ISSN 2090-3359 (Print)
ISSN 2090-3367 (Online)

Advances in Decision Sciences

Volume 24
Issue 2
June 2020

Michael McAleer

Editor-in-Chief
University Chair Professor
Asia University, Taiwan

Published by Asia University, Taiwan

ADS@ASIAUNIVERSITY

Measurement Error in a First-order Autoregression*

Philip Hans Franses**

Econometric Institute
Erasmus School of Economics
Erasmus University Rotterdam
The Netherlands

Revised: June 2020

***** Thanks are due to Bert de Bruijn for the simulations, and a reviewer for helpful comments and suggestions.

****** Correspondence: franses@ese.eur.nl

Abstract

The Ordinary Least Squares (OLS) estimator for the slope parameter in a first-order autoregressive model is biased when the variable is measured with error. Such an error may occur with revisions of macroeconomic data. This paper illustrates and proposes a simple procedure to alleviate the bias, and is based on Total Least Squares (TLS). TLS is, in general, consistent, and also works well in small samples. Simulation experiments and an empirical example show the usefulness of this method.

Keywords: Errors-in-variables, OLS, First-order autoregression, Total Least Squares.

JEL: C20, C51.

1. Introduction

Macroeconomic data, like Gross Domestic Product (GDP) and its components, are often revised, even many years after their first quotes. In a sense, revisions can be viewed as measurement errors, and the trajectory between first quotes to final quotes can be viewed as moving from quotes with measurement errors to factual data, but perhaps also the reverse is possible. One may thus view macroeconomic national accounts data as data with measurement error.

When data are observed with error, this has consequences for econometric modelling. In this short paper the focus is on a simple first-order autoregression, just to indicate the matter of interest and to illustrate a potentially useful and very simple estimation method. Alternative estimators are presented in Staudenmayer and Buonaccorsi (2005) and recently in Zeng et al. (2018).

The next section deals with the basic problem, which is that the Ordinary least Squares (OLS) estimator of the slope parameter in the first-order autoregressive model is inconsistent. A simple solution is now to resort to an alternative estimator, and here Total Least Squares (TLS) is advocated. Such a TLS estimator is known to be consistent, see Fuller (1980, 1987), and some simulation experiments below support this claim. Interesting and theoretically detailed accounts of TLS are provided in Golub and Van Loan (1980) and Van Huffel, et al. (1996).

A key factor is the variance of the measurement error, and for practical purposes it is recommended to use an upper bound value, which is easy to establish using available macroeconomic data. An illustration to quarterly observed GDP growth rates in the Netherlands shows how easy it is to arrive at a proper TLS-based estimate for the slope parameter in a first-order autoregression.

2. The Problem

Consider a variable y_t^*, $t = 1,2, \ldots, T$, which obeys a first-order autoregression:

$$y_t^* = \alpha + \beta y_{t-1}^* + \varepsilon_t \quad (1)$$

where ε_t is a standard white noise process with mean 0 and variance σ_ε^2, and it is assumed that $|\beta| < 1$. Suppose now that y_t^* is measured with error, and that the measurement equation is:

$$y_t = y_t^* + w_t \quad (2)$$

with w_t is a white noise process with mean zero and variance σ_w^2. The measurement error w_t is assumed to be mutually uncorrelated with y_t^* and ε_t.

The OLS estimator for β in the regression of y_t on an intercept and y_{t-1} is:

$$\hat{\beta}_{OLS} = \frac{\sum_{t=2}^{T}(y_{t-1} - \bar{y})(y_t - \bar{y})}{\sum_{t=2}^{T}(y_{t-1} - \bar{y})^2}$$

with

$$\bar{y} = \frac{1}{T-1}\sum_{t=2}^{T} y_t$$

In the absence of measurement error, this OLS estimator is consistent, see Heij et al. (2004, page 559). In contrast, in case of measurement error as in (2), the OLS estimator is inconsistent (ibid, page 268), namely:

$$plim\ \hat{\beta}_{OLS} = \frac{\beta\sigma_\varepsilon^2}{\sigma_\varepsilon^2 + (1-\beta^2)\sigma_w^2}$$

as the true variance of y_t is:

$$Variance\ (y_t^* + w_t) = \frac{\sigma_\varepsilon^2}{1-\beta^2} + \sigma_w^2$$

There are various possible solutions to the measurement error problem; see Staudenmayer and Buonaccorsi (2005) for an overview of various estimators. A simple solution is changing the estimation method and resorting to Total Least Squares (TLS). This method does not seek to minimize the sum of squared vertical distances to the regression line, but instead seeks to minimize the sum of squared orthogonal distances.

In the Appendix some key aspects of the TLS method for the simple regression model are given, and a few of those results will be relevant next.

As in the case of (A.5) for the first-order autoregression, define:

$$\delta = \frac{\sigma_w^2+\sigma_\varepsilon^2}{\sigma_w^2} = 1 + \frac{\sigma_\varepsilon^2}{\sigma_w^2} \tag{3}$$

With the expressions in the Appendix, it is easy to derive that for an AR(1) holds that:

$$\bar{y} = \bar{x} = \frac{1}{T-1}\sum_{t=2}^{T} y_t$$

$$S_y = S_x = \frac{1}{T-2}\sum_{t=2}^{T}(y_t - \bar{y})^2$$

$$S_{yx} = \frac{1}{T-2}\sum_{t=2}^{T}(y_t - \bar{y})(y_{t-1} - \bar{y})$$

Define the first-order autocorrelation as:

$$\hat{\rho}_1 = \frac{\sum_{t=2}^{T}(y_t - \bar{y})(y_{t-1} - \bar{y})}{\sum_{t=2}^{T}(y_t - \bar{y})^2}$$

the expression for $\hat{\beta}_{TLS}$ thus becomes:

$$\hat{\beta}_{TLS} = \frac{1-\delta}{2\hat{\rho}_1} + \sqrt{\frac{(1-\delta)^2}{4\hat{\rho}_1^2} + \delta} \tag{4}$$

From (3) it can be seen that $\delta = 1$ only in the case where $\sigma_\varepsilon^2 = 0$, and that otherwise $\delta > 1$. When $\delta = 1$, then (4) shows that $\hat{\beta}_{TLS} = 1$.

3. Simulations

In order to empirically illustrate that TLS yields a consistent estimator, consider the simulation results in Table 1. These concern 10000 replications of an AR(1) process with parameters 0.5, 0.8 and 0.95, for sample sizes 10, 25, 50, 100, 200, 1000, 10000 and 100000. It is clear from the column with the header TLS that the estimator for the slope as in (4) is consistent, and that the OLS estimator is inconsistent. It can also be seen that there is a small sample bias, which is larger for OLS than for TLS.

Table 2 focuses on the smaller sample sizes (here only 25 and 50). From this table it can be learned that when the ratio $\frac{\sigma_\varepsilon^2}{\sigma_w^2}$ gets larger, that then the bias of OLS gets smaller, although TLS is always better.

4. Empirical Illustration

As an illustration, consider the first differences of quarterly GDP growth rates for the Netherlands in Figure 1, that is, growth rate at *t* minus growth rate at *t*-1. The flash value is the quote given 45 days after the end of the quarter. The final value is here taken as the quote after 3 years. The difference between the two can be viewed as a measurement error.

A model for the first differences of the GDP final-value growth rates appears to follow an AR(1) process. The OLS estimator for β is 0.381, which equals $\hat{\rho}_1$.

The σ_w^2 is estimated by the variance of the differences between the two variables in Figure 1 and it appears to be 0.149. Next, to have a first guess of the variance σ_ε^2, one can see from (1) that the maximum value is obtained when $\beta = 1$. The variance of the twice differenced final-value growth rates is 1.796.

When all relevant values are substituted into (3), one obtains a δ value of approximately 13. Substituting all relevant estimates into (4) results in a TLS estimate of the slope that is equal to 0.395.

5. Conclusion

Total Least Squares can be useful to alleviate measurement issues in time series models. This short paper has presented only a concise discussion of a first-order autoregression, but of course, also higher order time series models should be considered. More accurate methods to estimate the key parameter concerning the ratio between model error and measurement error are important too.

Appendix: TLS in a Simple Regression Model: A Rejoinder

Consider the regression model for the variables y_i and x_i with $i = 1, \dots, n$:

$$y_i = \alpha + \beta x_i^* + \varepsilon_i \tag{A.1}$$

where the true observations x_i^* are measured with error, that is:

$$x_i = x_i^* + v_i$$

The measurement error, v_i , with variance, σ_v^2 , is independent of both x_i^* and ε_i. The regression model in practice would then read as:

$$y_i = \alpha + \beta(x_i - v_i) + \varepsilon_i = \alpha + \beta x_i + \varepsilon_i - \beta v_i \tag{A.2}$$

In this regression model, we see that:

$$Covariance\ (x_i, \varepsilon_i - \beta v_i) = -\beta\sigma_v^2$$

such that a key assumption for OLS estimation is violated. Indeed, we have:

$$plim\ \hat{\beta}_{OLS} = \frac{Covariance\ (x_i, y_i)}{Variance\ (x_i)}$$

$$= \frac{\beta\sigma_{x^*}^2}{\sigma_{x^*}^2 + \sigma_v^2} \neq \beta$$

The unobserved measurement error biases the OLS estimator for β in (A.1). There are various possible solutions to this problem: see Koopmans (1937), Fuller (1987) and Wansbeek and Meijer (2003), among many others.

Next, consider the regression model:

$$y_i^* = \alpha + \beta x_i^* + \varepsilon_i \qquad \text{(A.3)}$$

where now also the dependent variable is measured with error, that is:

$$y_i = y_i^* + w_i,$$

with w_i having variance σ_w^2 that is independent of both v_i and ε_i. The regression model in practice would now read:

$$y_i = \alpha + \beta(x_i - v_i) + \varepsilon_i + w_i = \alpha + \beta x_i + \varepsilon_i - \beta v_i + w_i \qquad \text{(A.4)}$$

and the same expression for the $\hat{\beta}_{OLS}$ appears as above.

An alternative least squares estimator for β is the TLS estimator, which seeks to minimize the squares of the orthogonal distances to the regression line, see Koopmans (1937), Deming (1943), Linnet (1990), and Carroll and Ruppert (1996), among others. Define:

$$\delta = \frac{\sigma_\varepsilon^2 + \sigma_w^2}{\sigma_v^2} \qquad \text{(A.5)}$$

and

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i \quad \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

$$S_y = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2$$

$$S_x = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

$$S_{yx} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$$

It follows that:

$$\hat{\beta}_{TLS} = \frac{S_y - \delta S_x + \sqrt{(S_y - \delta S_x)^2 + 4\delta S_{yx}^2}}{2S_{yx}}$$

(see Deming (1943), among others). Fuller (1980, 1987) proves that the TLS estimators are consistent.

Table 1

Average of estimated parameters for 10000 replications of an AR(1) process with slope parameters 0.5, 0.8, 0.95, for sample sizes 10, 25, 50, 100, 200, 1000, 10000, 100000, with true $\boldsymbol{\delta}$ = 2

$\boldsymbol{\beta}$	**Sample size**	**TLS**	**OLS**
0.5	10	0.203	0.064
	20	0.326	0.195
	50	0.413	0.242
	100	0.456	0.264
	200	0.477	0.274
	1000	0.495	0.283
	10000	0.500	0.286
	100000	0.500	0.286
0.8	10	0.257	0.185
	20	0.579	0.286
	50	0.699	0.484
	100	0.754	0.536
	200	0.778	0.561
	1000	0.796	0.583
	10000	0.800	0.588
	100000	0.800	0.588
0.95	10	0.348	0.267
	20	0.711	0.523
	50	0.838	0.665
	100	0.898	0.758
	200	0.927	0.813
	1000	0.946	0.855
	10000	0.950	0.865
	100000	0.950	0.866

Table 2

Average of estimated parameters across 10000 replications of an AR(1) process, with parameters 0.5, 0.8, 0.95, for sample sizes 25, 50, with true $\boldsymbol{\delta}$ = 2, 11, 101

$\boldsymbol{\beta}$	$\boldsymbol{\delta}$	**Sample size**	**TLS**	**OLS**
0.5	2	20	0.326	0.195
		50	0.413	0.242
	11	20	0.389	0.361
		50	0.475	0.441
	101	20	0.396	0.393
		50	0.475	0.471
0.8	2	20	0.579	0.386
		50	0.699	0.484
	11	20	0.641	0.608
		50	0.763	0.733
	101	20	0.649	0.646
		50	0.765	0.761
0.95	2	20	0.711	0.523
		50	0.838	0.665
	11	20	0.754	0.726
		50	0.903	0.886
	101	20	0.760	0.757
		50	0.903	0.901

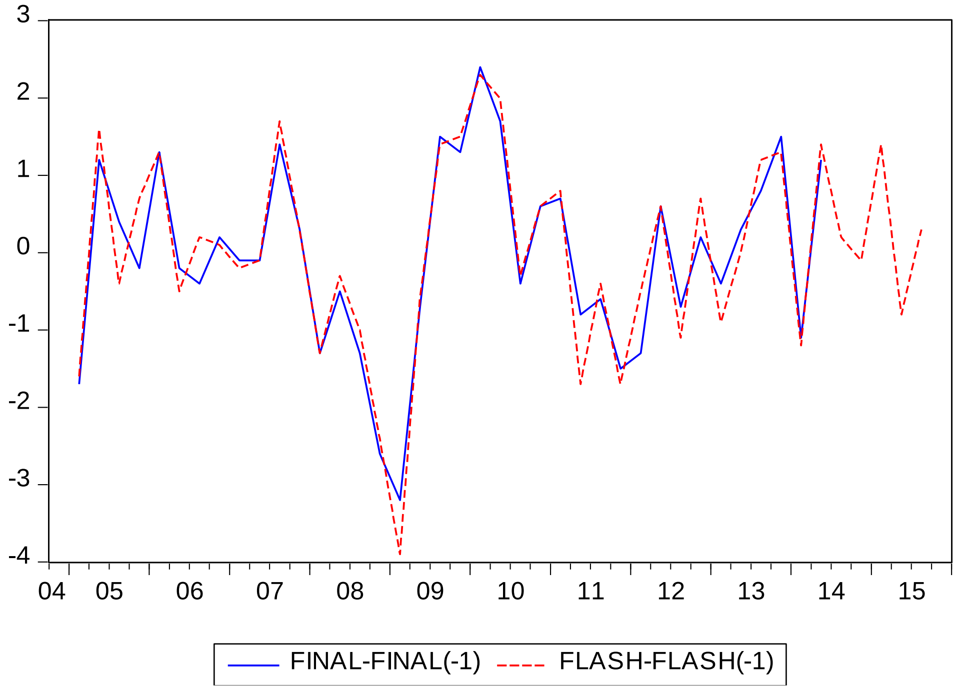

Figure 1

The first differences of the quarterly GDP growth rates, for the flash quotes (45 days after the quarter) and for the final quotes (3 years later). The sample runs from 2004Q4 to and including 2015Q3.

Source: Statistics Netherlands (retrieved November 2015)

References

Carroll, R.J. and D. Ruppert (1996), The use and misuse of orthogonal regression in linear errors-in-variables models, *The American Statistician*, 50, 1-6.

Deming, W.E. (1943), *Statistical Adjustment of Data*, London: Wiley.

Heij, C., P. de Boer, P.H. Franses, T. Kloek, and H.K. van Dijk (2004), *Econometric Methods with Application in Business and Economics*, Oxford: Oxford University Press.

Fuller, W.A. (1987), *Measurement Error Models*, New York: Wiley.

Golub, G.H and C.F. Van Loan (1980), An analysis of the Total Least Squares problem, *SIAM Journal of Numerical Analysis*, 17, 883-893.

Koopmans, T. (1937), *Linear Regression Analysis of Economic Time Series*, Haarlem the Netherlands: De Erven F. Bohn NV.

Linnet, K. (1990), Estimation of the linear relationship between the measurements of two methods with proportional errors, *Statistics in Medicine*, 9, 1463-1473.

Staudenmayer, J. and J.P. Buonaccorsi (2005), Measurement error in linear autoregressive models, *Journal of the American Statistical Association*, 100, 841-852.

Van Huffel, S., H. Park, and J.B. Rosen (1996), Formulation and solution of structured total least norm problems for parameter estimation, *IEEE Transactions of Signal Processing*, 44, 2464-2474.

Wansbeek, T. and E. Meijer (2003), Measurement error and latent variables, Chapter 8 in B.H. Baltagi (editor), *A Companion to Theoretical Econometrics*, London: Basil Blackwell.

Zeng, W., X. Fang, Y. Lin, X. Huang, and Y. Zhou (2018), On the total least-squares estimation for autoregressive model, *Survey Review*, 50 (359), 186-190.