

Thema	Abstract	Literatur
Updating Intercensal Indicators based on geospatial data	Censuses are fundamental building blocks of most modern-day societies, yet collected every 10 years at best. The thesis assesses different approaches for census updating techniques by incorporating auxiliary information in order to take ongoing subnational population shifts into account. The methods adds satellite imagery as additional data source to derive disaggregated estimates of poverty indicators in Mozambique. The performance of the methods is evaluated using data from two different census periods.	<ul style="list-style-type: none"> • T. Koebe, A. Arias-Salazar, N. Rojas-Perilla, and T. Schmid. Intercensal updating using structure-preserving methods and satellite imagery. <i>Journal of the Royal Statistical Society Series A</i>, 185: 170-196, 2022. • A.-K. Kreutzmann, S. Pannier, N. Rojas-Perilla, T. Schmid, M. Templ, and N. Tzavidis. The R package emdi for estimating and mapping regionally disaggregated indicators. <i>Journal of Statistical Software</i> 91(7): 1-33, 2019.
Multinomial Mixed Models in Small Area Estimation	The topic deals with small area estimation for the composition of categorical variables. A possible example is, for instance, the estimation of multidimensional poverty in small areas. From a methodological perspective, the thesis deals with small area estimation based on multinomial (mixed) models including the computation of the mean squared error of the estimators. A simulation study could be conducted to investigate the properties of the estimators. Finally, the methods can be applied to poverty data from Mexico.	<ul style="list-style-type: none"> • E. Lopez-Vizcaino, M. J. Lombardia, and D. Morales. Multinomial-based small area estimation of labour force indicators. <i>Statistical Modelling</i>, 13(2):153-178, 2013. • I. Molina, A. Saei, and M. J. Lombardia. Small area estimates of labour force participation under a multinomial logit mixed model. <i>Journal of the Royal Statistical Society: Series A</i>, 170(4):975-1000, 2007.
Nonparametric regression with application to small area problems	Within the field of small area estimation the best linear unbiased predictor (BLUP) or its empirical version, the empirical BLUP (short EBLUP), is a widely used tool to estimate statistics of small areas, even though it is sensitive to model misspecification. More robust estimators like the penalised spline regression, which avoids functional specification, are brought forward as alternatives. Opsomer et al. (2008) suggest an estimator for unit-level data that treats the penalisation coefficient as a random effect, which allows the use of standard EBLUP theory in further analysis. Their estimator is known as nonparametric EBLUP. The nonparametric EBLUP framework could be compared to the standard EBLUP and to Random forests within a simulation study and an application with real data.	<ul style="list-style-type: none"> • J. Opsomer, F. Breidt, G. Claeskens, G. Kauermann, and G. Ranalli. Nonparametric small area estimation using penalized spline regression. <i>Journal of the Royal Statistical Society. Series B (Statistical Methodology)</i>, 70(1):265-286, 2008. • D. Pfeiffermann. Small Area Estimation: New Developments and Directions. <i>International Statistical Review/ Revue Internationale de Statistique</i>, 70(1):125-143, 2002.
Poverty Mapping using ELL-Methodology	The World Bank Method, also called ELL (Elbers, Lanjouw and Lanjouw) method, is a small area estimation strategy used to estimate poverty indicators. This estimation technique is implemented in the World Bank software PovMap. In a thesis, a package could be programmed that also makes this estimation technique available to a broader audience via the R programming language. Furthermore, the ELL can be adjusted and analysed based on the comparison to other established methods in the field, such as the empirical best prediction method (EBP).	<ul style="list-style-type: none"> • C. Elbers, J. O. Lanjouw, and P. Lanjouw. Micro-level estimation of poverty and inequality. <i>Econometrica</i> 71(1): 355-364, 2003. • A.-K. Kreutzmann, S. Pannier, N. Rojas-Perilla, T. Schmid, M. Templ, and N. Tzavidis. The R package emdi for estimating and mapping regionally disaggregated indicators. <i>Journal of Statistical Software</i> 91(7): 1-33, 2019.
Quantile Regression in Complex Design Samples	The Quantile Regression (QR) is an extension of the (multiple) linear regression that specifies the change of conditional quantiles of a dependent variable, instead of the change the conditional mean. The additional information gathered form this method might uncover new insights in datasets such as the one of the Demographic and Health Survey (DHS) of Senegal. QR could be used to estimate the conditional quantiles of wealth indicators such as the Wealth Index Factor Score for women and men, which uses the total value	<ul style="list-style-type: none"> • R. Koenker. <i>Quantile Regression</i>. Cambridge University Press: Cambridge, New York, 2005. • ANSD. <i>Demographic and Health and Multiple Indicator Cluster Survey (EDS-MICS) 2010-2011</i>.

	of an individual's assets to determine wealth, instead of commonly used monetary variables	
Robust Small Area Estimation Methods in the Context of Spatial Correlations	The empirical best linear unbiased predictor (EBLUP), which is commonly used in small area statistics, uses the assumptions of normality and uncorrelated random effects. The robust version of EBLUP (REBLUP) can be used when a violation of normality is assumed. The SREBLUP is another extension that additionally considers spatial correlated area effects in the model. A possible thesis could compare the performance of these estimators for different scenarios according to each estimator's underlying assumptions in a model-based simulation study. Real world data that displays spatial dependency and deviation of normality can be used.	<ul style="list-style-type: none"> • T. Schmid, and R. Munnich. Spatial robust small area estimation. <i>Statistical Papers</i>, 55(3):653–670, 2014. • R. B. Papalia, C. Bruch, T. Enderle, S. Falorsi, A. Fasulo, E. Fernandez Vazquez, M. Ferrante, J.-P. Kolb, R. Münnich, S. Pacei, R. Priam, P. Righi, T. Schmid, N. Shlomo, F. Volk, and T. Zimmermann. Best practice recommendations on variance estimation and small area estimation in business surveys. BLUE-ETS. SP1-Cooperation-Collaborative Project - Small or medium-scale focused research project, 2013.
Statistical misuse and consequences	Statistical literacy: Selection from typical problem clusters of statistical literacy, such as confounding, lack of measures of significance and/or goodness of fit, correlation versus causality, relative versus absolute risk, various kinds of biases (e.g., self-selection, response bias, attrition, Hawthorne effect, social desirability bias), immortal time bias, etc.	<ul style="list-style-type: none"> • Milo Schield. Statistical literacy: Thinking critically about statistics. <i>Of Significance</i>, 1(1):15–20, 1999. • DJordje Kadjević & Max Stephens. Modern statistical literacy, data science, dashboards, and automated analytics and its applications. <i>Teaching of Mathematics</i>, 23(1):71–80, 2020. • Eric Soweay & Peter Petocz. A panorama of statistics: Perspectives, puzzles and paradoxes in statistics. John Wiley & Sons, 2017.
Convergence in Distribution: Examining the Gelman-Rubin Diagnostic	Convergence in distribution is factually impossible to prove. The GD diagnostic has become a widely-accepted measure to assess convergence in distribution. Investigate the Accuracy of the GD diagnostic using an extensive MC simulation study.	<ul style="list-style-type: none"> • Gelman, A., and D. B. Rubin. 1992. Inference from iterative simulation using multiple sequences. <i>Statistical Science</i> 7: 457–472.
Diagnostics for mass imputation: R package	To date none of the imputation packages in R (or Stata, SPSS;...) can visualize missing data in meaningful way if the number of observations/ variables is large.	<ul style="list-style-type: none"> • Meinfelder, F. (2013): Datenfusion: Theoretische Implikationen und praktische Umsetzung. In: Weiterentwicklung der amtlichen Haushaltsstatistiken. (eds. Riede, T.; Ott, N.; Bechthold, S.), pp. 83-100, 1st ed., Berlin, GWI Wissenschaftspolitik Infrastrukturentwicklung.
Inference for the Potential Outcome framework	Different matching strategies impact the variance term of the test statistic. With and without replacement matching as well as weighting affect how independent matched samples are.	<ul style="list-style-type: none"> • Bai, Yuehao; Romano, Joseph P.; Shaikh, Azeem M (2019) Inference in experiments with matched pairs, Econstor Working Paper
Multiple Imputation of categorical variables using an extension to Predictive Mean Matching	PMM has become a popular MI method and is the default imputation method in 'mice' for 'numeric'-type variables. However, no extension for non-ordered categorical (nominal-scale) variables is available. One problem is the identification of a nearest neighbour over k categories. This can be overcome by using the KL divergence to match via the empirical distribution of donor and recipient predictions. Your task is to implement this method in R and compare it with alternative approaches.	<ul style="list-style-type: none"> • Koller-Meinfelder, F. (2009): Analysis of Incomplete Survey Data – Multiple Imputation via Bayesian Bootstrap Predictive Mean Matching, Dissertation, Otto-Friedrich-Universität Bamberg.
Comparing Generalized Linear Mixed Models and Transformations	Two main approaches to deal with non normal and a non linear relationship between the independent and the dependent variable are generalized linear mixed models (GLM) and transformations. In a possible thesis the researcher could compare the pros and cons of both approaches in a simulations study.	<ul style="list-style-type: none"> • Morales, Domingo, et al. A course on small area estimation and mixed models. Methods, theory and applications in R (2021).

<p>in the SAE Context</p>		<ul style="list-style-type: none"> • Chandra, Hukum, and Ray Chambers. Small area estimation under transformation to linearity. (2008).
<p>Comparing the influence of different informative designs on small-area estimates</p>	<p>When the selection mechanism is informative, the selection probability of individuals is correlated with the outcome of the target variable and inference based on such a sample may be strongly biased without appropriate weighting. An extension of the EBP for estimation of non-linear parameters in SAE that handles informative selection has been proposed. This thesis could investigate the influence of different informative designs on the estimates by means of a simulation study. Different approaches to incorporating an informative design and the influence of the degree of informativeness could be the subject of the simulation study.</p>	<ul style="list-style-type: none"> • Pfeffermann, D., & Sverchkov, M. (2007). Small-Area Estimation Under Informative Probability Sampling of Areas and Within the Selected Areas • María Guadarrama, Isabel Molina, J.N.K. Rao, Small area estimation of general parameters under complex sampling designs. • Molina, Isabel, and J. N. K. RAO. Small Area Estimation of Poverty Indicators
<p>Inference for Nonprobability Samples. Quasi-randomization vs. Superpopulation-modeling</p>	<p>In a world of "big data" large amounts of data are available that are faster and easier to collect than are probability samples. Design-based inference, in which the distribution for inference is generated by the random mechanism used by the sampler, cannot be used for nonprobability samples. As two general approaches to inference are the pseudo-design-based Quasi-randomization approach and model-based Superpopulation-modeling The task is to conduct a simulation study, where you compare the performance of these two approaches.</p>	<ul style="list-style-type: none"> • Elliot M., Valliant R., Inference for Nonprobability Samples, In: Statistical Science, Vol. 32, No. 2, 249-264, 2017.