Comparison of the Spierdijk-Wansbeek test and its unweighted counterpart

Bastiaan Boland

13 January 2021

University of Groningen

Bachelor's Thesis Econometrics and Operations Research
Supervisor: T. J. Wansbeek
Second assessor: R. J. M. Alessie

Comparison of the Spierdijk-Wansbeek test and its unweighted counterpart

Bastiaan Boland

Abstract

The new Spierdijk-Wansbeek test promises to be a new entry in the standard toolbox for applied panel data econometrics. Testing the strict exogeneity assumption necessary for the consistency of the widely used within estimator could before only be done when instruments were already available.

In this paper the standard SW test is compared to a modified, so-called unweighted (SWU) variant. The SWU variant is investigated because it should give better results in the case of a singular (or nearly singular) covariance matrix, in contrast with the SW variant. A simulation study is run and shows the unweighted test to be less powerful overall. Also the unweighted test is not reliable when confronted with simultaneity when that relationship is weak. The standard SW test does not have this problem.

Both tests are applied to three research papers. In all three cases the standard SW test rejects the null hypothesis of consistency of the difference estimators, whereas the SWU test does not. Due to the presence of more evidence in favor of the existence of endogeneity we conclude that the unweighted test has not shown applicability in an empirical setting yet. Of special note is the third empirical case, where the covariance matrix is numerically singular and still the SW test (using a generalized inverse) points to endogeneity whereas the SWU test does not. Other evidence from the data also points to endogeneity.

This paper is the first in investigating the properties of the unweighted test statistic, results show the test performs worse than the standard SW test in most cases. More research will be necessary to conclude whether the use of the unweighted test has any benefit over the standard Spierdijk-Wansbeek test.

1 Introduction

The fixed effects estimator, also called the within estimator is one of the standard estimation methods widely taught and used in panel data econometrics. As is the case with all modelling of real-world (economic) phenomena, certain assumptions are made about unobservable aspects of those phenomena. For example due to a lack of ability to perform controlled experiments in economic settings, economists must, many cases, assume an underlying data generation process. These assumptions can then be used to create or apply procedures which are known (mathematically proven) to give results having certain statistical properties. Validity of these results, and the interpretation thereof, crucially depend on the validity of the underlying assumptions. It is therefore important to be able to test whether these assumptions are indeed valid. The new Spierdijk-Wansbeek (SW) test (Spierdijk and Wansbeek (2020)) gives practitioners a new way to test certain assumptions of the fixed effects estimator. This thesis will investigate properties of this test and compare them to situations where a slightly different test might be more useful.

The SW test is of the form of a Wald test, therefore it relies on the inverse of a weight matrix to be available. The weight matrix used is an estimator of a covariance matrix, which depends on the data. In the event of an ill-formed covariance matrix, namely when it is singular or when it is not positive definite, the test statistic might not behave as we would like it to. Such ill-formed covariance matrices are of course a possibility due to the nature of economic data. A modification of the SW test could then be to change the weight matrix into something which does not depend on the data and of which we can be sure that it behaves as we would like. The effect of this change influences the distribution of the test statistic, which turns out to be less simple than the χ^2-distribution of the SW statistic. This changed statistic, which will be called the unweighted variant of the SW test, or SWU test, promises to be a replacement for the SW test in the case of an ill-formed covariance matrix.

We will start with a short review of panel data models, the fixed effects model, and estimation of said model with focus on the within and first-difference estimators. After this we will elaborate on the consequences of violating the assumption of strict exogeneity on the consistency of the fixed effects estimators. We will review some existing ways to test for these violations, and explain where the new SW test fits into the literature. After this we will set out the research goals of this thesis. The distribution of the unweighted test statistic will be given. Next some methodological aspects of the new test are discussed. A simulation study is run comparing power and size of the standard Spierdijk-Wansbeek test and the unweighted variant. Both tests are then applied to three empirical research papers.

2 Literature review

This section will refreshes some basic knowledge about panel data. The fixed effects model and the within estimator are the focus, as these are the subject of the SW test. Existing tests for the assumption of strict exogeneity are touched upon.

2.1 The panel data model

The standard panel data regression model can be written as,

$$\begin{aligned} y_{it} &= x_{it}\beta + u_{it}, \qquad (1) \\ &\text{for } i = 1, \dots, N, \text{ and } t = 1, \dots, T, \end{aligned}$$

with i denoting individuals, countries, firms, etc., and t denoting time. Also where y_{it} is the dependent variable, x_{it} is the independent variable (perhaps including a constant term), β is the vector of coefficients to be estimated, and u_{it} the error term.

Since we are dealing with the same individuals, measured in different points in time, there will exist heterogeneity across those individuals. This heterogeneity is due to individual characteristics, such as variations in skill or preferences. In the case that these individual effects are correlated with the included regressors, i.e. $\mathrm{E}[\alpha_i \mid x_{it}] \neq 0$, pooled OLS will be inconsistent, which is why we need a different way of estimating the model.

2.2 Estimation of the fixed effects model

Since we assumed the unobserved heterogeneity α_i to be time-invariant, at least time-invariant for the duration of the data collection period, we can use several techniques to control for it. We will present the two estimation methods relevant to our case, namely the within estimator and the first difference estimator. Both these techniques work by removing the fixed effect α_i from the model and estimating the transformed model by OLS. This introduces a trade-off, without further assumptions, we cannot include time-constant factors in our independent variable x_{it}. These effects cannot be distinguished from the unobservable effect of α_i (Wooldridge (2010)).

One of the techniques to estimate the fixed effects model is the within-estimator. This estimator which has become the standard way of estimating fixed effects models with individual effects, is usually termed time-demeaning and is defined as (Croissant and Millo (2018a)):

$$(y_{it} - \bar{y}_{i.}) = (x_{it} - \bar{x}_{i.})\beta + (\varepsilon_{it} - \bar{\varepsilon}_{i.}), \qquad (2)$$

where $\bar{y}_{i.}$ and $\bar{x}_{i.}$ denote individual means of y and x. Note that since α_i is constant over time, $\alpha_i - \bar{\alpha} = 0$, meaning the individual effect is eliminated. Applying an OLS regression on (2) gives the fixed effect estimator $\hat{\beta}_{FE}$.

Another way of estimating the fixed effects model is to transform the data by taking first differences, i.e.

$$\begin{aligned} (y_{it} - y_{i,t-1}) &= (x_{it} - x_{i,t-1})\beta + (\alpha_i - \alpha_i) + (\varepsilon_{it} - \varepsilon_{i,t-1}), \qquad (3) \\ \Delta y_{it} &= \Delta x_{i,t}\beta + \Delta\varepsilon_{it}. \qquad (4) \\ &\text{for } i = 1, \dots, N, \text{ and } t = 2, \dots, T, \end{aligned}$$

Here again, due to the time-invariance of the α_i's they are removed after applying the transformation. We will note here that the choice of transformation in first differences over a period of 1 is arbitrary, the fixed effects α_i can be removed by the transformation $y_{it} - y_{i,t-2}$ as well as any other difference over a timespan $\Delta t = 1, \dots, T-1$. This insight will form the basis for the Spierdijk-Wansbeek test which is the subject of this thesis.

2.3 Consistency of fixed effects estimators

The first fixed effects assumption is strict exogeneity of the explanatory variables conditional of α_i:

$$\mathrm{E}[\varepsilon_{it}|\mathbf{x}_i, \alpha_i] = 0,\ t = 1, 2, \ldots, T. \tag{5}$$

When the first fixed effects assumption (5) does not hold, the within and first difference estimators are not consistent in the limit as $N \to \infty$. This assumption is violated for example due to the inclusion of a lagged dependent variable ($x_{i,t-1}$) or a feedback effect (where x_{it} depends on $y_{i,t-1}$). These inclusions are not always avoidable by a researcher. The existence of measurement error can render the within estimator inconsistent (see for example Wansbeek and Meijer (2000), section 6.9).

This possible inconsistency poses a problem for the validity of our inferences in panel data modeling. We will elaborate on tests for this inconsistency next, leading to the subject of this thesis, the Spierdijk-Wansbeek test.

2.4 Testing for consistency

The inconsistency of the estimators shows that the model we are trying to estimate is not the correct one, i.e. there is model misspecification. Several tests have been developed to detect this misspecification. When detected, steps can be undertaken to control for the endogeneity of certain regressors, a common way being the use of instrumental variables. Both the Sargan-Hansen test (Sargan (1958)) and the Durbin-Wu-Hausman test (Hausman (1978)) are used in practice. Both these tests however rely on the a-priori existence of an alternative model. The tests compare for example a model with instrumental variables to a model without those. As Spierdijk and Wansbeek (2020) put it: "Both tests are based on prior suspicions about certain covariates and rely on the availability of instrumental variables. In the absence of prior information or instruments, it is still important to test the validity of the within estimator's exogeneity assumption."

This leads us to the Spierdijk-Wansbeek specification test of Spierdijk and Wansbeek (2020). As mentioned above, both the within estimator (2) and first difference estimator (3) are inconsistent when the strict exogeneity assumption (5) is violated. Also mentioned above is the arbitrary choice of timespan in the first difference estimator. Observe that if the original model in (1) is well-specified, meaning there exists no endogeneity in the regressors, this must mean both the within and first difference estimator are consistent. Also note that all $T-1$ variants of the first difference estimator are consistent, hence will approach the true value of β in the limit. However if the original model contains some form of endogeneity, the $T-1$ variants of the first difference estimator are not consistent and therefore will not agree on the value of β. This disagreement in value is used by Spierdijk and Wansbeek to test for the existence of endogeneity, and hence for the potential misspecification of a model.

The same insight underlying the SW test is also used by others to create tests for the strict exogeneity assumption. The test proposed by Wooldridge (2010) is based on a regression of y_{it} on $x_{it}, x_{i,t+1}$, estimated by the fixed effects estimator,

$$y_{it} = \beta x_{it} + \gamma x_{i,t+1} + \alpha_i + \varepsilon_{it}, \tag{6}$$

where give the null hypothesis of strict exogeneity, $\gamma = 0$ should hold.

In an unpublished report by Mayer (2016), the author, according to whom the idea can be traced back Geweke (1981) (first in the context of time-series), uses the insight to build upon the test proposed by Wooldridge (2010). The test statistic here depends on the amount and size of lags and leads used, denoted by the parameters a and b respectively. The author discusses the effect of different choices of a and b, and notes the inclusion of small a and b is likely to be preferred.

A more comprehensive version of this test was independently developed by Su et al. (2016), using all leads and lags available they form a maximum test statistic. The test selects the largest value from all regressions of the form 6 which can be created using all combinations of leads and lags. A downside of this maximum statistic is that it has no simple probability distribution and its critical values must be obtained by bootstrap. Also the amount of calculations to be performed grows quickly as T grows large. None of these tests depend on the existence of instrumental variables and as such can be applied in a wide range of empirical situations.

3 Problem formulation

First we will explain the workings of the SW test of Spierdijk and Wansbeek (2020) and note why a slightly different test might give better results in certain scenarios. After this the research questions covered in this paper will be given.

3.1 The tests

The null hypothesis to be tested is that all first difference estimators over all timespans are consistent estimators of the true value of the coefficients of our model. The Wald test statistic of the Spierdijk-Wansbeek test corresponding to H_0 is given by

$$q_{\text{sw}} = \hat{\beta}' R \Big(\sum_i R' u_i u_i' R \Big)^{-1} R' \hat{\beta}, \tag{7}$$

where

$$u_i = \begin{pmatrix} (\sum_l X_l' \Delta_1 X_l)^{-1} X_i' \Delta_1 \hat{\varepsilon}_{i1} \\ \vdots \\ (\sum_l X_l' \Delta_{T-1} X_l)^{-1} X_i' \Delta_{T-1} \hat{\varepsilon}_{i,T-1} \end{pmatrix},$$
$$\hat{\varepsilon}_{ij} = y_i - X_i \hat{\beta}_j,$$

where $R = B_1 \otimes I_k$, B_1 being the matrix taking first differences, Δ_t the matrix taking differences over timespan t. The $\hat{\beta}$ estimates are calculated by OLS over the different timespan differences (j), i.e.

$$\hat{\beta}_j = \Big(\sum_i X_i' \Delta_j X_i \Big)^{-1} \sum_i X_i' \Delta_j y_i. \tag{8}$$

Under the null hypothesis of consistency of the first difference estimator, the statistic q_{sw} will be $\chi^2_{(T-2)k}$-distributed.

The test statistic is a Wald test statistic and hence a quadratic form in normal random variables, $a'\widehat{W}^{-1}a$, where $\widehat{W}$ is a consistent estimator of the covariance matrix W of a. In well behaved cases W will be positive definite and since $\widehat{W}$ is consistent it will also be positive definite. Problems could arise when the estimator $\widehat{W}$ has no inverse or the eigenvalues of $\widehat{W}$ are close to zero. The inverse of $\widehat{W}$ could be volatile in small samples. To circumvent these potential problems, we can replace the weighting matrix $\widehat{W}$ with the identity matrix $I_{(T-2)k}$.

$$q_{\text{swu}} = \hat{\beta}' R R' \hat{\beta} \tag{9}$$

The construction of this new, unweighted test statistic removes the inversion problem, but this change adds a different problem, the distribution of q_{swu} is no longer $\chi^2_{(T-2)k}$.

Since $\hat{\beta}$ is an OLS estimator is it normally distributed around its true value. The Wald test statistic is now a of the form $a'a$ where $a = R\hat{\beta}$. In the next section we will derive the distribution of the unweighted test statistic.

3.2 Research questions

The main questions this thesis will try to answer are,

1. How do the Spierdijk-Wansbeek and its unweighted counterpart tests work?
2. How do the SW and SWU tests compare in size and power in simulations?
3. How do the SW and SWU tests compare in an empirical setting?

The first question is of methodological nature and will be explained in the text. We would expect the SWU to perform worse as the choosing the inverse of the covariance matrix as the weighting matrix is optimal according to Wansbeek and Meijer (2000) (p. 234 and p. 241).

4 Distribution of test statistic

In the following sections we will use that $\hat{\beta} \sim \mathcal{N}(\beta_0, \Sigma)$, since $\hat{\beta}$ is the OLS estimator with true value β_0, and covariance matrix Σ. We will write $\hat{\Sigma}$ for a consistent estimator of Σ.

The distribution of the test statistic of the SWU test in Equation (9) is a generalized χ^2-distribution. This distribution is parameterized by λ, the vector of weights of the χ^2-components; m the vector of degrees of freedom of those χ^2-components; δ, the vector of non-centrality parameters of the components; and σ, the scale of the normal term.

For the SWU statistic the weights vector λ the eigenvalues of the matrix $\hat{\Sigma}_R = R'\hat{\Sigma}R$, where $\hat{\Sigma}$ is the estimated covariance matrix of $\hat{\beta}$. All degrees of freedom in parameter m are 1, and the non-centrality parameters δ are all 0. In this case the generalized χ^2-distribution is defined as the weighted sum of independent central χ^2_1 variables. We will now derive the parameters of this distribution.

Recall that the SWU test statistic is a quadratic form,

$$q_{\text{swu}} = \hat{\beta}' R R' \hat{\beta} = x' A x, \tag{10}$$

with $x = R'\hat{\beta} \sim \mathcal{N}(0, \hat{\Sigma}_R)$, R the restrictions matrix, and A the identity matrix. We will show $\mathrm{E}[R'\hat{\beta}] = 0$ in section 4.1. Following the derivation in Appendix B of Wansbeek and Meijer (2000), we start with the eigenvalue decomposition of $\Sigma_R = K\Delta K'$. So Δ is the diagonal matrix with the eigenvalues of Σ_R on its diagonal entries, and K is the matrix with as columns the eigenvectors of Σ_R. Since Σ_R is a covariance matrix it is symmetric and hence K is an orthonormal matrix, i.e. $K^{-1} = K'$. Define

$$\Upsilon \equiv \Delta^{\frac{1}{2}} K' A K \Delta^{\frac{1}{2}}, \tag{11}$$

which in the case of $A = I$ simplifies to,

$$\Upsilon = \Delta^{\frac{1}{2}} K' K \Delta^{\frac{1}{2}} = \Delta. \tag{12}$$

This matrix Δ is again decomposed as Λ, but since it is already a diagonal matrix, it equals its eigendecomposition. Defining then $P \equiv K\Delta^{-\frac{1}{2}}$, gives

$$P'\Sigma_R P = \Delta^{-\frac{1}{2}} K' K \Delta K' K \Delta^{-\frac{1}{2}} = I. \tag{13}$$

Let $z \equiv P'x$, then Equation (13) implies $z \sim \mathcal{N}(P'\mu, I)$. Using that $\mu = \mathrm{E}[R'\hat{\beta}] = 0$ yields

$$x'Ax = z'\Lambda z = \sum_{j=1}^{(T-2)k} \lambda_j z_j^2, \tag{14}$$

where $\Lambda = \Delta$ is the diagonal eigenvalue matrix of Σ_R, the λ_j's are the eigenvalues of Σ_R, and each z_j is a standard normal distributed random variable. Since in the case of a multivariate normal distribution, zero correlation implies independence, these z_j's are independent. Now it is easily seen that the distribution of q_{swu} in Equation (10) is a weighted sum of central χ^2-distributed random variables.

As the variance matrix Σ (and hence Σ_R) is unknown, we estimate it using White's heteroskedasticity-consistent estimator (White (1980)), which is given by

$$\hat{\Sigma} = \frac{1}{n}\Big(\frac{1}{n}\sum_{i=1}^{n} X_i'X_i\Big)^{-1}\Big(\frac{1}{n}\sum_{i=1}^{n} X_i'u_iu_i'X_i\Big)\Big(\frac{1}{n}\sum_{i=1}^{n} X_i'X_i\Big)^{-1}, \tag{15}$$

where $u_i = y_i - X_i\hat{\beta}$ are the residuals.

4.1 Mean of $R'\hat{\beta}$

We will now show $\mathrm{E}[R'\hat{\beta}] = 0$. Let $B_1 : (T-1) \times (T-2)$ be the matrix taking first differences. Let $R = B_1 \otimes I_k$, $R : (T-1)k \times (T-2)k$. Then

$$R = \begin{pmatrix} -1 & 0 & 0 & \dots \\ 1 & -1 & 0 & \\ 0 & 1 & -1 & \\ \vdots & & & \ddots \end{pmatrix} \otimes I_k = \begin{pmatrix} -I_k & 0 & 0 & \dots \\ I_k & -I_k & 0 & \\ 0 & I_k & -I_k & \\ \vdots & & & \ddots \end{pmatrix}.$$

Under the null hypothesis $H_0 : \beta_j = \beta_{j+1},\ j = 1, \ldots, T-1$, $\mathrm{E}[R'\hat{\beta}]$ then becomes,

$$\mathrm{E}[R'\hat{\beta}] = \begin{pmatrix} -I_k & I_k & 0 & \ldots \\ 0 & -I_k & I_k & \\ 0 & 0 & -I_k & \\ \vdots & & & \ddots \end{pmatrix} \mathrm{E} \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \vdots \\ \hat{\beta}_{T-1} \end{pmatrix} = \mathrm{E} \begin{pmatrix} \hat{\beta}_2 - \hat{\beta}_1 \\ \hat{\beta}_3 - \hat{\beta}_2 \\ \vdots \\ \hat{\beta}_{T-1} - \hat{\beta}_{T-2} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix},$$

so the expectation $\mathrm{E}[R'\hat{\beta}]$ under H_0 is the zero vector with dimensions $(T-2)k \times 1$.

5 Methods

This thesis investigates the differences between two similar tests, the Spierdijk-Wansbeek or SW test, and its unweighted counterpart or SWU test. The sequel will touch on some computational aspects of both tests.

5.1 Spierdijk-Wansbeek test statistic

To calculate the Spierdijk-Wansbeek test statistic we use the method outlined in the Appendix of Spierdijk and Wansbeek (2020). We note that the calculation as outlined can be approached in two ways, as given in the paper section 2.3, or as given in the Appendix section C. After implementing both approaches, we found the one in the Appendix to have a shorter run-time. We expect this to be due to taking first differences before starting to calculate $\hat{\beta}$. This aggregates values into matrix form which are faster to manipulate than separate vectors. Also there is a difference in the amount of times certain parts of the data matrix X are accessed, in favor of the second approach, although this may be due to a naive implementation. The resulting test statistics and estimated covariance matrices of both implementations were of course identical.

5.2 Unweighted test statistic

The unweighted test statistic can be readily calculated when the Spierdijk-Wansbeek statistic is also calculated.

The generalized χ^2-distribution of the SWU statistic has no simple closed form solution for its probability density function or cumulative distribution function. Therefore we rely on an implementation to calculate the cdf in the `CompQuadForm` `R` package by Duchesne and de Micheaux (2010). The algorithm used is by Imhof (1961), as this is an exact method in the sense that it is possible to bound the approximation error, which can be made arbitrarily small (Duchesne and de Micheaux (2010)). The Imhof method computes the survival function $\Pr[Q > q]$, where Q is a quadratic form in normal random variables. The parameters used are the point q at which to evaluate the survival function, and the eigenvalues of $\hat{\Sigma}_R = R'\hat{\Sigma}R$. A simulation of generalized χ^2 random variables showed excellent agreement with the Imhof method.

The fact that only the survival function is readily available to us has the effect that, when performing the hypothesis test, p-values are easily obtained, but an α-level bound on the test statistic can only be found by a root-finding algorithm. The use of this root-finding algorithm introduces inaccuracies and complications in which parameters to use, but we will not concern ourselves with that.

6 Simulation

To investigate the differences between the Spierdijk-Wansbeek test and the unweighted test a simulation study was performed. Two important measures for the behaviour of a statistical test are power and size. The (finite-sample) power of a test is defined as the probability of rejecting the null hypothesis when the alternative hypothesis is true, i.e. Pr[reject $H_0|H_1$ true]. Power can be described as one minus the probability of making a type II error. The size of a test is defined as the probability of rejecting the null hypothesis when the null hypothesis is in actuality true, i.e. Pr[reject $H_0|H_0$ true]. Size can be described as the probability of making a type I error. (Hayashi (2000))

Three common forms of endogeneity are investigated, namely classical measurement error, omitted variable bias, and simultaneity. For each of these forms, a Monte Carlo simulation is performed. We will first specify the three data generating processes used in the Monte Carlo simulation, and specify the coefficients used. Methods for simulating the different forms of endogeneity were taken from the Appendix, section B, of Spierdijk and Wansbeek (2020). After this we will discuss some generalities about the simulations, and the parameters used will be given. This section will end with a discussion of the results.

6.1 Measurement error

The first form of endogeneity we will simulate is measurement error.

If the "true" model is $y_{it} = \alpha_i + \beta x_{it} + \nu_{it}$ and what is observed is $x^*_{it} = x_{it} + \nu_{it}$, the estimated model writes: $y_{it} = \alpha_i + \beta(x^*_{it} - \eta_{it}) + \nu_{it}$, or $y_{it} = \alpha_i + \beta x^*_{it} + \varepsilon_{it}$ with $\varepsilon_{it} = \nu_{it} - \beta\eta_{it}$. Hence, ε_{it} is correlated with x^*_{it}, which is therefore endogenous. (Croissant and Millo (2018a))

To simulate measurement error we generate data according to the model given by

$$
\begin{aligned}
y_{it} &= \alpha_i + \beta\xi_{it} + \varepsilon_{it}, \\
x_{it} &= \xi_{it} + \nu_{it}, \\
\xi_{it} &= \rho\xi_{i,t-1} + \theta_{it}, \\
\nu_{it} &= \delta\nu_{i,t-1} + \eta_{it},
\end{aligned}
$$

where $i = 1, \ldots, n$, $t = 1, \ldots, T$, and ξ_{it} and ν_{it} are AR(1) processes.

In this form the data will exhibit measurement error, and can thus be used for power simulations. To simulate size we need the null hypothesis to be true, and so set $\sigma^2_\eta = 0$, so as to remove those shocks from the model.

6.2 Omitted variable bias

The second form of endogeneity we will simulate is omitted variable bias.

If the "true" model is $y_{it} = \alpha_i + \beta x_{it} + \gamma z_{it} + \nu_{it}$ and z_{it} is unobserved, the estimated model is $y_{it} = \alpha_i + \beta x_{it} + \varepsilon_{it}$, with $\varepsilon_{it} = \beta z_{it} + \nu_{it}$. The error of the estimated model then contains the influence of the omitted variable, and this error is correlated with x_{it} if x_{it} and z_{it} are correlated. Once again, the covariate x_{it} is then endogenous. (Croissant and Millo (2018a))

To simulate omitted variables we generate data according to the model given by,

$$
\begin{aligned}
y_{it} &= \alpha_i + \beta x_{it} + \gamma z_{it} + \varepsilon_{it},\\
x_{it} &= \rho x_{i,t-1} + \theta_{it},\\
z_{it} &= \delta z_{i,t-1} + \eta_{it},
\end{aligned}
$$

where $i = 1, \ldots, n$, $t = 1, \ldots, T$, and x_{it} and z_{it} are AR(1) processes.

We then estimate the model $y_{it} = \alpha_i + \beta x_{it} + \varepsilon_{it}$, which suffers from omitted variables. To remove the omitted variables and simulate for size, we set $\gamma = 0$, thereby removing the omitted variable from the model.

6.3 Simultaneity

The last form of endogeneity we simulate is simultaneity, this form occurs when variables at either side of a model equation influence each other at the same time.

We simulate the following

$$
\begin{aligned}
y_{it} &= \beta_i + \beta x_{it} + \varepsilon_{it},\\
x_{it} &= \alpha_i + \alpha y_{it} + u_{it},\\
u_{it} &= \rho u_{i,t-1} + \theta_{it},
\end{aligned}
$$

where $i = 1, \ldots, n$, $t = 1, \ldots, T$, and u_{it} is an AR(1) process.

After calculating the values in all timeperiods, we then regress y on x. To simulate size we set $\alpha = 0$, so there is no simultaneity anymore.

6.4 General simulation considerations

Empirical power was calculated as the proportion of simulations in which the test rejects the null hypothesis, while it was known to be false. Empirical size was calculated as the proportion of simulations in which the test rejects the null hypothesis, while it was known to be true. All tests were performed at α-level 0.05.

All simulations were performed in R (R Core Team (2020)). To speed up data generation and calculations we took advantage of parallel computing via the `doParallel` package Microsoft Corporation and Weston (2020). A small scale experiment showed a decrease in simulation run-time of a factor slightly less than 2, using 3 computing cores. Using three out of four CPU cores allowed for running small calculations next to the running simulation. It is important to note that by default (in R) each core uses the same random number stream, which can lead to duplicate random numbers among the different cores, and therefore can skew results in unexpected ways. To remedy this use L'Ecuyer Combined Multiple Recursive random number generators (L'Ecuyer (1999)), part of the `parallel` package (part of base R), which ensure independent random number streams for each computing core.

6.5 Simulation parameters

All simulations were repeated 500 times. Data was generated with number of individuals $n \in \{100, 500, 1000\}$ and timespan $T \in \{5, 10\}$. In the case of measurement error and omitted variable bias the coefficient δ of one of the AR(1) processes was kept fixed at $\delta = 0.3$, whereas for all forms the coefficient ρ was either $\rho = 0.6$ or $\rho = 0.9$. The parameter ρ, though used differently in the different cases, can be interpreted as a measure of how strong the effects of endogeneity are expected to be.

Measurement error

The AR(1) processes ξ_{it} and ν_{it} have coefficients $\rho \in \{0.6, 0.9\}$ and $\delta = 0.3$ respectively, and their shocks have zero mean and variances $\sigma_\theta^2 = 1.44$ and $\sigma_\eta^2 = 0.64$. The error term ε has mean zero and variance $\sigma_\varepsilon^2 = 1$, the coefficient $\beta = 1$. This implies an R^2 of 0.892 ($\rho = 0.9$) and 0.747 ($\rho = 0.6$). Also the reliability (ratio of variance of the true value and the observed value of the regressor) is 0.915 ($\rho = 0.9$) and 0.762 ($\rho = 0.6$). The noise-to-signal ratio (ratio of variance of true value and the variance of the error) is 0.09 ($\rho = 0.09$) and 0.312 ($\rho = 0.6$).

Omitted variable bias

The AR(1) processes x_{it} and z_{it} have coefficients $\rho \in \{0.6, 0.9\}$ and $\delta = 0.3$ respectively, and their shocks have zero mean and variances $\sigma_\theta^2 = 0.36$ and $\sigma_\eta^2 = 0.36$, the covariance between θ and η is $\sigma_{\theta\eta} = -0.216$. The error term ε has mean zero and variance $\sigma_\varepsilon^2 = 0.25$, the coefficient $\beta = 1$, and $\gamma = 1$. This implies an R^2 of 0.912 ($\rho = 0.9$) and 0.654 ($\rho = 0.6$).

Simultaneity

The AR(1) process u_{it} has coefficient $\rho \in \{0.6, 0.9\}$, the shocks have zero mean and variance $\sigma_\theta^2 = 1$. The error term ε has mean zero and variance $\sigma_\varepsilon^2 = 4$, the coefficient $\beta = 1$, and $\alpha = 2$. This implies an R^2 of 0.467 ($\rho = 0.9$) and 0.207 ($\rho = 0.6$).

6.6 Results

The results of the simulations for power and size for both tests can be found in Tables 1 and 2 respectively.

The most striking of results is found Table 1, where the power of both tests in the presence of measurement error is incredibly weak. One option is that both tests are not able to pick up on the effects of measurement error, another is that the method of simulation is not correct. Even though great care was taken in implementing the data generation for this kind of endogeneity, it seems the process was flawed.

We can see in Table 1 that the unweighted test is less powerful in all cases. Furthermore, while the SW test gets more powerful as n or T increases, for all kinds of simultaneity, the unweighted test seems to be more volatile. Especially when ρ, the coefficient of the AR(1) process describing the error term, is small the power of the unweighted test does not necessarily go up as n or T increases. It is expected for power to reduce as ρ decreases, though it is striking that the amount by which power reduces is this large.

In Table 2 the results obtained from the simulation pertaining to the size of both tests are shown. In the ideal scenario we would expect to see a test size of around 0.05, the used α-level for both tests. In most parameter combinations this is the case. We see size improves for the SW test as n increases, in most cases. Again, simultaneity is a form of endogeneity where the results are not what we would expect beforehand. Size can increase for both tests as n increases.

7 Empirical application

Three cases are presented where both the SW test and the SWU test are applied. Two of these cases use widely known datasets from econometrics literature. These datasets have been thoroughly tested and as such no hidden idiosyncrasies are expected to cause problems in the application of the tests. The other was used by Su et al. (2016) as an empirical application of their test. It can serve as a short comparison between the effectiveness of the tests in a practical setting.

7.1 Application to a wage equation

In this section, we apply both tests to a well known panel data set. Cornwell and Rupert (1988) model a wage equation and apply different methods to estimate it. They use a panel set drawn from the Panel Study of Income Dynamics (PSID) years 1976-1982, containing 595 heads of households between the ages of 18 and 65 (in 1976). We follow Baltagi (2005) in our replication, with whom our results align. Data is collected on wages, years of education (ed), weeks worked (wks), years of full-time work experience (exp), an dummy equaling one if the person is in a blue collar occupation ($bluecol$), a dummy equaling one if the person lives in the south ($south$), a dummy equaling one if the person lives in a standard metropolitan statistical area ($smsa$), a dummy equaling one if the person works in the manufacturing industry (ind), a dummy equaling one if the person is married ($married$), a dummy equaling one if the person is in a union ($union$), a dummy equaling one if the person is female ($female$), and a dummy equaling one if the person is black ($black$).

The logarithm of $wage$ is regressed on all the above variables.

Table 3 shows results of various estimation methods used. A Hausman test between the random effects GLS estimates and the fixed effects within estimates gives a χ^2_9-statistic value of 5075.3, which is highly significant. This leads us to prefer the fixed effects model for this problem. It is now worthwhile to run the SW tests to check whether the within estimator is indeed usable.

The data is well behaved enough that it is possible to calculate the regular inverse of the estimated covariance matrix. This means we expect no clear advantage in using the unweighted test. The Spierdijk-Wansbeek test statistic has a value of $\chi^2_{54} = 64.37$ (critical value 61.66) with a p-value of 0.031, and hence rejects the null hypothesis of identical difference estimators. The unweighted variant of the test gives a statistic with a value of 0.0122 (critical value 0.0380) with a p-value of 0.809, and hence does not reject the null hypothesis. Interestingly the tests do not agree.

Table 1: Simulation empirical power

	SW test			Unweighted test		
	$n = 100$	$n = 500$	$n = 1000$	$n = 100$	$n = 500$	$n = 1000$
		$T = 5$			$T = 5$	
$\rho = 0.9$						
ME	0.584	1.000	1.000	0.428	0.994	1.000
OVB	0.490	0.998	1.000	0.402	0.998	1.000
SI	0.800	1.000	1.000	0.334	0.840	0.904
$\rho = 0.6$						
ME	0.180	0.610	0.904	0.102	0.364	0.732
OVB	0.164	0.674	0.922	0.108	0.508	0.862
SI	0.496	0.992	1.000	0.186	0.306	0.186
		$T = 10$			$T = 10$	
$\rho = 0.9$						
ME	0.994	1.000	1.000	0.842	1.000	1.000
OVB	0.932	1.000	1.000	0.828	1.000	0.994
SI	1.000	1.000	1.000	0.328	1.000	1.000
$\rho = 0.6$						
ME	0.506	1.000	1.000	0.102	0.620	0.988
OVB	0.915	0.974	1.000	0.425	0.616	0.960
SI	0.944	1.000	1.000	0.093	0.360	0.173

For parameters see text.

Table 2: Simulation empirical size

	SW test			Unweighted test		
	$n = 100$	$n = 500$	$n = 1000$	$n = 100$	$n = 500$	$n = 1000$
		$T = 5$			$T = 5$	
$\rho = 0.9$						
ME	0.062	0.066	0.058	0.046	0.066	0.056
OVB	0.044	0.050	0.038	0.028	0.044	0.032
SI	0.032	0.064	0.080	0.024	0.072	0.080
$\rho = 0.6$						
ME	0.068	0.060	0.058	0.064	0.040	0.060
OVB	0.070	0.040	0.066	0.070	0.038	0.054
SI	0.064	0.072	0.088	0.048	0.072	0.080
		$T = 10$			$T = 10$	
$\rho = 0.9$						
ME	0.112	0.046	0.058	0.060	0.056	0.062
OVB	0.112	0.056	0.052	0.040	0.056	0.042
SI	0.120	0.080	0.040	0.072	0.064	0.072
$\rho = 0.6$						
ME	0.094	0.082	0.072	0.056	0.074	0.082
OVB	0.120	0.054	0.076	0.054	0.044	0.048
SI	0.064	0.072	0.064	0.032	0.072	0.072

For parameters see text.

The difference curves shown in Figure 1 seem to show definite patterns in the value of the difference estimator over different timespans.

The data collected includes variables which can be used as instruments. The authors estimate the model using the method of Hausman and Taylor (1981). The instruments used are $X_1 = (bluecol, south, smsa, ind)$, $X_2 = (exp, exp^2, wks, married, union)$, $Z_1 = (female, black$, $Z_2 = (ed)$. The third column of Table 3 shows the results of this estimation. Again running a Hausman test between the HT estimator and the within estimator, gives a test statistic value of $\chi^2_9 = 5.26$, which is not significant. This has the implication that the instruments chosen seem to be valid.

The reason for creating the unweighted test was that the covariance matrix might be non-singular in practice. This is not the case for this dataset. The 54 eigenvalues of the covariance matrix are small, ranging from $5 \cdot 10^{-2}$ to $4 \cdot 10^{-13}$, with 39 greater than 10^{-6}. Taking the inverse of such small numbers is well within the numerical accuracy of modern computers.

7.2 Application to an agricultural production function

Su et al. (2016) create a test similar to the Spierdijk-Wansbeek test. It tests the exogeneity assumption necessary for the fixed effects model using transformed linear panel regressions. They apply this test to a panel dataset from the Food and Agriculture Organization of the United Nations. The full dataset is not available but can be recreated, although it is reduced in size compared to Su et al.. We investigate 37 countries over 6 years from 2002 to 2007. The model to estimate is given by,

$$\log(output)_{it} = \beta_1 \log(area)_{it} + \beta_2 \log(pop)_{it} + \beta_3 \log(govexp)_{it} + \alpha_i + u_{it}, \qquad (16)$$

where $i = 1, \ldots, 37$, $t = 2002, \ldots, 2007$. The dependent variable $\log(output)_{it}$ is the logarithm of net production value for country i in year t. The independent variable $\log(area)_{it}$ denotes agricultural area; $\log(pop)_{it}$ denotes population working in agriculture; $\log(govexp)_{it}$ denotes government expenditure; α_i denotes country specific attributes; and u_{it} denotes the error term. Government expenditure is used as a proxy for investments in the agricultural industry. Table 4 shows the estimation results using different models. We can see large variation in the estimated value of coefficients among the different methods. Su et al. (2016) note the large difference is estimates between the within estimator and the first difference estimator, and say this is a this indicates a possible breakdown of the strict exogeneity assumption. A Hausman test is performed between the random effects and fixed effects model. The value of the test statistic is 49.47 with a p-value of 10^{-10}.

Again, the data permits us to calculate the regular inverse of the estimated covariance matrix. The SW test statistic has a value of $\chi^2_{12} = 89.23$ (critical value 21.03), with a p-value less than 10^{-13}. The SWU test statistic has a value of 0.793 (critical value 1.976), with a p-value of 0.272. Again we see that the SW test rejects the null hypothesis and the unweighted test does not. In Figure 2 we can clearly see a pattern in the estimated β's. The patterns in the difference curves give more evidence against the consistency of the within estimator. Adding to this is the fact that the test by Su et al. (2016) also points to a violation of the strict exogeneity assumption. With all this evidence against exogeneity, it is surprising that the unweighted test does not detect it.

The 15 eigenvalues of the estimated covariance matrix $\hat{\Sigma}$ were comparatively large, all were greater than 10^{-5}, the first two were had values of 5.60 and 0.127.

7.3 Application to investment data

One of the most widely used data sets in all of econometrics (Kleiber and Zeileis (2010)), the Grunfeld data, will be investigated next. This panel data set contains investment data of 11 large US firms, taken over the years 1935-1954. We use the subset provided by Croissant and Millo (2018b), containing only 10 of the firms.

Kleiber and Zeileis (2008) show the modeling done by Grunfeld (1958). We want to estimate the basic model

$$investment_{it} = \beta_1 value_{it} + \beta_2 capital_{it} \alpha_i + u_{it}, \tag{17}$$

where $i = 1, \ldots, 10$, $t = 1935, \ldots, 1954$. The dependent variable $investment_{it}$ denotes gross investment. The independent variables are $value_{it}$ denoting market value of the firm, and $capital_{it}$ denoting the stock of plant and equipment.

Estimation results of the model can be found in Table 5. An F-test was done to compare pooled OLS and fixed effects, this resulted in a statistic value of 49.177 which is far greater than the critical value of 1.929, indicating the existence of interfirm heterogeneity. Estimating a random effects model gives results close to the results of the fixed effects model, performing a Hausman test gives a p-value of 0.3119, both indicating that there exists no problematic endogeneity.

Applying the SW test gives a test statistic value of $\chi^2_{36} = 533.4$ (critical value 50.99), and a p-value within the computational tolerance of zero. The SWU test statistic has value of 0.0224 (critical value 0.0997), and p-value of 0.512. The difference curves for this application are shown in Figure 3, these definitely show less of a pattern than the previous empirical applications, though it is hard to not be suspicious of a pattern. The difference curves do no give evidence for or against the existence of endogeneity.

The 38 eigenvalues of the estimated covariance matrix $\hat{\Sigma}$ could be problematic for the SW test. Only 8 are larger than 10^{-6}, the other 30 are within 10^{-16} of zero, both positive and negative. This results in R not being able to compute the inverse of $\hat{\Sigma}$ needed for the SW test, as the numerical precision of R lies around 10^{-16}. It even hints at a situation where $\hat{\Sigma}$ is not positive semi-definite. This situation should be the one where the SWU test does better than the SW test.

7.4 Discussion of empirical results

In all three empirical cases the standard Spierdijk-Wansbeek test rejects the null hypothesis of identical difference estimators and hence gives evidence for a violation of the strict exogeneity assumption. Also in both cases, despite well-founded suspicions about endogeneity problems, the unweighted test does not reject the null hypothesis. The reasons for this discrepancy are as of yet unknown. From the simulation results we would expect both tests to behave quite similar. The power of the SW test was lower than that of the SWU test overall, but in most cases the difference was not overly large. Especially in the third application on the data of Grunfeld (1958) we would expect the SWU test to agree with the SW test. There is an opportunity for further research into what conditions affect the unweighted test, and what causes this discrepancy in acceptance of the null hypothesis.

8 Conclusion

The new Spierdijk-Wansbeek test (SW test) for strict exogeneity of regressors in panel data models accounting for unobserved individual heterogeneity gives researchers more options when faced with possible endogeneity. Especially in the case when no prior information or instruments are available. The authors advice researchers to routinely run the SW test on their data, and discard the within estimator for further inference, if it fails the test. But due to the structure of the test, practical application might prove difficult. We investigated an alternate formulation of the test in which the test statistic was not weighted by the inverse of the covariance matrix (SWU test), but by the identity matrix. The distribution of this unweighted statistic was found to be a generalized χ^2-distribution.

A simulation study was run to compare power and size of the two tests, given the (non-)existence of several kinds of endogeneity. The simulation results showed the SWU test to be less powerful but it had a lower size overall. The test has a higher probability of giving a false negative, but a lower probability of giving a false positive. In the presence of simultaneity due to joint determination of independent and dependent variables the SWU test did not perform well when the impact of lagged effects was small. The SW test did not suffer from this problem.

Both tests were applied to panel data models in two research papers. In both cases the SW test found evidence for the existence of endogeneity, while the SWU test did not. In both papers there was more evidence against the assumption of strict exogeneity than just the SW test. The assumption was therefore deemed invalid.

This paper is the first in investigating the properties of the SWU test, the results point in the direction that this test performs worse than the standard SW test in most cases. More research is necessary to definitively conclude whether the application of the SWU test has any benefit over the standard SW test.

It would be interesting to know what covariance structures imply problems for the SW test. This paper simulated data and calculated the covariance from that data, further research could focus on the reverse, by simulating a covariance matrix from for example the Wishart distribution and looking into what factors make the SWU test perform better. Another possibility would be to investigate the different tests from the literature which use the same underlying principle as the SW test, namely those by Su et al. (2016), and by Mayer (2016). These three tests differ in the choice of what leads and lags to include in the calculation of the coefficients used in the test statistic. It would be interesting to know what insights may be gained by comparing the three.

References

Baltagi, B. H. (2005). *Econometric Analysis of Panel Data* (3rd ed.). Wiley.

Cornwell, C. and P. Rupert (1988). Efficient estimation with panel data: An empirical comparison of instrumental variables estimators. *Journal of Applied Econometrics 3*(2), 149–155.

Croissant, Y. and G. Millo (2018a). *Panel data econometrics with R* (1st ed.). John Wiley and Sons Ltd.

Croissant, Y. and G. Millo (2018b). *Panel Data Econometrics with R: the plm package*. Wiley and Sons Ltd.

Duchesne, P. and P. Lafaye de Micheaux (2010). Computing the distribution of quadratic forms: Further comparisons between the liu-tang-zhang approximation and exact methods. *Computational Statistics and Data Analysis 54*, 858–862.

Geweke, J. (1981). A comparison of tests of the independence of two covariance-stationary time series. *Journal of the American Statistical Association 76*(374), 363–373.

Grunfeld, Y. (1958). *The Determinants of Corporate Investement*. Ph.D. Thesis, Departement of Economics, University of Chicago.

Hausman, J. A. (1978). Specification tests in econometrics. *Econometrica 46*(6), 1251–1271.

Hausman, J. A. and W. E. Taylor (1981). Panel data and unobservable individual effects. *Econometrica 49*(6), 1377–1398.

Hayashi, F. (2000). *Econometrics*. Princeton University Press.

Imhof, J. P. (1961). Computing the distribution of quadratic forms in normal variables. *Biometrika 48*(3), 419–426.

Kleiber, C. and A. Zeileis (2008). *Applied Econometrics with R*. Use R! Springer-Verlag.

Kleiber, Christian and Achim Zeileis (2010). The grunfeld data at 50. *German Economic Review 11*(4), 404–417.

L'Ecuyer, P. (1999). Good parameters and implementations for combined multiple recursive random number generators. *Operations Research 47*(1), 159–164.

Mayer, A. (2016). *On the Local Power of Some Tests of Strict Exogeneity in Linear Fixed Effects Models*. Number ID 2769747. Unpublished SSRN.

Microsoft Corporation and S. Weston (2020). *doParallel: Foreach Parallel Adaptor for the 'parallel' Package*. R package version 1.0.16.

R Core Team (2020). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.

Sargan, J. D. (1958). The estimation of economic relationships using instrumental variables. *Econometrica 26*(3), 393–415.

Spierdijk, L. and T. Wansbeek (2020). *Differencing as a Consistency Test for the Within Estimator*. Working paper, University of Groningen.

Su, L., Y. Zhang, and J. Wei (2016). A practical test for strict exogeneity in linear panel data models with fixed effects. *Economics Letters 147*, 27–31.

Wansbeek, T. and E. Meijer (2000). *Measurement error and latent variables in econometrics* (1st ed.). Advanced textbooks in economics, 0169-5568 ; vol. 37. Amsterdam: Elsevier.

White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica 48*(4), 817–838.

Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data* (2nd ed.). MIT Press.

A Appendix

Table 3: Estimation results for the wage equation

	lwage		
	RE	Within	HT
exp	0.082***	0.113***	0.113***
	(0.003)	(0.002)	(0.002)
exp^2	−0.001***	−0.0004***	−0.0004***
	(0.0001)	(0.0001)	(0.0001)
wks	0.001	0.001	0.001
	(0.001)	(0.001)	(0.001)
bluecol	−0.050***	−0.021	−0.021
	(0.017)	(0.014)	(0.014)
ind	0.004	0.019	0.014
	(0.017)	(0.015)	(0.015)
south	−0.017	−0.002	0.007
	(0.027)	(0.034)	(0.032)
smsa	−0.014	−0.042**	−0.042**
	(0.020)	(0.019)	(0.019)
married	−0.075***	−0.030	−0.030
	(0.023)	(0.019)	(0.019)
union	0.063***	0.033**	0.033**
	(0.017)	(0.015)	(0.015)
female	−0.339***		−0.131
	(0.051)	–	(0.127)
black	−0.210***		−0.286*
	(0.058)	–	(0.156)
ed	0.100***		0.138***
	(0.006)	–	(0.021)
Constant	4.264***		2.913***
	(0.098)	–	(0.284)
Observations	4,165	4,165	4,165
R^2	0.390	0.658	0.609

Note: *p<0.1; **p<0.05; ***p<0.01

Figure 1: Difference curves for the wage equation. Error bars denote 95% confidence interval.

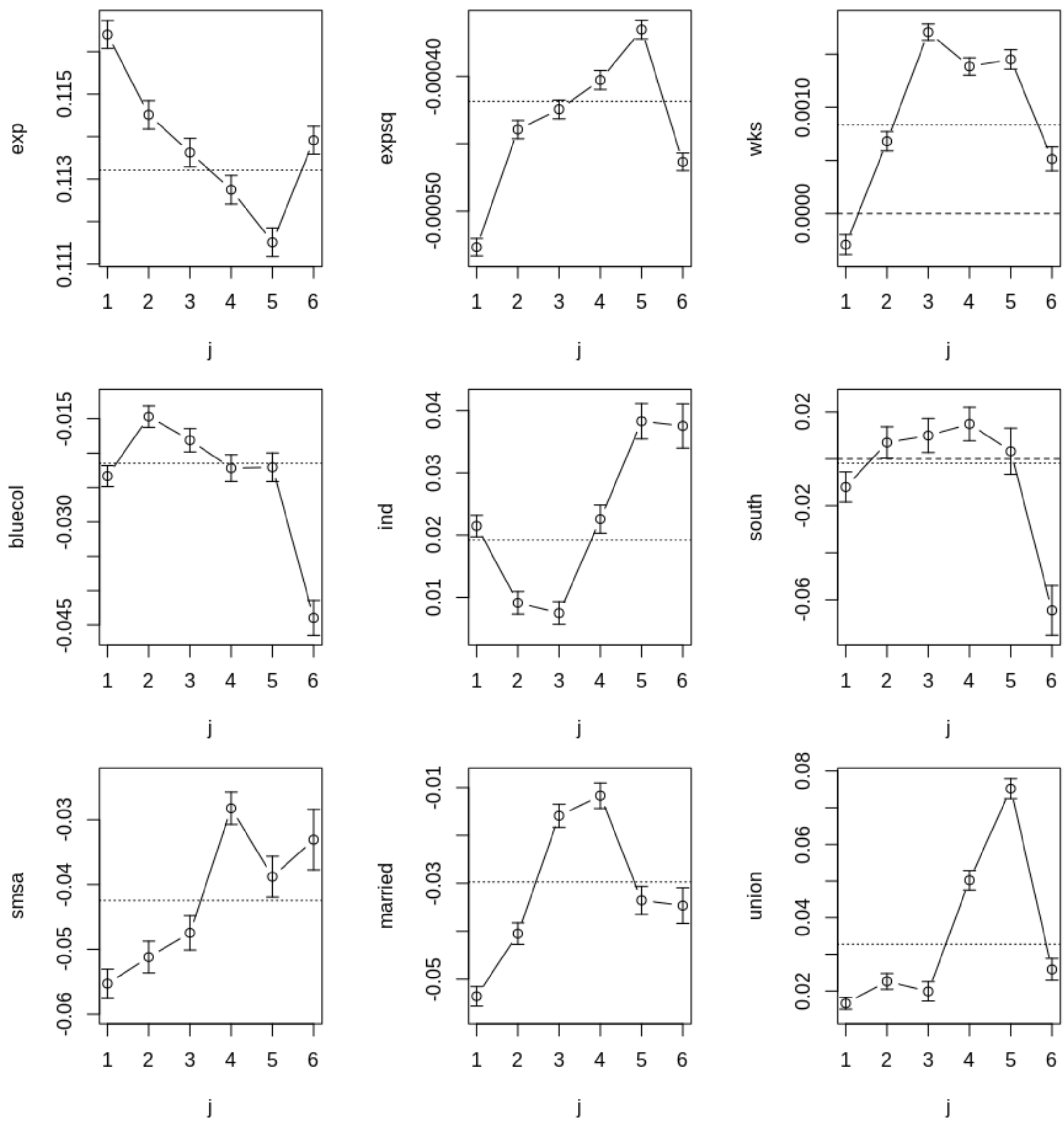

Table 4: Estimation results for the agricultural production function

	loutput			
	Pooled OLS	RE	Within	First difference
lagrarea	0.179***	0.393***	−0.504**	−0.100
	(0.024)	(0.062)	(0.222)	(0.254)
lpop	0.255***	−0.052	−0.272***	0.062
	(0.025)	(0.055)	(0.067)	(0.070)
lgovt	0.525***	0.428***	0.384***	0.126***
	(0.023)	(0.027)	(0.027)	(0.031)
Constant	1.912***	2.425***		0.050***
	(0.126)	(0.387)	–	(0.008)
Observations	408	408	408	374
R^2	0.873	0.516	0.435	0.044
Note:				*p<0.1; **p<0.05; ***p<0.01

Figure 2: Difference curves for the agricultural production function

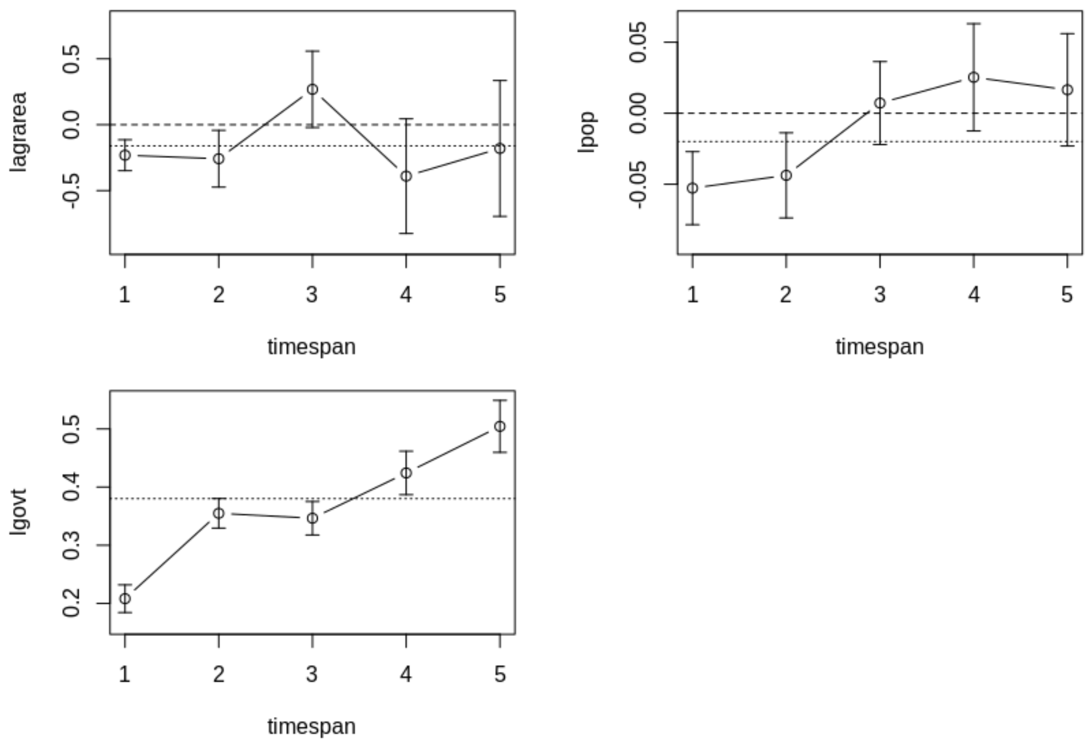

Table 5: Estimation results for Grunfeld investment data

	investment		
	Pooled OLS	Within	RE
value	0.116^{***}	0.110^{***}	0.110^{***}
	(0.006)	(0.012)	(0.010)
capital	0.231^{***}	0.310^{***}	0.308^{***}
	(0.025)	(0.017)	(0.017)
Constant	-42.714^{***}		-57.834^{**}
	(9.512)	–	(28.899)
Observations	200	200	200
R^2	0.812	0.767	0.770
Note:			*p<0.1; **p<0.05; ***p<0.01

Figure 3: Difference curves for Grunfeld investment data