

C | E | D | L | A | S

Centro de Estudios
Distributivos, Laborales y Sociales

Maestría en Economía
Facultad de Ciencias Económicas



UNIVERSIDAD NACIONAL DE LA PLATA

Tests for Normality in Linear Panel Data Models

Javier Alejo, Antonio Galvao, Gabriel Montes-Rojas y
Walter Sosa-Escudero

Documento de Trabajo Nro. 178
Febrero, 2015

ISSN 1853-0168

Tests for normality in linear panel data models

Javier Alejo
CEDLAS-FCE-UNLP-CONICET
La Plata/Argentina
javier.alejo@depeco.econo.unlp.edu.ar

Antonio Galvao
University of Iowa
Iowa City, Iowa/United States
antonio-galvao@uiowa.edu

Gabriel Montes-Rojas
CONICET-Universidad de San Andrés
Victoria/Argentina
gmontesrojas@udesa.edu.ar

Walter Sosa-Escudero
Universidad de San Andrés-CONICET
Victoria/Argentina
wsosa@udesa.edu.ar

Abstract. A new *Stata* command, `xtsktest`, is proposed to explore non-normalities in linear panel data models. The tests explore skewness and excess kurtosis allowing researchers to identify departures away from gaussianity in both error components of a standard panel regression, separately or jointly. The tests are based on recent results by Galvao, Montes-Rojas, Sosa-Escudero and Wang (2013), and can be seen as extending the classical Bera-Jarque normality test for the case of panel data.

Keywords: `st0001`, `xtsktest`, skewness, kurtosis, normality, panel data

1 Introduction

The need to check for non-normal errors in regression models obeys to both methodological and conceptual reasons. From a strictly methodological point of view, lack of Gaussianity sometimes harms the reliability of simple estimation and testing procedures, and calls for either better methods under alternative distributional assumptions, or for robust alternatives whose advantages do not depend on distributional features. Alternatively, whether errors should be more appropriately captured by skewed and/or leptokurtic distributions may be a statistical relevant question *per se*.

The normality assumption also plays a crucial role in the validity of inference procedures, specification tests and forecasting. In the panel data literature, Blanchard and Mátyás (1996) examine the consequences of non-normal error components for the performance of several tests. Montes-Rojas and Sosa-Escudero (2011) show that non-normalities severely affect the performance of the panel heteroskedasticity tests by Holly and Gardiol (2000) and Baltagi et al. (2006). Despite these concerns the Gaussian framework is widely used for specification tests in the one-way error components model; see, for instance, the tests for spatial models in panel data by Baltagi et al. (2003) and Baltagi et al. (2007).

Even though there is a large literature on testing for skewness and kurtosis in cross-sectional and time-series data, including Erickson and Whited (2010), Bai and Ng (2005), and Bera and Premaratne (2001), to cite a few of an extensive list results for panel data models are scarce. A natural complication is that, unlike their cross-section or time-series counterparts, in simple error-components models lack of Gaussianity may arise in more than one component. Thus, an additional problem to that of detecting departures away from normality is the identification of which component is causing it. Previous work on the subject include Gilbert (2002), who exploits cross-moments, and Meintanis (2011), who proposes an omnibus-type test for normality in both components jointly based on empirical characteristic functions.

Galvao et al. (2013) develop tests for skewness (lack of symmetry), kurtosis, and normality for panel data one-way error component models. These tests are important in practice because, in the panel data case, the standard Bera-Jarque test is not able to disentangle the departures of the individual and remainder components from non-Gaussianity. The tests are constructed based on moment conditions of the within and between transformations of the OLS residuals. These conditions are exploited to develop tests for skewness and kurtosis in the individual-specific and the remainder components, separately and jointly. The tests are particularly useful for the case where the number of individuals, N , goes to infinity, but the number of time periods, T , is fixed and might be small.

The proposed tests are implemented in practice using a bootstrap procedure. Since the tests are asymptotically normal, the bootstrap can be used to compute the corresponding variance-covariance matrices of the statistics of interest and carry out inference. In particular, the tests are implemented using a cross-sectional bootstrap. We formally prove the consistency of the bootstrap method applied to our case of short panels.

The new command `xtsktest` implements a battery of tests to identify non-normalities in standard error components panel models, based on recent results by Galvao et al. (2013). For standard regression models, the classical Bera-Jarque test (implemented in `Stata` with `sktest`), is a simple procedure that detects departures away from gaussianity in the form of skewness and excess kurtosis in the regression error term. A natural concern in the case of panel data models is to identify which error component (if not both) is the source of non-normalities. The proposed tests allows researchers to explore skewness and excess kurtosis in each component separately or jointly. In this context, the proposed procedure can be seen as extending the famous Bera-Jarque tests for the case of simple panel data models.

Section 2 reviews the results of Galvao et al. (2013) and presents the tests. Section 3 describes the `xtsktest` syntax. Next we illustrate the procedure by applying the new tests to an investment model studied by Fazzari et al. (1988). We conclude with practical suggestions on the proper use of the tests.

2 Skewness and kurtosis in the one-way error components model

Consider the following standard panel data one-way error components model

$$y_{it} = x_{it}b + u_i + e_{it}, \quad i = 1, \dots, N, \quad t = 1, \dots, T, \quad (1)$$

where b is a p -vector of parameters, and u_i , e_{it} , and x_{it} are copies of random variables u , e , and x , respectively (b does not contain a constant). As usual, the subscript i refers to individual, and t to time. Here u_i and e_{it} refer to the individual-specific and to the remainder error component, respectively, both of which have mean zero.

The quantities of interest are each component skewness,

$$s_u = \frac{E[u^3]}{(E[u^2])^{3/2}}, \quad \text{and} \quad s_e = \frac{E[e^3]}{(E[e^2])^{3/2}},$$

and kurtosis,

$$k_u = \frac{E[u^4]}{(E[u^2])^2}, \quad \text{and} \quad k_e = \frac{E[e^4]}{(E[e^2])^2}.$$

Galvao et al. (2013) construct statistics for testing for skewness and kurtosis in the individual-specific and the remainder components, separately and jointly. When the underlying distribution is normal, the null hypotheses of interest become $H_0^{s_u} : s_u = 0$ and $H_0^{s_e} : s_e = 0$ for skewness and $H_0^{k_u} : k_u = 3$ and $H_0^{k_e} : k_e = 3$ for kurtosis. Moreover, under normality, the null hypotheses for these cases are given by

$$H_0^{s_u \& k_u} : s_u = 0 \text{ and } k_u = 3,$$

$$H_0^{s_e \& k_e} : s_e = 0 \text{ and } k_e = 3.$$

Following Galvao et al. (2013), the statistics for symmetry are $\widehat{SK}_u^{(1)} = \widehat{E}[u^3]$ and $\widehat{SK}_e^{(1)} = \widehat{E}[e^3]$ and those for kurtosis are $\widehat{KU}_u^{(1)} = \widehat{E}[u^4] - 3(\widehat{E}[u^2])^2$ and $\widehat{KU}_e^{(1)} = \widehat{E}[e^4] - 3(\widehat{E}[e^2])^2$. Similar statistics can also be presented in a standardized way as $\widehat{SK}_u^{(2)} = \frac{\widehat{E}[u^3]}{(E[u^2])^{3/2}}$ and $\widehat{SK}_e^{(2)} = \frac{\widehat{E}[e^3]}{(E[e^2])^{3/2}}$ for symmetry and $\widehat{KU}_u^{(2)} = \frac{\widehat{E}[u^4]}{(E[u^2])^2} - 3$ and $\widehat{KU}_e^{(2)} = \frac{\widehat{E}[e^4]}{(E[e^2])^2} - 3$ for kurtosis. Each statistic is consistent and properly standardized follows a $N(0,1)$ asymptotic law under the corresponding null hypothesis. However, they may differ in small samples. Moreover, tests for joint symmetry and kurtosis are constructed using $(\widehat{SK}_u^{(j)})^2 + (\widehat{KU}_u^{(j)})^2$ and $(\widehat{SK}_e^{(j)})^2 + (\widehat{KU}_e^{(j)})^2$, $j = 1, 2$, each following a χ^2_2 asymptotic law under the corresponding null hypothesis. The tests implementation uses bootstrap to estimate the variances of the skewness and kurtosis test statistics in practice.

3 The `xtsktest` syntax

3.1 Syntax

The command syntax is:

```
xtsktest depvar [ indepvars ] [ if ] [ in ] [ , reps(#) seed(#) standard  
]
```

The simplest case is when it is executed only by invoking its name.

3.2 Options

`xtsktest` supports the following options:

`reps(#)` specifies the number of bootstrap replications. The default is `reps(50)`.

`seed(#)` specifies the seed for the random-number generator. See [R] `set seed`.

`standard` specifies if the skewness and kurtosis statistics are standardized by the estimated variance. The default is no standardization.

3.3 Remarks

`xtsktest` can be used both as a standard command or as a post estimation command after `ols` or random effects model (see [R] `regress` and [XT] `xtreg`). In the first setting the command requires at least one variable in the `varlist`, while in the post-regression background `varlist` is not required. Example 1 shows the first option; examples 2 and 3 use `xtsktest` as a post estimation command.

3.4 Saved results

`xtsktest` stores the following estimates saved to `e()`:

Matrices

`e(xtsk_text)` skewness and kurtosis test results, one per row; first column: point estimation, second: standard errors, third:p-values.

`e(joint)` joint skewness and kurtosis test results, one per row; first column: chi-squared statistics, second:p-values.

4 Empirical application: Investment equation

In this section, we apply the developed tests to Fazzari, Hubbard, and Petersen's (1988) investment equation model, where a firm's investment is regressed on an observed measure of investment demand (Tobin's q) and cash flow. This is one of the most well-known models in the corporate investment literature and we use this application as a way to illustrate our theoretical results. Following Fazzari et al. (1988), investment-cash-flow sensitivities became a standard metric in the literature that examines the impact of financing imperfections on corporate

investment (Stein 2003). These empirical sensitivities are also used for drawing inferences about efficiency in internal capital markets (Lamont 1999; Shin and Stulz 1998), the effect of agency on corporate spending (Hadlock 1998; Bertrand and Mullainathan 2005), the role of business groups in capital allocation (Hoshi et al. 1991), and the effect of managerial characteristics on corporate policies (Bertrand and Schoar 2003; Malmendier and Tate 2005).

Tobin's q is the ratio of the market valuation of a firm and the replacement value of its assets. Firms with a high value of q are considered attractive as investment opportunities, whereas a low value of q indicates the opposite. Investment theory is also interested in the effect of cash flow, as the theory predicts that financially constrained firms are more likely to rely on internal funds to finance investment (see e.g. Erickson and Whited (2000)). The baseline model in the literature is

$$I_{it}/K_{it} = \alpha + \beta q_{it-1} + \gamma CF_{it-1}/K_{it-1} + u_i + e_{it}, \quad (2)$$

where I denotes investment, K capital stock, CF cash flow, q Tobin's q , u is the firm-specific effect and e is the innovation term.

We check for skewness and kurtosis in both u and e using the proposed tests. We are interested in testing for skewness and kurtosis for at least three reasons. First, testing normality plays a key role in forecasting models at the firm level. Second, asymmetry in both components is used for solving measurement error problems in Tobin's q . The operationalization of q is not clear-cut, so estimation poses a measurement error problem. Many empirical investment studies found a very disappointing performance of the q theory of investment, although this theory has a good performance when measurement error is purged as in Erickson and Whited (2000). Their method requires asymmetry in the error term to identify the effect of q on firm investment. Third, skewness and kurtosis by themselves provide information about the industry investment patterns. Skewness in u determines that a few firms either invest or disinvest considerably more than the rest, while kurtosis in u determine that a few firms locate at both sides of the investment line, that is, some invest a large amount while others disinvest large amounts too. Skewness and/or kurtosis in e show that the large values of investment correspond to firm level shocks.

We follow Almeida et al. (2010), who considered a sample of manufacturing firms (SICs 2000 to 3999) over the 2000 to 2005 period with data available from COMPUSTAT's P/S/T, full coverage. Only firms with observations in every year are used, in order to construct a balanced panel of firms for the five year period. Moreover, following those authors, we eliminate firms for which cash-holdings exceeded the value of total assets and those displaying asset or sales growth exceeding 100%. Our final sample consists of 410 firm-years and 82 firms. Because we only consider firms that report information in each of the five years, the sample consists mainly of relatively large firms.

To demonstrate the use of `xtsktest` to this application, we must first open the dataset and declare it to be panel data. See [XT] `xtset`.

```
. use investment.dta
```

```
. xtset idcode time
      panel variable: idcode (strongly balanced)
      time variable: time, 2 to 6
      delta: 1 unit
```

Consider first an OLS estimation of the effect of Tobin's q and cash flows on investment.

```
. regress investment tobinq cashflow
```

Source	SS	df	MS			
Model	.536747282	2	.268373641	Number of obs =	410	
Residual	1.22632448	407	.003013082	F(2, 407) =	89.07	
Total	1.76307176	409	.004310689	Prob > F =	0.0000	
				R-squared =	0.3044	
				Adj R-squared =	0.3010	
				Root MSE =	.05489	

investment	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
tobinq	.0384663	.0094022	4.09	0.000	.0199834	.0569492
cashflow	.1117721	.0096142	11.63	0.000	.0928724	.1306718
_cons	.0669764	.0087876	7.62	0.000	.0497016	.0842512

Next, consider a one-way error components random effects model.

```
. xtreg investment tobinq cashflow, re
```

investment	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
tobinq	.0673706	.0129138	5.22	0.000	.04206	.0926812
cashflow	.0824715	.0115191	7.16	0.000	.0598944	.1050486
_cons	.0516002	.0127921	4.03	0.000	.0265281	.0766722
sigma_u	.0380806					
sigma_e	.03857635					
rho	.49353308	(fraction of variance due to u_i)				

Random-effects GLS regression		Number of obs =	410
Group variable: idcode		Number of groups =	82
R-sq: within =	0.1014	Obs per group: min =	5
between =	0.3583	avg =	5.0
overall =	0.2779	max =	5
corr(u_i, X) =	0 (assumed)	Wald chi2(2) =	84.09
		Prob > chi2 =	0.0000

The results show a positive and significant effect of both Tobin's q and cash flows on investment flows in both models. The random effects model also shows that there is considerable variation across firms in terms of unobservables. In fact, half of the variation is due to the firm specific component u_i and other half to the remainder component e_{it} . Note that the presence of firm specific effects determines that OLS standard errors are not correct, while the random effects are.

(output omitted)

The screen output shows in the first column the observed coefficients of the four statistics (without standarization, $\widehat{SK}_e^{(1)} = .0000387$, $\widehat{KU}_e^{(1)} = 9.33e - 06$, $\widehat{SK}_u^{(1)} = .0000511$ and $\widehat{KU}_u^{(1)} = 4.27e - 08$) used for symmetry and kurtosis for each error component. The next columns show the standard errors computed by bootstrap replications, the z statistics, p-values and the 95% confidence intervals using the normal approximation. Finally, the lower part of the results output show the joint test for normality on each component of the error term and their p-values. The model shows that both components are asymmetric (with right symmetry), while only the remainder component e has excess kurtosis. Thus, while we expect the occurrence of large positive investment shocks ($E[e^3] > 0$) these are systematic in some firms (i.e. $E[u^3] > 0$). Asymmetry thus produces the rejection of the null hypothesis of normality in both error components, although the rejection is stronger for the remainder than for the firm specific component.

We also evaluate symmetry and kurtosis in each component using the standardized statistics, $\widehat{SK}_e^{(2)}$, $\widehat{KU}_e^{(2)}$, $\widehat{SK}_u^{(2)}$ and $\widehat{KU}_u^{(2)}$. These can be implemented with the option `standard`.

```
. * Example 4, standardized coefficients
. xtstktest investment tobinq cashflow, reps(500) seed(123) standard
(running _xtstktest_calculations on estimation sample)
```

```
Bootstrap replications (500)
----- 1 ----- 2 ----- 3 ----- 4 ----- 5
..... 50
..... 100
..... 150
..... 200
..... 250
..... 300
..... 350
..... 400
..... 450
..... 500
```

```
Tests for skewness and kurtosis          Number of obs    =    410
                                           Replications      =    500
```

(Replications based on 82 clusters in idcode)

	Observed	Bootstrap			Normal-based	
	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
Skewness_e	.6040947	.1523494	3.97	0.000	.3054954	.902694
Kurtosis_e	3.645848	.6932803	5.26	0.000	2.287044	5.004653
Skewness_u	.9857612	.2176725	4.53	0.000	.5591309	1.412391
Kurtosis_u	.0220666	.4963144	0.04	0.965	-.9506917	.994825

```
Joint test for Normality on e:          chi2(2) = 43.38    Prob > chi2 = 0.0000
Joint test for Normality on u:          chi2(2) = 20.51    Prob > chi2 = 0.0000
```

Note: standardized coefficients

As expected, the results do not differ from those presented with the non-standardized statistics. The numeric results however provide an easier interpretation of the excess kurtosis in the remainder component with a value of $\widehat{KU}_e^{(2)} = 3.645848$ and firm-specific component $\widehat{KU}_u^{(2)} = .0220666$. The joint test for normality in u , however, provides a higher chi-squared value with a clearer rejection than in the previous examples using non-standardized coefficients.

5 Conclusion

This paper implements tests for skewness/symmetry and kurtosis of the error components in linear panel data random effects models. The `xtsktest` procedure allows the evaluation of each error component third and fourth moments. This can be used as an alternative to the Bera-Jarque test in panel data models.

As discussed in the Introduction, checking for skewness and kurtosis in the error components plays an important role in testing and estimation procedures in linear panel data models. Deviations from symmetry and kurtosis of 3 invalidate methods that are not robust to normality. Moreover, estimating third and fourth moments is also important for forecasting in panel data models (see Baltagi 2008, for a discussion).

6 References

- Almeida, H., M. Campello, and A. Galvao. 2010. Measurement Errors in Investment Equations. *Review of Financial Studies* 23: 3279–3328.
- Bai, J., and S. Ng. 2005. Tests for skewness, kurtosis, and normality for time series data. *Journal of Business and Economic Statistics* 23: 49–60.
- Baltagi, B., B. Bresson, and A. Pirotte. 2006. Joint LM test for homoskedasticity in a one-way error component model. *Journal of Econometrics* 134: 401–417.
- Baltagi, B., S. Song, B. Jung, and W. Koh. 2007. Testing for serial correlation, spatial autocorrelation and random effects using panel data. *Journal of Econometrics* 140: 5–51.
- Baltagi, B., S. Song, and W. Koh. 2003. Testing panel data regression models with spatial error correlation. *Journal of Econometrics* 68: 133–151.
- Baltagi, B. H. 2008. Forecasting with Panel Data. *Journal of Forecasting* 27(2): 153–173.
- Bera, A., and G. Premaratne. 2001. Adjusting the tests for skewness and kurtosis for distributional misspecifications. University of Illinois at Urbana-Champaign, http://www.business.uiuc.edu/Working_Papers/papers/01-0116.pdf.

- Bertrand, M., and S. Mullainathan. 2005. Bidding for Oil and Gas Leases in the Gulf of Mexico: A Test of the Free Cash Flow Model? Mimeo, The University of Chicago and MIT.
- Bertrand, M., and A. Schoar. 2003. Managing with Style: The Effect of Managers on Firm Policies. *Quarterly Journal of Economics* 118: 1169–1208.
- Blanchard, P., and L. Mátyás. 1996. Robustness of tests for error component models to non-normality. *Economics Letters* 51: 161–167.
- Erickson, T., and T. Whited. 2000. Measurement Error and the Relationship between Investment and Q . *Journal of Political Economy* 108: 1027–1057.
- . 2010. Conditional skewness, kurtosis, and density specification testing: Moment-based versus nonparametric tests. *Studies in Nonlinear Dynamics & Econometrics* 14: Article 5.
- Fazzari, S., R. G. Hubbard, and B. Petersen. 1988. Financing Constraints and Corporate Investment. *Brooking Papers on Economic Activity* 1: 141–195.
- Galvao, A., G. Montes-Rojas, W. Sosa-Escudero, and L. Wang. 2013. Tests for skewness and kurtosis in the one-way error component model. *Journal of Multivariate Analysis* 122: 35–52.
- Gilbert, S. 2002. Testing the distribution of error components in panel data models. *Economics Letters* 77: 47–53.
- Hadlock, C. 1998. Ownership, Liquidity, and Investment. *RAND Journal of Economics* 29: 487–508.
- Holly, A., and L. Gardiol. 2000. A score test for individual heteroskedasticity in a one-way error components model. In *Panel Data Econometrics: Future Directions*, ed. J. Krishnakumar and E. Ronchetti, 199–211. Elsevier/North-Holland, Amsterdam.
- Hoshi, T., A. Kashyap, and D. Scharfstein. 1991. Corporate Structure, Liquidity, and Investment: Evidence from Japanese Industrial Groups. *Quarterly Journal of Economics* 106: 33–60.
- Lamont, O. 1999. Cash Flow and Investment: Evidence from Internal Capital Markets. *The Journal of Finance* 52: 83–110.
- Malmendier, U., and G. Tate. 2005. CEO Overconfidence and Corporate Investment. *The Journal of Finance* 60: 2661–2700.
- Meintanis, S. 2011. Testing for normality with panel data. *Journal of Statistical Computation and Simulation* 81: 1745–1752.
- Montes-Rojas, G., and W. Sosa-Escudero. 2011. Robust tests for heteroskedasticity in the one-way error components model. *Journal of Econometrics* 160: 300–310.

Shin, H., and R. Stulz. 1998. Are Internal Capital Markets Efficient? *Quarterly Journal of Economics* 113: 531–552.

Stein, J. 2003. Agency Information and Corporate Investment. In *Handbook of the Economics of Finance*, ed. M. Harris and R. Stulz. Elsevier/North-Holland, Amsterdam.

About the authors

Javier Alejo: Senior researcher with specialization in labor economics and inequality.

Antonio Galvao: Associate Professor with specialization in econometric theory and statistics.

Gabriel Montes-Rojas: Senior researcher with specialization in econometric theory and statistics.

Walter Sosa-Escudero: Associate Professor with specialization in econometric theory and statistics.