Combining Time Series and Cross-sectional Data for Current Employment Statistics Estimates October 2015

Julie Gershunskaya
U.S. Bureau of Labor Statistics, 2 Massachusetts Ave NE, Suite 4985, Washington, DC, 20212

Abstract

Estimates from the Current Employment Statistics (CES) Survey are produced based on the data collected each month from the sample of businesses that is updated once a year. In some estimation cells, where the sample is not large enough, the Fay-Herriot model is used to improve the estimates. Under the current approach, the model combines information from a set of areas and is estimated independently every month. Given the design of the survey, it may be beneficial to borrow information not only cross-sectionally but also over time. This paper explores the feasibility of applying such a model. The results are evaluated based on historical "true" employment data available on a lagged basis.

Key Words: small area estimation, Fay-Herriot model, Current Employment Statistics Survey

1. Introduction

Estimation for domains where the traditional direct sample based estimator lacks precision requires strengthening the estimator by using modeling assumptions. In the past several decades, the methodology for estimation in such "unplanned" domains has grown into a field of Small Area Estimation (SAE). The literature on the subject is rich and it is still growing (see Rao 2003; Pfeffermann 2002, 2013)

The quality of the result in SAE depends on the amount and relevance of the information summoned by the model. Sometimes, the parsimoniousness of the model and the ability to include more dimensions of the available data are at odds.

This paper considers application of alternative models in estimation of employment from the Current Employment Statistics (CES) survey conducted by the U.S. Bureau of Labor Statistics (BLS). Given the design of the survey, it is reasonable to expect that it is beneficial to base the model on information available not only across areas but also over time. This paper explores the feasibility[1] of applying such a model. The results are evaluated based on historical "true" employment data, available to CES on a lagged basis. Contrary to our expectations, the empirical results show that, in the case of the CES series considered in our research, the classical Fay-Herriot model that borrows information across areas at a given point in time works about as well as a more sophisticated Rao-Yu model that combines information over areas and time. One reason the results were so close is that both the Fay-Herriot and Rao-Yu models used in this research included the same predictor that captured most useful information regarding the estimates. Still, we were perplexed by the

[1] Any opinions expressed in this paper are those of the author and do not constitute policy of the Bureau of Labor Statistics.

observation that in a number of cases the simpler Fay-Herriot model performed slightly better than the more complete Rao-Yu model. We investigated possible reasons of this phenomena using the simulation study.

The paper is organized as follows. In Section 2, we introduce the CES setup: the data and the CES estimator. We talk about the reasons why borrowing information across time might be beneficial and discuss the covariance structure of the sampling errors in the CES series. We introduce the models in Section 3. In Section 4, we present results from the real data analysis. In Section 5, we use simulated data to study the effect of various values of the model parameters on the results of the model fit. The data is generated from models similar to the ones that are assumed to govern the real data.

2. Employment estimator in CES

Every month, CES computes estimates of the relative change in employment from the previous to current month. The estimation is performed for various domains defined by intersections of industry and geography.

The estimator of the employment level $Y_{d,T}$ in domain d at month T has the following form:

$$\hat{Y}_{d,T} = Y_{d,0}\hat{R}_{d(0,T)}, \tag{1}$$

where $Y_{d,0}$ is a known "true" employment level at month 0 (also referred to as the "benchmark" level) and $\hat{R}_{d(0,T)}$ is an estimate of the relative employment change from the base period 0 to T, the latter being the product of estimates of monthly trends $\hat{R}_{d(t-1,t)}$, $t = 1,...,T$,

$$\hat{R}_{d(0,T)} = \hat{R}_{d(0,1)}\hat{R}_{d(1,2)}...\hat{R}_{d(T-1,T)}. \tag{2}$$

(To avoid hindering the narrative with unnecessary details, (1) and (2) present a slightly simplified version of the estimator compared to what actually is used in production.)

We note that the finite population parameters of interest in domain d are both the employment levels $Y_{d,t}$ at months $t = 1,...,T$ (it also can be viewed as the cumulative change from the base period to month t) and employment changes over m months, $\Delta Y_{d(t-m,t)} = Y_{d,t} - Y_{d,t-m}$. Specifically, at a given month t, the target finite population quantity of interest is the one-month relative change

$$R_{d(t-1,t)} = \frac{\sum_{j \in P^{(d)}} y_{jt}}{\sum_{j \in P^{(d)}} y_{j,t-1}}, \tag{3}$$

where y_{jt} is the employment of business j at time t; $P^{(d)}$ is the set of population units in the domain. The sample based estimator of $R_{d(t-1,t)}$ is

$$\hat{R}_{d(t-1,t)} = \frac{\sum_{j \in s_t^{(d)}} w_j y_{jt}}{\sum_{j \in s_t^{(d)}} w_j y_{j,t-1}}, \quad (4)$$

where w_j is the sampling weight of unit j and $s_t^{(d)}$ is a set of units sampled in the domain and used in the estimation at month t (generally, the sets of responding units used in the monthly estimation differ from month to month.)

The estimator of levels is considered approximately unbiased:

$$\hat{Y}_{d,t} = Y_{d,t} + e_{d,t},$$

where $e_{d,t}$ is the sampling error, uncorrelated across domains, with $E\left(e_{d,t}\right) = 0$.

Since the sets of respondents $s_t^{(d)}$ largely overlap during the estimation period, sampling errors are correlated over time. Let us assume the following stationary autoregressive model for the sampling errors:

$$e_{d,t} = \rho_e e_{d,t-1} + \varepsilon_{d,t}, \quad |\rho_e| < 1, t = 1,...,T \quad (5)$$

where $E\left(\varepsilon_{d,t}\right) = 0; Var\left(\varepsilon_{d,t}\right) = \sigma_d^2; E\left(\varepsilon_{d,t}\varepsilon_{d,s}\right) = 0$ for $t \neq s$.

The model implies that the variance of $\hat{Y}_{d,t}$ is

$$Var\left(\hat{Y}_{d,t}\right) = \sigma_d^2 \frac{1-\rho_e^{2t}}{1-\rho_e^2}$$

and for large t it nears $V_d = \frac{\sigma_d^2}{1-\rho_e^2}$.

Covariance between the level estimates at times $t-m$ and t is $\mathrm{cov}(\hat{Y}_{d,t}, \hat{Y}_{d,t-m}) = \rho_e^m V_d$. Previous research shows that correlations between the level estimates in consecutive months are high, in the vicinity of 0.8 to 0.9.

For estimates of monthly changes, $\Delta\hat{Y}_{d,t} = \hat{Y}_{d,t} - \hat{Y}_{d,t-1} = \Delta Y_{d,t} + \Delta e_{d,t}$, the variance is

$$Var\left(\Delta Y_{d,t}\right) = Var\left(e_{d,t} - e_{d,t-1}\right) = 2V_d\left(1-\rho_e\right)$$

and the covariance is

$$Cov\left(\Delta\hat{Y}_{d,t}, \Delta\hat{Y}_{d,t+1}\right) = E\left(e_{d,t} - e_{d,t-1}\right)\left(e_{d,t+1} - e_{d,t}\right) = -\left(1-\rho_e\right)^2 V_d \ .$$

Correlation between changes in the adjacent months is

$$Corr\left(\Delta\hat{Y}_{d,t}, \Delta\hat{Y}_{d,t+1}\right) = -\frac{1}{2}\left(1-\rho_e\right). \qquad (6)$$

(See empirical results in Scott et al. 2012, Scott and Sverchkov 2005.)

Barring the noise in the direct esti mates of correlations between sam pling errors in the estimates of changes, the previous resear ch, generally , supports the conclusion that correlations between the adjacent months are negative, approximately -0.1. Due to the noisy esti mates, it is even more difficu lt to discern a definitive pattern in correlations between periods that are more than 1 m onth apart. For this paper, we assu me that model (5) for the sampling errors holds.

3. The Rao-Yu Model for the CES Series

It is a common assumption that the relative over-the-month changes from the same month in previous years serve as good predictors for the current relative over-the-month changes. True values for historical employment counts are available fro m the Quarterly Census of Employment and Wages (QCEW), anot her BLS program. Auxiliary variable $X_{d,t}$ is the relative over-the-month change in employment at month t in cell d as forecasted from the historical QCEW data.

The m odels below are for mulated for relative monthly changes. Note that $R_{d(t-1,t)}$ is usually close to 1. Thus, we have the following approximate formulas.

Variance: $Var\left(\hat{R}_{d(t-1,t)}\right) \doteq \frac{1}{Y_{d,t-1}^2} Var\left(\Delta\hat{Y}_{d,t}\right).$

Covariance: $Cov\left(\hat{R}_{d(t-1,t)}, \hat{R}_{d(t,t+1)}\right) \doteq \frac{1}{Y_{d,t-1}Y_{d,t}} Cov\left(\Delta\hat{Y}_{d,t}, \Delta\hat{Y}_{d,t+1}\right).$

Correlation: $Corr\left(\hat{R}_{d(t-1,t)}, \hat{R}_{d(t,t+1)}\right) \doteq Corr\left(\Delta\hat{Y}_{d,t}, \Delta\hat{Y}_{d,t+1}\right) = -\frac{1}{2}\left(1-\rho_e\right).$

To simplify notation in the models formulation, we denote:

$y_{d,t} \equiv \hat{R}_{d(t-1,t)}.$

The Fay-Herriot (FH) model that is currently used for select CES series at the statewide industrial supersector level is formulated independently for each month. At month t, for domains $d = 1,...,M$,

$$y_{d,t} = \beta_t X_{d,t} + u_{d,t} + e_{d,t}, \tag{7}$$

where the random terms $u_{d,t}$ and $e_{d,t}$ are mutually independent and

$$u_{d,t} \overset{iid}{\sim} N\left(0, \sigma_{u,t}^2\right) \text{ and } e_{d,t} \overset{ind}{\sim} N\left(0, \sigma_d^2\right),$$

with variances σ_d^2 of the sampling errors considered known.

The Rao-Yu (RY) model for the CES case is formulated for domains $d = 1,...,M$ as

$$\begin{aligned} y_{d,t} &= \beta_t X_{d,t} + v_d + u_{d,t} + e_{d,t}, \\ u_{d,t} &= \rho u_{d,t-1} + \zeta_{d,t}, \qquad |\rho| < 1. \end{aligned} \tag{8}$$

where

random terms $v_d, e_{d,t}, \zeta_{d,t}$ are mutually independent;

$v_d \overset{iid}{\sim} N\left(0, \sigma_v^2\right)$ are random effects representing variation between areas;

$\zeta_{d,t} \overset{iid}{\sim} N\left(0, \sigma_u^2\right)$;

ρ is the correlation between random effects $u_{d,t-1}$ and $u_{d,t}$ at two consecutive time points.

The covariance matrix for the sampling errors is assumed known. It has the block-diagonal structure. The block corresponding to domain d has the following structure:

$$\mathrm{Cov}\left(\mathbf{e}_d\right) = \sigma_d^2 \mathbf{B},$$

where

$$\mathbf{e}_d = \left(e_{d,1},...,e_{d,t}\right)^T,$$

σ_d^2 is the variance for $e_{d,t}$,

$\mathbf{B}$ is a $T \times T$ symmetric matrix having 1 on the diagonal and $-0.5\rho_e^{|i-j|-1}\left(1-\rho_e\right)$ at the off-diagonal position $j, i \neq j$.

Parameter β_t reflects differences between the history-based movements $X_{d,t}$ and the current tendency. Besides serving as adjustment to historical movements based on the most current CES data, β_t also acts as the correction factor for the differences in seasonality between the CES and QCEW series. This is the main reason for having the month specific coefficient, as indicated by subscript t.

Covariance matrices for the time and area random effects $u_{d,t}$ and v_d depend on unknown parameters σ_u^2, σ_v^2, and ρ. As noted above, the covariance matrix of sampling errors is considered known. This is required for model to be identifiable. In practice, it is populated by variances and covariances obtained based on previous research (an approach often involves fitting a generalized variance function.) For surveys where the same sample or a portion of the sample is used repeatedly during the estimation period, as in CES, the sample based estimates in a given area are correlated over time. Ability to account for the correlated sampling errors is one point supporting the use of the Rao-Yu model instead of the cross-sectional model Fay-Herriot.

It is noted, based on the results of Rao and Yu (1994) simulation study, that the smaller the variance associated with the time random effect σ_u^2 and the larger the variance associated with the area random effect σ_v^2, the stronger the gains from using the Rao-Yu model over the cross-sectional Fay-Herriot model.

Given the structure of the CES data, the use of information both across time and domains looks appealing. On the other hand, the Rao-Yu model is more complicated: it contains more parameters that need to be estimated from the data; in addition, it has parameters that need to be used as known – in practice, this requires further assumptions. Motivated by results from the CES real data example, we are trying to explore some of the conditions justifying the use of the Rao-Yu model over a simpler, Fay-Herriot, model.

4. Results for the CES Series

States within different industries define the sets of domains to which we fit our models. The estimation is performed for each of the 12 months of the estimation period. For example, at month 5 after the starting point, we fit the Rao-Yu model to estimate relative change at month 5 based on the information available from all preceding months, 1 through 5; at month 12 after the starting point, we can use information available from months 1 through 12. Estimates for the first two months are obtained using only the Fay-Herriot model. We use "Small Area Estimation: Time-series Models" sae2 R package (http://cran.r-project.org/web/packages/sae2/sae2.pdf) to fit the Rao-Yu model. The true population values are available from QCEW program several months after the actual estimation. This enables us to compare results of estimation with the population target. Due to differences in seasonality between the CES series and the QCEW administrative data source, the most meaningful sets of comparison is after 12 months of estimation. Results from 4 years of estimation are presented in Tables 1-4.

Table 1: October 2010 - September 2011 estimation period

Industry NAICS code	*M*	*Mean Absolute Revision*		*RY Parameter Estimates and standard errors*					
		FH	RY	*rho*		*sig2_u*		*sig2_v*	
10000000	44	**1,019**	1,024	0.00	(0.13)	0.65	(0.11)	0.00	(0.02)
20000000	44	3,894	**3,046**	0.23	(0.09)	1.41	(0.16)	0.16	(0.09)
31000000	47	**3,550**	3,870	0.89	(0.08)	0.44	(0.09)	0.00	(1.03)
32000000	47	2,251	**1,774**	0.14	(0.16)	0.46	(0.10)	0.08	(0.04)
41000000	51	2,361	**1,674**	0.00	(0.32)	0.17	(0.07)	0.10	(0.03)
42000000	51	2,614	**2,442**	0.00	(0.21)	0.28	(0.08)	0.00	(0.02)
43000000	51	1,751	**1,625**	0.00	(0.74)	0.07	(0.07)	0.03	(0.02)
50000000	51	**1,283**	1,392	0.00	(0.11)	0.74	(0.11)	0.02	(0.03)
55000000	51	2,807	**2,625**	0.76	(0.19)	0.12	(0.07)	0.04	(0.12)
60000000	32	4,174	**3,296**	0.28	(0.14)	0.69	(0.14)	0.02	(0.05)
60540000	19	4,018	**3,703**	0.69	(0.72)	0.05	(0.11)	0.00	(0.09)
60550000	19	1,994	**1,990**	0.00	(0.13)	1.38	(0.24)	0.07	(0.08)
60560000	19	7,299	**5,598**	0.77	(0.40)	0.09	(0.10)	0.00	(0.17)
65610000	24	3,717	**2,560**	0.00	(0.13)	1.21	(0.20)	0.00	(0.05)
65620000	24	**3,182**	3,263	0.95	(133.12)	0.00	(0.05)	0.01	(2.15)
70710000	24	**2,124**	2,273	0.40	(0.36)	0.19	(0.12)	0.00	(0.04)
70720000	24	3,686	**3,380**	0.00	(0.66)	0.12	(0.10)	0.01	(0.02)
80000000	51	2,391	**2,105**	0.00	(0.25)	0.23	(0.08)	0.04	(0.02)

Table 2: October 2011 - September 2012 estimation period

Industry NAICS code	*M*	*Mean Absolute Revision*		*RY Parameter Estimates and standard errors*					
		FH	RY	*rho*		*sig2_u*		*sig2_v*	
10000000	44	970	**704**	0.73	(0.27)	0.09	(0.07)	0.00	(0.09)
20000000	44	3,386	**3,108**	0.04	(0.23)	0.28	(0.09)	0.03	(0.02)
31000000	47	2,819	**1,989**	0.86	(0.46)	0.03	(0.05)	0.00	(0.23)
32000000	47	1,296	**1,228**	0.87	(1.05)	0.01	(0.05)	0.00	(0.23)
41000000	51	**1,244**	1,305	0.69	(1.41)	0.02	(0.06)	0.05	(0.05)
42000000	51	**1,976**	2,004	0.00	(57.87)	0.00	(0.06)	0.02	(0.01)
43000000	51	1,808	**1,475**	0.94	(2.97)	0.00	(0.04)	0.00	(0.97)
50000000	51	1,007	**1,002**	0.00	(0.07)	1.68	(0.17)	0.00	(0.04)
55000000	51	1,773	**1,677**	0.00	(0.81)	0.06	(0.07)	0.02	(0.01)
60000000	32	3,515	**2,685**	0.92	(2.07)	0.01	(0.05)	0.00	(0.75)
60540000	19	**3,953**	4,045	0.74	(1.24)	0.02	(0.09)	0.00	(0.11)
60550000	19	2,254	**2,214**	0.98	(5.58)	0.00	(0.05)	0.00	(16.55)
60560000	19	8,556	**8,532**	0.00	(0.30)	0.35	(0.14)	0.00	(0.03)
65610000	24	3,933	**2,884**	0.00	(0.16)	0.75	(0.16)	0.00	(0.04)
65620000	24	**4,265**	4,654	0.00	(9.37)	0.01	(0.09)	0.08	(0.03)
70710000	24	**1,700**	2,344	0.00	(0.72)	0.10	(0.10)	0.00	(0.02)
70720000	24	2,256	**2,151**	0.00	(3.52)	0.02	(0.09)	0.00	(0.02)
80000000	51	1,409	**1,239**	0.00	(35.19)	0.00	(0.06)	0.01	(0.01)

Table 3: October 2012 - September 2013 estimation period

Industry NAICS code	*M*	*Mean Absolute Revision*		*RY Parameter Estimates and standard errors*					
		FH	RY	*rho*		*sig2_u*		*sig2_v*	
10000000	44	951	**811**	0.02	(507.46)	0.00	(0.07)	0.08	(0.03)
20000000	44	3,402	**2,837**	0.11	(0.17)	0.41	(0.10)	0.00	(0.02)
31000000	47	2,783	**2,008**	0.91	(0.55)	0.02	(0.04)	0.00	(0.61)
32000000	47	**1,303**	1,318	0.93	(1.72)	0.01	(0.04)	0.00	(0.80)
41000000	51	1,248	**1,181**	0.84	(133.09)	0.00	(0.05)	0.03	(0.15)
42000000	51	**2,409**	2,500	0.00	(42.77)	0.00	(0.06)	0.02	(0.01)
43000000	51	1,871	**1,579**	0.00	(3.39)	0.01	(0.06)	0.03	(0.02)
50000000	51	1,121	**1,101**	0.00	(0.08)	1.52	(0.16)	0.00	(0.04)
55000000	51	**1,436**	1,602	0.00	(0.45)	0.12	(0.07)	0.02	(0.02)
60000000	28	**2,139**	2,258	0.98	(36.83)	0.00	(0.04)	0.01	(12.99)
60540000	23	3,327	**3,151**	0.81	(227.41)	0.00	(0.07)	0.10	(0.19)
60550000	23	1,675	**1,617**	0.00	(1.35)	0.06	(0.10)	0.05	(0.03)
60560000	23	3,747	**3,708**	0.55	(0.61)	0.08	(0.11)	0.00	(0.05)
65610000	48	2,367	**1,736**	0.00	(0.14)	0.50	(0.10)	0.00	(0.02)
65620000	48	**4,989**	5,038	0.00	(0.33)	0.17	(0.08)	0.01	(0.02)
70710000	36	**782**	1,170	0.22	(512.93)	0.00	(0.08)	0.00	(0.02)
70720000	36	3,220	**2,177**	0.01	(0.24)	0.30	(0.10)	0.00	(0.02)
80000000	51	1,299	**1,268**	0.00	(462.81)	0.00	(0.06)	0.00	(0.01)

Table 4: October 2013 - September 2014 estimation period

Industry NAICS code	*M*	*Mean Absolute Revision*		*RY Parameter Estimates and standard errors*					
		FH	RY	*rho*		*sig2_u*		*sig2_v*	
10000000	44	598	**596**	0.00	(489.45)	0.00	(0.07)	0.01	(0.01)
20000000	44	**3,045**	3,049	0.00	(0.14)	0.60	(0.11)	0.03	(0.03)
31000000	47	1,699	**1,590**	0.98	(85.08)	0.00	(0.03)	0.01	(9.86)
32000000	47	993	**947**	0.98	(31.77)	0.00	(0.03)	0.02	(10.01)
41000000	51	1,015	**997**	0.00	(0.43)	0.12	(0.07)	0.03	(0.02)
42000000	52	**3,671**	3,783	0.00	(0.14)	0.48	(0.09)	0.00	(0.02)
43000000	52	1,324	**1,212**	0.00	(0.25)	0.23	(0.08)	0.02	(0.02)
50000000	51	**725**	978	0.00	(0.23)	0.25	(0.08)	0.01	(0.02)
55000000	51	**1,311**	1,467	0.00	(315.16)	0.00	(0.06)	0.04	(0.02)
60000000	19	2,131	**2,122**	0.00	(0.34)	0.30	(0.13)	0.00	(0.03)
60540000	33	2,912	**2,587**	0.00	(0.55)	0.12	(0.09)	0.02	(0.02)
60550000	33	1,297	**1,128**	0.17	(0.22)	0.35	(0.11)	0.00	(0.03)
60560000	33	5,036	**4,974**	0.00	(3.90)	0.02	(0.08)	0.01	(0.01)
65610000	48	1,875	**1,668**	0.00	(0.14)	0.50	(0.10)	0.00	(0.02)
65620000	48	3,902	**3,567**	0.00	(0.55)	0.10	(0.07)	0.02	(0.02)
70710000	39	**1,544**	1,550	0.00	(0.78)	0.07	(0.08)	0.00	(0.01)
70720000	39	2,894	**2,242**	0.06	(0.16)	0.50	(0.11)	0.00	(0.03)
80000000	51	1,777	**1,728**	0.00	(0.18)	0.36	(0.08)	0.00	(0.02)

The results show no clear advantage of using the Rao-Yu model over the Fay-Herriot model: mean absolute revisions after 12 months of estimation are generally close. There are industries where the Rao-Yu model results are somewhat better in all 4 years (e.g., Transportation, Education, Accommodation and Food Services, Other Services), in other industries, one model is better than the other in one year while the opposite is true in another year; in industry 70710000 (Arts, Entertainment, and Recreation), the Fay-Herriot model worked better in all 4 years.

One reason why there was no clear benefit from using the Rao-Yu model is that the variance of the area random effects was small relative to the sampling error or to the variance of the time effect. Possible misspecification of the sampling error matrix may also contribute to the result. Indeed, by the defined setup of the cross-sectional Fay-Herriot model case, sampling errors do not correlate over time. Thus the sampling variance matrix, the "known" component of the model, is diagonal, which is simpler than the block-diagonal structure of the "known" matrix when one decides to include the knowledge of the over-time correlation in the model.

To test the above conjectures, we performed simulations (presented in the next section).

5. Investigation Based on Simulated Data

In this section, we use simulated data to study the effect of the model parameters and errors in the sampling error variances on the results of the model fit.

As can be seen from the previous section, the variance of the area random effect is close to zero. This is the worst scenario if one counts on taking advantage from using information over time with the Rao-Yu model. Still, even in this case, it is possible to benefit from accounting for the sampling error correlation. Our simulations, indeed, show that this is the case. However, one must remember that the sampling error covariance structure is known only in theory. In practice, we use some estimated values and assumptions about the covariance structure as if they were true and known.

We generated data from the following model:

$$y_{d,t} = 2 + v_d + u_{d,t} + e_{d,t}, \qquad (9)$$

for $d = 1,...,20$ areas and $t = 1,...,12$ time points.

Random terms $v_d, e_{d,t}, u_{d,t}$ are generated independently:

$u_{d,t} \overset{iid}{\sim} N\left(0, \sigma_u^2\right)$ with $\sigma_u^2 = 0.25$

$v_d \overset{iid}{\sim} N\left(0, \sigma_v^2\right)$ with two choices for the values of σ_v^2

a. $\sigma_v^2 = 0$

b. $\sigma_v^2 = 0.25$

Sampling error structure:

$E\left(e_{d,t}\right)=0$

$Var\left(e_{d,t}\right)=1.$

The "employment level" error correlation between adjacent months is assumed to be $\rho_e=0.7$. Then "employment one-month change" error correlation is $-0.5\left(1-\rho_e\right)=-0.15$; the covariance matrix for errors of "employment changes" is block-diagonal; each block is $T\times T$ symmetric matrix having 1 on the diagonal and $-0.5\rho_e^{|i-j|-1}\left(1-\rho_e\right)$ at off-diagonal positions $j,i\neq j$.

We consider several versions of the assumed error structure as used at the time we fit the model. First, we may erroneously assume that the sampling errors are independent over time; second, we may use the true, correct variance structure, the same as was used to generate the model. In addition, we consider the situation where the variances of the sampling errors are estimated with error. To model this, we assume that the variance estimates are gamma-distributed $Gamma\left(k,\theta\right)$ with shape $k=1/3$ and scale $\theta=3.$ Thus, this corresponds to the unbiased variance estimates (the expectation is 1) with the variance of the variance estimates equal 3. The situation where variances are estimated with sizable errors is plausible with the employment data. The employment numbers have a highly skewed distribution; the employment changes are concentrated around zero with smaller proportion of businesses having significant changes in employment while yet smaller proportion having extreme large positive or negative changes.

The simulation study is based on 500 simulation runs for $t=1,...,T$, where $T=3,...,12.$ We present results for models using $T=6$ and $T=12$ points of "history". To fit the Rao-Yu model, we used the method of moments as given in Rao and Yu (1994). This method provided approximately the same results as the REML-based sae2 R package that we used for the real data. The advantage of using this method rather than REML was that it works significantly faster. Instead of estimating the model correlation parameter, we assumed it to be 0, i.e., equal to the true model parameter, which in the case of simulation is known to us.

Since all the areas are equally distributed, the empirical mean squared error was calculated by both averaging the errors across areas and simulations. Thus, the simulation error is based on the actual simulation size of $20\times 500=10,000$ trials:

$$MSE\left(E\right)=\sum_{s=1}^{500}\sum_{d=1}^{20}\left(E_{d,s}-\theta_{d,s}\right)^2 \quad \text{for } E=\text{Direct, FH, or RY based estimate.}$$

The relative efficiency of RY over FH was computed as

$$RE=100\%\frac{MSE\left(RY\right)-MSE\left(FH\right)}{MSE\left(FH\right)}.$$

As can be seen from Table 5, when there is no error in the variance estimates, the Rao-Yu model is more efficient than the Fay-Herriot model. This is true even for the case where the area random effect is absent ($\sigma_v^2 = 0$), even for the case where the sampling errors are wrongly assumed to be independent. With the existing area random effect, the efficiency of Rao-Yu over Fay-Herriot increases to over 30%.

Table 5: Mean squared error based on 500 simulation runs for different model parameters and assumptions on covariance structure of the sampling errors

Sampling Error Correlation		*Error in Sampling Variances*	*Direct*		*FH*		*RY*		*RE,%*	
True	*Assumed*		*T=6*	*T=12*	*T=6*	*T=12*	*T=6*	*T=12*	*T=6*	*T=12*
$\sigma_u^2 = 0.25, \quad \sigma_v^2 = 0$										
-0.15	0	None	0.998 1	.023 0	.284 0	.288	0.259 0	.252	-8.7	-12.5
-0.15	-0.15	None	0.998 1	.023 0	.284 0	.288	0.256 0	.249	-10.0	-13.6
-0.15	0	Gamma	0.998 1	.023 0	.606 0	.631	0.662 0	.693	9.3	9.9
-0.15	-0.15	Gamma	0.998 1	.023 0	.606 0	.631	0.687 0	.708	13.3	12.2
$\sigma_u^2 = 0.25, \quad \sigma_v^2 = 0.25$										
-0.15	0	None	1.016 1	.026 0	.410 0	.415	0.313 0	.274	-23.7	-33.8
-0.15	-0.15	None	1.016 1	.026 0	.410 0	.415	0.289 0	.258	-29.6	-37.9
-0.15	0	Gamma	1.016 1	.026 0	.689 0	.698	0.728 0	.682	5.6	-2.2
-0.15	-0.15	Gamma	1.016 1	.026 0	.689 0	.698	0.726 0	.705	5.3	1.0

The situation is drastically different when the "known" sampling error variances are generated from the $Gamma(1/3, 3)$ distribution. This results in the increase of the mean squared error in both Rao-Yu and Fay-Herriot based estimates; yet the MSE of the FH-based estimates is lower than the MSE of the RY-based estimates. It is also interesting to note that the assumption of the diagonal sampling error covariance structure leads to lower MSE in the RY-based results as compared with the results based on the correct assumption that the matrix is block-diagonal.

6. Summary

We explored advantages of using the Rao-Yu model that utilizes information from time as well as cross-sectionally, as compared to the cross-sectional-only Fay-Herriot model. The empirical results showed that, in the case of the CES data, there is no clear advantage from applying the Rao-Yu model. In the attempt to understand the nature of these mixed results, we performed the simulation study. We showed that misspecification in the estimated sampling variances, ordinarily considered fixed and known in both models, affects the results in such a way that the Fay-Herriot-based model may become more efficient compared to the Rao-Yu model.

References

Fay, R.E., and Herriot, R.A. (1979), Estimates of Income for Small Places: An Application of James-Stein Procedures to Census Data. *Journal of the American Statistical Association*, 74, 269-277.

Pfeffermann, D. (2002). Small area estimation - new developments and directions. *Int. Statist. Rev.* 70 125-143.

Pfeffermann, D. (2013). New Important Developments in Small Area Estimation. *Statistical Science*.28. 40–68.

Rao, J.N.K. (2003), *Small Area Estimation,* John Wiley & Sons, Hoboken, NJ.

Rao, J.N.K. and Yu, M. (1994), Small Area Estimation by Combining Time Series and Cross-Sectional Data. *Canadian Journal of Statistics,* 22, 511-528.

Scott, S. and Sverchkov, M. (2005), Variance Measures for X-11 Seasonal Adjustment: A Summing Up of Empirical Work. *ASA Proceedings of the Joint Statistical Meetings.*

Scott, S., Pfeffermann, D., and Sverchkov, M. (2012). Estimating Variance in X-11 Seasonal Adjustment. In *Economic Time Series: Modeling and Seasonality*, edited by William R. Bell, Scott H. Holan, and Tucker S. McElroy, 185–210. London: Chapman and Hall.