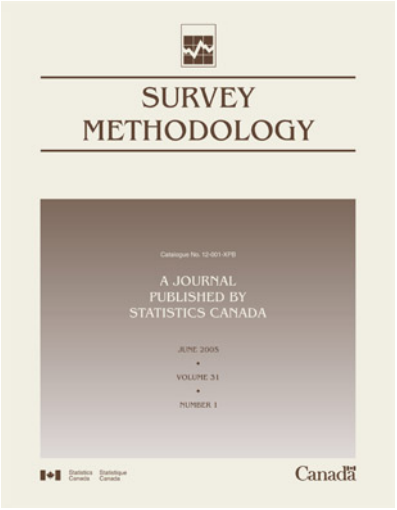
Catalogue no. 12-001-XIE

Survey Methodology

2005



Statistics Canada Statistique Canada

Canada

How to obtain more information

Specific inquiries about this product and related statistics or services should be directed to: Business Survey Methods Division, Statistics Canada, Ottawa, Ontario, K1A 0T6 (telephone: 1 800 263-1136).

For information on the wide range of data available from Statistics Canada, you can contact us by calling one of our toll-free numbers. You can also contact us by e-mail or by visiting our website.

National inquiries line	1 800 263-1136
National telecommunications device for the hearing impaired	1 800 363-7629
Depository Services Program inquiries	1 800 700-1033
Fax line for Depository Services Program	1 800 889-9734
E-mail inquiries	infostats@statcan.ca
Website	www.statcan.ca

Information to access the product

This product, catalogue no. 12-001-XIE, is available for free. To obtain a single issue, visit our website at www.statcan.ca and select Our Products and Services.

Standards of service to the public

Statistics Canada is committed to serving its clients in a prompt, reliable and courteous manner and in the official language of their choice. To this end, the Agency has developed standards of service that its employees observe in serving its clients. To obtain a copy of these service standards, please contact Statistics Canada toll free at 1 800 263-1136. The service standards are also published on www.statcan.ca under About Statistics Canada > Providing services to Canadians.

Statistics Canada
Business Survey Methods Division

Survey Methodology

2005

Published by authority of the Minister responsible for Statistics Canada

November 2005

Catalogue no. 12-001-XIE
ISSN 1492-0921

Frequency: semi-annual

Ottawa

Cette publication est disponible en français sur demande (nº 12-001-XIF au catalogue).

Note of appreciation

Canada owes the success of its statistical system to a long-standing partnership between Statistics Canada, the citizens of Canada, its businesses, governments and other institutions. Accurate and timely statistical information could not be produced without their continued cooperation and goodwill.

Vol. 29, No. 1, pp. 63-70
Statistics Canada, Catalogue No. 12-001

Evaluating the Fundamentals of a Small Domain Estimator

Rachel Harter, Michael Macaluso and Kirk Wolter [1]

Abstract

The Illinois Department of Employment Security is using small domain estimation techniques to estimate employment at the county/industry division level. The estimator is a standard synthetic estimator, based on the ability to match Current Employment Statistics sample data to ES202 administrative records and an assumed model relationship between the two. This paper is a case study; it reviews the steps taken to evaluate the appropriateness of the model and the difficulties encountered in linking the two data sources.

Key Words: Small domain; Employment; Labor market; Generalized regression model; Auxiliary data.

1. Introduction

The Current Employment Statistics (CES) program of the U.S. Bureau of Labor Statistics (BLS) is a federal-state cooperative survey of employers used for estimating employment, women workers, production workers, production worker hours, and production worker earnings on a monthly basis. The estimates are among America's leading economic indicators. The sample was designed to support estimates at the national, state, and large metropolitan statistical area (MSA) levels. CES is roughly comparable to Statistics Canada's monthly Survey of Employment, Payroll and Hours (SEPH).

The Illinois Department of Employment Security (IDES), and similar agencies in other states across the nation, participates with the BLS in the collection, tabulation, and publication of the CES data. The state agencies have considerable customer demand for employment estimates at smaller sub-state levels than the CES sample was intended to support. In particular, IDES needs monthly employment estimates at the county/industry division level, and it formed a partnership with the National Opinion Research Center (NORC) to find a solution to this small domain estimation problem.

In a prior paper (Harter, Wolter and Macaluso 1999), we discussed some simulations done to test various small domain estimators. In this paper, we focus on the practical aspects of finding suitable auxiliary data, determining an appropriate model, merging the data sources, and monitoring the estimation process.

2. Evaluating Auxiliary Data

Purcell and Kish (1980), Ghosh and Rao (1994), and Singh, Gambino and Mantel (1994) provide excellent overviews of many small domain estimators. Most small domain estimators improve on direct sample-based estimators by (1) taking advantage of known auxiliary data, and (2) assuming and fitting a model relationship between the auxiliary data and the sample data. In this section we describe the auxiliary data for Illinois' small domain estimation problem and our evaluation of the data for this purpose.

The CES has a sister federal-state cooperative program – known as the Covered Employment and Wages (or ES202) program – in which employment and wage data are collected quarterly from all employers that participate in states' unemployment insurance programs. The employment figures from the ES202 are available approximately five months following the reference quarter. The ES202 records provide the sampling frame for the CES program. Furthermore, since the ES202 data are available for essentially all employers in the sampling frame, ES202 employment figures are considered "truth" for practical purposes.

CES monthly estimates are regularly benchmarked to ES202 figures. While they are revised several times as more complete information becomes available, the first release of CES data occurs on the first Friday of the month following the reference month. Although the ES202 employment figures lag behind the initial CES estimates by several months, ES202 employment is an obvious candidate for auxiliary data in our small domain estimation project.

A good auxiliary variable should be highly correlated with the estimation variable. In this case, ES202 employment is measuring the same concept as CES employment, except for minor scope and coverage differences, such as student workers at colleges and universities. Therefore, we expect ES202 employment and CES employment to be highly correlated.

Illinois data for a matched sample of employers from 1995 and 1996 shows that, indeed, ES202 employment and CES employment are highly correlated, regardless of the

1. Rachel Harter, National Opinion Research Center, 55 East Monroe, Suite 4800, Chicago, IL 60603; Michael Macaluso, Illinois Departement of Employment Security, Economic Information and Analysis, 401 South State Street, 7 North, Chicago, IL 60605; Kirk Wolter, Interdisciplinary Research Institute for Survey Science, 218 Snedecor Hall, Ames, IA 50010.

time lag between the two. Table 1 shows simple correlation coefficients for various industries and time lags. The correlations are slightly higher for shorter lags in growing in dustries, such as Finance, Insurance, and Real Estate, and for 12–month lags in seasonal industries, such as Construction. Nevertheless, we conclude from these statistics that any recent period of ES202 data is likely to serve successfully as auxiliary data for CES estimation.

Table 1
Mean Correlations of CES Employment with ES202 Employment*

Industry Division	ES202 lagged 12 months from CES	Most recent March ES202 available for CES month	Average monthly ES202 for most recent available quarter to CES month
Mining	0.951	0.965	0.980
Construction	0.936	0.909	0.909
Manufacturing	0.983	0.984	0.985
Transportation & Utilities	0.978	0.981	0.982
Trade	0.979	0.979	0.979
Finance, Insurance,& Real Estate	0.982	0.985	0.987
Services	0.975	0.966	0.966
Government ownership	0.996	0.995	0.993

* Within 2–digit Standard Industrial Classification (SIC) codes, we computed correlations for pairs of CES and ES202 months with the lagged relationships shown. We averaged the correlations across reference months and across SICs within the industry divisions shown.

We reviewed the scope and coverage differences between CES and ES202 to determine where the use of ES202 data may require special attention. The student worker example cited above was one such difference. Railroad workers do not participate in state unemployment insurance programs, so this industry is one in which ES202 data are not likely to be helpful. We reviewed the processing schedules for both CES and ES202 to help us determine which period of ES202 data would be available for estimation on the CES schedule. We reviewed the edits applied in both programs to see where differences may affect outcomes. For both of these programs, many anomalies in the data are explained through the use of comment variables containing standard coded values for various business conditions. We reviewed these comment variables to see how special cases are handled. All of these background checks were necessary to identify potential pitfalls in using ES202 data as an auxiliary variable for the small domain estimation problem.

Finally, we needed some indication that CES and ES202 data could be successfully linked for individual employers. To examine this issue, we matched and plotted CES and ES202 data. See Figures 1–3 for examples of statewide plots by 2–digit SIC (Standard Industrial Classification). The plots immediately alert us to potential matching problems in individual cases (Points considerably off the straight line signify potential matching or data problems), but assure us that most observations can be successfully matched. We discuss this issue in greater detail in section 4.

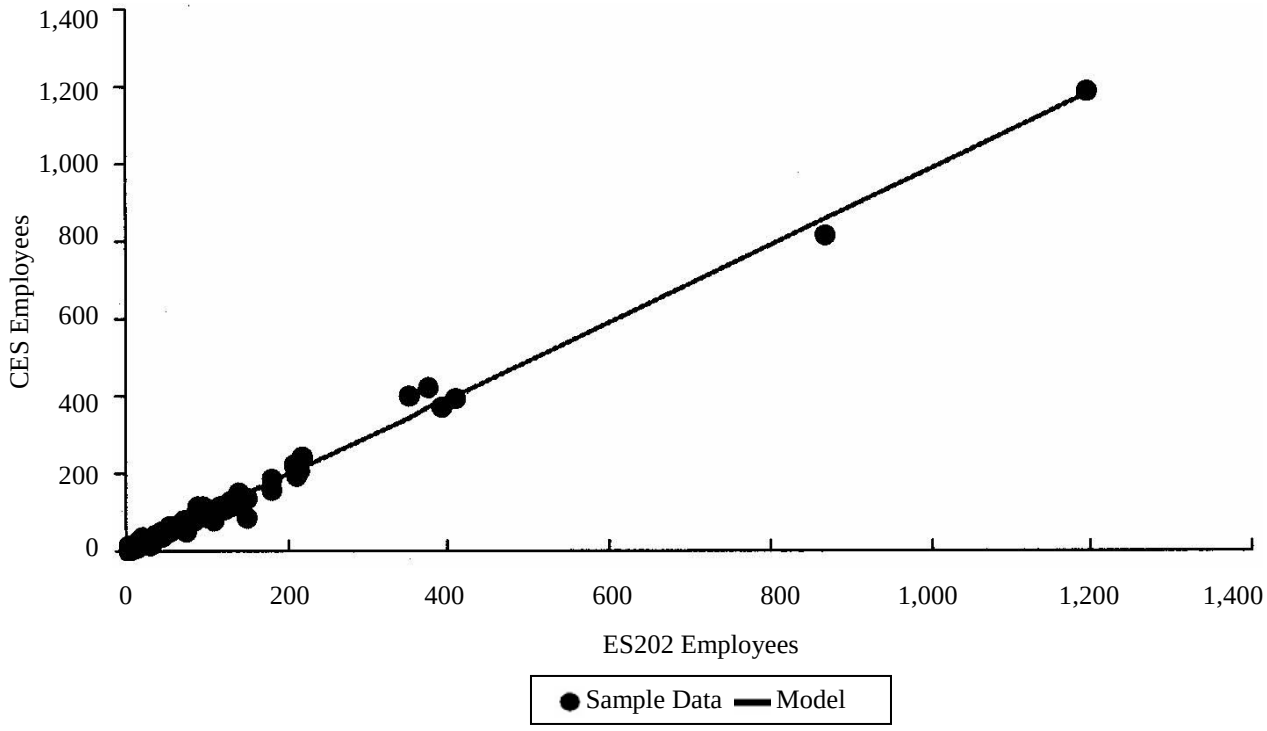


Figure 1. CES Versus ES202 Employment for a Sample of 103 Illinois Employers Classified in the Primary Metal Manufacturing Industry.

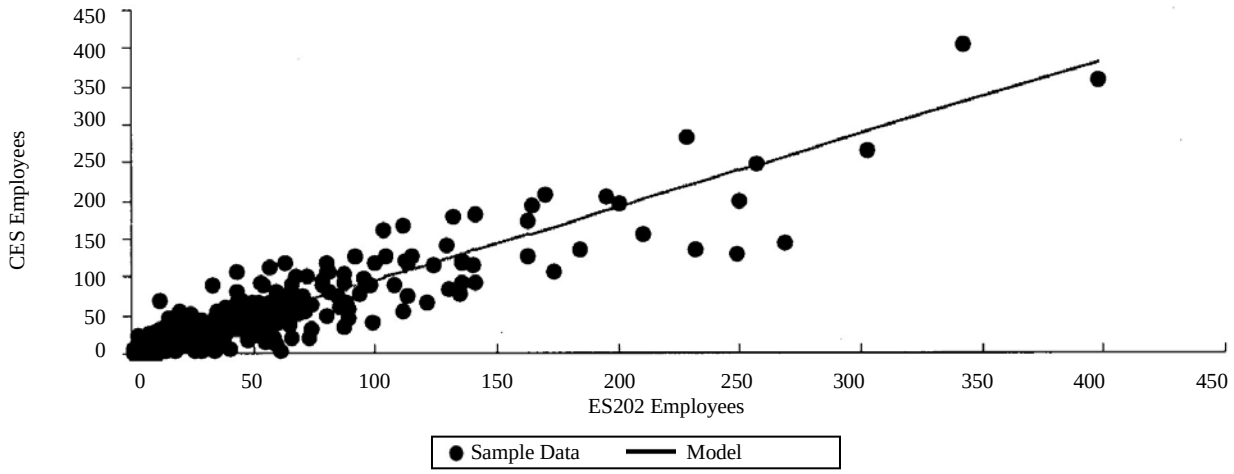


Figure 2. CES Versus ES202 Employment for a Sample of 701 Employers Classified in the Trade Contractors Industry.

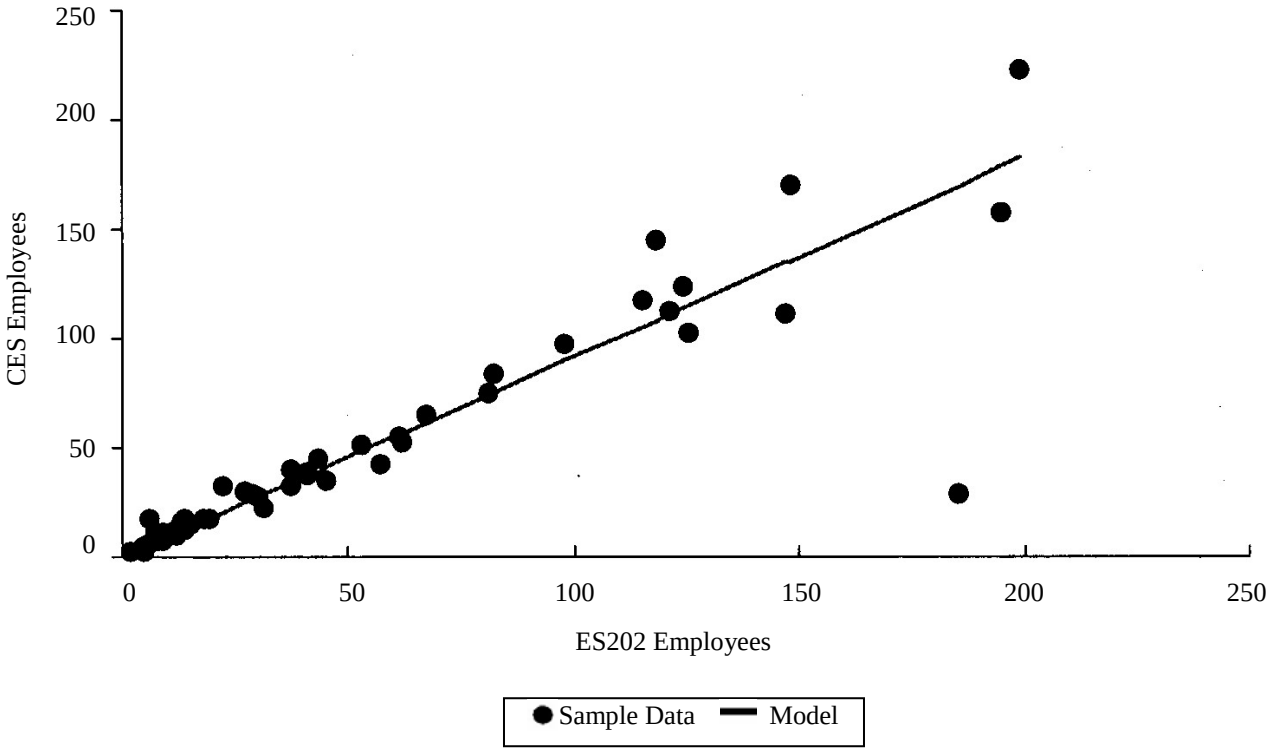


Figure 3. CES Versus ES202 Employment for a Sample of 50 Employers Classified in the Apparel Manufacturing Industry.

3. Evaluating the Model

Since the CES and ES202 programs are both measuring employment, we expect the relationship between the two to be linear with intercept zero and slope close to one. The plots in Figures 1–3, and the many other similar plots we produced and reviewed, indicate that this is generally true. Industries with changes over time or differences in scope and coverage sometimes display slopes other than one. The plots also indicate variability in the linear relationships, and some industries exhibit more variability about the linear relationship than others. Generally, the residual variance about the line increases with employment.

The standard "ratio" model adequately describes most of our data. Let y_j be the current month CES employment for employer j, and let x_j be the ES202 employment for the

same employer at some previous time period. Then the assumed model relationship is

$$y_j = x_j\beta + \varepsilon_j, \varepsilon_j \sim \text{NID}(0, \sigma^2 x_j). \quad (1)$$

The model parameter β can be estimated by generalized least squares, resulting in the ratio estimator $\hat{\beta} = \bar{y}/\bar{x}$, where $\bar{y}$ and $\bar{x}$ are the means of the observed current-month and auxiliary data, respectively (Sampling weights may or may not be employed in the analysis depending on many considerations beyond the scope of this article).

If Model (1) is true, then the best linear unbiased predictor of current month employment for sub-state domain D_{kl} (industry division *k* and geographic area *l*) is

$$\hat{Y}(D_{kl}) = \sum_{j \in U} x_j \hat{\beta}_{kl} \delta_j(D_{kl}), \quad (2)$$

where $\delta_j(D_{kl})$ indicates whether unit *j* is in small domain D_{kl}; the summation is over all employers *j* within the state (or universe *U*); and $\hat{\beta}_{kl}$ is the ratio estimator within D_{kl}. With insufficient sample data to estimate the model parameters reliably at the small domain level, we instead estimate the parameters for model cell *m* (typically a 2–digit SIC at the state level), and apply the estimated model parameters to each of the small domains within the state. The resulting synthetic estimator is of the form

$$\hat{Y}(D_{kl}) = \sum_{m \in k} \sum_{j \in U_m} x_j \hat{\beta}_m \delta_j(D_{kl}), \quad (3)$$

where the first summation is over all model cells that overlap with domain D_{kl} and the second summation is over all employers within the model cell. The estimator is a simple sum of predicted employment over all employers in the universe within the domain.

We tried an intercept in the model and verified that it was not significantly different from zero, in most cases. We tested that the slope was significantly different from zero. We plotted the residuals to verify that they were suitably well behaved. We checked the R^2 values to quickly assess the explanatory power of the model.

To illustrate this work, Table 2 gives summary statistics for models in Trade using January 1996 CES and January 1995 ES202 data. All of the R^2 values in Table 2 are quite high, ranging from 0.87 to 0.96. Only two of the intercepts are significantly different from zero. Except for Retail Trade, Apparel, where the intercept is significantly different from zero, all of the slopes are between 0.9 and 1.1.

The largest employers are selected into the sample with certainty. Because they are so influential and not necessarily typical, we decided to exclude them from the estimation of the model parameters.

We also tried Estimator (3) corresponding to large sub-state model cells. This approach loses sample size (and thus precision) relative to the statewide model cells, but presumably gains some greater ability to target local economic conditions (thus reducing bias, if any). Yet in comparing the resulting small domain estimates with "true" values in simulations, we found the estimators from statewide model cells to have the smaller mean squared errors.

Following the work of Battese, Harter, and Fuller (1988), we fit a components-of-variance model of the form

$$y_{ij} = x_{ij}\beta + v_i + \varepsilon_{ij}, v_i \sim \text{NID}(0, \sigma_v^2), \varepsilon_{ij} \sim \text{NID}(0, \sigma_e^2 x_{ij}) \quad (4)$$

and tested the homogeneity of the county-level variance components, v_i. While there was some indication of heterogeneity, the variability in the variance component estimates actually increased the mean squared errors of the small domain estimates in our simulations. We decided that the variance components approach was not superior to the simple synthetic estimator.

Table 2
Generalized Regression Models for CES All Employment on ES202 Year-Ago Employment: Trade Industries

Industries Defined by 2–Digit SIC Code	n	R^2	Intercept		Slope	
Wholesale trade, durable goods	700	0.96	–0.061		1.015	**
Wholesale trade, nondurable goods	381	0.95	–0.032		0.978	**
Retail trade, building and garden supplies	189	0.96	0.420		0.918	**
Retail trade, general merchandise	42	0.95	–1.325		1.081	**
Retail trade, food stores	156	0.95	0.410		0.934	**
Retail trade, automobiles	379	0.97	0.130		0.971	**
Retail trade, apparel	112	0.90	1.320	**	0.750	**
Retail trade, furniture	110	0.95	0.242		0.931	**
Retail trade, eating & drinking establishments	460	0.89	0.382		0.968	**
Miscellaneous retail trade	332	0.87	0.810	**	0.915	**

* Significant at 0.05 level ** Significant at 0.01 level

We evaluated the synthetic estimator and other small domain estimators in a simulation study using Illinois data. The study included the simple unbiased estimator, the link relative estimator (Madow and Madow 1978, and West 1983, 1984), raked estimators using CES estimates at higher aggregations as marginal totals, two variations of generalized regression estimators (Särndal and Hidiroglou 1989), and three variations of synthetic estimators. For some of the simulations, the data were restricted to cases for which the CES and ES202 data could be cleanly linked. We then drew repeated samples from this "universe" and tested the results against "truth". For later simulations, the data files included non-matches with rules for special handling based on likely causes of the mismatches. The handling of non-matches is described in the next section.

In the simulations, we used all the samples and the known truth to compute bias, relative bias, mean squared error, and relative mean squared error of estimated total employment and month-to-month change in employment. We also plotted the 5^{th}, 50^{th}, and 95^{th} percentiles of the distribution of the estimators and examined the distributions in relation to the true values.

Results of the simulation study are reported in Harter *et al.* (1999). In general, we found that estimators that used ES202 as auxiliary data performed better than the direct sample-based estimator, the link relative estimator, and the raked estimators that used only sample data. The estimator that performed best overall was a variation of the synthetic estimator, derived from the prediction theory approach to survey sampling (Royall 1970, 1988, and Royall and Cumberland 1981a, 1981b). This estimator

$$\begin{aligned}\hat{Y}(D_{kl}) &= \sum_{m\in k}\sum_{j\in s_m} y_j\,\delta_j(D_{kl}) \\ &\quad + \sum_{m\in k}\sum_{j\notin s_m} x_j\,\hat{\beta}_m\,\delta_j(D_{kl}) \\ &= \sum_{m\in k}\sum_{j\in U_m} x_j\,\hat{\beta}_m\,\delta_j(D_{kl}) \\ &\quad + \sum_{m\in k}\sum_{j\in s_m} (y_j - x_j\,\hat{\beta}_m)\,\delta_j(D_{kl}) \qquad (5)\end{aligned}$$

is intuitively appealing to non-statisticians because the sample data are used directly for sample employers, while the model predictions are used only for nonsample employers. It is the synthetic estimator plus a sample-based correction for any lack of fit in the models.

4. Merging the Data

The success of the small domain estimator depends, in part, on the ability to accurately match the CES and ES202 data. We can match CES and ES202 records by unemployment insurance number (UI) and establishment or reporting unit number (RU). When the CES reporter is an aggregate of establishments, such as a multi-site employer reporting all employees together without distinguishing individual work sites, the corresponding ES202 records must be aggregated to match. Figure 3 demonstrates an isolated instance of a bad aggregate match.

Plots of the kind presented in Figures 1–3 enabled us to identify many miscoded observations. For example, an aggregate reporter coded in the files as containing all the company's work sites, but that actually covers only a single work site, should have been coded as a single establishment. The process of checking outliers in all the plots was time-consuming, but resulted in major improvements in the micro data, which in turn improved the estimated model parameters.

Several situations make the match process problematic. First, the ES202 data contain employers that have gone out of business. Conversely, the CES data contain new employers that were not in existence at the time the ES202 data were collected, although difficulty in identifying new businesses in a timely fashion makes this scenario less common. Births and deaths of businesses, then, cause real mismatches in the data.

Second, nonresponse to either the CES or ES202 causes mismatches. Missing or delinquent reporters to the ES202 are usually imputed for a time. At present, imputation is not done for missing CES cases. A key difficulty with both programs is distinguishing nonresponse from a death.

Third, businesses often reorganize, merge, acquire other businesses, divest divisions, and so on. Any of these status changes can cause states to assign new unemployment insurance numbers. The predecessor businesses and successsor businesses are treated as deaths and births. Alternatively, if a single predecessor can be linked to a single successor, their records could be joined to form one unified record. Unfortunately, the linkages are often not one-to-one. In many instances, predecessors are indistinguishable from deaths and delinquent CES reporters, and successors are indistinguishable from births and missing ES202 data.

For the initial implementation of our small domain estimator, we treat missing CES units as nonsample units; that is, we use their ES202 data and the model to predict their current month values. Since we cannot distinguish deaths and predecessors from missing CES data, we predict their current month employment using their ES202 data and the model. We use imputed ES202 data as real observations. Because it is relatively rare for a new business to appear in the CES sample data before it appears in the ES202, we treat CES records without ES202 counterparts as successor records. That is, in the small domain estimator, we treat them as nonmembers of the CES sample and predict their employment from the unmatched predecessor records in the ES202 file and the model. All of these decisions or judgments were based on IDES' experience.

Even if the UI and RU numbers match, the CES and ES202 records may differ in their industry or geographic

codes due to differences in the programs' update cycles. Discrepancies might represent errors or legitimate changes. Originally, our thought was to use the CES codes in the small domain estimator, assuming CES codes were the more current. However, as the small domain estimator was being implemented, more and more of the CES data collection operations were being transferred from Illinois' control to central data collection centers operated by the BLS. IDES felt this loss of control could compromise the quality of the CES codes and thus they decided to use the ES202 codes instead. In actual production, we use these classification codes for all purposes, including definition of model cells, estimation of the slope parameters, and calculation of the small domain estimates.

Sometimes a well-matched sample unit experiences employment shifts that are not typical of the industry or the region as a whole. Both the CES and ES202 systems allow for comment codes in the data files so that anomalies and their reasons can be flagged. We developed an extensive set of rules for determining when a matched sample record may be used in the estimation of model parameters, and when this would be unwise. For example, a drop in employment due to weather or climate conditions, such as flooding along the Mississippi River, is a situation likely to be common to other businesses in the area. A record with a code for this type of anomaly should probably be included in the estimation of model parameters. A fire, on the other hand, is likely to affect one and only one business, and a drop in employment due to the fire could be very misleading if applied to nonsample businesses. In this case, the sample unit with the fire stands for itself, but it is not part of the calculation of the model parameters.

All the potential data problems and potential mismatches led us to modify the estimator slightly. The revised estimator is

$$\hat{Y}(D_{kl}) = \sum_{m \in k} \sum_{j \in s_m} y_j \, \delta_j(D_{kl}) + \sum_{m \in i} \sum_{j \notin s_m} x_j \, \hat{\beta}_m \, \delta_j(D_{kl}) + A_{kl}, \qquad (6)$$

where A_{kl} is an additive adjustment for known data deficiencies. This concession to practical realities was originally intended for situations such as the addition of railroad workers, where Illinois' CES manager obtains information on railroad employment from the Railroad Retirement Board because railroad workers are not covered by the state unemployment insurance program, and thus are missing from the ES202 data file. Clergy and summer youth workers are often added the same way. The CES manager and affiliated local economists scattered throughout the state have found the adjustment option useful for other known problems, such as employees that are reported at headquarters when they are really located around the state. Employees whose location is unknown are usually assigned to a nonspecific county "999" for inclusion in statewide estimates, but traditionally have been omitted from sub-state estimates. With the adjustment option, the CES manager can allocate the county 999 employment to individual counties in proportion to other employees in the same industry. Major births and deaths can be reflected in the estimates through the adjustments until the CES and ES202 files can catch up.

The danger of this adjustment capability is that it can be used to force small domain estimates to conform to the CES manager's or economists' judgments, rather than letting the data and models speak for themselves. The best possible model is useless if it is ignored or "fudged".

Despite the danger, Estimator (6) is the one that we have actually moved into production in Illinois. All matched respondent records contribute to the first term. All matched records not designated as atypical or certainty contribute to the estimated slope in the second term. The summation in the second term includes nonmatched ES202 cases and missing sample cases – all cases that are treated as nonsample cases that month. If we have a CES record that does not match anything in ES202, it is dropped altogether. At present, all data adjustments, A, are coordinated and approved through the CES manager, who operates under strict guidelines, including a requirement to maintain consistency with the CES estimates published by the BLS. Within the guidelines, the manager is granted discretion to determine when adjustments are in the best interest of the estimation process.

5. Monitoring the Process

It is preferable to discover and fix data problems prior to estimation rather than rely on the adjustment capability in estimation. Illinois has developed several tools for monitoring the data that feed the monthly estimation process. Many of these tools reside in Illinois' software that pre-processes and matches the data prior to estimation.

Matching proceeds as a by-product of CES' daily processing activities. The editing and registry maintenance of CES records involves review of ES202 records, which are available to CES staff through simple "point and click" tools. The CES staff designates a match between CES and ES202 records by a special code manually applied to the CES record and later read by the pre-processing software. Those CES records so indicated as matched are subsequently checked for ES202 congruence and uniqueness on the combination of UI, RU, industry, ownership type, county, and delinquency status. The clean matches are added to a *matched file*, which is available for further review through special diagnostic or exception reports. We developed and implemented an extensive set of rules for the staff to follow in resolving the messy matches – the one to many and many to one matches. The pre-processing software executes the rules and prints all cases of a certain type in a table for staff review. After applying all the rules and

resolving the match statuses of the cases in the printed tables, we write remaining non-matching records to a separate *nonmatched file* for diagnostic reports and additional staff review.

From the matched file, we develop diagnostic or exception reports for CES staff. For instance, the pre-processing software generates a report of sample records whose CES and ES202 data differ more than one might expect. The basis for this exception report is a statistic derived from information theory. See Theil (1967), Strobel (1982), and Harter (1987). The statistic is computed for each sample observation as follows:

$$E_j = \frac{(y_j - x_j)^2}{(y_j + x_j)/2}. \qquad (7)$$

It is a Taylor series approximation of a measure of entropy and under the null hypothesis has a χ^2 distribution with 1 df. The statistic provides a way of ranking data differences, and balancing absolute differences, dominated by larger employers, and relative differences, dominated by smaller employers. The CES manager can evaluate the cases with the largest values of *E*, identifying and correcting miscoded data prior to small domain estimation.

Other exception reports display duplicate CES records that were removed from the files. Duplicates are rare but can happen, for example, if two respondents from the same company each file CES reports. The exception reports display for review single establishment records in CES incorrectly matched to an aggregation in the ES202 that were dropped by the pre-processing software. Also displayed for review are unmatched CES records that could represent a successor or a birth employer. Other specialized diagnostics check the sums of ES202 records at county, MSA, and statewide levels for comparison with their respective CES counterparts.

After going through these exception reports and making changes where appropriate, CES staff may decide to rerun the pre-processing software using the newly updated data, if the production schedule permits.

The software that computes the small domain estimates has a final data check built in. The input data values and the estimated model parameters are checked against tables of "sanity values" for reasonableness. This is a gross check only, designed to signal when something very unexpected has occurred.

The estimation system produces tables of matched sample data and tables of nonsample data at the individual reporting unit level. The authorized users of the small domain estimation software – the CES manager and the affiliated local economists, among others – can review the micro data as well as the computed estimates. Based on their review, they can provide useful guidance regarding specification of the adjustment term A_{kl}.

The CES manager and local economists review the estimates themselves along with historical estimates to see whether the trends and seasonality in the observed time series are reasonable. For instance, Construction, Retail Trade, and Education Services all have strong seasonal patterns. Deviation from such patterns would suggest to the analyst that further review is needed. Manufacturing employment is thought to be trending downward over the long term, and there is a natural tendency to examine its time series in this context.

Finally, the CES manager and local economists summarize all of the labor market areas into one large entity. The larger employment numbers allow sharper delineation of seasonal and trend expectancies. They also allow for subsequent comparison with statewide estimates.

6. Conclusion

Many aspects of small domain estimation must be checked and rechecked in production on a monthly basis. The auxiliary variable must be investigated carefully with respect to its correlation with the survey variable and its reliability, compatibility, and availability. The record linkage process is challenging (but highly rewarding) and requires vigilance. The models and assumptions underlying the estimator must be checked and verified for reasonableness. The estimates themselves must be scrutinized regularly. Development of the small domain estimator forcefully shows that even with the most ideal auxiliary variable and a textbook model, practical issues can intrude and require that flexibility be built into the estimation process.

References

Battese, G.E., Harter, R.M. and Fuller, W.A. (1988). An error-components model for prediction of county crop areas using survey and satellite data. *Journal of the American Statistical Association*, 83, 28-36.

Ghosh, M., and Rao, J.N.K. (1994). Small area estimation: An appraisal. *Statistical Science*, 9, 55-93.

Harter, R. (1987). Exception reporting: judging what is significant. *Nielsen Marketing Trends*, January. 20-23.

Harter, R., Wolter, K. and Macaluso, M. (1999). Small domain estimation of employment using CES and ES202 data. In *Statistical Policy Working Paper 30, 1999 Federal Committee on Statistical Methodology Research Conference*: *Complete Proceedings, Part 1 of 2*. Washington DC: Statistical Policy Office, Office of Information and Regulatory Affairs, Office of Management and Budget.

Madow, L., and Madow, W. (1978). On link relative estimators. *Proceedings of the Section on Survey Research Methods*, American Statistical Association, 534-539.

Purcell, N.J., and Kish, L. (1980). Postcensal estimates for local areas (or domains). *International Statistical Review*, 48, 3-18.

Royall, R.M. (1970). On finite population sampling theory under certain linear regression models. *Biometrika*, 57, 377-387.

Royall, R.M. (1988). The prediction approach to sampling theory. In *Handbook of Statistics*, (Eds. P.R. Krishnaiah and C.R. Rao). New York: North Holland. 6, 399-413.

Royall, R.M., and Cumberland, W.G. (1981a). An empirical study of the ratio estimator and estimators of its variance. *Journal of the American Statistical Association*, 76, 66-77.

Royall, R.M., and Cumberland, W.G. (1981b). The finite population linear regression estimator and estimators of its variance – an empirical study. *Journal of the American Statistical Association*, 76, 924-930.

Särndal, C.-E., and Hidiroglou, M.A. (1989). Small domain estimation: A conditional analysis. *Journal of the American Statistical Association*, 84, 266-275.

Singh, M.P., Gambino, J. and Mantel, H.J. (1994). Issues and strategies for small area data (with discussions). *Survey Methodology*, 20, 3-22.

Strobel, D. (1982). Determining outliers in multivariate surveys by decomposition of a measure of information. *Proceedings of Section on Business and Economic Statistics*, American Statistical Association.

Theil, H. (1967). Economics and Information Theory. *Studies in Mathematical and Managerial Economics*, (Ed. H. Theil). Amsterdam: North Holland.

West, S. (1983). A comparison of different ratio and regression type estimators for the total of a finite population. *Proceedings of the Section on Survey Research Methods*, American Statistical Association, 388-393.

West, S. (1984). A comparison of estimators for the variance of regression-type estimators in a finite population. *Proceedings of the Section on Survey Research Methods*, American Statistical Association, 170-175.