Computation of standard errors in DAD

This section shows how the standard errors of DAD's estimators of distributive indices and curves are computed. The methodology is based on the asymptotic sampling distribution of such indices and curves. All of DAD's estimators are asymptotically normally distributed around their true population value. As will be discussed below, we expect this methodology to provide a good approximation to the true sampling distribution of DAD's estimators for relative large samples.

ESTIMATORS OF THE DISTRIBUTIVE INDICES

Estimators of distributive indices (such as poverty and inequality indices) take the following general form:

$$\hat{\theta} = g(\hat{\alpha}_1, \hat{\alpha}_2, \cdots \hat{\alpha}_K) \text{ with } \alpha_k \text{ asymptotically expressible as } \alpha_k = \sum_{j=1}^{m} y_{k,j}$$

where θ can be expressed as a continuous function g of the α's, m is the number of sample observations and $y_{k,j}$ is usually some transform of the living standard of individual or household j. We use Rao's (1973) linearization approach [1] to derive the standard error of these distributive indices. This approach says that the sampling variance $\hat{\theta}$ equals the variance of a linear approximation of $\hat{\theta}$:

$$\mathrm{Var}(\hat{\theta}) = \mathrm{Var}\left(\frac{\partial\theta}{\partial\alpha_1}(\hat{\alpha}_1 - \alpha_1) + \frac{\partial\theta}{\partial\alpha_2}(\hat{\alpha}_2 - \alpha_2) + \cdots + \frac{\partial\theta}{\partial\alpha_K}(\hat{\alpha}_K - \alpha_K) \right)$$

In matrix format, the variance of $\hat{\theta}$ is given by

$$\mathrm{Var}(\hat{\theta}) = \mathrm{Var}(V'MV)$$

with M the covariance matrix of the $\hat{\alpha}$ and V the gradient of θ:

$$V = \begin{pmatrix} \frac{\partial\theta}{\partial\alpha_1} \\ \frac{\partial\theta}{\partial\alpha_2} \\ \vdots \\ \frac{\partial\theta}{\partial\alpha_K} \end{pmatrix}$$

The gradient elements $\left(\frac{\partial\theta}{\partial\alpha_1}, \frac{\partial\theta}{\partial\alpha_2}, \cdots \right)$ can be estimated consistently using estimates $\left(\frac{\partial\hat{\theta}}{\partial\hat{\alpha}_1}, \frac{\partial\hat{\theta}}{\partial\hat{\alpha}_2}, \cdots \right)$ of the true derivatives. The covariance matrix is defined as

[1] Rao,C.R. (1973). *Linear Statistical Inference and Its Application*. New York: Wiley.

$$M = \begin{vmatrix} \mathrm{Var}(\alpha_1) & \mathrm{Cov}(\alpha_1,\alpha_2) & & \mathrm{Cov}(\alpha_1,\alpha_K) \\ \mathrm{Cov}(\alpha_2,\alpha_1) & \mathrm{Var}(\alpha_2) & \cdots & \mathrm{Cov}(\alpha_2,\alpha_K) \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{Cov}(\alpha_K,\alpha_1) & \mathrm{Cov}(\alpha_K,\alpha_2) & \cdots & \mathrm{Var}(\alpha_K) \end{vmatrix}$$

The elements of the covariance matrix are again estimated consistently using the sample data, replacing for instance $\mathrm{Var}(\hat{\alpha})$ by $\hat{\mathrm{V}}\mathrm{ar}(\hat{\alpha})$. It is at the level of the estimation of these covariance elements that the full sampling design structure is taken into account.

FINITE-SAMPLE PROPERTIES OF ASYMPTOTIC RESULTS

It may be instructive to compare the results of the above asymptotic approach to those of a numerical simulation approach like the bootstrap. The bootstrap (BTS) is a method for estimating the sampling distribution of an estimator which proceeds by re-sampling repetitively one's data. For each simulated sample, one recalculates the value of this estimator and then uses that BTS distribution to carry out statistical inference. In finite samples, neither the asymptotic nor the BTS sampling distribution is necessarily superior to the other. In infinite samples, they are usually equivalent.

BOOTSTRAP AND SIMPLE RANDOM SAMPLING

The following steps the BTS approach for a sample drawn using Simple Random Sampling:
1- Draw *with replacement* m observations from the initial sample.
2- Compute the distributive estimator from this new generated sample.
3- Repeat the first two steps N times.
4- Compute the variance or the BTS distributions using these N generated estimators.

BOOTSTRAP AND COMPLEX SAMPLING DESIGN

The steps here are similar to those above with Simple Random Sampling. Only the first step differs to take into account the precise way in which the original sample was drawn. Suppose for example that:

- The data were drawn from two strata, with m1 observations in stratum 1 and m2 observations in stratum 2.
- Observations in every stratum were selected randomly with equal probabilities
- The first step will then consist in selecting randomly and with the same probability m1 observations from stratum1 and (independently) m2 observations from stratum2. Aggregating these two sub samples will yield the new generated sample. Repeating this N times will generate the BTS sampling distribution.

ILLUSTRATIONS

The following table presents the sampling design information of a hypothetical sample of 800 observations.

Sampling Design Information

Number of observations	**800**
Sum of weights	**6200.0**
Number of strata	**2 strata in the Sampling Design**

CODE	STRATA	PSU	LSU	OBS	P(strata)	FPC (f_h)
1	1	30	300	300	0,193548	0.0
2	2	50	500	500	0,806452	0.0
Total	2	80	800	800	1.0	--

The following tables present estimates of the standard errors of some distributive indices using asymptotic theory (DAD) and the BTS procedure.

Atkinson Index (ε =0.5) = 0,09131119						
W	Strata	Psu	Lsu	Size =psu	St.err. DAD	St.err. BTS
✕					0,00403011	0,00404464
✕	✕				0,00396117	0,00391402
✕		✕			0,00479089	0,00473645
✕	✕	✕			0,00414549	0,00412479
✕	✕	✕		✕	0,00455368	0,00454454

FGT (α =1; z=3000) = 566.47774194						
W	Strata	Psu	Lsu	Size =psu	St.err. DAD	St.err. BTS
✕					30,15130207	30,31106186
✕	✕				29,76615787	29,82831383
✕		✕			34,90968660	34,49846649
✕	✕	✕			31,21606735	31,36449814
✕	✕	✕		✕	40,20904414	40,10400009

Lorenz (p=0.5) =0,26371264						
W	Strata	Psu	Lsu	Size =psu	St.err. DAD	St.err. BTS
✕					0,00618343	0,00617247
✕	✕				0,00612036	0,00614563
✕		✕			0,00695073	0,00697490
✕	✕	✕			0,00632417	0,00636899
✕	✕	✕		✕	0,00726710	0,00724934

Gini (ρ =2) = 0,42403734						
W	Strata	Psu	Lsu	Size =psu	St.err. DAD	St.err. BTS
✕					0,00801557	0,00809321
✕	✕				0,00786047	0,00781983
✕		✕			0,00964692	0,00964823
✕	✕	✕			0,00820847	0,00827642
✕	✕	✕		✕	0,00949502	0,00946204

Notes:

W	Sampling weight
✕	Sampling-design feature is used