.

Cases	$\mathbf{D}$
Adjacent	$\mathbf{D}_a$
Adjacent Sorted	$\mathbf{D}_a\mathbf{S}(\mathbf{x})$
Full-pairwise	$\mathbf{D}_F$
Full-pairwise Sorted	$\mathbf{D}_F\mathbf{S}(\mathbf{x})$

Table 2: Definitions of $\mathbf{D}$ based on different construction of pairwise parameters

Given the vector of pairwise parameters $\boldsymbol{\beta}_1$, the estimator of β_1 can be written as

$$\hat{\beta}_1 = \left(\mathbf{w}_1\mathbf{i}'\right)^{-1}\mathbf{w}'_1\boldsymbol{\beta}_1\,, \tag{31}$$

where $\mathbf{w}_1$ is the vector of weights associated with each $\beta_{1,(i,j)}$ in $\boldsymbol{\beta}_1$. Following from the analysis above, some of the choices include $\mathbf{w}_1 = (|\Delta x_1|, \ldots, |\Delta x_n|)'$ and $\mathbf{w}_1 = (\Delta x_1, \ldots, \Delta x_n)'$.

It is straightforward to generalise Equation (30), and subsequently Equation (31), to the case when $\mathbf{X}$ is a $n \times K$ matrix with K regressors. Consider

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u} \tag{32}$$

$$= \mathbf{x}_k\beta_k + \mathbf{X}_{-k}\boldsymbol{\beta}_{-k} + \mathbf{u} \tag{33}$$

$$\mathbf{M}_k\mathbf{y} = \mathbf{M}_k\mathbf{x}_k\beta_k + \mathbf{u}\,, \tag{34}$$

where $\boldsymbol{\beta}_{-k}$ contains the same elements as $\boldsymbol{\beta}$ with the k^{th} element removed. Following Equation (30) the pairwise parameter of β_k is

$$\boldsymbol{\beta}_k = \mathbf{diag}\left(\mathbf{M}_k\mathbf{x}_k\right)^{-1}\mathbf{D}\mathbf{M}_k\mathbf{y}. \tag{35}$$

Similarly, Equation (31) can be generalised as

$$\hat{\beta}_k = \left(\mathbf{w}_k\mathbf{i}'\right)^{-1}\mathbf{w}_k'\boldsymbol{\beta}_k \tag{36}$$

and

$$\hat{\boldsymbol{\beta}} = \begin{bmatrix} \left(\mathbf{w}_1\mathbf{i}'\right)^{-1}\mathbf{w}_1' & \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \ddots & \vdots & \dots & \mathbf{0} \\ \mathbf{0} & \dots & \left(\mathbf{w}_k\mathbf{i}'\right)^{-1}\mathbf{w}_k' & \dots & \mathbf{0} \\ \vdots & \dots & \vdots & \ddots & \dots \\ \mathbf{0} & \dots & \dots & \dots & \left(\mathbf{w}_K\mathbf{i}'\right)^{-1}\mathbf{w}_K' \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta}_1 \\ \vdots \\ \boldsymbol{\beta}_k \\ \vdots \\ \boldsymbol{\beta}_K \end{bmatrix}. \tag{37}$$

As shown in previous results, suitable choice of $\mathbf{w}_k$ leads to consistency of $\hat{\beta}_k$ and therefore $\hat{\boldsymbol{\beta}} - \boldsymbol{\beta} = o_p(1)$ under the same conditions as those discussed above for all $k = 1, \dots, K$.

5 Testing for Endogeneity

There are multiple endogeneity tests available for empirical use, but all rely on some kind of additional or external information, mostly in the form of instrumental variables (see e.g., Hausman (1978), or Wooldridge (2002), pp. 118-122).

The aim here is to develop two tests for endogeneity based on the estimator of the form

$$\hat{\beta} = \left(\sum_{i=2}^{n}\sum_{j=1}^{i-1} w_{ij}\right)^{-1}\left(\sum_{i=2}^{n}\sum_{j=1}^{i-1} w_{ij}\beta_{ij}\right)$$

which solely relies on data already in use for the estimation.

5.1 Residuals Test

The residuals test is particularly useful when $\beta_0 = 0$. Consider $w_{ij} = |\Delta x_{ij}|$ which implies

$$\hat{\beta} = \beta + \delta_n\,, \tag{38}$$

where

$$\delta_n = \left(\sum_{i=2}^{n}\sum_{j=1}^{i-1} |\Delta x_{ij}|\right)^{-1}\left(\sum_{i=2}^{n}\sum_{j=1}^{i-1} \operatorname{sgn}\left(\Delta x_{ij}\right)\Delta u_{ij}\right) \tag{39}$$

and under $x_i \perp u_i$, $\delta_n = o_p(1)$. Now given the following specification

$$y_i = x_i\beta + u_i \tag{40}$$

the estimated residual from the EwPO estimator is

$$\begin{aligned}\hat{u}_i =& y_i - x_i\hat{\beta} \\ =& x_i(\beta - \hat{\beta}) + u_i \\ =& -x_i\delta_n + u_i.\end{aligned}$$

It is straightforward to show that

$$\begin{aligned}n^{-1}\sum_{i=1}^{n}\hat{u}_i =& -n^{-1}\sum_{i=1}^{n}x_i\delta_n + u_i \\ =& -\delta_n n^{-1}\sum_{i=1}^{n}x_i + n^{-1}\sum_{i=1}^{n}u_i\,.\end{aligned}$$

Now as $n \to \infty$ the last line above is

$$n^{-1}\sum_{i=1}^{n}\hat{u}_i = -\delta_n\mu_x + o_p(1) \tag{41}$$

under the assumption that $\mathbb{E}(u_i) = 0$. There are two cases when $n^{-1}\sum_{i=1}^{n}\hat{u}_i$ is $o_p(1)$ namely, $x_i \perp u_i$ or $\mu_x = 0$. Assuming $\mu_x \neq 0$ which is always possible by appropriate transformations on the regressors, it is therefore possible to *directly* test for endogeneity by testing if the mean of the estimated residuals is statistically different from 0 using standard testing procedure, such as the t-test.[2]

5.2 Covariance Test

Note that $\beta_0 = 0$ is often too restrictive in practice. One way to alleviate this is to remove the intercept by considering the model in difference. Consider $\Delta\hat{u}_{pq} = \Delta y_{pq} - \Delta x_{pq}\hat{\beta}$ i.e., the 'estimated' residuals in the form

$$\begin{aligned}\Delta\hat{u}_{pq} =& \Delta y_{pq} - \Delta x_{pq}\hat{\beta} \\ =& \Delta x_{pq}\left(\beta - \hat{\beta}\right) + \Delta u_{pq}.\end{aligned} \tag{42}$$

In the case of a consistent estimator $\hat{\beta}$ can be expressed in the form of $\hat{\beta} = \beta + \delta(\mathbf{w}, \mathbf{u})$ where $\mathbf{w}$ and $\mathbf{u}$ denote the vectors of the weights i.e., w_{ij} and the residuals

[2]The argument here also applies to the usual Ordinary Least Squares (OLS) estimator. That is, when $\beta_0 = 0$, the estimated residuals do not have 0 mean. Thus, it also provides a test of endogeneity in this special case.

u_i, respectively, for $i, j = 1, \dots, n$ and B vanishes to 0 asymptotically under ideal conditions, Therefore

$$\Delta\hat{u}_{pq} = \Delta x_{pq}\delta(\mathbf{w}, \mathbf{u}) + \Delta u_{pq}.$$

So consider the test statistics

$$S(\mathbf{w}) = n^{-2}\sum_{p=2}^{n}\sum_{q=1}^{p-1}\Delta x_{pq}\Delta\hat{u}_{pq}. \tag{43}$$

5.2.1 Case 1: $\mathbf{w} = \{\Delta x_{ij}\}_{i,j=1}^{n}$

In this case,

$$\Delta\hat{u}_{pq} = \Delta x_{pq}\left(\sum_{i=2}^{n}\sum_{j=1}^{i-1}\Delta x_{ij}\right)^{-1}\left(\sum_{i=2}^{n}\sum_{j=1}^{i-1}\Delta u_{ij}\right) + \Delta u_{pq}$$

and the test statistics

$$\begin{aligned} S(\mathbf{w}) =& n^{-2}\left(\sum_{p=1}^{n}\sum_{q=1}^{p-1}\Delta x_{pq}^2\right)\left(\sum_{i=2}^{n}\sum_{j=1}^{i-1}\Delta x_{ij}\right)^{-1}\left(\sum_{i=2}^{n}\sum_{j=1}^{i-1}\Delta u_{ij}\right) \\ &+ n^{-2}\left(\sum_{p=2}^{n}\sum_{q=1}^{p-1}\Delta x_{pq}\Delta u_{pq}\right). \end{aligned} \tag{44}$$

Under the null that $x_i \perp u_i$, $S(\Delta x_{ij})$ has mean 0. The asymptotic behaviour of EwPO estimator as defined in Equation (11) with $w_{ij} = \Delta x_{ij}$ and that of the test statistics as defined above can be found in Propositions 1 and 2, respectively.

Before presenting the theoretical results, consider the following assumptions

Assumption 1. $\mathbb{E}(x_i) = \mu_x < \infty\ \forall i = 1, \dots, n$ and $\exists \sigma_x^2 < \infty$ such that $n^{-1}\sum_{i=1}^{n}(x_i - \mu_x)^2 - \sigma_x^2 = o_p(1)$.

Assumption 2. $\mathbb{E}(u_i) = 0\ \forall i = 1, \dots, n$ and $\exists \sigma_u^2 < \infty$ such that $n^{-1}\sum_{i=1}^{n} u_i^2 - \sigma_u^2 = o_p(1)$.

Assumption 3. $\mathbb{E}(u_i|x_j) = 0$ for $i, j = 1, \dots, n$.

Assumption 4. Δx_{ij} and Δu_{ij} are mixingales as defined in McLeish (1975) and satisfy all the conditions for Theorem 3.8 in McLeish (1975).

Assumptions 1 and 2 are standard in the sense that the existence of second moments are required for both the regressor and the residuals. The exogeneity assumption as presented in Assumption 3 is required for the consistency of the estimator. It also stipulates the null underlying the distribution of the test statistics as asserted in Proposition 2. Assumption 4 is required to ensure that the various partial sums converge to the Brownian Motion process using the *Functional Central Limit Theorem.* It is required due to the partial sums from the full-pair construction of the estimator. See also Wooldridge and White (1988) for potentially more general assumptions on x_i and u_i.

Asymptotic distribution of the full-pairwise EwPO with $w_{ij} = \Delta x_{ij}$ can be found in Proposition 1 below

Proposition 1. *Under Assumptions 1 to 4 with $\hat{\beta}$ as defined in Equation (11) and $\mathbf{w} = \{\Delta x_{ij}\}_{i,j=1}^{n}$:*

$$\hat{\beta} - \beta \xrightarrow{d} \frac{W(1) - 2\int_0^1 W(\lambda)d\lambda}{B(1) - 2\int_0^1 B(\lambda)d\lambda}. \tag{45}$$

Proof. See Appendix. □

The proposition below gives the asymptotic distribution of the covariance test statistics generated by the estimator considered in Proposition 1 above.

Proposition 2. *Under the assumptions of Proposition 1*

$$\begin{aligned}(\sigma_x\sigma_u)^{-1}S(\mathbf{w}) \xrightarrow{d} & B(1)W(1) + \int_0^1 B(\lambda)W(\lambda)d\lambda - \int_0^1 B(\lambda)dW(\lambda) \\ & - \int_0^1 W(\lambda)dB(\lambda) - B^2(1)\frac{W(1) - 2\int_0^1 W(\lambda)d\lambda}{B(1) - 2\int_0^1 B(\lambda)d\lambda},\end{aligned} \tag{46}$$

where $B(\lambda)$ and $W(\lambda)$ are independent standard Brownian motion processes.

Proof. See Appendix C □

Remark 1. It can be shown that both terms on the right hand side of Equation (46) are negative on expectation. Under the alternative that u_i and x_i are correlated, then the values of the last two stochastic integrals in Equation (46) increase as the correlation increases.

Remark 2. The asymptotic distribution as stated in Equation (46) is obviously non-standard and must be simulated in order to obtain the critical value(s) for the purpose

α	Lower CV	Upper CV
0.01 %	-4.129	3.913
0.05 %	-3.021	2.912
0.10 %	-2.530	2.498

Table 3: Simulated Critical Values: $w_{ij} = \Delta x_{ij}$

of inference. Since $B(\lambda)$ and $W(\lambda)$ are independent Brownian processes, each of the stochastic integrals can be simulated as

$$\int_0^1 B(\lambda)d\lambda \approx \sum_{p=1}^{n} z_{1p} \tag{47}$$

$$\int_0^1 W(\lambda)d\lambda \approx \sum_{p=1}^{n} z_{3p} \tag{48}$$

$$\tag{49}$$

where $z_{ip} = z_{ip-1} + \varepsilon_{ip}$, $\boldsymbol{\varepsilon}_p = (\varepsilon_{1p}, \varepsilon_{2p})' \sim NID(0, \mathbf{I})$ and

$$\int_0^1 W(\lambda)dB(\lambda) \approx \sum_{p=1}^{n} z_{2p}\varepsilon_{1p} \tag{50}$$

$$\int_0^1 B(\lambda)dW(\lambda) \approx \sum_{p=1}^{n} z_{1p}\varepsilon_{2p}. \tag{51}$$

Under the null, other quantities in the distribution namely, σ_x and σ_u can be estimated consistently from the data and estimated residuals, respectively. Thus the asymptotic distribution can be simulated by repeated computation of Equations (47) - (51) and the substitution their values into Equation (46) for desired number of replications. For further details on the simulation of stochastic integrals, see for example, Johansen (1995).

By way of demonstration, the critical values from the simulated sample of Equation (46) using Equations (47) - (51) with $\sigma_x = \sigma_u = 1$ and $n = 10000$ can be found in Table 3 and the respective histogram in Figure 3. Note that this is simulated with $w_{ij} = \Delta x_{ij}$ and may not be appropriate when $w_{ij} = |\Delta x_{ij}|$. In this case, for example, the jackknife procedure may provide more accurate approximation.

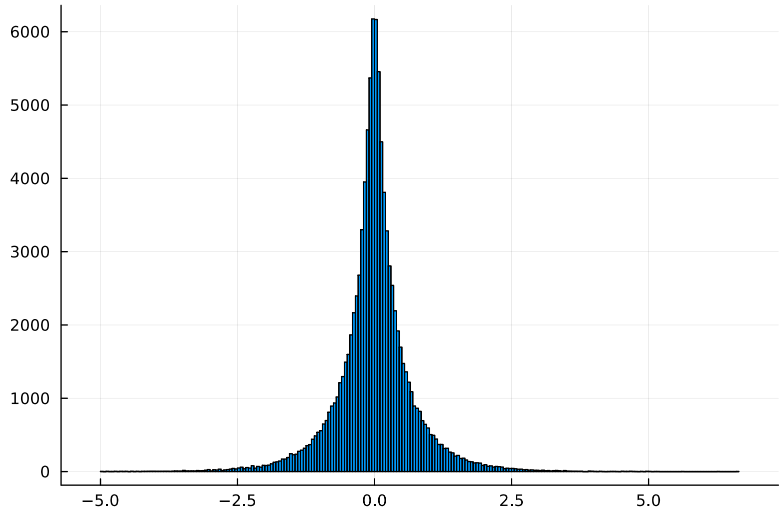

Figure 3: Simulated Sample of the Distribution

5.2.2 Case 2: $w_{ij} = |\Delta x_{ij}|$

Similar derivation as above yields

$$\begin{aligned} S(\mathbf{w}) =& n^{-2} \left(\sum_{p=1}^{n} \sum_{q=1}^{p-1} \Delta x_{pq}^2 \right) \left(\sum_{i=2}^{n} \sum_{j=1}^{i-1} |\Delta x_{ij}| \right)^{-1} \left(\sum_{i=2}^{n} \sum_{j=1}^{i-1} \operatorname{sgn}(\Delta x_{ij}) \Delta u_{ij} \right) \\ &+ n^{-2} \left(\sum_{p=2}^{n} \sum_{q=1}^{p-1} \Delta x_{pq} \Delta u_{pq} \right) . \end{aligned} \tag{52}$$

Given the absolute value of the weight, the asymptotic distribution of the test statistics is more challenging to obtain. However, the jackknife procedure similar to those proposed in Section 3.4.3 can easily be used to obtain the critical value of the test statistics in this case as well.

Appendix B contains some Monte Carlo simulations of the test statistics with the two different weighting schemes under the null and selected alternatives. As shown in Appendix B, the mean test statistics deviates from zeros as the correlation between the regressor and the error term grows. Further investigations would suggest that the covariance test has the right size with relatively good power, even in small sample.

5.3 Implications

Basically the results can be used in two ways. First, we could calculate the appropriate critical values for a given data set and perform the testing procedure formally.

However, this requires the simulation of the critical values which is not always convenient in practice.

Another way, however, to look at this is based on the fact that the value of the test statistic is a monotonically increasing function (in absolute value) of the correlation between x_i and u_i, i.e., the degree of endogeneity. Therefore, the following selection procedure can be performed to reduce or minimise the risk of estimation bias due to endogeneity:

a) Estimate the model, e.g., $y_i = \beta x_i + u_i$ with EwPO;

b) Estimate the model with EwPO after being transformed by the potential instrumental variable (IV) candidate, $z_i = \beta w_i + v_i$ where $z_i = g_i y_i$, $w_i = g_i x_i$ and $v_i = g_i u_i$ with g_i an IV candidate that satisfies the same assumptions as x_i;

c) Select the model for inference with the lowest statistics value. This will minimise the risk of endogeneity bias.

Note that the procedure above does not require the simulation of any distribution but a repeat application of the EwPO estimator. Thus, it is potentially a simpler but effective procedure.

6 Consistent Estimation with Endogeneity

From Equation (41), it is straightforward to show that

$$\delta_n = -(n\mu_x)^{-1} \sum_{i=1}^{n} \hat{u}_i + o_p(1) \tag{53}$$

and substitute the right hand side into Equation (38) yields

$$\hat{\beta} + (n\mu_x)^{-1} \sum_{i=1}^{n} \hat{u}_i - \beta = o_p(1)\,, \tag{54}$$

which means that

$$\hat{\beta}_1 = \hat{\beta} + (n\mu_x)^{-1} \sum_{i=1}^{n} \hat{u}_i \tag{55}$$

is in fact a consistent estimator of β.

Equation (55) can be interpreted as a bias correction for the EwPO estimator in the presence of endogeneity. The results from the Monte Carlo simulation (see Appendix C) suggests it works remarkably well. Specifically, Tables 11 and 12 contains the mean of the estimated residuals and the mean of the β estimates, respectively. In each case when $\rho \neq 0$, applying the bias correction as defined in Equation (55) produce an estimate that is indeed very close to the true value (see, e.g., Table 13).

Note that the reduction in the variance as the sample size increases is an implication of the consistency property. To illustrate this, consider

$$
\begin{aligned}
y_i =& \beta x_i + u_i \\
u_i =& \rho(x_i - \mu_x) + v_i.
\end{aligned}
$$

Since $\hat{\beta}$ is consistent, this means $\hat{\beta} - \beta - \rho = o_p(1)$, with the bias induced by ρ, i.e., the linear relation between u_i and x_i. This means $\hat{u}_i + \rho\mu_x - v_i = o_p(1)$. Given $\mathbb{E}(v_i) = 0$, one would expect that the sample mean of $\hat{u}_i$ is converging to $-\rho\mu$ with decreasing variance as n increases.

The confidence interval of the bias-corrected estimator as defined in Equation (55) can be obtained by modifying the jackknife procedure proposed in Section 3.4.3. Specifically, by replacing Step 6 with the following:

Step 6a. Compute $\hat{\beta}_1$ using the jackknife sample J_i, call it $\hat{\beta}_{1i}$.

The bias-corrected EwPO may also potentially be useful for testing endogeneity in the form of a Hausman's 1978 test. Since the bias-corrected EwPO is consistent under the null and the alternative, it is a valid candidate for a Hausman's type test. This could be an interesting area of future research.

7 Conclusion

Endogeneity has been an enduring problem in econometrics. The practice to compensate for the information loss due to the correlation between the explanatory variable(s) and the disturbance terms has been to use additional information in the form of instrumental variables or moment conditions for testing and consistent estimation. The problem with this approach is that the results hinge on the "quality" of this information. The common wisdom to the day is that testing for endogeneity and consistent estimation of a regression model with endogeneity is not possible based exclusively on the information provided by the observations of the variables a model. This paper through the introduction of a new estimation method, the so-called estimation with pairwise observations (EwPO), demonstrates that this belief is misconceived and shows how to test directly for endogeneity. It also derives a simple new consistent estimation procedure for the linear regression model with endogeneity, which can easily be implemented in practice.

Appendix A: Monte Carlo Simulations of EwPO Estimation with Selected Weights

This Appendix presents some Monte Carlo (MC) simulation results to examine the properties of some EwPO estimators with selected weights. Additional results can be found in the Online Supplement (see Chan et al. (2023)).

The data generating process for the MC simulations is based on the model

$$y_i = \beta_0 + \beta_1 x_i + u_i$$

and the MC experiments consider two possible distributions for u_i, namely

1. $u_i \sim N(0, 1)$,
2. $u_i \sim$ skewed normal distribution,

where the skewed normal distribution is generated as

$$u_i = \xi + \lambda|v_i| + z_i,$$

with $\xi = -\lambda\sqrt{\frac{2}{\pi}}$, $v_i \sim N(0, 1)$ and $z_i \sim N(0, \sigma^2)$ such that v_i and w_i are independently distributed.

The MC experiments consider uniform distribution U(-10,10) for the regressor x_i.

Two parameter vectors have been considered $(\beta_0, \beta_1) = (1, 0.5)$ and $(\beta_0, \beta_1) = (1, 1.5)$ for purposes of robustness checking. The number of MC replications was 1000. For the sake of brevity, only the results for $(\beta_0, \beta_1) = (1, 0.5)$ are reported here. The other results can be found in the Online Supplement (Chan et al. (2023)).

Sorted MC	Estimates and MC standard errors			
	Parameter	Estimate/S.e.	OLS	pairwise
n=50	$\hat{\beta}_0$	Estimate	0.9866	0.9866
		S.e.	0.1609	0.1609
	$\hat{\beta}_1$	Estimate	0.4999	0.4999
		S.e.	0.2928	0.2928
n = 500	$\hat{\beta}_0$	Estimate	1.0001	1.0001
		S.e.	0.0504	0.0504
	$\hat{\beta}_1$	Estimate	0.4971	0.4971
		S.e.	0.0948	0.0948
n = 5000	$\hat{\beta}_0$	Estimate	0.9992	0.9992
		S.e.	0.0169	0.0169
	$\hat{\beta}_1$	Estimate	0.4997	0.4997
		S.e.	0.0281	0.0281

Table 4: Sorted – full-pairwise MC, $x_i \sim U(-10, 10)$, $u_i \sim$ skewed normal distribution, Δx weighted estimator

Sorted MC	Estimates and MC standard errors			
	Parameter	Estimate/S.e.	OLS	pairwise
n=50	$\hat{\beta}_0$	Estimate	0.9866	0.9866
		S.e.	0.1609	0.1609
	$\hat{\beta}_1$	Estimate	0.4999	0.4999
		S.e.	0.2928	0.2928
n = 500	$\hat{\beta}_0$	Estimate	1.0001	1.0001
		S.e.	0.0504	0.0504
	$\hat{\beta}_1$	Estimate	0.4971	0.4971
		S.e.	0.0948	0.0948
n = 5000	$\hat{\beta}_0$	Estimate	0.9992	0.9992
		S.e.	0.0169	0.0169
	$\hat{\beta}_1$	Estimate	0.4997	0.4997
		S.e.	0.0281	0.0281

Table 5: Non-sorted full-pairwise MC, $x_i \sim U(-10, 10)$, $u_i \sim$ skewed normal distribution, Δx weighted estimator

Note: No mistake, the sorted and non-sorted full-pairwise estimation results given in Tables 4 and 5 are exactly the same.

Non-sorted MC	Estimates and MC standard errors			
	Parameter	Estimate/S.e.	OLS	pairwise
n=50	$\hat{\beta}_0$	Estimate	1.0009	0.9967
		S.e.	0.1408	0.1997
	$\hat{\beta}_1$	Estimate	0.5007	0.5018
		S.e.	0.0251	0.0291
n = 500	$\hat{\beta}_0$	Estimate	0.9967	0.9966
		S.e.	0.0446	0.0640
	$\hat{\beta}_1$	Estimate	0.5001	0.4998
		S.e.	0.0081	0.0095
n = 5000	$\hat{\beta}_0$	Estimate	1.0001	1.0004
		S.e.	0.0144	0.0192
	$\hat{\beta}_1$	Estimate	0.4999	0.4999
		S.e.	0.0025	0.0031

Table 6: Non-sorted adjacent MC, $x_i \sim U(-10, 10)$, $u_i \sim N(0, 1)$, $|\Delta x|$ weighted estimator

Full-pairwise MC	Estimates and MC standard errors			
	Parameter	Estimate/S.e.	OLS	pairwise
n=50	$\hat{\beta}_0$	Estimate	1.0009	1.0009
		S.e.	0.1408	0.1408
	$\hat{\beta}_1$	Estimate	0.5007	0.5007
		S.e.	0.0251	0.0251
n = 500	$\hat{\beta}_0$	Estimate	0.9967	0.9967
		S.e.	0.0446	0.0446
	$\hat{\beta}_1$	Estimate	0.5001	0.5001
		S.e.	0.0081	0.0081
n = 5000	$\hat{\beta}_0$	Estimate	1.0001	1.0001
		S.e.	0.0144	0.0144
	$\hat{\beta}_1$	Estimate	0.4999	0.4999
		S.e.	0.0025	0.0025

Table 7: Sorted full-pairwise MC, $x_i \sim U(-10, 10)$, $u_i \sim N(0, 1)$, Δx weighted estimator

Appendix B: Monte Carlo Simulations for the Covariance Test

The MC setup considers again sample sizes $n = 50$, 500 and 5000 with 1000 replications.

Step 1. Generated the same model as above with one explanatory variable:

$$y_i = \beta_0 + \beta_1 x_i + u_i$$

with $\beta_0 = 1$, $\beta_1 = 0.5$ to star with, and u_i was generated as $N(0, 1)$. Finally, x was generated as $N(0, 5)$ and also as $U(-5, 5)$.

The simulations of x_i and u_i were carried out under four different correlations, namely $\rho = 0$ (benchmark exogeneity), $\rho = 0.2$ (small), $\rho = 0.5$ (medium), and $\rho = 0.8$ (large).

Step 2. Estimate the model with EwPO with $w_{ij} = \Delta x_{ij}$ and $w_{ij} = |\Delta x_{ij}|$. In each case, calculate the test statistics as defined in Equation (43).

For further simulation results please refer to the Online Supplement (Chan et al. (2023)).

Full-pairwise MC	Average test statistics				
	Parameter	Pairwise	Standard deviation	Skewness	Kurtosis
n=50	Exogen	-0.0003	0.7842	-0.1804	3.1856
	$\rho = 0.2$	-0.5478	0.8647	0.0273	3.1118
	$\rho = 0.5$	-0.6175	0.5483	0.0593	3.0018
	$\rho = 0.8$	-1.2625	0.4008	0.0122	3.1264
n = 500	Exogen	-0.0058	0.2175	-0.0497	3.0011
	$\rho = 0.2$	-0.4241	0.2075	-0.1293	3.0698
	$\rho = 0.5$	-1.1112	0.2017	0.0756	3.0814
	$\rho = 0.8$	-1.5581	0.1311	-0.0891	2.8930
n = 5000	Exogen	-0.0031	0.0666	0.0101	2.8834
	$\rho = 0.2$	-0.3995	0.0684	0.0123	3.1566
	$\rho = 0.5$	-0.9737	0.0604	-0.0610	2.9180
	$\rho = 0.8$	-1.5681	0.0421	0.0637	3.0883

Table 8: Average test statistics, full-pairwise MC, $x_i \sim N(0, 5)$, Δx weighted estimator

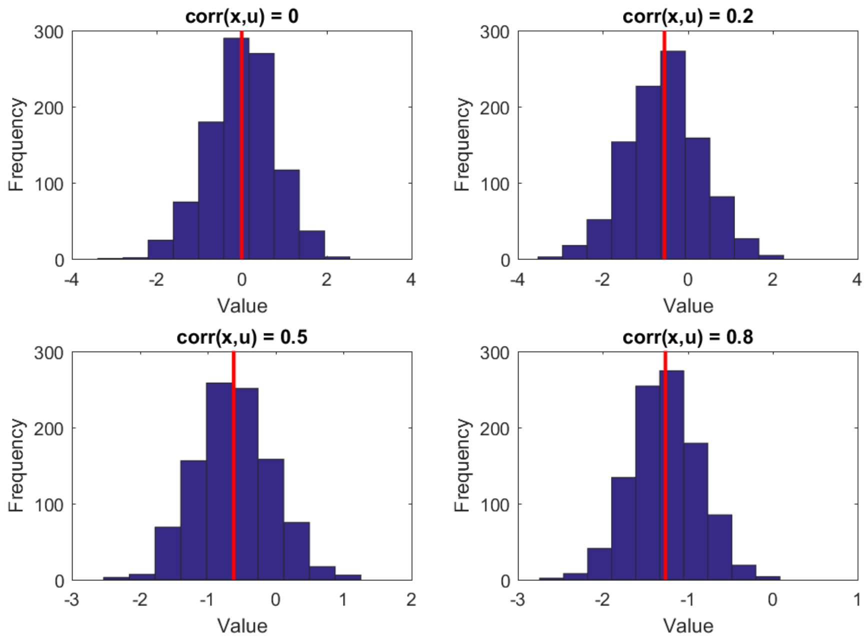

Figure 4: Distribution of the test-statistics, $x_i \sim$ N(0,5), Δx weighted full-pairwise estimator, $n = 50$

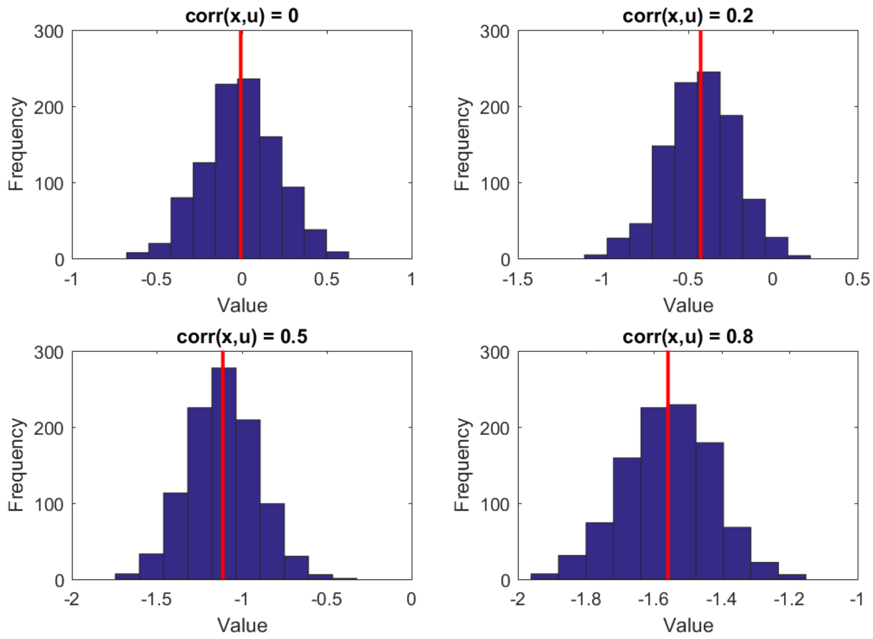

Figure 5: Distribution of the test-statistics, $x_i \sim$ N(0,5), Δx weighted full-pairwise estimator, $n = 500$

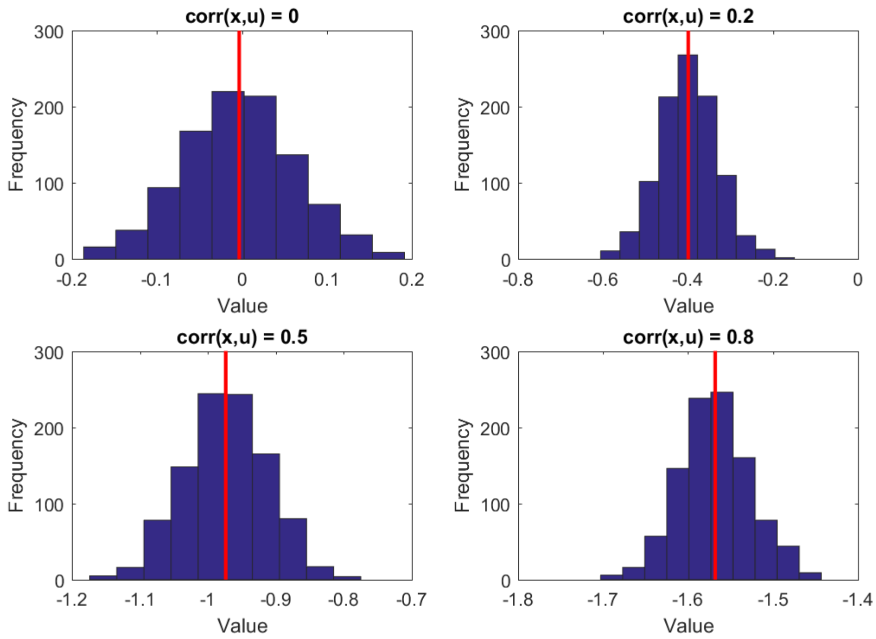

Figure 6: Distribution of the test-statistics, $x_i \sim$ N(0,5), Δx weighted full-pairwise estimator, $n = 5000$

Full-pairwise MC	Average test statistics				
	Parameter	Pairwise	Standard deviation	Skewness	Kurtosis
n=50	Exogen	0.0060	0.8890	0.0382	3.2387
	$\rho = 0.2$	-0.9247	0.7920	-0.0244	2.9620
	$\rho = 0.5$	-2.6616	0.7437	0.0121	3.2542
	$\rho = 0.8$	-5.9118	0.5729	-0.0305	3.0115
n = 500	Exogen	-0.0234	0.2750	-0.1429	3.1852
	$\rho = 0.2$	-1.0846	0.2471	0.0120	2.7804
	$\rho = 0.5$	-2.9665	0.2309	0.1625	3.2859
	$\rho = 0.8$	-4.2625	0.1494	0.0345	3.1253
n = 5000	Exogen	-0.0036	0.0785	0.0088	2.8812
	$\rho = 0.2$	-1.1328	0.0786	-0.0625	3.1951
	$\rho = 0.5$	-2.8321	0.0710	-0.0394	2.9287
	$\rho = 0.8$	-4.5415	0.0494	0.1022	3.0687

Table 9: Average test statistics, full-pairwise MC, $x_i \sim U(-5, 5)$, Δx weighted estimator

Full-pairwise MC	Average test statistics				
	Parameter	Pairwise	Standard deviation	Skewness	Kurtosis
n=50	Exogen	0.0191	0.8476	-0.1264	3.1170
	$\rho = 0.2$	-1.2884	0.8520	-0.1054	2.8218
	$\rho = 0.5$	-2.8584	0.7488	-0.0452	2.9483
	$\rho = 0.8$	-6.9515	0.6224	-0.0069	3.2675
n = 500	Exogen	0.0010	0.2608	-0.0092	2.9270
	$\rho = 0.2$	-1.2375	0.2588	0.0632	2.7668
	$\rho = 0.5$	-2.7598	0.2218	-0.0266	2.9173
	$\rho = 0.8$	-4.3437	0.1479	0.0420	2.9936
n = 5000	Exogen	-0.0003	0.0784	0.1196	2.8769
	$\rho = 0.2$	-1.1162	0.0790	-0.1253	2.8993
	$\rho = 0.5$	-2.8175	0.0703	-0.1483	3.0588
	$\rho = 0.8$	-4.5073	0.0490	-0.0011	2.9966

Table 10: Average test statistics, full-pairwise MC, $x_i \sim U(-5, 5)$, $|\Delta x|$ weighted estimator

Appendix C: Monte Carlo Results of the Residual Test and the Consistent Estimation

The DGP in this case is

$$y_i = 0.5x_i + u_i$$

where $x_i \sim N(5, 2)$ and $u_i \sim N(0, 1)$. The MC experiments considered no correlation and $\rho = 0.2, 0.5$ and 0.8 with sample sizes ranging from 50 to 5000 for each case.

Sample size	Correlation	Mean	Variance	Skewness	Kurtosis
n=50	Exogen	0.0044	0.1604	-0.0176	0.1476
	$\rho = 0.2$	-0.4930	0.1546	-0.0546	0.0601
	$\rho = 0.5$	-1.2506	0.1180	-0.0113	0.1834
	$\rho = 0.8$	-2.0033	0.0606	0.0223	0.1587
n=500	Exogen	0.0032	0.0149	0.0839	0.0023
	$\rho = 0.2$	-0.5005	0.0142	0.0264	-0.0169
	$\rho = 0.5$	-1.2517	0.0118	0.023	0.1509
	$\rho = 0.8$	-2.0015	0.0051	-0.0321	-0.2207
n=1000	Exogen	0.0009	0.0075	-0.0792	0.0818
	$\rho = 0.2$	-0.5027	0.0073	-0.0534	0.1680
	$\rho = 0.5$	-1.2474	0.0056	0.0687	0.0894
	$\rho = 0.8$	-2.0003	0.0028	0.0432	-0.0320
n=5000	Exogen	0.0003	0.0015	-0.0341	0.1900
	$\rho = 0.2$	-0.4993	0.0014	0.0541	0.1994
	$\rho = 0.5$	-1.2507	0.0012	0.0573	0.1016
	$\rho = 0.8$	-2.0003	0.0005	0.0722	-0.0429

Table 11: MC Results on the Distributions of $\hat{u}_i$.

Sample size	Correlation	Mean	Variance	Skewness	Kurtosis
$n = 50$	Exogen	0.4993	0.0056	0.0584	0.2268
	$\rho = 0.2$	0.5984	0.0054	0.0641	0.067
	$\rho = 0.5$	0.7505	0.004	0.013	0.0595
	$\rho = 0.8$	0.9001	0.0021	-0.0344	0.3254
$n = 500$	Exogen	0.4994	0.0005	-0.0126	0.0119
	$\rho = 0.2$	0.5998	0.0005	0.0287	-0.1062
	$\rho = 0.5$	0.7505	0.0004	-0.0401	0.0729
	$\rho = 0.8$	0.9003	0.0002	0.0101	-0.1041
$n = 1000$	Exogen	0.4998	0.0003	0.0331	0.0474
	$\rho = 0.2$	0.6006	0.0003	0.0736	0.2013
	$\rho = 0.5$	0.7498	0.0002	-0.0369	-0.0212
	$\rho = 0.8$	0.9001	0.0001	-0.0594	-0.0057
$n = 5000$	Exogen	0.5000	0.0001	0.0240	0.0152
	$\rho = 0.2$	0.5999	4.877e-5	-0.0424	0.1421
	$\rho = 0.5$	0.7502	4.055e-5	-0.0478	0.2321
	$\rho = 0.8$	0.9000	1.850e-5	-0.0246	-0.0227

Table 12: Monte Carlo Results on $\hat{\beta}$

Sample size	Correlation	Mean	Variance	Skewness	Kurtosis
$n = 50$	Exogen	0.5002	0.0008	0.0503	-0.0639
	$\rho = 0.2$	0.4998	0.0008	0.0861	0.2677
	$\rho = 0.5$	0.5004	0.0006	0.0015	-0.0670
	$\rho = 0.8$	0.4995	0.0003	0.0377	0.0951
$n = 500$	Exogen	0.5000	8.031e-5	-0.0401	0.0146
	$\rho = 0.2$	0.4997	7.297e-5	-0.1208	0.0535
	$\rho = 0.5$	0.5002	6.249e-5	0.0534	0.0321
	$\rho = 0.8$	0.5000	2.946e-5	0.0825	0.1877
$n = 1000$	Exogen	0.5000	3.975e-5	-0.0374	-0.1315
	$\rho = 0.2$	0.5000	3.871e-5	0.0496	-0.0301
	$\rho = 0.5$	0.5003	2.94e-5	-0.1190	0.0958
	$\rho = 0.8$	0.5001	1.435e-5	0.0450	0.1941
$n = 5000$	Exogen	0.5000	8.225e-6	0.0703	-0.0280
	$\rho = 0.2$	0.5000	7.724e-6	0.0096	0.0661
	$\rho = 0.5$	0.5000	5.731e-6	-0.0170	-0.0944
	$\rho = 0.8$	0.5000	2.933e-6	0.0195	0.0369

Table 13: Biased Corrected $\hat{\beta}$

Appendix D: Mathematical Appendix

Proof of Proposition 2.

Define

$$B_n(\lambda) = n^{-1}\sum_{i=1}^{\lfloor \lambda n \rfloor} \frac{x_i - \mu_x}{\sigma_x} \tag{D.1}$$

$$W_n(\lambda) = n^{-1}\sum_{i=1}^{\lfloor \lambda n \rfloor} \frac{u_i}{\sigma_u} \tag{D.2}$$

$$\tag{D.3}$$

where $1/n \leq \lambda \leq 1$ and $\lfloor x \rfloor$ denotes the largest integer contains in x. Under Assumptions (1) to (4) $B_n(\lambda) \xrightarrow{d} B(\lambda)$ and $W_n(\lambda) \xrightarrow{d} W(\lambda)$ following the arguments in McLeish (1975).

The proof first identifies the asymptotic distribution of each term in Equation (43) and the result follow by an application of the Continuous Mapping Theorem as shown in Billingsley (1999). From Equation (16), it is straightforward to show that

$$n^{-\frac{3}{2}}\sigma_x^{-1}\sum_{i=2}^{n}\sum_{j=1}^{i-1}\Delta x_{ij} \xrightarrow{d} B(1) - 2\int_0^1 B(\lambda)d\lambda \tag{D.4}$$

and follow the same argument, it can also be shown that

$$n^{-\frac{3}{2}}\sigma_u^{-1}\sum_{i=2}^{n}\sum_{j=1}^{i-1}\Delta u_{ij} \xrightarrow{d} W(1) - 2\int_0^1 W(\lambda)d\lambda. \tag{D.5}$$

Now, note that $\Delta x_{pq}^2 = [(x_p - \mu_x) - (x_q - \mu_x)]^2$ and

$$\begin{aligned}
\sum_{i=2}^{n}\sum_{j=1}^{i-1}\Delta x_{pq}^2 =& \sigma_x^2 \sum_{i=2}^{n}\sum_{j=1}^{i-1}\left(\frac{x_i - \mu_x}{\sigma_x} - \frac{x_j - \mu_x}{\sigma_x}\right)^2 \\
=& \sigma_x^2\left[\sum_{i=2}^{n}\sum_{j=1}^{i-1}\left(\frac{x_i - \mu_x}{\sigma_x}\right)^2 + \sum_{i=2}^{n}\sum_{j=1}^{i-1}\left(\frac{x_j - \mu_x}{\sigma_x}\right)^2 - \sum_{i=2}^{n}\sum_{j=1}^{i-1}\frac{(x_i - \mu_x)(x_j - \mu_x)}{\sigma_x^2}\right] \\
=& \sigma_x^2 n^2\left(B_n^2(1) - 2\int_0^1 B_n^2(\lambda)d\lambda\right)
\end{aligned}$$

The last line follows from a similar argument as Equations (D.4) and (D.5). Therefore,

$$n^{-2}\sum_{p=2}^{n}\sum_{q=1}^{p-1}\Delta x_{pq}^2 \xrightarrow{d} \sigma_x^2\left(B^2(1) - 2\int_0^1 B^2(\lambda)d\lambda\right). \tag{D.6}$$

For the last term, repeat the argument above and note that $u_p = W_n(\frac{p}{n}) - W_n(\frac{p-1}{n})$

$$\begin{aligned}\sum_{p=2}^{n}\sum_{q=1}^{p-1} \Delta x_{pq} \Delta u_{pq} =& \sum_{p=2}^{n}\sum_{q=1}^{p-1} x_p u_p + x_q u_q - x_p u_q - x_q u_p \\ =& n^2 \sigma_x \sigma_u \left(B_n(1) W_n(1) - \int_0^1 W_n(\lambda) dB_n(\lambda) - \int_0^1 B_n(\lambda) dW_n(\lambda) \right).\end{aligned}$$

Therefore

$$n^{-2} \sum_{p=2}^{n}\sum_{q=1}^{p-1} \Delta x_{pq} u_{pq} \xrightarrow{d} \sigma_x \sigma_u \left(B(1)W(1) - \int_0^1 W(\lambda) dB(\lambda) - \int_0^1 B(\lambda) dW(\lambda) \right). \tag{D.7}$$

Combining Equations (D.4), (D.5), (D.6) and (D.7) and apply the Continuus Mapping Theorem gives the result. This completes the proof. ■

REFERENCES

Billingsley, P. (1999). *Convergence of Probability Measures, Second Edition*. Wiley Inter-Science.

Chan, F., L. Mátyás, and K. Reizinger (2023). Estimation with pairwise observations – online supplement. *Github Repository*. Click here to navigate to the online supplement.

Hausman, J. A. (1978). Specification tests in econometrics. *Econometrica*, 1251–1271.

Johansen, S. (1995). *"Likelihood-based Inference in Cointegrated Vector Auto-regressive Models*. Oxford University Press.

McLeish, D. L. (1975). Invariance Principles for Dependent Variables. *Zeitschrift Fur Wahrscheinlichkeitstheorie Und Verwandte Gebiete 32*(3), 165–178.

Wooldridge, J. M. (2002). *Econometric Analysis of Cross Section and Panel Data*. MIT Press.

Wooldridge, J. M. and H. White (1988). Some invariance principles and central limit theorems for dependent heterogeneous processes. *Econometric Theory*, 210–230. Publisher: JSTOR.