Are Two Cheap, Noisy Measures
Better Than One Expensive, Accurate One?

Martin Browning
University of Oxford and IFS
Department of Economics, University of Oxford, Manor Road,
Oxford OX1 3UQ, England.
Phone: +44-1865-281487.
Fax: +44-1865-271094.
Martin.Browning@economics.ox.ac.uk

Thomas F. Crossley (Corresponding Author)
University of Cambridge and Institute for Fiscal Studies
Faculty of Economics, University of Cambridge,
Cambridge CB39DD, England.
Phone: +44-1223-335225.
Fax: +44-1223-335475.
Thomas.Crossley@econ.cam.ac.uk

Session: Data Watch
Session Chair: Martin Browning
Discussant: James Smith

Are Two Cheap, Noisy Measures Better Than One Expensive, Accurate One?

Martin Browning

Thomas Crossley[*]

Survey responses are always subject to measurement error. This is the case even for such well defined concepts as age, earnings, expenditures, net wealth and market hours; henceforth the *target variable*. In general surveys (and especially longitudinal surveys), there are severe constraints on the time that can be spent eliciting a less noisy response for any target variable. In this paper we consider when it may be better to consider multiple noisy measures of the target measure rather than improving the reliability of a single measure.

The value of multiple measures for means and regression coefficients is familiar to most empirical economists (for example, in twins studies). However the use of multiple measures is much more general and can aid in recovering the full (conditional) distribution of the target variable. Our inspiration in the analysis below is the Kotlarski result (see B.L.S Praskasa Rao, 1992). This states that if the measurement errors in two measures of the same target variable are mutually independent and independent of the true value then we can recover the entire distribution of the quantity of interest, up to location.

The Kotlarski result has been used in recent contributions for dealing with measurement error (see Tong Li and Quang Vuong, 1998, Susan M. Schennach, 2004, and Aurore Delaigle, Peter Hall and Alexander Meister, 2008). These papers follow the standard route of taking measurement errors with specific properties and then devising estimators that can take account of such measurement errors. Our interest is in survey design. Consequently we propose turning the usual procedure on its head and designing surveys to deliver measurement error with desirable properties. As we shall see, the emphasis then shifts from reliability (the signal to noise ratio for any given measure) to the joint properties of the multiple measures.[1] Using an illustration of asking about total expenditure, we shall show

how a mixture of economic theory and analysis of auxiliary data sets can provide insights into the design of survey questions. Although we do not consider it in this paper, this analysis also suggests complementary pre-testing and use of focus groups to further enhance the utility of the survey questions.

I. An example.

To make things concrete, we consider a specific measurement problem: estimating the variance of the log of consumption (total expenditure) in a population. There is a large literature that investigates consumption inequality (David Cutler and Lawrence Katz, 1992, Richard Blundell and Ian Preston, 1998, Dirk Krueger and Fabrizio Perri, 2006) and the log of the variance is common measure of inequality in this literature. Our choice of an inequality measure as our parameter of interest is intended to reinforce the point that the ideas sketched in this paper are not limited to the estimation of means.

Let C be (true) log consumption with variance σ_c^2; σ_c^2 is the statistic of interest. We suppose that we have three potential measures: Z, X_1 and X_2. Z is the 'expensive' measure and X_1 and X_2 are the two 'noisy' measures. Let c, z, x_1 and x_2 be deviations from the respective means. *Define* the error associated with each method as:

(1) $$u = z - c$$

(2) $$\varepsilon_i = x_i - c$$

These errors have variances σ_u^2, σ_1^2 and σ_2^2 and covariances with true consumption of σ_{cu}, σ_{c1} and σ_{c2}. The quantities $\frac{\sigma_c^2}{\sigma_z^2}$ and $\frac{\sigma_c^2}{\sigma_{x_i}^2}$ are usually referred to as the 'reliability' of z and x_i respectively.

The variance of the single measure z is given by:

(3) $$E[z^2] = \sigma_c^2 + 2\sigma_{cu} + \sigma_u^2$$

We take the sample analogue of the variance $E[z^2]$ as one estimate of σ_c^2. Improving the precision of this estimate involves reducing σ_{cu} and/or σ_u^2. For example, expenditures on individual items might be collected using diaries kept for one week. This delivers high quality data but at high cost and with substantial respondent burden. The quality of the data (as well as the cost and the respondent burden) would be increased by asking respondent to fill out the diary for longer than one week. Alternatively z might be the sum of a detailed list of recall questions. Asking more (and finer) categories of expenditure leads to better data, (Joachim Winter, 2004, Menno Pradham, 2001) but this requires more interview time, and thus implies greater cost and respondent burden. The 'cheap' measures (x_1 and x_2) might be a 'one shot' recall question about total expenditure and a recall question about a single category of expenditure (food at home, for example).

When we consider using two measures to estimate the variance of c we take the covariance between the two measures:

$$(4) \qquad E[x_1 x_2] = \sigma_c^2 + \sigma_{c1} + \sigma_{c2} + \sigma_{12}$$

Then the estimate of σ_c^2 is given by the sample estimate of the covariance between the two measures:

$$(5) \quad \widehat{E[x_1 x_2]} = \frac{1}{n}\sum_{i=1}^{n} x_{1i} x_{2i}$$

Note that the asymptotic bias $\left(E[x_1 x_2] - \sigma_c^2\right)$ does not depend on the individual variances σ_1^2 and σ_2^2 and is hence not dependent on the reliability of the multiple measures. It is in this sense that we can allow them to be 'noisy' measures.[2] Our interest is in designing measures that reduce the asymptotic bias, $\sigma_{c1} + \sigma_{c2} + \sigma_{12}$.

A very common assumption for measurement errors is that they have *classical* properties. Although this term is used to denote different things for different people, in the current context the classical assumptions are $\sigma_{cu} = 0$ and $\sigma_{c1} = \sigma_{c2} = \sigma_{12} = 0$. In this case the asymptotic bias of the single measure is given by:

(6) $E[z^2] - \sigma_c^2 = \sigma_u^2$

This only reduces to zero if we eliminate measurement error altogether. On the other hand, if we assume classical measurement errors for the multiple measures then the bias $\left(E[x_1 x_2] - \sigma_c^2\right)$ is zero. This is an illustration of the Kotlarski result in a very specific context.[3] The classical assumptions are widely invoked in psychometrics and other disciplines and hence there is a widespread use in these disciplines of multiple measures.. In many contexts in economics, however, the assumption that measurement errors are independent of each other and of the true value are more expressions of hope than realistic assessments (John Bound, Charles Brown and Nancy Mathiowetz, 2001). In the next section we shall consider how to craft multiple measures which at least reduce the bias.

II. Multiple measures with nonclassical measurement error.

As we have seen, if we assume classical properties for the measurement errors in our multiple measures then it would always be best to use two (or more) noisy measures rather than one less noisy one. We now consider how we might design multiple survey questions for a single target variable that induce measurement errors that come close to the classical assumptions. Any such exercise will be very specific to the target variable and will ideally involve extensive and judicious pre-testing and use of focus groups as well as analysis of other data sources and considerations from economic theory.

Consider again the basic set up:

(7) $$E[z^2] = \sigma_c^2 + 2\sigma_{cu} + \sigma_u^2$$

(8) $$E[x_1 x_2] = \sigma_c^2 + \sigma_{c1} + \sigma_{c2} + \sigma_{12}$$

Both cognitive theories of response behaviour and economic theory can be informative about the sources of bias (σ_{cu}, σ_u^2, σ_{c1}, σ_{c2} and σ_{12}). This can help us to design good measures (questions) or choose wisely from an available set of measures.

III. Multiple measures for total expenditure.

We consider again the issue of finding out about total expenditure in a given period. There is considerable evidence that well informed respondents in a household can provide reasonably accurate recall information about expenditures on specific groups of goods (such as food at home) in, say, the last month (Naeem Ahmed, Matthew Brzozowski and Thomas F. Crossley, 2006.) Suppose that we only have survey time to ask about a small number of items; which should we ask about? Martin Browning, Thomas F. Crossley and Guglielmo Weber (2003) recommended:

- *Always ask a 'total expenditure on non-durables and services' question…..there is a great deal of idiosyncratic behaviour in demand and sometimes households spend a good deal on sub-items that we would never think to ask about……*
- *Always ask a 'food at home' and a 'food outside the home' question with the same time period as for total expenditure….. respondents can report food at home accurately….being a large budget item, it is very useful in imputation…….*
- *Ask about utilities such as fuel and telephones…..*

Subsequently, evidence has piled up that the first 'one-shot' question is very unreliable and takes a lot of survey time. Moreover, recent cognitive testing we have undertaken was particularly discouraging for this question. The idea behind the third recommendation was that these items could be measured reliably and contained variation that was orthogonal to food in/out. In making these recommendations, we very much had in mind to capture a large share of the total and/or a 'prediction' approach (Jonathan Skinner, 1987).

The multiple measures analysis above suggests a quite different approach. Think of log consumption of specific goods (food, clothing, telephone, recreation) as our cheap error ridden measures (x_1, x_2) of total consumption. We then use demand theory and analysis of expenditure surveys to choose goods so that the measurement errors have desirable properties. An Engel curve relates consumption on specific items to the target variable, total expenditure. Consider a linear in logs approximate Engel curve:

(9) $x_i = \alpha_i c + \eta_i(c) + e_i$

The parameter α_i captures the income elasticity of good i if the double log form is correct; luxuries have $\alpha_i > 1$ and necessities have $\alpha_i < 1$. The variable $\eta_i(c)$ is the approximation error from using the log-log form. The variable e_i captures heterogeneity in tastes. Define the measurement error for good i as:

(10) $$\varepsilon_i \equiv x_i - c = (a_i - 1)c + \eta_i(c) + e_i$$

This allows us to relate economic theory concepts to the decomposition given in equation (8). Thus $\sigma_{ci} \approx 0$ for goods with unit income elasticity and a small log-log approximation error ($a_i \approx 1$ and $\eta_i(c) \approx 0$), with $\sigma_{ci} > 0$ for luxuries and $\sigma_{ci} < 0$ for necessities. Equation (8) then implies that it might be better to have one luxury and one necessity rather than either two luxuries or two necessities. In terms of σ_{12}, the best choice gives $\sigma_{12} \simeq 0$. Complementary goods (coffee and cream) will tend to have $\sigma_{12} > 1$ and substitutes (coffee and tea) will tend to have $\sigma_{12} < 0$. Adding up implies $\sigma_{12} < 0$ on average, especially for highly aggregated goods.

Based on this, we can make the following recommendations for the choice of two goods to ask about. We should choose goods that:

- Respondents can readily report,
- have close to unit income elasticities (or a luxury and a necessity), and not too much approximation error,
- are not strong complements or substitutes.

Note that large budget shares not necessary (and may be undesirable since adding up induces $\sigma_{12} < 0$). Moreover, reliability (a low variance for measurement error) is helpful but not paramount, in contrast to a single measure approach.

IV. A Simulation Study using the Canadian FAMEX.

To illustrate the ideas described above, we conducted a small simulation study. This experiment is based on data from the 1996 Canadian Family Expenditure Survey (FAMEX). This is an intensive, high quality budget survey based on annual recall. The nature of this survey is attractive for our purposes for two reasons. First, the responses do not suffer from infrequency. Second, recall questions are what we imagine are feasible in a general survey. From this data we selected a sample of couples without children. We treat this sample as our population of interest and take the logarithm of total nondurable consumption as the "true" value of the target variable (c) for each household. The variance of the logarithm of total nondurable consumption in this sample is then the population parameter we wish to estimate. We consider two possibilities for z (the high cost measure). One is simply the logarithm of total nondurable consumption ($z = c$) as we observe it in the data (this corresponds to observing the target variable with reliability of one). Alternatively, we take the logarithm of total nondurable consumption as observed in the data and add classical measurement so that the reliability of the measure is 0.6. For (x_1, x_2) we take pairs of goods guided by the advice summarized in the previous section. We chose to use food (the sum of food at home and food in restaurants) as one of the two "cheap" measures (x_1) both because of the evidence (noted above) that it is well measured by recall questions and because it seems likely that questions about food will always be included in surveys that collect expenditure information. The choice of the second good was informed by subsidiary analysis of Engel curves estimated on the FAMEX budget data. On the basis of fit, income elasticity and error correlations, we concluded that recreation (or "leisure") nondurable/semidurable goods and services would be a good choice for the second good (x_2). Evidence from Denmark (Jens Bonke and Martin Browning, 200X) suggests that households can and will answer recall questions on this category of expenditure. Finally, we consider another option. Following Richard Blundell, Luigi Pistaferri and Ian Preston (2004) we also estimate the logarithm of total nondurable consumption by the inverse food Engel curve: $\widehat{\left(1/\alpha_f \right)} x_f$.

These measures of the logarithm of total nondurable consumption then imply the following estimators of the variance of the logarithm of nondurable consumption:

i) The sample variance of of z

ii) The sample variance of $\widehat{\left(1/\alpha_f \right)} x_f$

iii) The sample covariance of (x_1, x_2)

We resample repeatedly from our sub-sample of the FAMEX (making 1000 draws with replacement) and study the bias and variance of our estimators. The results are presented in Table 1 and Figure 1. The 'true' value of the variance of the logarithm of total nondurable expenditure in our 'population' (the initial sub-sample of the 1996 FAMEX) is 0.189. The first row of Table 1 indicates that if we observe the target variable with reliability 1, and simple calculate the sample variance of the logarithm of total nondurable expenditure, our estimates are, of course, centred on the true value (these estimates differ from the true value only because of sampling variability.) Note however, that when the reliability of our 'expensive' measure falls to 0.6 (row 2), substantial bias is introduced in our estimate of variance. This simply illustrates Equation (6): even classical measurement error biases the traditional estimator of the variance. The next row of Table 1 (Row 3) reports the result of estimating the variance of the logarithm of total nondurable expenditure with the sample variance of an imputation of the target variable. The imputation is from food expenditure, via an estimated Engel curve. This estimate of the variance is biased as well, for the same reason as the estimator in the row above: the imputation measures the target variable with error, and those measurement errors inflate the sample variance of the imputation. Blundell, Pistaferri and Preston (2004) provide evidence that this bias may be reasonably constant through time, so that changes in the variance can be recovered in this way, although the level of the variance cannot.

Finally, Row 4 of Table 1 reports the distribution of the estimate based on the sample covariance of two ‘cheap and noisy’ measures: in this case the logarithm of expenditure on food and the logarithm of expenditure on recreation. This estimator does very well. In our re-sampling experiment, it is centred on the true value of the parameter of interest, and not too widely dispersed. Thus our choice of goods has delivered two noisy measures whose respective measurement errors have the properties necessary to allow a good estimate of the parameter of interest.

TABLE 1: Simulation Results			
Estimator	Mean	Std. Dev.	Mean % Bias
Direct, Reliability = 1	0.189	0.0058	-0.11
Direct, Reliability = 0.6	0.320	0.0092	64.4
BPP (inverse Engel curve imputation)	0.328	0.0107	73.1
2 Good (food, recreation)	0.188	0.0102	-0. 87
True value = 0.189 1000 replications (re-sampling with replacement) Population size = sample size = 2379			

V. Discussion

Designing survey questions to eliminate measurement error is very difficult – perhaps impossible. However, with the right kind of measurement errors, two error ridden measures can tell you a lot about the distribution of a quantity of interest. Our suggestion is therefore that it may be easier to design survey questions to get close (or closer) to the right kinds of measurement error.

Going forward, our research agenda is to investigate ways in we can introduce multiple measures of target variables into household surveys, and in particular, to investigate ways in which we might induce the errors in those measures to have desirable properties. Internet panels have recently been developed in both the U.S. and Europe to support social science research. These seem the natural platform on which to further explore the suggestion made in this short paper.

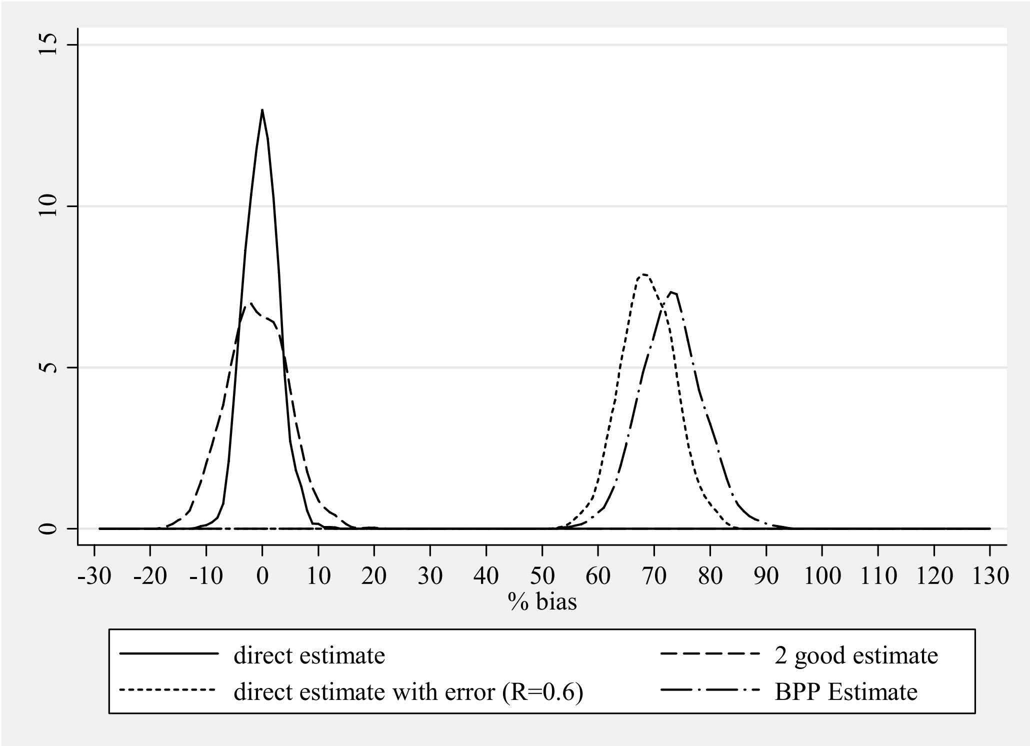

Figure 1: Resampling Distribution of Alternative Estimators

REFERENCES

Ahmed, Naeem, Matthew Brzozowski and Thomas F. Crossley. 2006. "Measurement Errors in Recall Food Consumption Data," The Institute for Fiscal Studies, WP06/21.

Battistin, Erich, Richard Blundell and Arthur Lewbel. 2007. "Why is Consumption More Log Normal than Income," The Institute for Fiscal Studies, WP08/07.

Blundell, Richard and Ian Preston. 1998. "Consumption Inequality and Income Uncertainty." *Quarterly Journal of Economics,* 113(2):603-40.

Blundell, Richard, Luigi Pistaferri and Ian Preston. 2004. "Imputing Consumption in the PSID Using Food Demand Estimates from the CEX," The Institute for Fiscal Studies, WP04/27.

Bonke, Jens and Martin Browning.

Bound, John, Charles Brown and Nancy Mathiowetz. 2001. "Measurement Error in Survey Data." In *Handbook of Econometrics Volume 5,* ed. Heckman, James J. and Edward E. Leamer, XXXX:Elsevier Science.

Browning, Martin, Thomas F. Crossley and Guglielmo Weber. 2003. "Asking Consumption Questions in General Purpose Surveys." *Economic Journal,* 113:F540-F567.

Cutler, David and Lawrence Katz 1992. "Rising Inequality? Changes in the Distribution of Income and Consumption in the 1980s." *American Economic Review*, *Papers and Proceedings*, 82(2):546-51.

Delaigle, Aurore, Peter Hall and Alexander Meister. 2008. "On Deconvolution with Repeated Measurements." *The Annals of Statistics*, 36(2):665–685.

Krueger, Dirk and Fabrizio Perri. 2006. "Does Income Inequality Lead to Consumption Inequality? Evidence and Theory." *Review of Economic Studies*, 73:163-193.

Li, Tong and Quang Vuong. 1998. "Nonparametric Estimation of the Measurement Error Model Using Multiple Indicators." *Journal of Multivariate Analysis*, 65:139-165.

Pradhan, Menno. 2001. "Welfare Analysis with a Proxy Consumption Measure: Evidence from a repeated experiment in Indonesia." Cornell Food and Nutrition Policy Program Working Paper No. 126.

Rao, B.L.S. Prakasa. 1992. *Identifiability in Stochastic Models*. New York: Academic Press.

Schennach, Susan M. 2004. "Estimation of Nonlinear Models with Measurement Error." *Econometrica*, 72(1):33–75.

Skinner, Jonathan. 1987. "A Superior Measure of Consumption from the Panel Study of Income Dynamics." *Economic Letters*, 23(2):213–6.

Winter, Joachim. 2004. "Response Bias in Survey-based Measures of Household Consumption." *Economics Bulletin,* 3(9):1−12

* Martin Browning. University of Oxford and Institute for Fiscal Studies. Department of Economics, University of Oxford, Manor Road, Oxford OX1 3UQ, England. Phone: +44-1865-281487. Fax: +44-1865-271094. Martin.Browning@economics.ox.ac.uk

Thomas F. Crossley, University of Cambridge and Institute for Fiscal Studies. Faculty of Economics, University of Cambridge, Cambridge CB3 9DD, England. Phone: +44-1223-335251. Fax: +44-1223-335475. Thomas.Crossley@econ.cam.ac.uk

The authors are grateful to the ESRC-funded Centre for Microeconomic Analysis of Public Policy at the Institute for Fiscal Studies (grant number RES-544-28-5001) for funding this work.

[1] An alternative approach to designing surveys to allow tractable methods for dealing with measurement error is given in Chen, Hong and Tamer, 2005. This involves using a validation subsample.

[2] In a finite sample the noise will matter for the precision of the estimator.

[3] This is an exact result if everything is lognormal.