Supplementary Material To "A Simple Method to Account for Measurement Errors in Revealed Preference Tests"
Not For Publication

Per Hjertstrand*

Research Institute of Industrial Economics (IFN)

Stockholm, Sweden

December 2014

Abstract

This supplementary material contains two parts: First, it shows how the measurement error procedure can accomodate Berkson additive and multiplicative error models as well as errors in prices. Second it presents a Monte Carlo study of the classical measurement error model.

1 Alternative measurement error models

In this section, I show how to apply Berkson error models in FW's measurement error procedure. I begin with the Berkson additive error model and then turn to Berkson multiplicative error structures. I end with a brief discussion of how to incorporate errors in prices instead of quantities. See the main paper for definitions and notation.

1.1 Berkson measurement error models

Berkson additive error structure. Suppose that the 'true' quantities $\mathbf{q}$, are predicted (or caused) by the observed quantities $\mathbf{x}$ via the following Berkson additive measurement error model:

$$\mathbf{q}_t = \mathbf{x}_t + \eta_t, \tag{BA}$$

where η_t denotes the measurement errors. In contrast to classical measurement errors, a Berkson structure thus assumes that $\mathbf{q}$ fluctuates around $\mathbf{x}$ such that $E[\mathbf{q}_t] = \mathbf{x}_t$, which means that the observed quantities predict the 'true' quantities. Given the Berkson additive error structure (BA), and by defining $\varepsilon_t = -\eta_t$ (so that $\mathbf{x}_t = \mathbf{q}_t + \varepsilon_t$), we can directly apply Theorem 3 and show that it under H_0 in (HYP) holds that $\widehat{F} \leq \max_{s,t\in T}\{\mathbf{p}_t \cdot (\eta_s - \eta_t)\}$. Assuming normally distributed errors with variance σ^2, FW's measurement error procedure takes the following steps.

FW's measurement error procedure with Berkson additive errors

*Research Institute of Industrial Economics (IFN), P.O. Box 55665, SE-102 15 Stockholm, Sweden. E-mail: Per.Hjertstrand@ifn.se.

1. *Calculate F using Algorithm 1. Denote the solution $\widehat{F}$.*

1. *Choose σ^2 and M and set $m = 0$.*

3. *Draw random numbers $\eta_{kt} \sim N\left(0, \sigma^2\right)$ for all $k \in K$ and $t \in T$.*

4. *If $\widehat{F} > \max_{s,t \in T} \{\mathbf{p}_t \cdot (\eta_s - \eta_t)\}$, then set $m = m + 1$.*

5. *Repeat steps 3 and 4 M times.*

2. H_0 in (HYP) (i.e., that $\{\mathbf{p}_t, \mathbf{q}_t\}_{t \in T}$ satisfies `GARP`) is rejected if $100 \times (m/M) > (100 - \alpha)$.

Berkson multiplicative error structure. Finally, I consider a Berkson multiplicative error structure,

$$\mathbf{q}_t = \mathbf{x}_t \odot (\mathbf{1} + \eta_t), \tag{BM}$$

where, as above, η_t denotes the measurement errors. This is the error structure considered by Varian (1985) and is the most commonly used model in empirical applications (See for example, Epstein and Yatchew, 1985; Jones and de Peretti, 2005; Elger and Jones, 2008; Hjertstrand, 2007). Assuming that (BM) holds and defining $\varepsilon_t = -\mathbf{x}_t \odot \eta_t$, we can once more apply Theorem 3 and show that it holds under H_0 in (HYP) that $\widehat{F} \leq \max_{s,t \in T} \{\mathbf{p}_t \cdot (\mathbf{x}_s \odot \eta_s - \mathbf{x}_t \odot \eta_t)\}$. Jones and Edgerton (2009, p. 225-227) provides a detailed discussion on how to implement FW's measurement procedure given the Berkson multiplicative error model (BM) with normally distributed errors. However, note that they suggest to calculate the test statistic, $\widehat{F}$, by solving (op_AI$_\mathrm{F}$) using non-linear optimization techniques, whereas I recommend calculating it using the much faster Algorithm 1.

1.2 Distribution of errors and errors in prices

An important property of the measurement errors procedures described in the previous and this section are that they can be implemented with any imaginable (parametric) error distribution. More precisely, although we have assumed normal errors, they can, in fact, take on any distribution chosen by the researcher. For example, FW assumed that the errors were uniformly distributed in their original procedure. In the empirical application in the main part of the paper I implement FW's procedure with normal, log-normal and uniformly distributed errors to investigate the robustness of the procedures with respect to different error distributions. Interestingly, I find that the procedures seem rather robust to the choice of error distribution.

Finally, I stress that the procedures described above are not pertained to errors in the quantities. Jones and Edgerton (2009, p. 219-220), for example, describe how FW's procedure can be implemented assuming a classical additive error specification in the prices, i.e., $\mathbf{p}_t = \widetilde{\mathbf{p}}_t + \varepsilon_t$, where $\widetilde{\mathbf{p}}_t$ denotes the 'true' unobserved price-data. It is straightforward to modify FW's measurement error procedure to account for errors in the prices in classical and Berkson measurement error models.

2 Monte Carlo simulations

This Monte Carlo study investigates size and power for the classical multiplicative measurement error model.

2.1 Size study

I follow Gross (1995) and Flessig and Whitney (2005) and generate data from Cobb-Douglas demand functions:

$$q_{kt} = \alpha_k \frac{\mathbf{p}_t \mathbf{q}_t}{p_{kt}}, \tag{1}$$

for $k = 1, ..., K$ and $t = 1, ..., N$, where α_k are the preference parameters (i.e., expenditure shares). Following Flessig and Whitney (2005), I use 5 goods (i.e., $K = 5$) and 40 time observations (i.e., $N = 40$). The prices, $\mathbf{p}_t$, are drawn from independent uniform distributions with support $(50, 150)$, i.e., $p_{kt} \sim U_{(50;150)}$ for all k and t, while total expenditure, $\mathbf{p}_t \mathbf{q}_t$, are drawn from independent uniform distributions with support $(10,000; 12,000)$, i.e., $\mathbf{p}_t \mathbf{q}_t \sim U_{(10,000;12,000)}$ for all t. I follow Fleissig and Whitney (2005) and use two sets of preference configurations. The first, denoted α^A, sets one expenditure share relatively large relative to the other goods, while the another, denoted α^B, sets this share relatively small:

Preference set	α_1	α_2	α_3	α_4	α_5
α_A	0.60	0.25	0.10	0.04	0.01
α_B	0.40	0.30	0.15	0.10	0.05

After having calculated the 'true' quantities from (1) I generate the 'observed' quantites from a classical multiplicative measurement error model:

$$x_{kt} = q_{kt} (1 + \varepsilon_{kt}),$$

for all k and t. Following Fleissig and Whitney (2005), the errors, ε_{kt}, are drawn from independent uniform distributions with support $(-\kappa, \kappa)$, i.e., $\varepsilon_{kt} \sim U_{(-\kappa;\kappa)}$. The scalar κ controls the amount of measurement errors. For example, measurement errors of 5% corresponds to $\kappa = 0.05$ and gives minimum and maximum error demands of 0.95 $\times q_{kt}$ and 1.05 $\times q_{kt}$. I use 1%,5% and 10% measurement errors in the simulations, i.e., $\kappa = \{0.01, 0.05, 0.1\}$.

I set the number of simulations in the calculation of the critical values to $M = 500$. All hypothesis tests are carried out on the 5% significance level. I use 1,000 replications, and the size of the test is the fraction of the 1,000 Monte Carlo replications for which the test rejects the null (HYP).

Table 1: Results from size study

Amount of assumed errors	Amount of errors in generated data					
	α_A			α_B		
	1%	5%	10%	1%	5%	10%
1%	0	0	0	0	0	0
5%	0	0	0	0	0	0
10%	0	0	0	0	0	0

The results from the size study are presented in table 1. In this table, the heading 'amount of errors in generated data' refers to the %-amount of errors in the 'observed' quantity data while the heading 'amount of assumed errors' refers to the amount of measurement errors assumed by the researcher. We see that the procedure does not reject the null of utility maximization in any case; thus, the procedure is undersized (i.e., it is conservative). This is expected and clearly in line with what the theory predicts, since the procedure is by construction conservative. In other words, according to the theory, the test should reject the null less frequently than the nominal size which is exactly what table 1 shows. It is interesting to note that the procedure does not reject even when there are only 1% errors in the data but the researcher assumes 10%. Thus, the procedure seem robust in terms of the assumed amount of errors, which is a good property.

2.2 Power study

To calculate the power of the test I follow Fleissig and Whitney (2005) and generate data from Bronars (1987) uniform random consumption model. This model has become the standard choice of irrational consumption behavior in the literature, and is widely used to calculate the power of revealed preference tests (See Andreoni, Gillen and Harbaugh, 2013, for a discussion). To generate quantities uniformly distributed over the budget hyperplane one first generates budget share uniformly distributed on the unit simplex by drawing random numbers from the Dirichlet distribution with parameters set equal to one. Denote the K random numbers drawn from the Dirichlet distribution in time t as $(D_{1t}, ..., D_{Kt}) \sim Dir\,(1, ..., 1)$. The uniformly random quantities are then calculated as:

$$x_{kt} = D_{kt}\frac{\mathbf{p}_t\mathbf{q}_t}{p_{kt}},$$

for all k and t. Prices and expenditure are generated as in the size study. As before, I use $K = 5$ goods and $N = 40$ time observations. I use $1,000$ replications, and the power of the test against uniform random behavior is the fraction of the $1,000$ Monte Carlo replications for which the test rejects the null (HYP).

Figure 1: Power curve

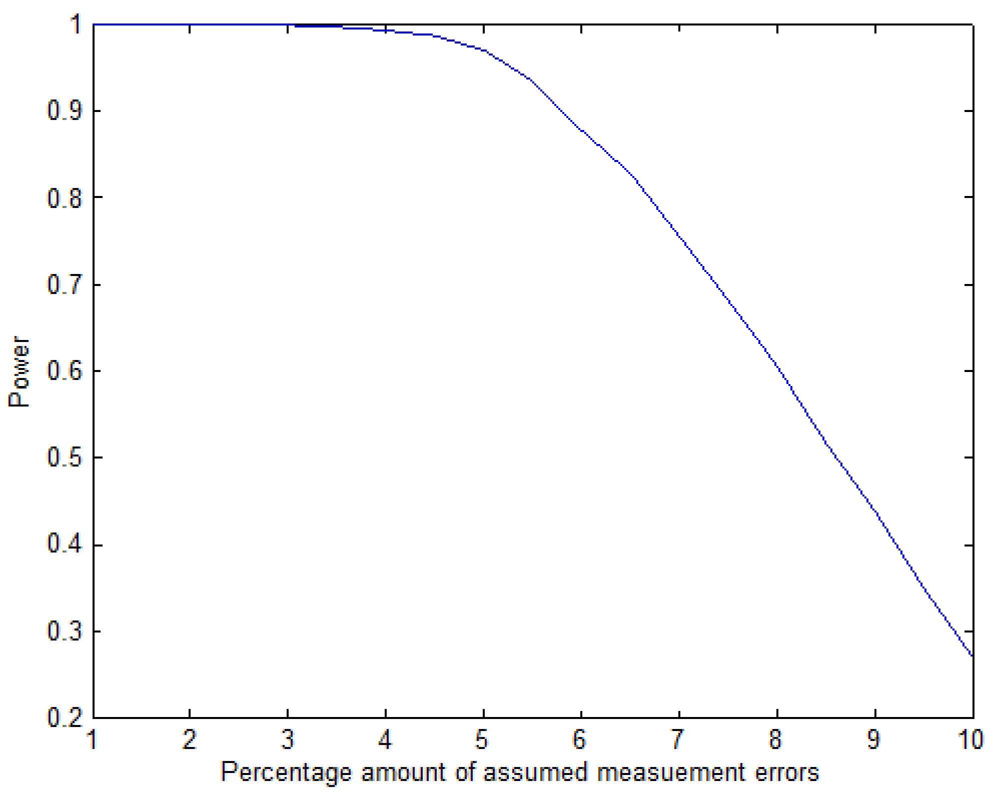

Figure 1 traces out the power curve for different amounts of measurement errors (assumed by the researcher). For example, if the researcher assumes measurement errors of 6% then the procedure has a power of about 88% against irrational behavior. In fact, the figure shows that the procedure has very good power even for rather large amounts of errors: The power is above 75% for assumed measurement errors of 7% and larger, and it does not drop below 50% until the researcher assumes errors of 9% and above. Thus, based on these results, it seem that the procedure has good power against irrational behavior even when the researcher assumes that that the data is shocked with rather large amounts of errors.

References

[1] Andreoni, J., B. Gillen, and W.T. Harbaugh (2013). The power of revealed preference tests: Ex-post evaluation of experimental design. Working paper.

[2] Bronars, S.G. (1987). The power of nonparametric tests of preference maximization. *Econometrica* **55**, 693–698.

[3] Elger, T. and B.E. Jones (2008). Can rejections of weak separability be attributed to random measurement errors in the data? *Economics Letters* **99**, 44-47.

[4] Epstein, L.G. and A.J. Yatchew (1985). Nonparametric hypothesis testing procedures and applications to demand analysis. *Journal of Econometrics* **30**, 149–169.

[5] Fleissig, A.R. and G.A. Whitney (2005). Testing the significance of violations of Afriat's inequalities. *Journal of Business and Economic Statistics* **23**, 355-362..

[6] Gross, J. (1995). Testing data for consistency with revealed preference. *The Review of Economics and Statistics* **77**, 701-710.

[7] Hjertstrand, P. (2007). Food demand in Sweden: A non-parametric approach. In W. A. Barnett & A. Serletis (Eds.), Functional Structure Inference (pp. 157–182). International Symposia in Economic Theory and Econometrics (ISETE). London: Elsevier Science.

[8] Jones, B.E. and D.L. Edgerton (2009). Testing utility maximization with measurement errors in the data. *Advances in Econometrics* **24**, *Measurement error: Consequences, Applications and Solutions*, 199-236.

[9] Jones, B.E. and P. de Peretti (2005). A comparison of two methods for testing the utility maximization hypothesis when quantity data is measured with error. *Macroeconomic Dynamics* **9**, 612-629.

[10] Varian, H.R. (1985). Non-parametric analysis of optimizing behavior with measurement error. *Journal of Econometrics* **30**, 445-458.