The demand for reliable business statistics at disaggregated levels such as NACE classes increased considerably in recent years. Owing to small sample sizes for some of the domains, design-based methods may not provide estimates with adequate precision. Hence, model-based small area estimation techniques which increase the effective sample size by borrowing strength are called for.

Business data are frequently characterised by skewed distributions, with a few large enterprises that account for the majority of the total for the variable of interest, e.g. turnover. Moreover, the relationship between the variable of interest and the auxiliary variables is often non-linear on the original scale. In many cases, a lognormal mixed model provides a reasonable approximation of this relationship. Assuming that the sampling design is non-informative, Berg and Chandra (2014) derived the empirical best predictor (EBP) under the lognormal mixed model.

In this work, we extend the approach of Berg and Chandra to compensate for informative sampling by incorporating design information among the covariates via an augmented modelling approach. This gives rise to the EBP under the augmented model. An important question relates to the choice of the augmenting variable. We propose to select the augmenting variable based on a joint assessment of residual plots, a measure of predictive accuracy and a check of the normality assumptions. We compare our approach with the EBP of Berg and Chandra in a model-based simulation study under different informative sampling mechanisms.

Reference:
Emily Berg and Hukum Chandra (2014): Small area prediction for a unit-level lognormal model, Computational Statistics & Data Analysis, Pages 159-175.