Mismeasured Household Size and Its Implications for the Identification of Economies of Scale: On-Line Appendix

Timothy J. Halliday*

University of Hawai'i at Mānoa

November 8, 2012

Abstract

In this note, we correct an error from Halliday (2010) in which we explored the degree to which household size is measured with errors. After our correction, we find that an upper bound on the variance of these measurement errors increases by a factor of four. In fact, our calculations suggest that these errors could be as high as 20% of the total variation in household size. However,

*E-mail: halliday@hawaii.edu. Address: 2424 Maile Way; Saunders Hall 533; Honolulu, HI 96822. Tele: (808) 956-8615. All errors are my own.

this dramatic increase is still not enough to resolve the puzzle posed in Deaton and Paxson (1998).

JEL Classification: J12, C14

Keywords: migration, measurement error, semi-parametric bounds, economies of scale

1 The Empirical Model

We begin by restating some of the basics from Section 7 of the original paper and correcting a few typographical errors. The base empirical model is

$$\omega_f^* = \alpha + \beta\left(x^* - n^*\right) + \gamma n^* + \varepsilon^*.$$

This is the true model in which the residual, ε^*, is orthogonal to the right-hand side variables: $x^* - n^*$, log per capita expenditures, and n^*, log household size. Household size is measured with error *i.e.* $n = n^* - e$ where $E[e] = 0$. Because we observe n and not n^*, from the perspective of the econometrician, the model is

$$\omega_f^* = \alpha + \beta\left(x^* - n\right) + \gamma n + \upsilon$$

where the residual is

$$\upsilon = \varepsilon^* + (\gamma - \beta)\, e.$$

This is not a linear projection and so OLS will deliver biased estimates.

To derive the bias, we project the measurement error onto the right-hand side regressors and obtain

$$e = \kappa + \phi\,(x^* - n) + \lambda n + u. \tag{1}$$

Note that the published version used x^* in lieu of $x^* - n$ in equation (1) which was a typographical error. This then implies that

$$\omega_f^* = \widetilde{\alpha} + \widetilde{\beta}\,(x^* - n) + \widetilde{\gamma} n + \widetilde{\upsilon} \tag{2}$$

where $\widetilde{\upsilon} \equiv \varepsilon^* + (\gamma - \beta)\, u$, $\widetilde{\alpha} \equiv \alpha + \kappa\,(\gamma - \beta)$, $\widetilde{\beta} \equiv \beta + \phi\,(\gamma - \beta)$, and $\widetilde{\gamma} \equiv \gamma + \lambda\,(\gamma - \beta)$. Now, the residual in this model is orthogonal to the explanatory variables. Once again, in the published version, we erroneously used expenditure instead of *per capita* expenditure in equation (2).

The OLS estimate of the economies of scale parameter will converge to $\widetilde{\gamma}$ and so, we will have that

$$p\lim \widehat{\widetilde{\gamma}} = \gamma + \lambda\,(\gamma - \beta) = \gamma\,(1 + \lambda) - \lambda\beta.$$

So, we can write

$$\gamma = \frac{\widetilde{\gamma} + \lambda\beta}{1+\lambda}.$$

To construct a "back-of-the-envelope" bound on γ, we can derive an upper bound on $\sigma_e^2 = E[e^2]$ which we will denote by $\overline{\sigma}_e^2$. To simplify matters, we suppose that $\phi = 0$ and obtain that

$$\lambda = \frac{\sigma_{n,e}}{\sigma_n^2} = \frac{\sigma_{n^*,e} - \sigma_e^2}{\sigma_n^2} \geq -\frac{\sigma_e^2}{\sigma_n^2} \geq -\frac{\overline{\sigma}_e^2}{\sigma_n^2} \equiv \underline{\lambda}$$

provided that $\sigma_{n^*,e} \geq 0$ which should obtain if larger households are more prone to errors. We can bound γ by noting that

$$\gamma \leq \frac{\widetilde{\gamma} + \underline{\lambda}\beta}{1+\lambda}$$

since $\beta < 0$, $\underline{\lambda} < 0$ and $\lambda \in (-1, 0)$. Thus, noting that since $\widetilde{\gamma} < 0$, if

$$\widetilde{\gamma} + \underline{\lambda}\beta > 0 \Leftrightarrow |\underline{\lambda}| > \frac{\widetilde{\gamma}}{\beta} \tag{3}$$

then a positive value of γ is possible. This will happen when there is a large degree of measurement error in household size. In the published version, we did not properly compute $\overline{\sigma}_e^2$ and, hence, $\underline{\lambda}$.

2 Bounding the Measurement Error

We now fix that error. First, we note that

$$\sigma_e^2 = V(e) = V(E[e|W]) + E(V[e|W])$$

where $W \equiv (N, M, D, B)$ which are the number of household members, migrants, deaths, and births during the survey period. Each of these quantities is in levels and is observable in the data. The first component is commonly referred to as the between component whereas the second component is typically called the within component in the literature on inequality. Next, we recall that $E(n^*|W) \in [l(W), u(W)]$ where we computed the bounds in previous work. Because $E[e] = 0$, we note that

$$\begin{aligned} V(E[e|W]) &= E\left(E[e|W]^2\right) \\ &= E\left([E(n^*|W) - n]^2\right) \end{aligned}$$

and because

$$[E(n^*|W) - n]^2 \leq \max\left\langle (l(W) - n)^2, (u(W) - n)^2 \right\rangle,$$

we will obtain that

$$V\left(E\left[e|W\right]\right) \leq \sum_{W} \max\left\langle \left(l\left(W\right)-n\right)^{2}, \left(u\left(W\right)-n\right)^{2} \right\rangle p\left(W\right).$$

This was the formula that we gave for $\overline{\sigma}_e^2$ in the published version, but of course this was incomplete as it ignored the within component which is potentially greater.

To correct this omission, we write

$$\sigma_e^2 = V\left(e\right) = E\left[E\left[e^2|W\right]\right].$$

Next, we note that our primitive assumptions on N^* from the original paper imply similar bounds on its log so that

$$\begin{aligned}
n^* &\in \left[\log\left(N-B\right), \log\left(N+D\right)\right] \text{ for } \Delta M = 0 \\
n^* &\in \left[\log\left(N-B\right), \log\left(N+D+j\right)\right] \text{ for } \Delta M = j > 0 \\
n^* &\in \left[\log\left(\max\left\langle N-B+j, 1\right\rangle\right), \log\left(N+D\right)\right] \text{ for } \Delta M = j < 0.
\end{aligned}$$

Each of these bounds on n^* can be indexed by (j,k) where $\Delta M = j$ and $W = k$ and

so we can write these more compactly as

$$n^* \in [l_{j,k}, u_{j,k}] \text{ for } \Delta M = j, W = k.$$

This implies that

$$[n^* - n]^2 \leq \max \left\langle (l_{j,k} - n)^2, (u_{j,k} - n)^2 \right\rangle \text{ for } \Delta M = j, W = k.$$

Therefore, we will have that

$$\begin{aligned} E\left[e^2 | W = k\right] &= E\left[(n^* - n)^2 | W = k\right] \\ &\leq \sum_j \max \left\langle (l_{j,k} - n)^2, (u_{j,k} - n)^2 \right\rangle p\left(\Delta M = j | W = k\right) \end{aligned}$$

which then implies that

$$\sigma_e^2 \leq \sum_k \sum_j \max \left\langle (l_{j,k} - n)^2, (u_{j,k} - n)^2 \right\rangle p\left(\Delta M = j | W = k\right) p\left(W = k\right) \equiv \overline{\sigma}_e^2.$$

This bound accounts for both the between and within components of the variance.

When we compute this bound using the data discussed in the published paper, we obtain that $\overline{\sigma}_e^2 = 0.0624$, whereas we had obtained $\overline{\sigma}_e^2 = 0.0157$ when we ignored within variation. So, the within component is indeed more substantial than the

between component. However, given that we had that $\sigma_n^2 = 0.3236$ in the data, we obtain that $\underline{\lambda} = -0.1928$. Next, we recall that in the original paper, $\widehat{\widehat{\gamma}} = -0.0796$ and $\widehat{\beta} = -0.0882$ and we can see that the condition in 3 is not met despite the larger value for $\overline{\sigma}_e^2$. In fact, we would have to have $\overline{\sigma}_e^2 > 0.2920$ in order for positive values of γ to be possible which would suggest a large degree of measurement error in household size *i.e.* at least 468% more.

Finally, although these errors are not sufficient to resolve the paradox, it is important to point out that they may be non-trivial. The parameter $|\underline{\lambda}|$ is an upper bound on the ratio of the variance of measurement errors to the total variance of household size. Our calculations suggest that these errors could be as high as 20% of the total variation in household size.

References

[1] Deaton, Angus and Christina Paxson. 1998. "Economies of Scale, Household Size and the Demand for Food." *Journal of Political Economy.* 106 (5): 897-930.

[2] Halliday, Timothy J. 2010. "Mismeasured Household Size and Its Implications for the Identification of Economies of Scale" *Oxford Bulletin of Economics and Statistics.* 72(2): 246-262.