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

Martin Browning
University of Oxford and Institute for Fiscal Studies

Thomas F. Crossley
University of Cambridge and Institute for Fiscal Studies

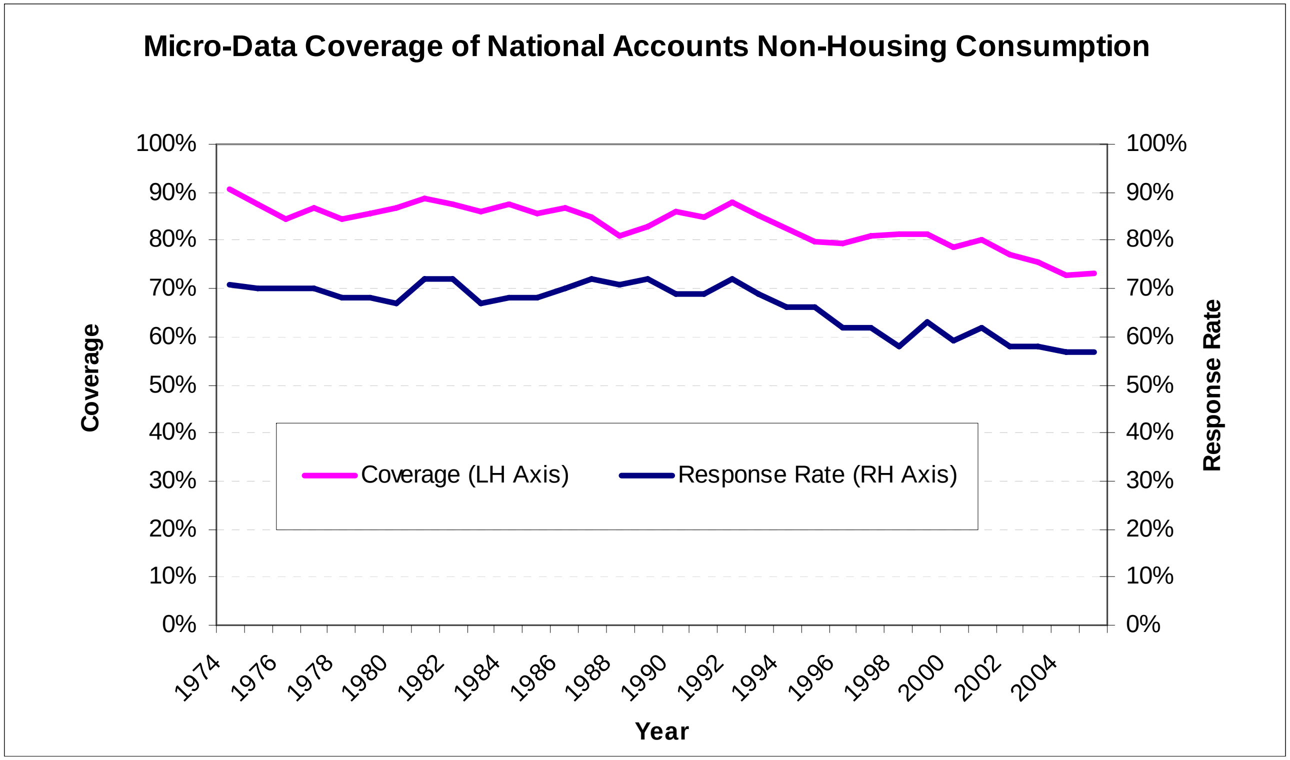
March 2009

1. Motivation

Measurement error in surveys

- Survey responses are always subject to measurement error
- Even true for well defined quantities (eg., age, earnings, expenditure, not health)
- In general surveys (and especially longitudinal surveys), severe constraints on time that can be spent measuring any target variable

- Some important variables getting harder to measure, even by intensive methods



The value of multiple measures

- The value of multiple measures for means, regression coefficients, is familiar to most economists (eg., twin studies).
- The value of multiple measures is much more general.
- If the measurement error in two measures is mutual independent and independent of the true value, we can recover the entire distribution of the quantity of interest, up to location (Kotlarski).
- Basis for some recent econometrics.

Our Suggestion

- Rather than design estimators for given measurement error characteristics, design surveys to deliver measurement error with desirable properties.
- May sometimes be better to go for two `noisy' measures than attempting to increase the accuracy and precision of a single measure.

2. An Example

- To make things concrete, we consider a specific problem: estimating the population variance of the log of consumption.
- There is a large literature on consumption inequality (see for example Cutler and Katz (1992), Slesnick (1993), Blundell and Preston (1998), Krueger, D. and F. Perri, (2006).)
- The variance of the log is common measure of inequality.

- Let C be (true) log consumption with variance σ_c^2. (σ_c^2 is what we want to estimate.)
- We conceive of three measures: Z, X_1 and X_2. Z is the `expensive` measure and X_1 and X_2 are the `noisy' measures.
- Let c, z, x_1 and x_2 be de-meaned versions.

- *Define* the error associated with each method as (note no loss of generality – additivity here has no content, will be the independence etc. that matters):

(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}.

- While more expensive, z is also more precise ($\sigma_u^2 < \sigma_i^2, \quad i = 1,2$)
- We imagine that σ_u^2 (and possibly σ_{cu}) can be reduced, but at greater cost and respondent burden.
    - Longer diary period.
    - More (and finer) expenditure categories for recall questions.

- The `cheap' measures ($x_i$) might be a `one shot' recall question about total expenditure, or a recall question about a single category of expenditure (food, or other).
- The quantities $\frac{\sigma_c^2}{\sigma_z^2}$ and $\frac{\sigma_c^2}{\sigma_{x_i}^2}$ are the reliability of z and x_i respectively.

- The single measure z has variance:

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

- Thus if we use the sample variance of z as an estimate of σ_c^2, the asymptotic bias is $E[z^2] - \sigma_c^2 = 2\sigma_{cu} + \sigma_u^2$.

- The two measures (x_1, x_2) have covariance:

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

- If we use the sample covariance of (x_1, x_2) as estimate of σ_c^2, the asymptotic bias is $E[x_1 x_2] - \sigma_c^2 = \sigma_{c1} + \sigma_{c2} + \sigma_{12}$.

- Note that this does not depend on σ_1^2 and σ_2^2 (and hence on the reliability of the noisy measures.)

- Suppose measurement errors are ***classical*** (independent of the true value and each other.) Equations (3) and (4) reduce to:

(5) $E[z^2] = \sigma_c^2 + \sigma_u^2$

(6) $E[x_1 x_2] = \sigma_c^2$

- $\frac{1}{N}\sum x_1 x_2$ is a consistent estimator of σ_c^2, *regardless* of the reliability of these measures. Of course, the precision of this estimator depends on the reliability of two measures.

- Measurement errors are typically ***non-classical.***

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

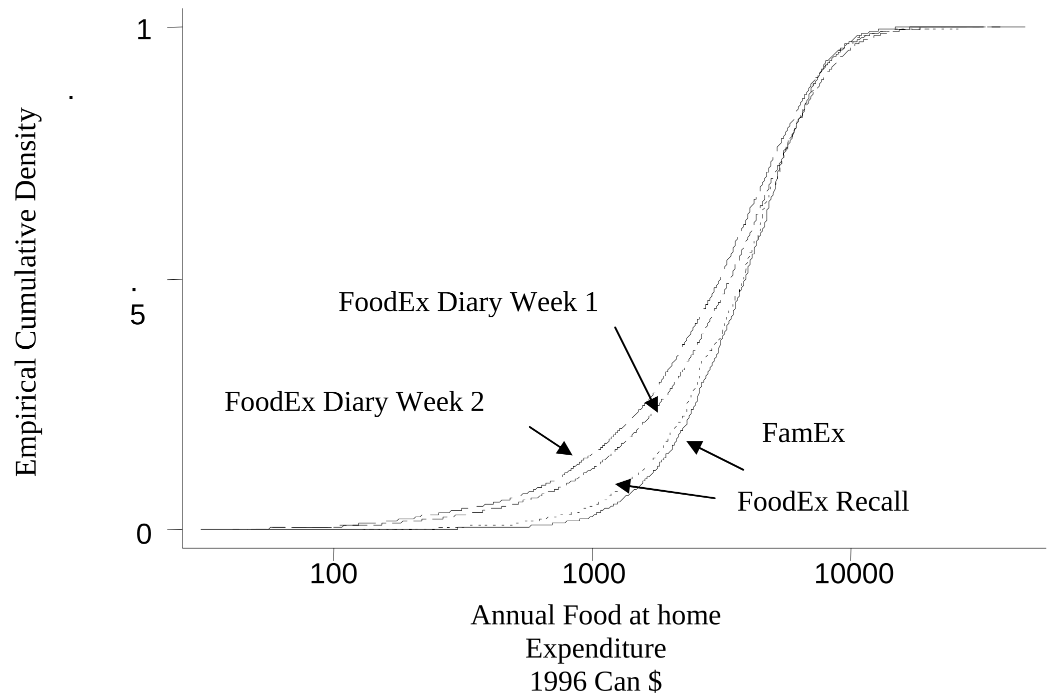
(8) $$E[x_1 x_2] = \sigma_c^2 + \sigma_{c1} + \sigma_{c2} + \sigma_{12}.$$

- Cognitive theories of response behaviour, economic theory, and pre-testing can be informative about the sources of bias (σ_{cu}, σ_u^2, σ_{c1}, σ_{c2} and σ_{12})
- Help us to design better questions/choose wisely from available measures

Measuring Consumption (Inequality) With a Subset of Goods

- Survey respondents seem able to answer recall questions about some categories of expenditure (eg. food.)
- If we can only ask about a subset of goods, which ones should we ask?

Food Expenditure, Empirical CDFs



Source: Ahmad, Brzozowski and Crossley, 2006

Here is what we said in 2003 (Browning, Crossley, Weber, *EJ*):

- *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…..*

- The idea here was that these could be measured reliably and contained variation that was orthogonal to food in/out. We very much had in mind to capture a large share of the total and/or a "prediction" approach (Skinner, 1987).

Some doubts:

- The `one-shot' total expenditure question performed poorly in recent experiments for the UKHLS.
  - Good response rate (>10%)
  - Significant under-reporting (~40%) relative to EFS (previous experience of about 30% in Canada and Italy).
  - Cognitive testing was particularly discouraging.

- Multiple measure framework suggests a different approach:
- Think of log consumption of specific goods (food, clothing, telephone, recreation) as our cheap error ridden measures (x_1, x_2)
- Choose goods so that the measurement errors have desirable properties.
- Use theoretical and empirical knowledge of consumer behaviour to choose.

- An Engel curve relates consumption of specific items to total consumption.

- Consider a linear in logs approximate Engel curve:

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

Heterogeneity

Approximation Error

Income Elasticity

- Now define the measurement error in each measure as:

$$\begin{aligned}\varepsilon_i &= x_i - c \\ &= \alpha_i c + \eta_i(c) + e_i - c \\ &= (a_i - 1)c + \eta_i(c) + e_i\end{aligned} \qquad (10)$$

- This allows us to relate properties of the measurement errors to theoretical and empirical knowledge of demands.

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

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

- Consider:
  - Income elasticities (luxury/necessity)
  - Substitute compliment
  - Adding up
  - Approximation error

- *Therefore*, choose goods that:

   i. Respondents can readily report,
   ii. Have close to unit income elasticities (or a luxury and a necessity), and not too much approximation error,
   iii. Are not strong complements or substitutes,

- Note that:
  - large shares not necessary (and perhaps undesirable), and
  - reliability helpful but not paramount (in contrast to a single measure approach.)

3. Simulation Study

- 1996 Canadian FAMEX:
- Intensive, high quality, annual recall (little infrequency)
- Sample of couples without children
- Take the log of total nondurable consumption as “true” target variable (c) for each household.
- Take the sample variance of log total nondurable consumption as the population parameter we with to estimate.

- For z we take total nondurable consumption ($z = c$), or total nondurable consumption with classical measurement error added ($z = c + \upsilon$).
- For (x_1, x_2) we take pairs of goods guided by the above advice.
- Following BPP 2004 we also consider consumption imputed by the (inverted) food Engel curve: $\widehat{\left(1/\alpha_f\right)} x_f$.

- We resample repeatedly from the Famex (1000 draws with replacement)
- Estimators:

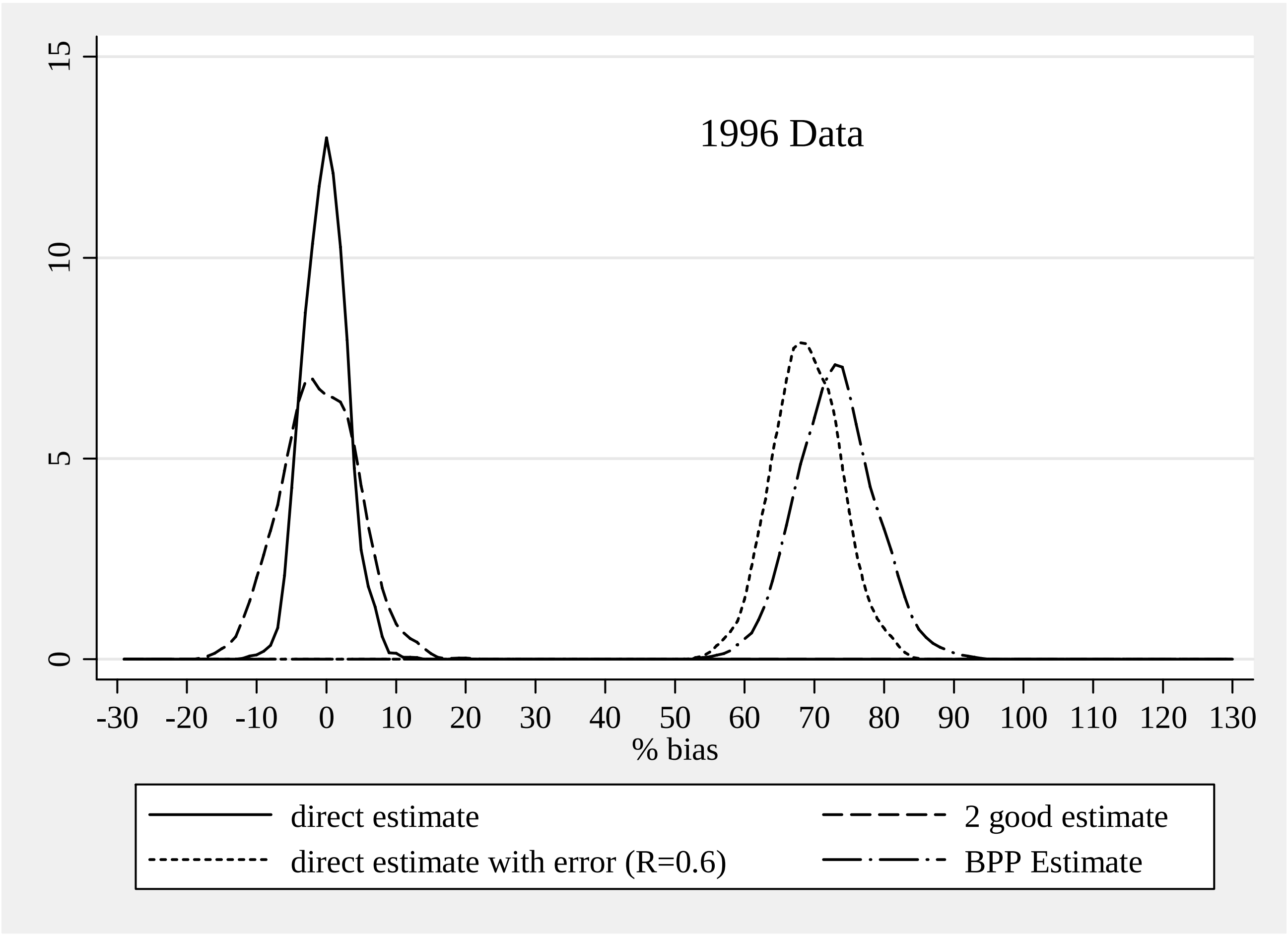
  i) The sample variance of of z

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

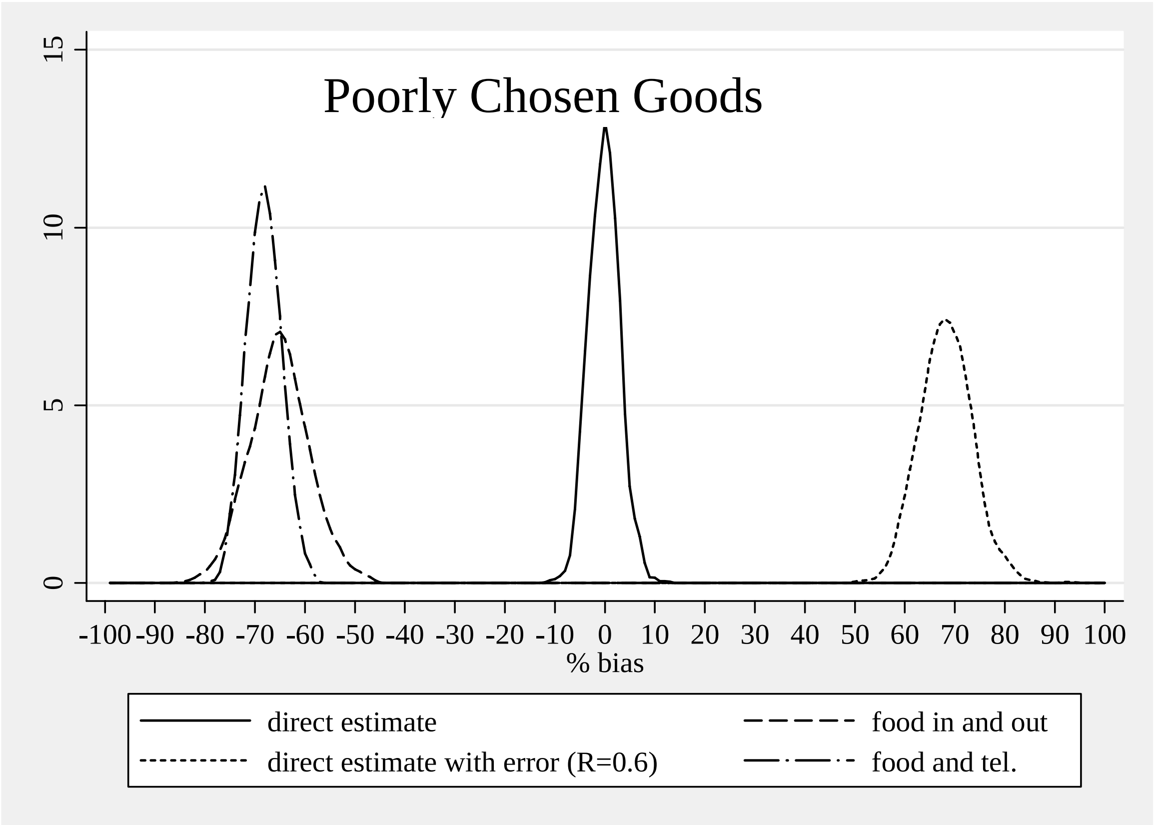
  iii) The sample covariance of (x_1, x_2)

- We study the bias and variance of our estimators.

- Use food as x_1 (seems well measured and likely always to be included in surveys that collect expenditure information.)
- Preliminary analysis: estimated Engel curves. Examine fit, income elasticities and error correlations.
- Suggests using recreation (or “leisure”) nondurable/semidurable goods and services as x_2.
- Danish evidence suggests that recall questions about recreation work well (Browning and Gortz, 2006).


1996 Data
0
5
10
15
-30 -20 -10 0 10 20 30 40 50 60 70 80 90 100 110 120 130
% bias
direct estimate
2 good estimate
direct estimate with error (R=0.6)
BPP Estimate

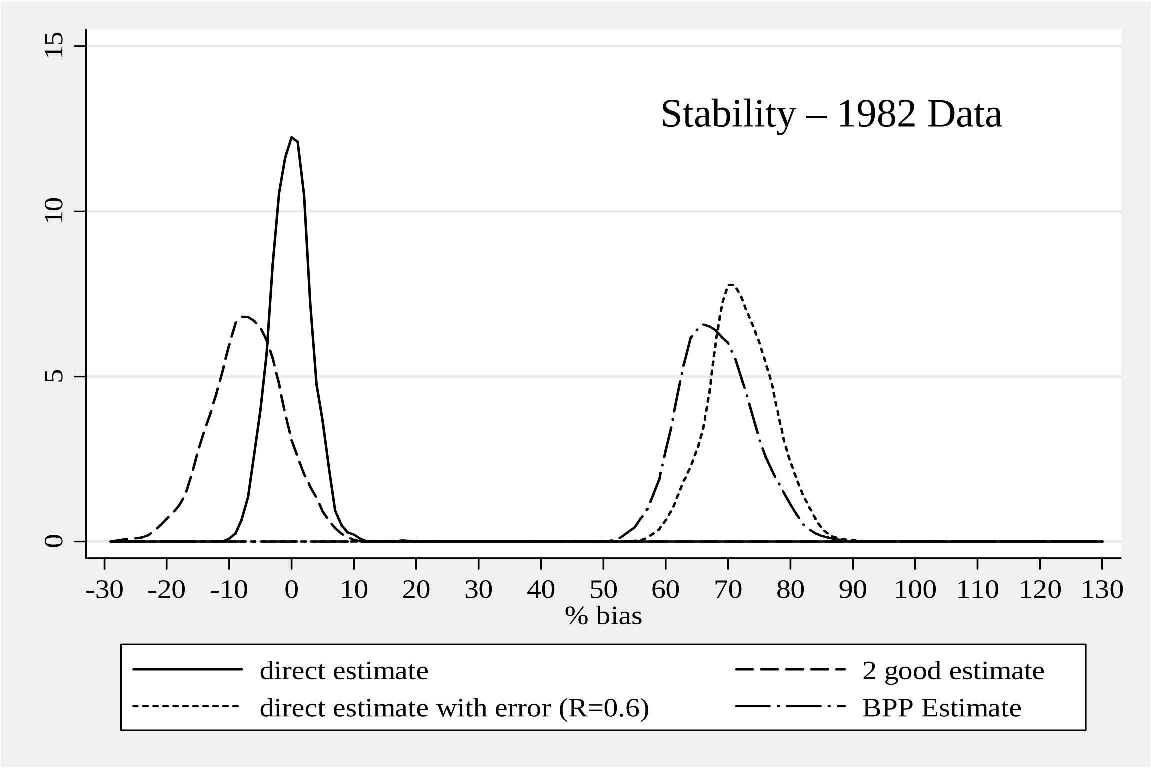
Estimator	Mean Estimate	Standard Deviation	95^{th} - 5^{th}	Mean % Bias
Direct, R=1	0.189	0.0058	0.199 - 0.180 = 0.019	-0.11
2 Good (food, recreation)	0.188	0.0102	0.204 - 0.171 = 0.033	-0. 87
Direct, R=0.6 (classical ME)	0.320	0.0092	0.335 - 0.305 = 0.030	64.4
BPP (inverse Engel curve imputation)	0.328	0.0107	0.345-0.310 = 0.035	73.1
True value = 0.189 1000 replications Population size = sample size = 2379				



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

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

- Note that approach is not completely general if $\alpha \neq 1$
- Also worry about stability in face of changing relative prices
- One solution is to combine the BPP approach (inverting Engel curves) with a 2 measure approach. But this requires more information.



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

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

IV. Discussion

- Designing survey questions to eliminate measurement error is very difficult (impossible?)
- With the right kind of measurement errors, two error ridden measures can tell you a lot about the distribution of a quantity of interest – indeed one can recover the entire distribution.
- Maybe it is (relatively) easier to design survey questions to get close (or closer) to the right kinds of measurement error.

- Research agenda: How can we get multiple measures of quantities of interest in household surveys?
    - Multiple proxies (two goods as measures for total nondurable consumption)?
    - Ask the same thing two different ways (the sum of expenditure in different categories, then, income minus saving)?
    - Ask multiple household members?

Guidance from: survey response theory, economics (theory and empirical evidence) and pre-testing.