DE GRUYTER OPEN

Journal of Official Statistics, Vol. 34, No. 2, 2018, pp. 543–555, http://dx.doi.org/10.2478/JOS-2018-0025

A Study of Small Area Estimation for Italian Structural Business Statistics

Orietta Luzi[1], *Fabrizio Solari*[2], *and Fabiana Rocci*[2]

The Frame SBS is a statistical register which has been developed at the Italian National Statistical Institute to support the annual estimation of structural business statistics (SBS). Actually, a number of core SBS are estimated by combining microdata directly supplied by different administrative sources. In this context, more accurate estimates for those SBS that are not covered by administrative sources can be obtained through small area estimation (SAE). In this article, we illustrate an application of SAE methods in the framework of the Frame SBS register in order to assess the potential advantages that can be achieved in terms of increased quality and reliability of the target variables. Different types of auxiliary information and approaches are compared in order to identify the optimal estimation strategy in terms of precision of the estimates.

Key words: Small area estimation; statistical register; administrative data.

1. Introduction

The availability of high quality detailed information on businesses is essential for assessing the competitiveness and performance of modern economic systems, and facilitating the development of policy measures focused on guaranteeing productivity and employment growth.

In this context, European Regulations require that National Statistical Institutes (NSIs) collect information on businesses in order to provide data on wealth creation, investments and labor input in different economic activities. Proper information on businesses across the EU can allow to analyze structural shifts and possible country specializations in particular activities, sectoral productivity and profitability. EU policymakers and analysts are particularly interested in statistical data at enterprise size level of detail in terms of number of employees, allowing in-depth analysis of entrepreneurship and its role in the economic system. In fact, small and medium-sized enterprises are often referred to as the backbone of the European economy, providing a potential source of jobs and economic growth. In this context, the EU and the National Statistical Institutes can further strengthen their role as data providers of official and business-relevant data. They can aim to be able to clearly identify and consistently track the business characteristics and performance of subpopulations of businesses that are of primary interest for business analysts and policymakers. To this end, very detailed

[1] Istat, Via C. Balbo 39, Rome 00184, Italy. Email: luzi@istat.it
[2] Istat, Via C. Balbo 16, Rome 00184, Italy. Emails: solari@istat.it and rocci@istat.it

analyses are needed to detect vulnerabilities, evaluate adjustment capability and structural change in an economic system, and assess the effectiveness of policy measures (Garda and Zieman 2014; Caldera et al. 2015) that have an effect on identified sets (subdomains) of enterprises.

However, over the past decade, it is recognized that established economic statistics do not completely fulfill the required utility for conducting comprehensive and detailed economic analyses, especially in complex and heterogeneous economic systems. Statistical production systems that exploit directly collected survey data have shown some limits in terms of costs and effectiveness. Information from secondary data sources is recognized as a powerful instrument for achieving more reliable and detalied results. In particular, administrative archives may considerably expand the potential for wide-ranging economic analyses, while simultaneously reducing the respondents' burden, as well as costs for the administering statistical institute. In fact, the direct use of administrative data to produce official economic statistics has spread out. It is based on different strategies that involve supporting existing processes with additional sources of information, or designing new statistical production processes with administrative data as a primary source of information, possibly complemented by statistical survey data. As many administrative data sources provide incomplete coverage of the variables/population of interest, these data may need to be integrated with other administrative data sources. Indeed, different statistical methodologies may need to be implemented within the same statistical program to account for varied features and sources of data. A number of complex methodological issues are to be managed, such as the assessment of the statistical usability and quality of input data, the micro-integration of data at unit level (Jang 2016; Di Zio 2016), the harmonization of administrative and statistical concepts, and variables integration (Laitila et al. 2011). Deterministic corrections and derivation rules are usually used for such applications, as some form of micro-editing may be needed to achieve consistency between different target variables observed in different sources (Di Zio and Luzi 2014). For an analysis of potential benefits and risks deriving from the use of administrative data, see Wallgren and Wallgren (2007) and the outcomes of European projects (e.g., Memobust 2014; ESSnet AdminData 2013; ESSnet on Data Integration 2011).

In the Italian context, the Italian National Statistical Institute (Istat) is progressively moving away from a production model essentially based on directly-collected survey data complemented by secondary information towards a new mixed estimation strategy based on the implementation of a system of statistical registers. With the mixed estimation strategy, administrative data are extensively exploited; directly-collected survey data are used for estimating subpopulations or variables that are not available from the fully-assembled administrative data. This approach is expected to overcome some of the limitations related to both the quality and the usefulness of statistical products that are entirely based on directly-collected survey data.

In this context, as highly reliable administrative data are available on businesses populations and economic variables, Istat has developed a new statistical register called *Frame*. This register contains data from several administrative archives and supports the annual production of structural business statistics required by the Eurostat's Regulation (EC) No 295/2008. It also represents a source of information for the majority of the Istat structural business statistics that play a central role in the analysis of business productivity and competitiveness.

The *Frame* register consists of population microdata on a number of key profit-and-loss accounts (*core*) variables for Italian small and medium enterprises, comprising enterprises with less than 99 persons employed (Luzi et al. 2014). Domain estimates of the *subcomponents* of *core* items are obtained based on a design-based/model-assisted approach that exploits the structural constraints between the two groups of variables, such as additivity constraints (subcomponent to core total) and correlation structure. Nowadays, there are other important loss-and-profit accounts variables for which neither the available administrative sources directly supply suitable information, nor the methodology adopted for the *subcomponents* of the *core* variables could be used. At present, these variables are still estimated entirely from the directly-collected sample data. However, many of these variables measure rare characteristics, and the sample data may not provide estimates of adequate precision. In order to achieve quality gains for highly detailed domains, alternative estimation methodologies are being studied for these items, varying the considered methods by item according to their particular features and potentially related available information. In this framework, small area estimation (SAE) methods (Rao and Molina 2015) could play a central role in developing more efficient estimates than the corresponding estimates from directly-collected survey estimates (hereafter referred to as direct estimates).

In this article, we present a study that assesses the usability and the potential of applying small area estimation to measure the profit-and-loss accounts variable *Total depreciation of fixed assets*. The ultimate objective is to improve the current estimation procedure for this variable, for the required estimation domains, with a more efficient estimation approach that exploits the increased amount of auxiliary information from registers. Indeed, different variables drawn from distinct registers are used as auxiliary information in order to assess the improvement that the new *Frame* register can provide. As the SAE estimates whose model assumptions are violated could be biased, we consider applying a variety of different SAE methods to the selected variable. The small area models that we consider in this article rely on explicit models that relate the individual unit data to area level covariates obtained from auxiliary data, effectively "borrowing strength" from the larger-area estimates. The utilized empirical best linear unbiased predictor estimators simultaneously account for the variance of the direct estimates and the model variance, weighting the contribution from both accordingly.

The article is structured as follows. Section 2 describes the main features of the new Structural Business Statistics estimation strategy involving the *Frame* register. The evaluated SAE methods are presented in Section 3. Section 4 presents the case study and the obtained results. We provide our conclusions and discuss planned future work in Section 5.

2. The Italian Strategy for Structural Business Statistics Estimation on Small and Medium Enterprises

Until 2012, Istat estimated structural business statistics for small and medium enterprises using directly-collected data from the Survey on Small and Medium Enterprises (SME).

The survey annually collects information on profit-and-loss accounts variables, employment, and investments (among others) in the industrial, construction, trade and nonfinancial services sectors, as requested by the above mentioned structural business statistics Eurostat Regulation. A large number of secondary variables is also collected, mainly for National Accounts estimation purposes. The target population is represented by the active establishments in the reference year (about 4.4 million enterprises). A sample of approximately 100,000 establishments is selected each year via a one-stage stratified simple random sample design (i.e., equal probability of selection for units within strata). The strata are defined by the combination of economic activity given by classification NACE Rev. 2 (Eurostat 2008), size class in terms of persons employed, and administrative region. Direct estimates of the target parameters (totals) are obtained at the required levels of detail by calibrating to four-digit NACE code, three-digit NACE code by seven size classes (in terms of number of employed persons), and two-digit NACE code by Italian Regions.

Beginning in 2012, Istat has implemented a mixed estimation strategy supported by the new statistical register *Frame* and complemented by the survey data to estimate these statistics. The *Frame* is a multi-source register based on the use of information on businesses' economic accounts supplied by five different administrative and fiscal archives, namely *Financial Statements, Sector Studies survey, Tax returns, Regional Tax on Productive Activities, Social Security Data* (Luzi et al. 2014; Curatolo et al. 2016). The combination of the archives covers about 95% of the target population of small and medium enterprises (full coverage of the entire population is achieved through statistical imputation to compensate for the sources' incompleteness and/or under-coverage with respect to specific subpopulations). The register contains firm-level information on a number of key profit-and-loss accounts *core* variables, including *production value, turnover, intermediate costs, value added, wages, labor cost*. The estimate totals for these variables are not subject to sampling errors, as they are obtained by summing up the unweighted microdata at any level of detail (economic sector, size, territory) and for every specific subpopulation of enterprises (e.g., exporters, subcontractors, micro-enterprises, etc.).

The profit-and-loss account's *subcomponents* of the *core* statistics cannot be directly derived from the available administrative sources, they are estimated using the *projection estimator* (Kim and Rao 2011). In this approach, each *subcomponent* is estimated via a weighted regression model that includes nonresponse adjusted *core* aggregates as covariates (Righi 2016).

The only data available for estimating the remaining structural business statistics are directly collected from the SME survey. These items include a set of relevant loss-and-profit accounts variables such as *Total depreciation of fixed assets* and *Provisions for liabilities and other provisions.* These items differ from the above-described *subcomponents* of the *core* variables due to their weak statistical relationships with the other structural business statistics. As a consequence, the *projection estimator* will not yield efficient estimates at microdata level. However, it may be possible to develop efficient estimates of these variables via SAE, which would allow the exploitation of further auxiliary information and of the relationships among the estimates along different domains at given levels of aggregation.

3. Small Area Estimation for Business Statistics

Model-based SAE techniques use explicit modelling to relate individual unit data or area-level direct estimates to a set of auxiliary variables. Although unit-level auxiliary data were available, we preferred to adopt an area-level mixed model (Fay and Herriot 1979) for two reasons. First, the target variable is right-skewed and contains a considerable proportion of zero-valued observations, making area-level modelling more tractable than unit-level modelling. See Chandra and Chambers (2011) and Karlberg (2014) for examples of unit-level models for zero-inflated skewed distributions and discussion of resultant challenges. Second, area-level models avoid the bias caused by the presence of area means in the unit-level population model as contextual effects that are not explicitly included in the unit-level model specification (Namazi-Rad and Steel 2011).

The most widely used class of small area estimation models is linear mixed models, which include area random effects to account for between-area variation beyond that explained by auxiliary information. Let $\theta = (\theta_1, \ldots, \theta_D)^T$ be the parameter to be estimated for each domain $d = 1, \ldots, D$. Assume the following linking model between θ_d and a set of covariates X whose values are known for each domain of interest:

$$\theta = X\beta + u \tag{1}$$

where X is the covariate matrix and $u = (u_1, \ldots, u_D)^T =$ is the area effects vector, assumed to be independently distributed with mean zero and variance σ_u^2. Furthermore, assume that a design unbiased direct estimator $\hat{\theta}_d$ is available for (at least a subset of) the domains, that is,

$$\hat{\theta} = \theta + e \tag{2}$$

where $e = (e_1, \ldots, e_D)^T$ is the vector of sampling errors associated with the direct estimators, $E[e_d|\theta_d] = 0$ and $V[e_d|\theta_d] = \phi_d$. To avoid identifiability problems, the variances ϕ_d are supposed to be known.

The following linear mixed model is obtained by combining Equations (1) and (2):

$$\hat{\theta} = X^T\beta + u + e \tag{3}$$

Given Model (3), the empirical best linear unbiased predictor (EBLUP) for each domain d is:

$$\hat{\theta}_d^{EBLUP} = \gamma_d\hat{\theta}_d - (1 - \gamma_d)X_d\hat{\beta}$$

where $\gamma_d = \hat{\sigma}_u^2/\left(\hat{\sigma}_u^2 + \varphi_d\right)$ is the weight of the direct estimator, $\hat{\beta} = (X^T\hat{V}^{-1}X)^{-1} \times (X^T\hat{V}^{-1}\hat{\theta})$ is the generalized least square (GLS) estimator of the regression coefficient vector, $\hat{V} = \hat{\sigma}_u^2(I_D + \Phi)$ is the estimate of the model variance matrix of $\hat{\theta}$ and $\Phi = diag\,(\phi_1, \ldots, \phi_D)$. The parameter estimates for σ_u^2 and β are obtained either iteratively (e.g., via maximum likelihood or restricted maximum likelihood estimation, under normality assumptions for the random effects) or by the method of fitting constants. Computational details can be found in Rao and Molina (2015, sect. 6.1.1 and 6.1.2).

If unit level information is available, then the variance ϕ_d can be estimated from a unit-level model under the hypothesis of homoscedasticity, (Rao and Molina 2015, 132) or a generalized variance function (see for instance Wolter 2007, ch. 7). Including these revised variance estimates would affect the MSE of the predicted domain values (Bell 2008). For details about MSE estimation, see Rao and Molina (2015, sec. 6.2).

Under the classic Fay-Herriot model specification given by (3), the area random effects are assumed to be independent. This hypothesis means that no correlation structure of the data is considered. Instead, it is reasonable to assume the random effects between the neighboring areas (defined, for example, by a contiguity criterion) are correlated, with the correlation decaying to zero as distance between areas increases. Petrucci et al. (2005) extended Model (3) to allow for correlated area effects. In this Spatial Fay-Herriot model, the uncorrelated vector of random effects u is substituted with a correlated vector of random effects v.

Let $v = (v_1, \ldots, v_D)$ follow a Simultaneously Autoregressive (SAR) process with proximity matrix W, unknown autoregression parameter ρ (Cressie 1993, ch. 6). Let u be defined as before, so that the correlated vector of random effects is defined as $v = \rho W v + u$. If the matrix $(I_D - W)$ is assumed to be non-singular, then v can be expressed as $v = (I_D - W)^{-1}u$, so that $\boldsymbol{v}$ has mean vector 0 and covariance matrix G equal to $G = \sigma_u^2[(I_D - \rho W)^T(I_D - \rho W)]^{-1}$. Incorporating this substitution into (3) yields

$$\hat{\theta} = X^T\beta + [(I_D - \rho W)]^{-1}u + e$$

whose EBLUP of the quantity of interest θ_d is

$$\hat{\theta}_d^{SEBLUP} = X_d\hat{\beta} + b_d^T\hat{G}\hat{V}^{-1}(\hat{\theta} - X\hat{\beta})$$

Here $\hat{G}$ is obtained by replacing variance components σ_u^2 and ρ with their estimates $\hat{\sigma}_u^2$ and $\hat{\rho}$, $\hat{V} = \hat{G} + \Phi$, and $\hat{\beta} = (X^T\hat{V}^{-1}X)^{-1}(X^T\hat{V}^{-1}\hat{\theta})$ is the GLS estimator of regression parameter β, and b_d is D-dimensional vector (0, . . ., 0, 1,. 0, .., 0) with 1 in the dth position. The vector of regression coefficients β and the variance components σ_u^2 and ρ can be estimated either by ML or REML methods. Details for the estimation of the model parameters and the MSE can be found in Petrucci et al. (2005).

4. The Empirical Study

In the empirical study, we examine the variable *Total depreciation of fixed assets* (*depreciation* hereafter). This variable represents the gradual and systematic decrease in the economic value of an asset, either through physical depreciation, obsolescence or changes in the demand for the services of the asset in question. *Depreciation* is quite relevant in Italian Official Statistics for many reasons, with particular reference to the estimates compiled in the framework of National Accounts, for example, in the context of the estimation of capital stock of enterprises, as well as in the estimation of the amount of the black economy (Istat 2014). Moreover, together with investments and other variables related to statement of the assets, liabilities, and capital of the enterprises, *depreciation* contributes to the determination of relevant economic indicators, such as productivity and technical efficiency of enterprises (Istat 2017), playing an important role as proxy of the *Capital Stock*.

The study presented in this article uses SME data from the 2013 reference year. The size of the observed sample consists of 79,056 respondent units. The four-digit NACE Rev. 2 levels (538 small domains) are considered in the analysis. About 25% of the considered domains contained between 50 to 100 enterprises, about 50% contained between 100 to 500 enterprises, and only three percent contained more than 500 enterprises.

For small and medium enterprises, the variable *depreciation* generally has a semi-continuous skewed distribution, with high rates of zeros.

The auxiliary information used in the study is obtained from two different registers: the Italian Business Register *(ASIA)*, and the *Frame*. We use the *Proxy of the Turnover* item from the *ASIA* and the *Value Added* variable from the *Frame* as auxiliary information. Both are highly positively correlated with the study variable, as the correlation coefficients are equal to 0.70 and 0.75 respectively. The *Proxy of Turnover* variable has been used in previous structural business statistics processes in calibration procedures, correcting for nonresponse. However, this variable does not correspond to the SME survey definition of *Turnover*, and no attempts to harmonize these variables are made. Consequently, the two variables have different expected values at the unit level and are not expected to yield equivalent aggregates. In contrast, the *Value added* variable is fully harmonized with the *Depreciation* variable, with both variables being subjected to concurrent statistical validation processes.

To this extent, comparing the results obtained using different auxiliary information as input also allows us to assess the potential efficiency gains from exploiting the more accurate auxiliary information available in the *Frame*.

Four different types of area-level SAE estimators are considered, cross-classifying the two models (Fay-Herriot and Spatial Fay-Herriot) with the two auxiliary data sources (*ASIA* and *Frame*). As mentioned above, the Fay-Herriot model assumes no correlation between the area estimates. The Spatial Fay-Herriot model assumes the existence of such a correlation structure, in which economic activities that share the same first three digits are considered to be neighbors in the proximity matrix.

For each four-digit NACE domain, we compared the estimated coefficients of variation (CVs) of the direct estimates to the corresponding MSE of the SAE estimates to obtain relative measures of precision. Table 1 reports summary statistics on the distribution of CVs for all the considered estimation methods. It must be emphasized that this comparison is primarily descriptive. A rigorous analysis of relative reliability in terms of CVs, or MSEs, should imply a common comparison framework, such as the use of replicated

Table 1. Summary statistics on coefficient of variation for depreciation over all study domains expressed as percentages.

		ASIA		Frame	
Evaluation statistic	Direct estimate	Fay-Herriot	Spatial Fay-Herriot	Fay-Herriot	Spatial Fay-Herriot
Minimum	0.00	0.00	0.00	0.00	0.00
Q1	37.94	33.10	32.56	30.62	30.49
Median	57.14	50.45	49.43	47.72	47.53
Mean	83.34	74.00	72.62	70.68	70.43
Q3	97.06	87.06	86,06	85.29	84.16
Maximum	755.19	702.34	678.18	667.15	660.79

Table 2. Model diagnostics for considered SAE estimators by source of auxiliary data.

	ASIA		Frame	
Evaluation statistic	Fay-Herriot	Spatial Fay-Herriot	Fay-Herriot	Spatial Fay-Herriot
Max log-likelihood	−2249.40	−2264.63	−2177,38	−2173.59
AIC	4604.81	4537.27	4360.75	4355.18
BIC	4616.91	4553.41	4372.86	4371.32

methods on one or more pseudo populations, especially since design-based and model-based estimators are being compared.

In order to perform a meaningful comparative analysis between the estimation methods, all the CVs in Table 1 are computed using the same direct estimates $\hat{\theta}_d$ as the denominator, that is, $CV\left(\hat{\theta}_d^i\right) = \sqrt{MSE\left(\hat{\theta}_d^i\right)}/\hat{\theta}_d$, where the i superscript indicators denote the SAE estimator. Although the values of CVs are high, notice that all the model-based methods display lower value of the estimated CVs of the corresponding direct estimates. Finally, the CVs obtained using the Spatial Fay-Herriot models are consistently smaller than those obtained from the standard Fay-Herriot model with the same auxiliary data.

These results provide indications of strongly improved performance with some form of SAE over direct estimation for *depreciation*. Moreover, it appears that the *Frame* auxiliary data are more predictive than the *ASIA* data in this case, and that there may be some advantage in using the Spatial Fay-Herriot model over the Fay-Herriot model in this application. To further explore this, Table 2 reports model-diagnostic measures for the four corresponding SAE estimates, specifically the maximum of the log-likelihood function and the AIC and BIC statistics.

The statistics in Table 2 provide further support of (a) the better performance of the models that use of *Frame* auxiliary variables and (b) the superiority of the spatial Fay-Herriot model with respect to the standard Fay-Herriot model for this variable, regardless of input data source for modeling.

Clearly, the SAE estimates display advantages in terms of precision over their direct estimate counterparts. However, that might be of limited interest if the corresponding

Table 3. Summary statistics on the ratio of the SAE estimate of depreciation to the direct estimate of depreciation over all study domains expressed as percentages.

	ASIA		Frame	
Evaluation statistic	Fay-Herriot	Spatial Fay-Herriot	Fay-Herriot	Spatial Fay-Herriot
Minimum	4.8	3.9	6.5	6.6
Q1	91.1	93.8	94.1	94.4
Median	101.3	100.7	100.9	100.7
Mean	97.3	97.4	97.7	97.8
Q3	105.6	104.5	104.5	103.4
Maximum	269.0	258.8	367.8	350.8

estimates are similar, as the additional modeling/processing effort might not be justified by the increased precision. Table 3 shows how different the SAE estimates are from their direct estimate counterparts.

Table 3 shows a consistent pattern, regardless of the form of SAE estimator or auxiliary data. When the estimate level is low (smaller areas), then the SAE estimates tend to be smaller than their directly estimated counterparts. As the estimate level increases, the pattern reverses.

Next, we compare the performances of the model-based methods in terms of bias. To this aim, we use the diagnostic tools proposed in Brown et al. (2001). The first tool is based on the assumption of the unbiasedness of the direct estimates. Then, in the absence of model bias, a scatter plot of model-based estimates against the direct estimates should be balanced around the $y = x$ line. This can likewise be evaluated by comparing the slope of the fitted no-intercept regression line of the direct estimates on the model-based estimates to $\beta = 1$ (the slope of the $y = x$ line).

The plots in Figure 1 and Figure 2 provide some indications of negative bias induced by the model assumptions for all the methods under this assumption when using either *ASIA* or *Frame* auxiliary information, respectively.

This underestimation is most pronounced with the Fay-Herriot model that utilizes *ASIA* variables (Figure 1) and is less evident with the Spatial Fay-Herriot model utilizing the *Frame* variables (Figure 2).

Brown et al. (2001) also propose a diagnostic that measures the amount of scaling required to calibrate the aggregated SAE estimates to reliable direct estimates from the

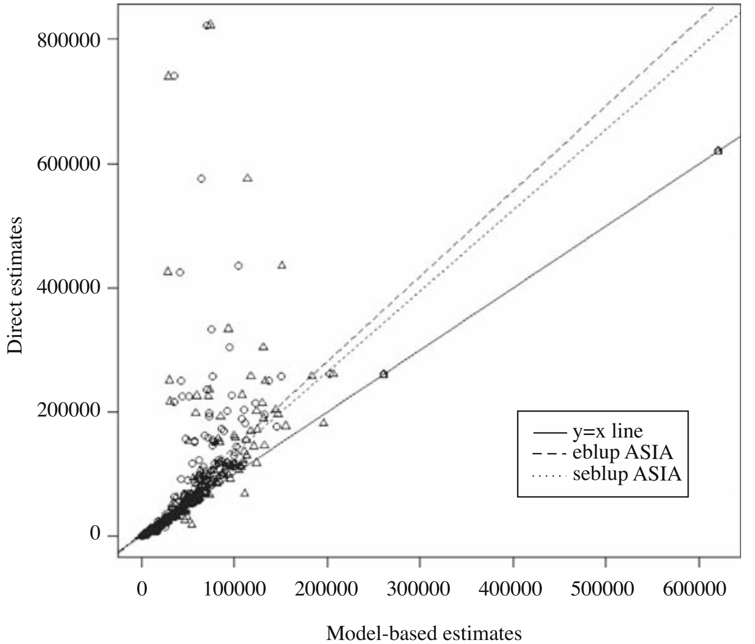

Fig. 1. Regression lines of direct estimates versus model-based estimates with ASIA as auxiliary variable.

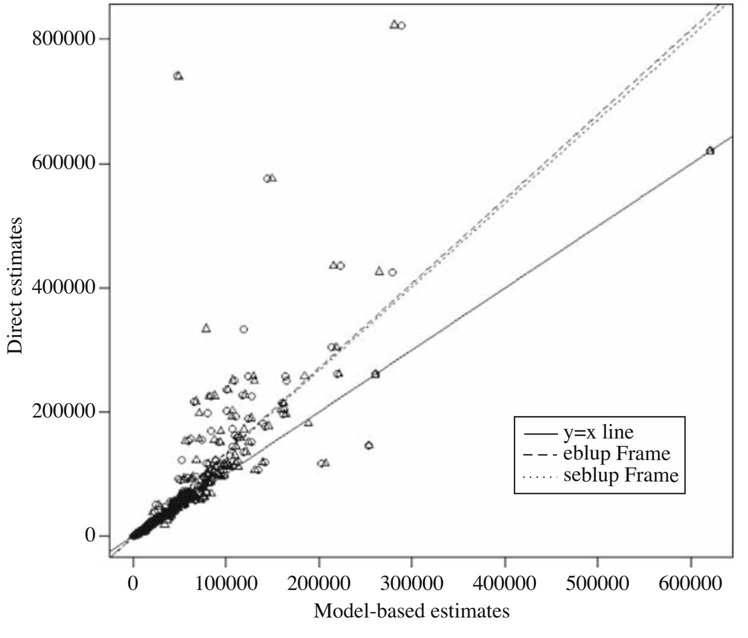

Fig. 2. Regression lines of direct estimates versus model based estimates with Frame as auxiliary variable.

larger domains. The authors suggest averaging the relative absolute difference between aggregated model-based and direct estimates over the higher level domains.

From Table 4 it can be seen that the Spatial Fay-Herriot model yields estimates with smaller absolute relative differences than those obtained using the Fay-Herriot model, again regardless of auxiliary data. And again, the best results in terms of absolute relative difference are obtained using the Spatial Fay-Herriot model with the *Frame* auxiliary data.

5. Conclusions and Future Work

The results presented in this article show methodological developments in exploiting administrative data that improve estimation of structural business statistics, which in turn will lead to the release of more reliable statistics to Istat's internal and external users. We focused on the *Total depreciation of fixed assets* variable, a profit-and-loss accounts variable that cannot be directly estimated from administrative sources and which has weak

Table 4. Average relative absolute differences at one-digit, two-digit, and three-digit NACE Rev. 2 levels.

	ASIA		Frame	
	Fay-Herriot	Spatial Fay-Herriot	Fay-Herriot	Spatial Fay-Herriot
1-digit NACE Rev. 2	13.51	8.81	9.18	8.52
2-digit NACE Rev. 2	15.20	9.90	9.99	8.58
3-digit NACE Rev. 2	15.12	10.08	9.82	8.64

statistical relationships with other directly-collected survey items. Instead, we propose the use of a SAE approach for this variable. By jointly exploiting survey estimates and proper auxiliary information, this approach yielded very promising results, producing more efficient estimates of *depreciation* in small domains than the corresponding direct estimates. These results are sufficiently encouraging to warrant further exploration and refinements to the SAE models for the study variable in order to eventually enhance its current estimation procedure output at all the required estimation domains.

We believe that the studied methods can be applied to other, similar, structural business statistics (e.g., the profit-and-loss accounts variable *Provisions for liabilities and other provisions*, the variable *Investments*, etc.), taking into account their specific features and available auxiliary information. For the latter, additional sources of information could be investigated, such as the *Notes of Financial Statements* and the *VAT* fiscal archive, taking into account possible weaknesses in terms of both coverage and usability for the target population.

The next steps of our research are further refinements of proposed estimation methods, ultimately extending them to other small domains. From this perspective, future activities will be focused on different areas. First, an assessment of the available information is needed in order to implement the best SAE model on a case-by-case basis for other items of interest. The *Frame* variables are a promising start. However, there may be additional auxiliary useful variables from new administrative or fiscal sources. Other innovative applications include *benchmarking* (ensuring for coherence between direct estimates and aggregated SAE estimates in large domains), the introduction of *multi-domain sampling designs* in Istat enterprise surveys, and the ongoing development of new strategies to manage consistency and confidentiality constraints in the new information context.

6. References

Bell, W.R. 2008. "Examining Sensitivity of Small Area Inferences to Uncertainty About Sampling Error Variances." In Proceedings of Survey Research Methods Section, Denver, August 4, 2008. 327–334. Alexandria, VA: American Statistical Association.

Brown, G., R. Chambers, P. Heady, and D. Heasman. 2001. "Evaluation of Small Area Estimation Methods – An Application to Unemployment Estimates from the UK LFS." In Proceedings of Statistics Canada Symposium 2001. Achieving Data Quality in a Statistical Agency: A Methodological Perspective. Hull, October 17, 2011. Ottawa: Statistics Canada. Available at: http://www.statcan.gc.ca/access_acces/alternative_alternatif.action?l=eng&loc=2001001/session6/6247-eng.pdf (accessed October 2017).

Caldera, A., M. Rasmussen, and O. Röhn. 2015. "Economic Resilience: What Role for Policies?" *OECD Economics Department Working Papers* 1251. Paris: OECD Publishing. DOI: http://dx.doi.org/10.1787/5jrxhgf61q5j-en.

Chandra, H. and R. Chambers. 2011. "Small Area Estimation for Skewed Data in Presence of Zeros." *The Bulletin of Calcutta Statistical Association* 63: 249–252.

Cressie, N. 1993. *Statistics for Spatial Data*. Revised ed. New York: John Wiley & Sons.

Curatolo, S., V. De Giorgi, F. Oropallo, A. Puggioni, and G. Siesto. 2016. "Quality Analysis and Harmonization Issues in the Context of the Frame SBS." *Rivista di*

Statistica Ufficiale 2016(1): 15–46. Available at: https://www.istat.it/it/files/2016/11/RSU_1_2016_Testointegrale.pdf (accessed October 2017).

Di Zio, M. 2016. "Estimating Population Size from Multisource Data with Coverage Unit Errors." In Proceedings of the 5th International Conference on Establishment Surveys (ICES). Geneva, June 23, 2016. American Statistical Association.

Di Zio, M. and O. Luzi. 2014. "Theme: Editing Administrative Data." In *Memobust Handbook on Methodology for Modern Business Statistics*. Luxembourg: Eurostat. Available at: https://ec.europa.eu/eurostat/cros/system/files/Statistical%20Data%20Editing-07-T-Administrative%20Data%20v1.0_0.pdf (accessed October 2017).

ESSnet AdminData. 2013. Project website: https://ec.europa.eu/eurostat/cros/content/use-administrative-and-accounts-data-business-statistics_en (accessed October 2017).

ESSnet on Data Integration. 2011. "Report on WP2 – Methodological Developments." Available at: https://ec.europa.eu/eurostat/cros/system/files/WP2.pdf (accessed October 2017).

Eurostat. 2008. *NACE Rev. 2. Statistical Classification of Economic Activities in the European Community*. Luxembourg: Office for Official Publications of the European Communities. Available at: http://ec.europa.eu/eurostat/documents/3859598/5902521/KS-RA-07-015-EN.PDF (accessed October 2017).

Fay, R.E. and R.A. Herriot. 1979. "Estimates of Income for Small Places: an Application of James-Stein Procedures to Census Data." *Journal of the American Statistical Association* 74(366): 269–277.

Garda, P. and V. Ziemann. 2014. "Economic Policies and Microeconomic Stability: a Literature Review and Some Empirics." *OECD Economics Department Working Papers* 1115. OECD Publishing. Doi: http://dx.doi.org/10.1787/5jz417mn2443-en.

Istat. 2014. "I Nuovi Conti Nazionali in SEC 2010 – Innovazioni e Ricostruzione Delle Serie Storiche (1995–2013)." *Nota informativa*. Rome: Istat.

Istat. 2017. *Rapporto Sulla Competitività Dei Settori Produttivi – Edizione 2017*. Rome: Istat. Available at: http://www.istat.it/storage/settori-produttivi/2017/Rapporto-competitivita-2017.pdf (accessed October 2017).

Jang, L. 2016. "Resolving Differences in Statistical Units: Statistics Canada's Experiences with Using Administrative Data in Economic Programs." In Proceedings of the 5th International Conference on Establishment Surveys (ICES). Geneva, June 23, 2016. American Statistical Association.

Karlberg, F. 2014. "Small Area Estimation for Skewed Data in the Presence of Zeros." *Statistics in Transition new series and Survey Methodology Joint Issue: Small Area Estimation 2014* 16(4): 541–562. Available at: https://stat.gov.pl/download/gfx/portalinformacyjny/en/defaultlistaplikow/3454/11/1/3d_karlberg_16_4_25_i_s541-562.pdf (accessed October 2017).

Kim, J.K.K. and J.N.K. Rao. 2011. "Combining Data from Two Independent Surveys: a Model-Assisted Approach." *Biometrika* 8: 1–16. Doi: https://doi.org/10.1093/biomet/asr063.

Laitila, T., A. Wallgren, and B. Wallgren. 2011. "Quality Assessment of Administrative Data." *Quality Assessment of Administrative Data. Research and Development – Methodology Reports from Statistics Sweden* 2011: 2. Statistics Sweden. Available at:

http://www.scb.se/statistik/_publikationer/OV9999_2011A01_BR_X103BR1102.pdf (accessed October 2017).

Luzi, O., U. Guarnera, and P. Righi. 2014. "The New Multiple-Source System for Italian Structural Business Statistics Based on Administrative and Survey Data." European Conference on Quality in Official Statistics (Q2014). Vienna, June 3, 2014.

Memobust. 2014. "Theme: Collection and Use of Secondary Data." In *Memobust Handbook on Methodology for Modern Business Statistics*. Luxembourg: Eurostat. Available at: https://ec.europa.eu/eurostat/cros/system/files/Data%20Collection-07-T-Secondary%20Data%20Collection%20v1.0.pdf (accessed October 2017).

Namazi-Rad, M.-R. and D.G. Steel. 2011. "Contextual Effects in Modeling for Small Domain Estimation." In Proceedings of the 4th Applied Statistics Education and Research Collaboration (ASEARC) Conference. Sidney, February 17, 2011. 12–14. Wollongong: University of Wollongong. Available at: http://ro.uow.edu.au/cgi/viewcontent.cgi?article=1049&context=smartpapers (accessed October 2017).

Petrucci, A., M. Pratesi, and N. Salvati. 2005. "Geographic Information in Small Area Estimation: Small Area Models and Spatially Correlated Random Area Effects." *Statistics in Transition* 7(3): 609–623.

Rao, J.N.K. and I. Molina. 2015. *Small Area Estimation*. 2nd ed. New York: John Wiley & Sons.

Righi, P. 2016. "Estimation Procedure and Inference for Component Totals of the Economic Aggregates in the New Italian Business Frame." *Rivista di Statistica Ufficiale* 2016(1): 83–97. Available at: https://www.istat.it/it/files/2016/11/RSU_1_2016_Testointegrale.pdf (accessed October 2017).

Wallgren, A. and B. Wallgren. 2007. *Register-Based Statistics: Administrative Data for Statistical Purposes*. New York: John Wiley & Sons.

Wolter, K.M. 2007. *Introduction to Variance Estimation*. 2nd ed. New York: Springer-Verlag.

Received October 2016
Revised November 2017
Accepted November 2017