<u>Prof.ssa Maria Ferrante</u>

Comparison of Bootstrap Confidence Interval Methods Small Area Estimation

The importance of developing reliable sub-population indicator estimates has increased significantly. Surveys usually provide information for broader areas, such as countries or administrative divisions. However, there is a growing need for estimates at a finer level of detail. Due to financial constraints that prevent expanding sample sizes, alternative methods, known as Small Area Estimation (SAE), are used. SAE methods encompass various statistical techniques to obtain reliable estimates for small sub-populations or geographic regions. These techniques are necessary when the variability of the direct estimator, like the Horvitz-Thompson estimator, is too large to produce reliable results (for a comprehensive review, see Rao and Molina, 2015 and Tzavidis et al., 2018).

The SAE literature has produced a large number of papers describing different techniques to

estimate the MSE (see among others Field and Welsh, 2007; Gonzalez-Manteiga et al., 2008; Liu

et al., 2022). The most common method is a parametric bootstrap. Even though the amount of

literature about bootstrap in SAE is considerable, less attention has been given to the estimation of bootstrap confidence intervals. There are various ways to compute bootstrap confidence intervals (Efron and Tibshirani, 1994; Chernick, 2011; and Jung et al., 2019), but only one has been usually applied in SAE (Liu et al., 2022).

In this project, we propose to develop simulation studies in which various methods of bootstrap

confidence intervals are compared to define the best possible choice concerning SAE models.

A first approach will be to test the three methods reported in Jung et al. (2019) for parametric

bootstrap on the baseline models in SAE (Rao and Molina, 2015, Ch. 6 and 7). The research
could then be generalized to more complex models (i.e., non-linear models) and also to the nonparametric bootstrap.

References

- Chernick, M. R. (2011). Bootstrap methods: A guide for practitioners and researchers. JohnWiley & Sons.
- Efron, B., & Tibshirani, R. J. (1994). An introduction to the bootstrap. Chapman; Hall/CRC.
- Field, C. A., & Welsh, A. H. (2007). Bootstrapping clustered data. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 69(3), 369–390.
- Gonzalez-Manteiga, W., Lombardia, M. J., Molina, I., Morales, D., & Santamaria, L. (2008). Bootstrap mean squared error of a small-area EBLUP. *Journal of Statistical Computation and Simulation*, 78(5), 443–462.
- Jung, K., Lee, J., Gupta, V., & Cho, G. (2019). Comparison of bootstrap confidence interval methods for GSCA using a monte carlo simulation. *Frontiers in psychology*, 10, 2215.
- Liu, Y., Liu, X., Pan, Y., Jiang, J., & Xiao, P. (2022). An empirical comparison of various MSPE estimators and associated prediction intervals for small area means. *Journal of Statistical Computation and Simulation*, 1–27.
- Rao, J., & Molina, I. (2015). Small area estimation. John Wiley & Sons, Inc.
- Tzavidis, N., Zhang, L.-C., Luna, A., Schmid, T., & Rojas-Perilla, N. (2018). From start to finish: A framework for the production of small area official statistics. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 181(4), 927–979.

Prof.ssa Maria Ferrante

Construct a synthetic population for unit-level small area estimation models

Small Area Estimation (SAE) methods encompass a variety of techniques designed to obtain reliable estimates for small sub-populations when the sample size is too low to yield accurate results using a classical Horvitz-Thompson estimator (for a comprehensive review, see Rao and Molina, 2015 and Tzavidis et al., 2018). SAE models leverage strengths from neighboring areas and auxiliary information. When the auxiliary information is at the individual level, these are referred to as unit-level SAE models.

One of the main limitations in the development of SAE methods, especially at the unit level, is the access to individual data necessary for simulation studies, which are often used to test the efficacy of a method. These data are frequently subject to privacy restrictions and are often replaced with synthetic populations (Ferrante and Pacei, 2017). Synthetic populations are simulated datasets from which it is possible to extract samples of various sizes, and for which all parameters are known. In other words, one possible solution to overcome the problem of the confidentiality constraints is synthetic data, which mimics the original observed data and preserves the relationships between variables without containing any disclosive records. Techniques to produce synthetic populations are well summarized in Taylor et al. (2016), Nowok et al. (2016) and Templ et al. (2017).

In this project, we propose to create a synthetic population to be published and made directly

accessible to the international statistical community. This population will be based on real data
from European surveys and will be generated using appropriate statistical approaches to create a
realistic representation that closely mirrors the original population.

References

- Ferrante, M. R., & Pacei, S. (2017). Small domain estimation of business statistics by using multivariate skew normal models. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 180(4), 1057–1088.
- Nowok, B., Raab, G. M., & Dibben, C. (2016). Synthpop: Bespoke creation of synthetic data in R. *Journal of statistical software*, 74, 1–26.
- Rao, J., & Molina, I. (2015). Small area estimation. John Wiley & Sons, Inc.
- Taylor, J., Moon, G., & Twigg, L. (2016). Using geocoded survey data to improve the accuracy of multilevel small area synthetic estimates. *Social Science Research*, 56, 108–116.
- Templ, M., Meindl, B., Kowarik, A., & Dupriez, O. (2017). Simulation of synthetic complex data: The R package simpop. *Journal of Statistical Software*, 79(10), 1–38.
- Tzavidis, N., Zhang, L.-C., Luna, A., Schmid, T., & Rojas-Perilla, N. (2018). From start to finish: A framework for the production of small area official statistics. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 181(4), 927–979.

<u>Prof. Daniele Ritelli</u>

Methods of solving differential equations with Lie symmetries using computer algebra

A very practical approach is followed on the subject, given that the existence of Lie symmetries, which allow the simplification of a given differential equation, consists of solving accessory differential equations, which are treated using computer algebra.

Computer algebra in the symbolic treatment of nonlinear differential equations using special functions

After introducing the student to the use of software for symbolic calculus, the internship aims at using it for the treatment of differential equations describing the behaviour of nonlinear oscillators. Tutorials illustrating the theoretical approach underlying the treatment of the models will be an integral part of the training. The candidate must be willing to acquire the mathematical tools necessary for the study of the models.

<u>Prof. Michele Scagliarini</u>

Methods for Monitoring Time Between Events and Amplitude Data

While many control charts have been developed for monitoring the time interval (T) between the occurrences of an event, many other charts are employed to examine the magnitude (X) of the event (E). These two types of control charts have usually been investigated and applied separately.

Time Between Events and Amplitude (TBEA) control charts are a combined scheme for monitoring the time interval T of an event E as well as its amplitude X.

The aim of this project is to study the implementation of such monitoring algorithms in the R environment. The developed methodology will be applied on both simulated and real data.

The internship will be divided into three phases:

- first phase dedicated to study Shewhart Time-Between-Events-and-Amplitude Control Charts and their implementation in R;
- second phase dedicated to study the effect of the correlation between T and X;
- third phase aimed to implement a non-parametric EWMA control chart for Monitoring TBEA.

References

D. Rahali, P. Castagliola, H. Taleb, and M.B.C. Khoo. Evaluation of Shewhart Time-Between-Events-and-Amplitude Control Charts for Several Distributions. Quality Engineering, 31(2):240-254, 2019. doi: 10.1080/08982112.2018.1479036.

D. Rahali, P. Castagliola, H. Taleb, and M.B.C. Khoo. Evaluation of Shewhart Time-Between-Events-and-Amplitude Control Charts for Correlated Data. Quality and Reliability Engineering International, 37(1):219-241, 2021. doi: 10.1002/qre.2731.

S. Wu, P. Castagliola, and G. Celano. A Distribution-Free EWMA Control Chart for Monitoring Time-Between-Events-and-Amplitude Data. Journal of Applied Statistics, 48(3):434-454, 2021. doi:10.1080/02664763.2020.1729347.

Z. Wu, J. Jiao, and H. Zhen. A Control Scheme for Monitoring the Frequency and Magnitude of an Event. International Journal of Production Research, 47(11):2887-2902, 2009.

<u>Prof.ssa Silvia Pacei</u>

Misura della resilienza

Le crisi globali del 2008 e 2020 hanno determinato una situazione di incertezza attorno ai sistemi economici e finanziari. La ripresa poi, com'è noto, è stata particolarmente rallentata in Europa, rispetto agli USA, benché con una certa variabilità nella velocità della ripresa tra i paesi europei. I dati delle indagini sulle famiglie evidenziano la loro preoccupazione per il periodo di incertezza economica che stanno vivendo e per la loro capacità di recuperare eventuali perdite. Tali preoccupazioni influiscono su diversi comportamenti umani.

Ad oggi pochi autori si sono proposti di misurare la resilienza (Asheim et al. 2020; Cissé e Barrett, 2018). L'obiettivo di questo progetto è proporre una definizione di resilienza e nuove misure di resilienza calcolabili a livello di individuo e in grado di soddisfare alcune proprietà auspicabili. Queste misure possono essere calcolate utilizzando i dati delle indagini sulle famiglie EU-SILC o Banca d'Italia. Le misure di resilienza possono poi essere impiegate in modelli per le determinanti di diversi comportamenti sociali.

Riferimenti bibliografici

- Asheim, G. B., Bossert, W., D'Ambrosio, C., & Vögele, C. (2020). The measurement of resilience. *Journal of Economic Theory*, *189*, 105104.
- Cissé, J. D., & Barrett, C. B. (2018). Estimating development resilience: A conditional moments-based approach. *Journal of Development Economics*, *135*, 272-284.

Prof. Luca Trapin

Analysis of illiquidity risk premium

Liquidity is a fundamental property of a well-functioning market, and lack of liquidity is generally at the heart of many financial crises and disasters. The financial economics literature hypothesizes the existence of an illiquidity premium in the market, i.e. investors require higher returns to hold illiquid stocks (Amihud and Mendelson, 1986). Liquidity is an elusive concept. It is not observed directly and cannot be captured in a single measure (Amihud and Mendelson, 1991). Using several proxies of liquidity, numerous studies have documented the existence of a positive relationship between stock returns and stock illiquidity, thus confirming empirically the existence of an illiquidity premium. Amihud (2002) shows that the existence of a premium is not only in cross-section but also in time series, i.e. future expected stock returns are increasing in expected illiquidity .

This project aims at investigating the conclusions on the illiquidity premium using ‘’liquidity factors’’ instead of ‘’liquidity proxies’’ (Hallin et al., 2011). The research student will have to: (i) build a large dataset of low-frequency liquidity proxies (Goyenko, 2009) for a large set of U.S. stocks using Eikon Refinitiv; (ii) extract liquidity factors from the liquidity proxies using factor models (Stock and Watson, 2002); (iii) run regression analysis for the identification of the illiquidity premium (Amihud, 2002).

References

1. Amihud, Y., & Mendelson, H. (1986). Asset pricing and the bid-ask spread. Journal of financial Economics, 17(2), 223-249.
2. Amihud, Y., & Mendelson, H. (1991). Liquidity, asset prices and financial policy. Financial Analysts Journal, 47(6), 56-66.
3. Amihud, Y. (2002). Illiquidity and stock returns: cross-section and time-series effects. Journal of financial markets, 5(1), 31-56.

4. Goyenko, R. Y., Holden, C. W., & Trzcinka, C. A. (2009). Do liquidity measures measure liquidity?. Journal of financial Economics, 92(2), 153-181.
5. Hallin, M., Mathias, C., Pirotte, H., & Veredas, D. (2011). Market liquidity as dynamic factors. Journal of econometrics, 163(1), 42-50.