THE EFFECTIVENESS OF THE DURBIN SELECTION METHOD FOR VARIANCE REDUCTION

Thanh Lê, Ismael Flores-Cervantes, John Burke, J. Michael Brick, Westat
Thanh Lê, Westat, 1650 Research Boulevard, Rockville, Maryland 20850

Key Words: Durbin's method; multistage design; selection with and without replacement; variance estimation

1. Introduction

The Early Childhood Longitudinal Study: Kindergarten Class of 1998-99 (ECLS-K) is sponsored by the U.S. Department of Education, National Center for Education Statistics (NCES). It provides national data on children's early school experience beginning with kindergarten and as they progress through the fifth grade.

The ECLS-K collected base year data on a nationally representative sample of 21,260 children in fall- and spring-kindergarten (fall 1998 and spring 1999). Beyond the kindergarten year, two rounds of data have been collected (fall- and spring-first grade) and two additional rounds are planned (spring-third and spring-fifth grades). Data collection consists of direct assessments of children, interviews with their parents, and abstracts of school records. Teachers and school administrators complete self-administered questionnaires.

In the base year, the sample of children was selected using a multi-stage probability design. The first-stage or primary sampling units (PSUs) were geographic areas that are counties or groups of counties. PSUs were selected with probability proportional to measures of size that take into account the desired oversampling of Asians and Pacific Islanders (APIs). The area sampling frame for the ECLS-K consisted of 1,335 PSUs. PSUs with the largest measure of size were designated as self-representing (SR) PSUs and were included in the ECLS-K with certainty. The remaining PSUs were partitioned into strata of roughly equal measure of size. From each of these strata, two PSUs were selected using Durbin's Method. The other stages of sampling are described later, but most of the results in this paper deal with the first-stage sampling.

This paper summarizes the theory of Durbin's method, describes pertinent aspects of the sample design and application of Durbin's method to the ECLS-K, and investigates the reduction in variance estimates associated with this method using a variety of statistics from the ECLS-K.

2. Durbin Selection Method

When the selection of first-stage units or PSUs is carried out with replacement, variance estimates can be computed using simple methods that are functions of PSU level estimates only. But sampling with replacement is not always practical or efficient. Durbin (1967) developed methods to select two first-stage units per stratum without replacement, with probability proportional to size and known joint probability of inclusion, in such a way to allow variances to be estimated as if the units had been selected with replacement. One such method is known as Method I, and it selects two units with unequal probabilities without replacement.

Method I requires two passes of the frame with a different selection probability at each pass to obtain the desired probabilities of inclusion and joint probabilities of inclusion. In the first pass, one PSU is selected in the stratum with probability p_i that is proportional to the measure of size of unit *i*. In the second pass, the selected PSU, unit *i*, is excluded and another PSU is selected with probability proportional to

$$p_j\left[\frac{1}{1-2p_i}+\frac{1}{1-2p_j}\right] \tag{1}$$

where $p_i = M_i/M$ and $p_j = M_j/M$. M_i is the measure of size of unit *i*, M_j the measure of size of the unit *j*, and *M* the measure of size of the stratum. The total probability of selection unit *k* is $2p_k$.

The joint probability of inclusion of the first and second units is

$$\pi_{ij} = \left[2p_i p_j\left(\frac{1}{1-2p_i}+\frac{1}{1-2p_j}\right)\right] \div \left(1+\sum_{k=1}^{N}\frac{p_k}{1-2p_k}\right) \tag{2}$$

In multi-stage sampling, an unbiased estimate of the variance of the linear estimator $x = \sum_h (y_{hi} + y_{hj})$, where y_{hi} and y_{hj} are the contributions from units *i* and *j* selected from the h^{th} stratum is

$$\hat{V}(x) = \sum_h \left(\frac{\pi_{hi}\pi_{hj}}{\pi_{hij}} - 1 \right) \left(y_{hi} - y_{hj} \right)^2 + \sum_h \left(\pi_{hi} s_{hi}^2 + \pi_{hj} s_{hj}^2 \right) \quad (3)$$

where π_{hi} and π_{hj} are the probabilities of selection of the i^{th} and j^{th} units from the h^{th} stratum, π_{hij} is the joint probability of selection of the pair, and s_{hi}^2 and s_{hj}^2 are unbiased estimates of the variances of y_{hi} and y_{hj} due to sampling at subsequent stages. This formula is much more complicated than is needed if the sampling had been carried out with replacement, in which case the unbiased estimate of variance of the same estimator is $\hat{V}(x) = \sum_h \left(y_{hi} - y_{hj} \right)^2$.

Durbin (1967) shows that if PSUs are selected using Method I, variance estimation can be simplified by dividing the strata into two groups: (1) SR strata whose PSUs are included with certainty, and (2) non-SR strata. Additionally, among the non-SR strata, some are selected with probability proportional to $\min\left(\frac{\pi_{hi}\pi_{hj}}{\pi_{hij}} - 1, 1 \right)$ and treated as SR strata for the purpose of variance estimation. The rule is to (1) compute the variance as if sampling had been with replacement for strata in which selected units come from different groups, and (2) compute the variance as if each stratum consists only of the group containing the units for strata in which the units come from the same group.

3. Sampling in the ECLS-K

In the base year, children were selected for the ECLS-K using a multi-stage probability design. PSUs were geographic areas that are counties or groups of counties, selected with probability proportional to size. The measure of size took into account the amount of oversampling of APIs required to meet precision goals. The 24 PSUs with the largest measure of size, identified from the frame of 1,335 PSUs, were included in the ECLS-K as SR PSUs. The remaining non-SR PSUs were partitioned into 38 strata of roughly equal measure of size. From each non-SR stratum, two PSUs were selected using Durbin's Method I.

In the second stage, public and private schools offering kindergarten programs were selected. For each PSU, a frame of public and private schools offering kindergarten programs was constructed using existing school universe files: the 1995-96 Common Core of Data (CCD), the 1995-96 Private School Universe Survey (PSS). Most schools run by the Bureau of Indian Affairs (BIA) and U.S. domestic schools run by the Department of Defense (DOD) are not included on the CCD. For this reason, the 1995-96 Office of Indian Education Programs Education Directory was consulted in order to complete the list of BIA schools in the CCD file. For the DOD schools, a 1996 list of schools was obtained directly from the DOD. The school frame was freshened in the spring of 1998 to include newly opened schools that were not included in the CCD and PSS, and schools that were in CCD and PSS but did not offer kindergarten programs according to those sources. The selection of schools was systematic, with probability proportional to the measure of size. As with the PSU sample, a weighted measure of size was constructed taking into account the oversampling of APIs. Public and private schools constituted distinct sampling strata. Within each stratum, schools were sorted to ensure good sample representation across other characteristics. In total, 1,280 schools were sampled from the original frame, and 133 from the freshened frame. Of these, 953 were public schools and 460 were private schools.

The third stage sampling units are children of kindergarten age, selected within each sampled school to obtain an approximately self-weighting sample of students by type of school (public/private), and at the same time to achieve a minimum required sample size for the API group. Because of the API oversampling and the differences between the estimated school sizes used for sampling schools and the actual school sizes used for sampling students, some variations in the weights within school types did occur.

4. Weighting the ECLS-K Data

Weighting the ECLS-K data is necessary to compensate for differential probabilities of selection at each sampling stage and to adjust for the effects of nonresponse. In each round of data collection of the ECLS-K, several sets of weights were computed. In the base year, weights were computed separately for children, teachers and schools.

There are several sets of child level weights: (1) weights to be used for the analysis of direct child assessment data, alone or in conjunction with a set of limited set of child characteristics such as age, gender, and race-ethnicity; (2) weights to be used for the analysis of parent interview data to be used alone or in combination with child assessment data, and (3) weights to be used for the analysis of direct child assessment data combined with parent interview and teacher data.

Longitudinal weights were also computed for children with complete data from several combinations of data collection rounds. For each set of weights computed, a set of replicate weights was created to be used in variance estimation using replication methods.

5. Variance Estimation in the ECLS-K

Variance estimation in the ECLS-K was done using the paired jackknife replication (JK2) method, taking into account the clustered, multistaged characteristics of sampling and the use of differential sampling rates to oversample APIs. For the ECLS-K, in which the first-stage SR sampling units were selected with certainty and the first-stage non-SR sampling units were selected with two units per stratum, the second stage units in the SR strata were paired and then combined to form two units per combined stratum. This combining into pairs makes the two sampled per stratum JK2 method an appropriate method of variance estimation.

In JK2, a survey estimate of interest is calculated from the full sample. Replicates of the full sample are then selected to calculate replicate estimates of the same parameter. The variability of the replicate estimates about the full sample estimate is used to estimate the variance of the full sample estimate. The variance estimator is computed as the sum of the squared deviations of the replicate estimates from the full sample estimate:

$$v\left(\hat{\theta}\right) = \sum_{g=1}^{G} \left(\hat{\theta}_{(g)} - \hat{\theta}\right)^2 \qquad (4)$$

where θ is the population quantity of interest, $\hat{\theta}$ is the estimate of θ based on the full sample, G is the number of replicates formed, and $\hat{\theta}_{(g)}$ is the g^{th} replicate estimate of θ based on the observations included in the g^{th} replicate.

Each replicate weight was calculated using the same adjustment steps as the full sample weight but using only the subsample of cases that constitute each replicate. For the ECLS-K, replicate weights were created taking into account the Durbin's method of PSU selection. Among the 38 non-SR strata, 11 strata were identified as Durbin strata (selected with probability proportional to $\left(\frac{\pi_{hi}\pi_{hj}}{\pi_{hij}} - 1\right)$), and treated as SR strata for variance estimation. This brings the number of SR PSUs to 46. The remaining 54 non-SR PSUs are in 27 non-SR strata; thus 27 replicates were formed, each corresponding to one non-SR stratum. All schools within a non-SR PSU were assigned to the same variance unit and variance stratum. Sampled schools in the 46 SR PSUs were grouped into 63 variance strata. The 90 replicates are used for variance estimation and yield approximately 76 degrees of freedom for calculating confidence intervals for many national estimates.

In the SR PSUs, schools were treated as first stage sampling units. Public schools had been sampled from within PSU while private schools had been pooled into one sampling stratum and selected systematically (except in the SR PSUs identified through the Durbin method where private schools were treated as if they were sampled from within PSU). To form PSUs for variance estimation, schools were sorted by sampling stratum, type of school (from the original sample or newly selected as part of freshening), type of frame (for new schools only), and their original order of selection (within stratum). From this sorted list, schools were grouped into pairs within each sampling stratum; the last pair in the stratum may be a triplet if the number of schools in the stratum is odd. This operation resulted in a number of ordered preliminary variance strata of two or three units each. The first ordered 63 strata were then numbered sequentially from 1 to 63; the next ordered 63 strata were also numbered sequentially from 1 to 63, and so on until the list was exhausted, thus forming the desired 63 variance strata.

In variance strata with two units, a unit being a PSU in the case of non-SR PSUs and a school in the case of SR PSUs, the base weight of the first unit was doubled to form the replicate weight, while the base weight of the second unit was multiplied by zero. In strata with three units, two variance strata were created: in the first variance stratum, the base weight of two of the three units was multiplied by 1.5 to form the replicate weight and the base weight of the last unit was multiplied by zero; in the second variance stratum, the base weight of a different group of two units was multiplied by 1.5, and the base weight of the third unit was multiplied by zero (Rust, 1986).

For the analysis covered in this paper, we created another set of replicate weights not taking into account the fact that Durbin's Method I was used in sample selection. Two sets of estimates were then computed, one set using the Durbin weights (taking into account Durbin's without replacement sampling method) and the other set using the non-Durbin weights (ignoring Durbin's method and treating the non-SR PSUs as if they were selected with replacement). We kept the number of replicate weights for the two sets of estimates the same, namely 90. The 90 replicate weights are split 63:27 for SR:non-SR in the case of Durbin weights, and 52:38 in the case of non-Durbin weights. The split for Durbin weights is explained above. For non-Durbin weights, the 38 non-SR strata constitute 38 non-SR replicates while the 24 SR strata were distributed among 52 replicates for a total of 90 replicates.

6. Durbin vs. Non-Durbin Variance Estimates

In the ECLS-K, a large number of data items were collected from children, parents, teachers and schools. Variance estimates were computed for a number of these data items. To evaluate the effectiveness of Durbin's Method I, we select more than 50 data items from spring-kindergarten including child assessment scores, child and parent characteristics as reported by parents, and school characteristics of schools from which the children were sampled. We computed the variance estimates for these items for all children and by subgroup, using both Durbin replicate weights and non-Durbin replicate weights. To compare the estimates, we computed the ratio of Durbin standard error to the non-Durbin standard error for each estimate. Note that the ratio is not a comparison of the variance estimate from PSUs that were selected without replacement to the variance estimate from PSUs that were selected with replacement.

The ratios of standard errors are shown in figure 1 by type of estimate, and in figures 2 and 3 by type of estimate and subgroup. The type of weights used depends on the type of estimate. For child assessment scores, child characteristics not coming from the parent data (such as child's age at assessment), and school characteristics, we used the child weights. For child and parent characteristics coming from the parent data, we used the child level parent weights.

Since the purpose of Durbin's Method is to allow computation of variance estimates as if sampling with replacement was done, the finite population correction (*fpc*) factor does not play a role in the variance estimate. When using non-Durbin weights to compute the variance estimates, essentially ignoring the *fpc* while sampling was done without replacement, we expect a slight overestimation of the variance. The ratios of standard errors in figure 1 are generally less than 1, confirming the overestimation of the variance if non-Durbin weights were used. Since the variance of the variances is large, we do not expect to see all ratios to be greater than 1. For estimates of totals, it is easier to see that large totals mostly have ratios of less than 1, and that the method may not work as well for totals for small subgroups. The overall average ratio of variance estimates is 0.93.

In figures 2 and 3, we plot the same ratios but for estimates of subgroups so that we can compare them with overall estimates. At the overall level, almost all ratios are less than 1. The one that is greater than 1 is for an estimate of proportions that is very large, over 90 percent. For gender, level of urbanicity and public schools, most of the ratios are less than 1. The effect of the random selection of Durbin strata among the noncertainty strata can be seen in the plots for regions and private schools. For example, if more strata from one region were identified as Durbin SR strata than from another, then this will increase variation in the variance estimates. The next graph shows the same comparison for estimates of totals. The pattern follows that of estimates of means and proportions, but the ratios are more variable.

7. Variance Decomposition

If the Durbin method of sampling and variance estimation had not been implemented, then it is likely that the variances of the estimates for the ECLS-K would have been computed as if the sample was selected with replacement. The complexity of computing variances at multiple stages of selection and including adjustments for the first stage unequal probabilities of selection makes other options unattractive. The Durbin variance estimator can be decomposed as

$$V\left(\hat{y}_D\right) = V_{SR} + B_{NSR,D} + W_{NSR} \tag{5}$$

where V_{SR} is the variance within SR PSUs, $B_{NSR,D}$ is the variance between non-SR PSUs using the Durbin definition of SR and non-SR, and W_{NSR} is the variance within non-SR PSUs.

If we had sampled PSUs in non-SR strata with replacement, then the with replacement variance estimator would be

$$V\left(\hat{y}_{WR}\right) = V_{SR} + B_{NSR,WR} + W_{NSR} \tag{6}$$

where V_{SR} and W_{NSR} are as in equation (5), and $B_{NSR,WR}$ is the variance between non-SR PSUs not using the Durbin definition.

Durbin (1953) shows the bias of the with-replacement variance estimator is two times the difference between the with-replacement variance and the Durbin variance:

$$\begin{aligned} Bias_{NSR}(v(\hat{y}_{WR})) &= 2\left[V\left(\hat{y}_{WR}\right) - V\left(\hat{y}_D\right)\right] \\ &= 2\left[B_{NSR,WR} - B_{NSR,D}\right] \end{aligned} \tag{7}$$

So the bias is totally in the between-PSU variance component from the non-SR strata.

To apply this approximation in the ECLS-K we estimated the components of variance due to the sampling of PSUs and the subsampling within-PSU using a different replication scheme. To estimate the within-PSU variance (W_{NSR}) from the non-SR PSUs, replicates in the non-SR strata were redefined using the second stage sampling units as if they were the first stage sampling units. The replicates for the SR strata were not altered. Using these redefined replicates, within-PSU variance estimates were computed as $w_{NSR} = v(\hat{y}_{WR}) - v_{SR}$, where $v(\hat{y}_{WR})$ is the with-replacement variance estimate using the replicate scheme described immediately above and v_{SR} is the with-replacement variance estimate within SR PSUs (i.e., setting the replicate weights of the non-Durbin NSR PSUs equal to the full sample weight). The estimated between-PSU component is the difference between the estimated overall variance and the within-PSU variance. Negative estimates of the between-PSU component were set equal to zero (the variance of the variance is relatively large and negative variance estimates of the components are not uncommon).

On average, the components of equation (5) are 26 percent for v_{SR}, 20 percent for $b_{NSR,D}$ and 54 percent for w_{NSR}. Hence, the bias of the with-replacement variance estimate only applies to 20 percent of the Durbin variance estimate. It is worth noting that the contribution of w_{NSR} is different depending on the type of estimates: 48 percent for estimates of means/proportions and 60 percent for estimates of totals. Since the contribution of v_{SR} from the SR PSUs does not change, the bias of the with-replacement variance estimate applies to 26 percent of the Durbin variance estimate in the case of means/proportions, and 14 percent in the case of totals.

8. Conclusion

Durbin's method of selecting first-stage units is useful when the contribution to the variance of the SR strata and within the non-SR strata is small proportion of the variance of the estimate, and when the sampling fraction for non-SR strata is large. Even when it is effective, the stability of the variance estimate for subgroups is a potential problem because of the random grouping of non-SR strata into Durbin strata.

9. References

Cochran, W. (1977). *Sampling Techniques* (3rd edition). New York: Wiley.

Durbin, J. (1953). Some results in sampling theory when the units are selected with unequal probabilities. *Journal of the Royal Statistical Society* B, 15, 262-269.

Durbin, J. (1967). Design of multi-stage surveys for the estimation of sampling errors. *Journal of the Royal Statistical Society*, C, 16, 152-164.

Rust, K. (1986). Efficient replicated variance estimation. *Proceedings of the Survey Research Methods Section of the American Statistical Association*, 81-87.

Figure 1. ECLS-K Spring-Kindergarten: Ratios of the Durbin to the non-Durbin standard errors

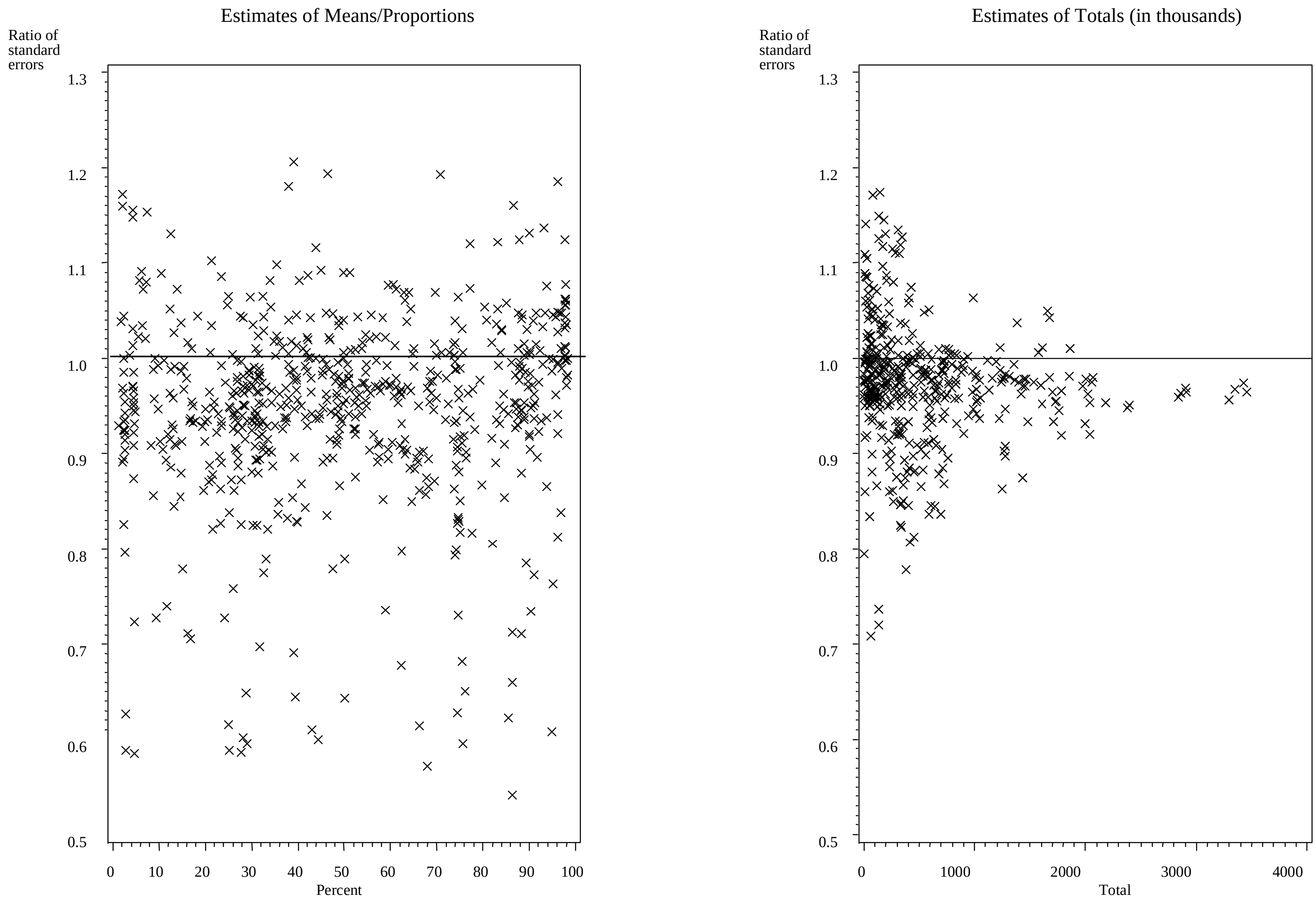

Figure 2. ECLS-K Spring-Kindergarten: Ratios of the Durbin to the non-Durbin standard errors, by subgroup – Means and proportions

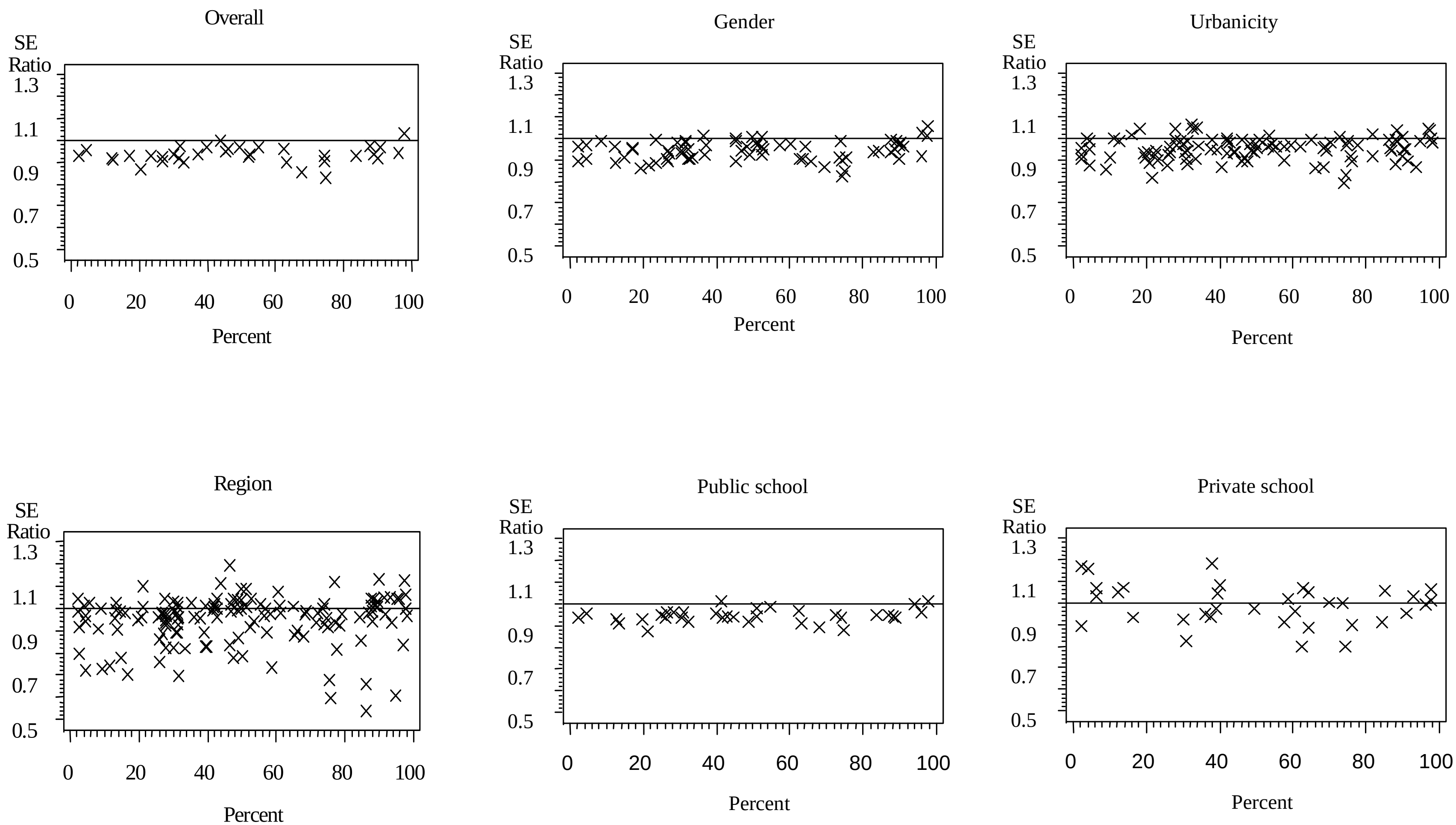

Figure 3. ECLS-K Spring-Kindergarten: Ratios of the Durbin to the non-Durbin standard errors, by subgroup – Totals (in thousands)