DIPARTIMENTO DI
SCIENZE ECONOMICHE

Working Paper Series
Department of Economics
University of Verona

Negative variance estimates in panel data models

Laura Magazzini, Giorgio Calzolari

WP Number: 15 October 2010

ISSN: 2036-2919 (paper), 2036-4679 (online)

Negative variance estimates in panel data models

Laura Magazzini[*] Giorgio Calzolari[†‡]

This version: October 26, 2010

Abstract

Negative values for estimated variances can arise in a panel data context. Empirical and theoretical literature dismisses the problem as not serious and a practical solution is to replace negative variances by its boundary value, i.e. zero. While this is not a concern when the individual variance components is "small" with respect to idiosyncratic variance component (making it indistinguishable from zero in practice), we claim that a negative estimated variance can also arise with a "large" individual variance component, when the orthogonality condition between the individual effects and regressors fails. Estimation problems are considered in the (feasible) generalized least squares and maximum likelihood frameworks.

Keywords: Panel data, random effect estimation, negative variances, maximum likelihood.

[*]Department of Economics, Università di Verona, Viale dell'Università 4, 37129 Verona, Italy, <laura.magazzini@univr.it>

[†]Department of Statistics, Università di Firenze, Viale Morgagni 59, 50134 Firenze, Italy, <calzolar@ds.unifi.it>

[‡]We gratefully acknowledge suggestions from D. Lubian, A. Sembenelli, but retain full responsibility for the contents of this paper.

1 Background

The possibility of estimating a negative variance in the random effect (RE) model for panel data is an established problem in the empirical and theoretical literature. Econometrics textbooks acknowledge this possibility and refer to the work of Maddala (1971), who considers maximum likelihood (ML) estimation in the case of normally distributed error components and sets forth the condition under which the estimated variance of the individual component is negative.[1] If non-negativity constraints are imposed for estimation, a boundary solution at zero can therefore arise. Analogously, when estimation is performed by two-step feasible generalized least squares (FGLS), the estimated variance of the individual component is not guaranteed to be non-negative.[2]

Maddala and Mount (1973) explores the performance of the ML and alternative FGLS estimators of the variance components by means of Monte Carlo simulations, showing that the problem of negative variances exists within a range of methods proposed in the literature. However, in their setting, an estimated negative variance arises when the individual effect variance is "small" relative to the variance of the idiosyncratic component (of the order of 1/400). In this case, OLS estimation (that sets to zero the variance of the individual component) provides results that are "just as reliable" (Maddala and Mount, 1973).[3] Accordingly, negative variance estimates are deemed as "not serious" in the literature, and the problem is solved by replacing

[1]Maddala (1971) also shows that the estimate of the variance of the idiosyncratic component cannot be lower than zero.

[2]Note that, under normality, (F)GLS and ML estimators are asymptotically equivalent. See e.g. Hsiao (2003, §3.3); Greene (2003, §13.4); Cameron and Trivedi (2005, §21.7); Baltagi (2008, §2.3).

[3] Under the assumptions of the RE model, the choice between FGLS and OLS estimation is only a matter of efficiency, where the FGLS estimator allows to fully exploit the information about the covariance structure of the error terms.

non-positive estimates with the boundary value, i.e. zero.[4]

Interest in the issue has waned over time, even though the explanation provided by the literature is incomplete (and sometimes potentially misleading).

In this paper, we show that a negative value for the estimated variance of the individual component (both within the FGLS and ML framework) can arise if the orthogonality condition between the individual effects and the regressors is not satisfied. The assumption is required for the consistency of the FGLS (and ML) estimators, and the problem is related to the well-known property of the estimated variance of a model with endogenous regressors: if endogeneity is not accounted for, the estimated variance of the error term is downward biased. Indeed, the estimated variance of the individual component is obtained as the difference between (i) the average of the squared OLS residuals (or the residuals from the between regression, both affected by the endogeneity problem in case the orthogonality condition is not satisfied), and (ii) the average squared residuals from the within regression (robust to any pattern of correlation between the regressors and the individual effect). As a result, if the orthogonality condition is not satisfied, the estimated value of the individual variance will be smaller than the true value, possibly also lower than zero.

2 Why a negative estimated variance?

The RE panel data model estimated on N units (firms, individuals, households,...) over T time periods ($i = 1, ..., N$; $t = 1, ..., T$) can be written as:

$$y = X\beta + u \tag{1}$$

[4]Maddala and Mount (1973); Cameron and Trivedi (2005, §21.7); Hsiao (2003, §3.3); Baltagi (2008, §2.3).

where y is a $NT \times 1$ vector, X is a $NT \times k$ matrix on k observable strictly exogenous variables, β is a $k \times 1$ vector that contains the parameters of interest, and u is an $NT \times 1$ vector of disturbances.

We decompose the disturbance terms into two independent components:[5]

$$u_{it} = \mu_i + \nu_{it} \tag{2}$$

where μ_i represents the individual effects assumed to be $IN(0, \sigma_\mu^2)$, and v_{it} are assumed $IN(0, \sigma_\nu^2)$.

Note that σ_μ^2 is not uniquely interpreted as the variance of the individual effect (bounded to be greater than 0), but it can also be interpreted as the covariance between u_{it} and u_{is} $(t \neq s)$.

2.1 Maddala (1971)'s condition

Maddala (1971) considers the following reparametrization: $\sigma^2 = \sigma_\mu^2 + \sigma_\nu^2$, and $\rho = \sigma_\mu^2 / \sigma^2$.

In his seminal paper, the likelihood equations are solved by a two-step procedure and a necessary and sufficient condition is identified for the occurrence of a boundary solution at $\rho = 0$.[6] This is:

$$T_{yy} - \alpha' T_{xx} \alpha > T[B_{yy} - 2\alpha' B_{xy} + \alpha' B_{xx} \alpha] \tag{3}$$

[5] More generally time effects can be also considered, leading to $u_{it} = \mu_i + \tau_t + \nu_{it}$ with the additional assumption that τ_t, the time effects, are distributed as $IN(0, \sigma_\tau^2)$. As standard panel dimensions allow the inclusion of time dummies, time effect are omitted from the error term (i.e. $\sigma_\tau^2 = 0$).

[6] Note that the reformulation of the model is only valid if the covariance between the two error components is equal to zero, as it is customarily assumed in panel data applications. Berzeg (1979) relaxes this assumption letting $\mathrm{cov}(\mu_i, \nu_{it}) = \sigma_{\mu\nu}$. Accordingly the sum of the two variance components is $\sigma^2 = \sigma_\mu^2 + \sigma_\nu^2 + 2\sigma_{\mu\nu}$, and $\rho = (\sigma_\mu^2 + 2\sigma_{\mu\nu})/\sigma^2$, which admits negative values if $\sigma_\mu^2 < -2\sigma_{\mu\nu}$.

where $\alpha = T_{xx}^{-1}T_{xy}$, $T_{yy} = \sum_{i=1}^{N} y_i' y_i$, $T_{xx} = \sum_{i=1}^{N} X_i' X_i$, $B_{xx} = \frac{1}{T}\sum_{i=1}^{N}(X_i' ee' X_i)$, $B_{yy} = \frac{1}{T}\sum_{i=1}^{N}(y_i' ee' y_i)$, and $B_{xy} = \frac{1}{T}\sum_{i=1}^{N}(X_i' ee' y_i)$, with X_i is a $T \times k$ matrix containing observations on unit i, y_i is the $T \times 1$ vector of observations for the dependent variable (unit i), and e is a $T \times 1$ vector of ones.

In its original form (3), Maddala's condition is simply an algebraic inequality involving OLS estimated coefficients and transformation of the data. With some algebra, condition (3) reduces to:[7]

$$\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s\neq t} \hat{u}_{is}\hat{u}_{it} < 0 \tag{4}$$

where $\hat{u}_{it}$ denotes OLS residuals.

The formulation that uses OLS residuals (4) shows us that Maddala (1971)'s condition in equation (3) does not provide any hints on the reasons behind a (meaningless) negative estimated variance. Equation (3) and (4) simply state "tautologically" that the ML estimate of the individual variance is negative if it can be estimated as a negative value. Indeed, the within group correlation of estimated residuals is exploited in order to build FGLS/ML estimates of the individual variance component.[8]

[7]Computations exploit the orthogonality between OLS estimated residuals and the explanatory variables included in the model. Detailed algebra is available from the authors upon request.

[8]When discussing the FGLS estimation of the random components model, Wooldridge (2002, §10.4) exploits the (degree of freedom adjusted average) within-group correlation of pooled OLS residuals (4) in order to get an estimate of σ_μ^2. Indeed, under the assumption of uncorrelated random components, $\mathrm{E}[\sum_{t=1}^{T}\sum_{s\neq t} u_{it}u_{is}] = \sigma_\mu^2 T(T-1)$. Based on this expression, Wooldridge (2002, §10.4) claims that a negative $\hat{\sigma}_\mu^2$ can be explained in the case of negative serial correlation in u_{it}. If this is the case, $\mathrm{E}\Big[\sum_{t=1}^{T}\sum_{s\neq t} u_{it}u_{is}\Big] = \sigma_\mu^2 T(T-1) + \sigma_\nu^2 \sum_{t=1}^{T}\sum_{s\neq t}\rho_{ts}$, with ρ_{ts} denoting the correlation between ν_{it} and ν_{is}. For the expected value to be negative, we therefore need: $\sigma_\mu^2/\sigma_\nu^2 < -\sum_{s\neq t}\rho_{ts}/T(T-1)$ (coherently a negative variance can only arise in the case of negative serial correlation). As an example, let us consider $T = 5$ and let ν_{it} follow an autoregressive process of order 1 with $\rho = -0.9$, i.e. $\rho_{ts} = -.9^{|t-s|}$. In order to have an expected negative value, we would need $\sigma_\mu^2/\sigma_\nu^2 < 0.2$, i.e. also in this case σ_μ^2 needs to be "small" with respect to σ_ν^2.

2.2 Failure of the orthogonality condition

Starting from model (1) with $u_{it} = \mu_i + \nu_{it}$, alternative ways of estimating the variance components have been proposed in the literature, exploiting different sums of squared residuals (Wallace and Hussain, 1969; Amemiya, 1971; Maddala, 1971; Swamy and Arora, 1972; Fuller and Battese, 1974). As an example, the estimate of σ_μ^2 can be recovered by taking the difference between (a) the residuals from OLS regression whose average sum of squares is exploited to estimate $\sigma_\mu^2 + \sigma_\nu^2$, and (b) the residuals from the within transformed regression (LSDV model) whose average sum of squares is exploited to estimate σ_ν^2 (Greene, 2003, §13.4). The estimate of σ_μ^2 is therefore expected to be non-negative as "the sum of squares in the LSDV model cannot be larger than that in the simple regression with only one constant term" (Greene, 2003, §13.4). Despite that, negative estimated variances can arise in applications.[9]

Odd at a first glance, the fact can be reconciled as OLS needs an additional restriction with respect to LSDV for consistency, i.e. the orthogonality condition. If this is not satisfied, the OLS (restricted) model suffers of the standard omitted variable bias (due to the omission of the unit specific components), and the estimate of the sum of the variance components is downward biased.[10] On the contrary, the estimated variance of the idiosyncratic component will rely on the residuals from the within regression, that is not affected by the presence of correlated individual effects. As a result, the estimate of the individual variance component $\hat{\sigma}_\mu^2$ will be downward biased, possibly also lower than zero.[11]

[9] When non-negativity is imposed, a boundary solution at zero arises.

[10] Bounded to lie above zero as computed as the sum of squared OLS residuals.

[11] As $u_{it} = \mu_i + \nu_{it}$, also the presence of correlation between ν and x would cause a downward bias in the estimate of the sum of the two variance components that rely on OLS residuals. However, this effect would also bias the estimated variance of the within regression (b), making a negative value less likely to appear.

Within a ML framework, computations analogous to the FGLS formulas are employed, without the degree of freedom correction for the estimated variance components (Greene, 2003, §13.4). By relying on standard textbook treatment, the ML estimate of σ_μ^2 is obtained by considering the difference between the variance of between residuals $\widehat{\sigma_\nu^2/T + \sigma_\mu^2}$ and the variance of within residuals $\hat{\sigma}_\nu^2/T$. It is possible to show that the difference is negative if and only if

$$\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s\neq t}\hat{\hat{u}}_{it}\hat{\hat{u}}_{is} < 0 \tag{5}$$

with $\hat{\hat{u}}_{it}$ denoting the ML (FGLS) residuals.[12] Still, the between-regression provides a consistent estimate of β if the orthogonality condition is satisfied, otherwise the model suffers of an omitted variable bias and the estimated sum of squared residuals is downward biased. Again, the estimate of the variance of within residuals $\hat{\sigma}_\nu^2/T$ is not affected by the presence of correlation between x and μ_i, and therefore the difference between the two, that is the estimated value of the variance of the individual component, will be downward biased, maybe also lower than zero.

3 A Simulation Experiment

In this section we focus on the correlation structure between x and the individual effect μ, and explore how this can affect the estimated value of σ_μ^2.

The model is simulated as follows ($i = 1, ..., N$; $t = 1, ..., T$):

$$y_{it} = x_{it} + \mu_i + \nu_{it}$$

[12]See e.g. Arellano (2003, §3.2) and Greene (2003, §13.4). Note the analogy with equation (4) obtained by developing Maddala's condition (3). Recall that OLS and FGLS only differ in terms of efficiency.

	Value of γ						
	0	0.1	0.25	0.5	0.75	0.95	0.99
Static DGP for x							
Mean of $\hat{\sigma}^2_{bet}$	1.199	1.151	.9483	.5733	.3333	.2203	.2034
Mean of $\hat{\sigma}^2_{\nu}$	.9996	.9996	.9996	.9996	.9996	.9996	.9996
% of negative $\hat{\sigma}^2_{\mu}$	0%	0%	0%	0%	0%	2.6%	37.8%
Dynamic DGP for x							
Mean of $\hat{\sigma}^2_{bet}$	1.198	1.134	.8886	.5076	.3034	.2159	.2031
Mean of $\hat{\sigma}^2_{\nu}$	1.001	1.001	1.001	1.001	1.001	1.001	1.001
% of negative $\hat{\sigma}^2_{\mu}$	0%	0%	0%	0%	0%	7.1%	39.5%

Table 1: Results of Monte Carlo experiments

with $\mu_i \sim N(0, \sigma^2_\mu)$, $\nu_{it} \sim N(0, \sigma^2_\nu)$, and, in order to allow for correlation between x and μ, we let $x_{it} = \gamma\mu_i + \sqrt{1-\gamma^2}\xi_{it}$ with $\xi_{it} \sim N(0,1)$ independent of μ_i. As the variables in a panel dataset may exhibit strong patterns of autocorrelation over time, we also consider a dynamic specification for x where we let $x_{it} = \gamma\mu_i + \sqrt{1-\gamma^2}\xi_{it} + 0.5x_{it-1}$.[13] We let $\sigma^2_\mu = \sigma^2_\nu = 1$.

The data generating process is simulated for different values of γ, selected in the interval $[0, 1)$.[14] We considered $N = 1,000$ and $T = 5$.

Standard formulas (Baltagi, 2008, pag. 19) are applied to compute $\hat{\sigma}^2_\mu$ and $\hat{\sigma}^2_\nu$. The sum of squared residuals of the between regression is used to estimate $\sigma^2_{bet} = \sigma^2_\nu/T + \sigma^2_\mu$, and the sum of squared residuals from LSDV estimation provides an estimate of σ^2_ν. The estimated value of σ^2_μ is therefore obtained by considering $\hat{\sigma}^2_{bet} - \hat{\sigma}^2_\nu/T$.

Results of the Monte Carlo experiments are reported in Table 1 as a function of the value of γ employed during the simulations.

For $\gamma > 0$, the estimated value of the variance from the between regression is

[13]In the dynamic specification, we let $x_{i0} = \gamma\mu_i/(1-0.5) + \xi_{i0}$. We simulate $10 + T$ values for x_{it} and then disregard the first 10.

[14]10,000 Monte Carlo replications are considered. The seed is reset after each simulation set, so that the differences among the distributions of the estimates are only driven by differences in γ.

increasingly biased toward zero (bias is increasing with γ), whereas the estimated variance of the idiosyncratic component (based on the FE estimator of β) is not affected by the increased correlation. As a result, the variance of the individual component, estimated as the difference between estimated variance of the between regression and the estimated variance of the idiosyncratic component is biased toward zero, and can also assume negative values. When $\gamma = 0.95$, in 263 cases out of 10,000 Monte Carlo replications with static x the variance of the individual components is estimated to be lower than zero (717 cases in the case of autocorrelated x) and the number increases to 3782 when $\gamma = 0.99$ (3956 in the dynamic specification for x).

3.1 Further extensions

This problem has also implications for the computation of the Hausman statistics that is used to discriminate between the RE and FE approaches (Hausman, 1978). In applications, the estimated variance covariance matrix involved in the computation of the Hausman statistic can be not positive definite, and in more extreme cases the value of the statistic is negative! Textbook explanation for the negative result relies on a small sample problem, as the distribution of the Hausman statistic is chi-squared *asymptotically.* We claim that it is possible to observe a (meaningless) negative value for the Hausman statistics (or an estimated variance covariance matrix that is not positive definite) when the orthogonality condition is not satisfied. As shown, the correlation between the individual effect and the regressors causes a downward biased estimate of the sum of the two individual error components, that can, in turn, lead to a non-positive definite estimate of the variance covariance matrix of the difference of the FE and RE estimate of β. Recent research shows that the Hausman test can be negative even asymptotically if the alternative hypothesis is correct, i.e., in our context, if the orthogonality conditions is not satisfied (Schreiber, 2008). Schreiber (2008) also shows that in some cases, the pitfall can lead to misleading positive test statistics.

References

[1] Amemiya, T.: 1971, The Estimation of the Variances in a Variance-Components Model, *International Economic Review* **12**(1), 1–13.

[2] Arellano, M.: 2003, *Panel Data Econometrics*, Oxford University Press.

[3] Cameron, A.C. and Trivedi, P.K.: 2005, *Microeconometrics*, Cambridge University Press.

[4] Baltagi, B.: 2008, *Econometric analysis of panel data*, John Wiley and Sons.

[5] Berzeg, K.: 1979, The Error Component Model. Conditions for the Existence of the Maximum Likelihood Estimates, *Journal of Econometrics* **10**, 99–102.

[6] Fuller, W. and Battese, G.: 1974, Estimation of Linear Models with Crossed-Error Structure, *Journal of Econometrics* **2**(1), 67–78.

[7] Greene, W.: 2003, *Econometric analysis*, 6th ed., Prentice Hall.

[8] Hausman, J.A.: 1978, Specification Tests in Econometrics, *Econometrica* **46**(6), 1251–1271.

[9] Hsiao, C.: 2003, *Analysis of Panel Data*, Cambridge University Press.

[10] Maddala, G.: 1971, The Use of Variance Components Models in Pooling Cross Section and Time Series Data, *Econometrica* **39**(2), 341–358.

[11] Maddala, G. and Mount, T.: 1973, A comparative study of alternative estimators for variance components models used in econometric applications, *Journal of the American Statistical Association* **68**(342), 324–328.

[12] Schreiber, S.: 2008, The Hausman Test Statistic can be Negative even Asymptotically, *Journal of Economics and Statistics*, **228**(4), 394–405.

[13] Swamy, P. and Arora, S.: 1972, The exact finite sample properties of the estimators of coefficients in the error components regression models, *Econometrica* **40**(2), 261–75.

[14] Wallace, T. and Hussain, A.: 1969, The Use of Error Components Models in Combining Cross Section with Time Series Data, *Econometrica* **37**(1), 55–72.

[15] Wooldridge, J.: 2002, *Econometric Analysis of Cross Section and Panel Data*, MIT Press.