Jean Monnet Chair

Analysis of European Data by Small Area Methods

Lecture 3: Classification of Small Area Estimation models; EBLUP estimator

http://sampleu.ec.unipi.it

Figure 2.1: How an SAE unit-level model works

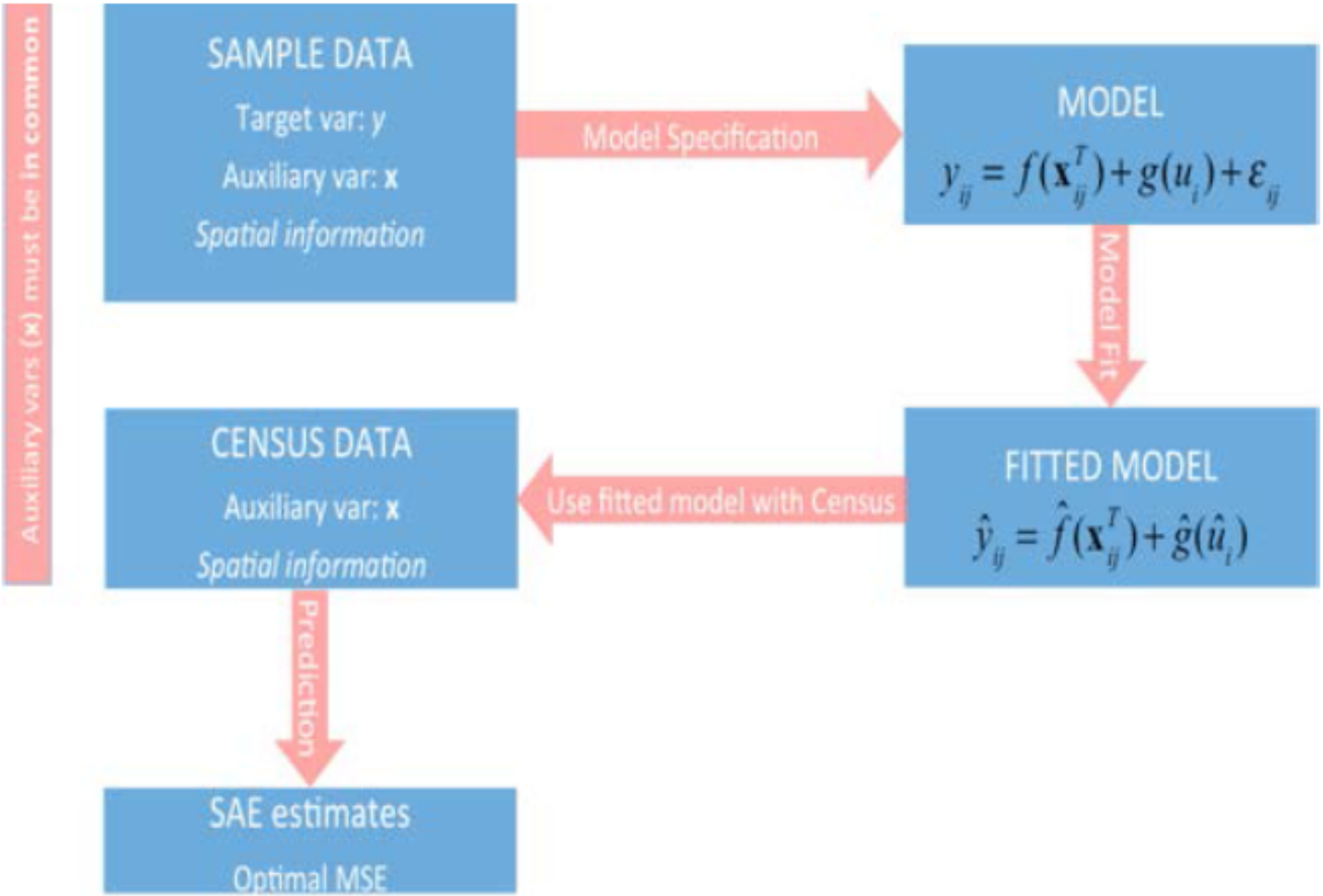
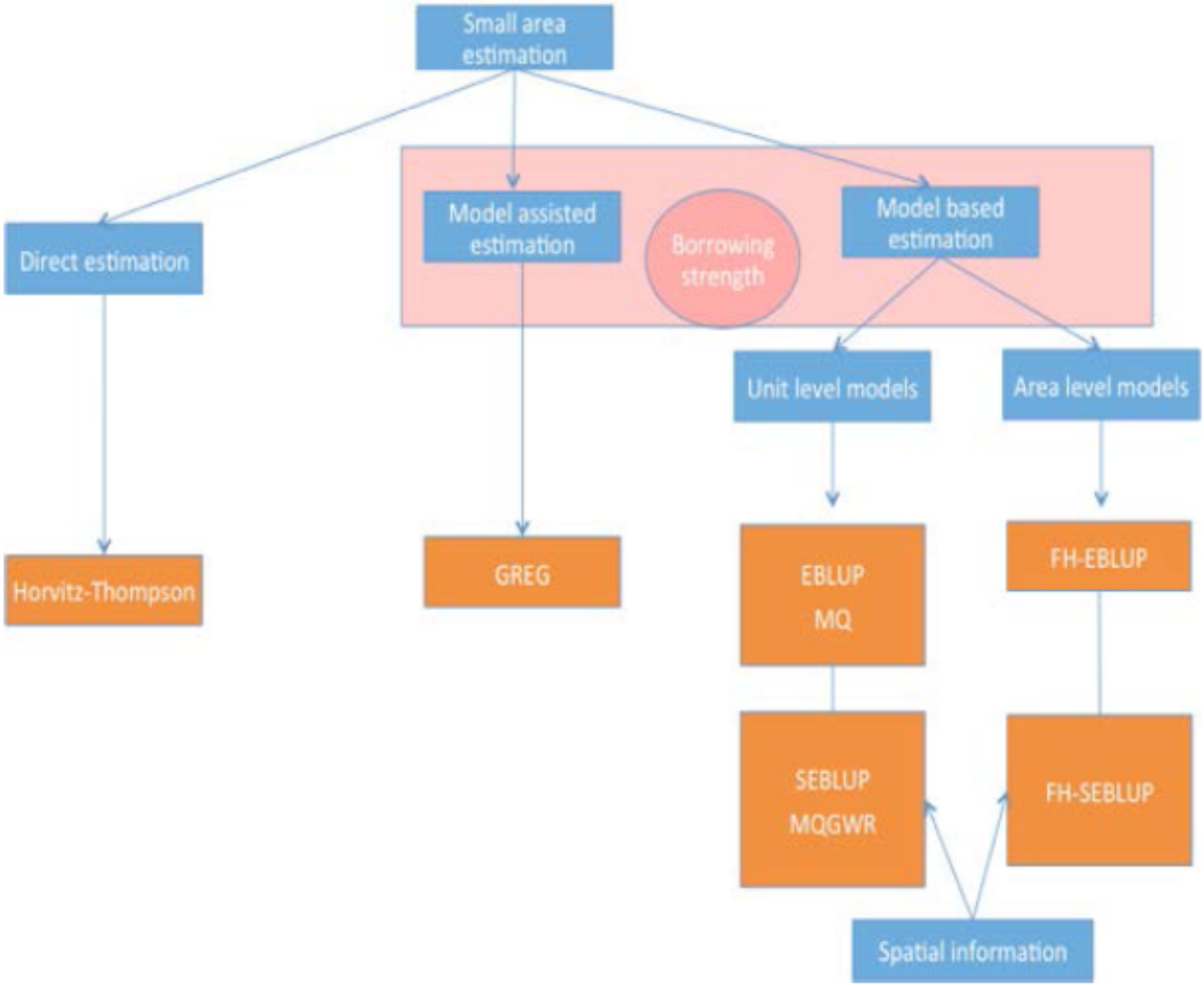


Figure 2.2: A classification of SAE estimation methods



2.4 Model-Based estimators: area-level

The most popular methods used for model-based SAE employ LLMs. Publications dealing with LMMs include Searle *et al*., (1992), Longford (1995), McCullogh and Searle (2001) and Demidenko (2004).

Model-dependent estimators that rely on linear-mixed or random-effects models have gained popularity (Rao, 2003; Jiang and Lahiri, 2006a) because they enable the inclusion of a random-area effect to explain inter-area variation in addition to that explained by fixed-effect covariates. The reliability of these methods depends on the validity of model assumptions, however, a criticism often raised in design-based research (Estevao and Särndal, 2004).

2.4.1 FH-EBLUP

The FH-EBLUP is the most popular method for producing small-area estimates from area-level data. The model can be extended to include correlated random area effects, the FH–Spatial EBLUP.

ϑ_i is a composite estimate of the form:

$$\hat{\vartheta}_i^{FH}(\hat{\sigma}_u^2) = \hat{\gamma}_i \hat{\vartheta}_i + \left(1 - \hat{\gamma}_i\right) \mathbf{x}_i^T \hat{\boldsymbol{\alpha}},$$

where $\hat{\gamma}_i = \hat{\sigma}_u^2 / \left(\hat{\sigma}_u^2 + \varphi_i\right)$ and $\hat{\boldsymbol{\alpha}}$ is the weighted least squares estimate of with weights $\left(\hat{\sigma}_u^2 + \varphi_i\right)^{-1}$ obtained by regressing $\hat{\vartheta}_i$ on $\mathbf{x}_i$, and $\hat{\sigma}_u^2$ is an estimate of the variance component σ_u^2.

The EBLUP estimate gives more weight to the synthetic estimate when the sampling variance, φ_i, is large or where $\hat{\sigma}_u^2$ is small, and moves towards the direct estimate as φ_i decreases or $\hat{\sigma}_u^2$ increases.

Table 2.6: EBLUP under area level specification: advantages, disadvantages and extensions

Properties	Advantages	Disadvantages	Extensions
Model assumptions	Efficiency under the assumption of Normality of LMM	Linearity of the relation with fixed effects aux variables Incorrelation between the random area effects	Non-parametric extension EBLUP (Giusti *et al.*, 2012) SEBLUP (Petrucci and Salvati, 2006; Pratesi and Salvati, 2008)
Design consistency	Design consistent		
Robustness to outliers		Not robust against outliers	
Out-of-sample predictions		Prediction not inclusive of spatial information	SEBLUP (Petrucci and Salvati, 2006; Pratesi and Salvati, 2008, 2009)

Model assumptions

The EBLUP is popular and is efficient under the assumption of normality of LLMs. It is specified under the assumption of linearity of the relation between the study variable and the auxiliary variables. Giusti *et al*. (2012) extended it, however, with a semi-parametric specification obtained by Psplines, which allows non-linearities in the relationship between the response variable and the auxiliary variables (see section 2.6.1). The correlation between random-area effects is introduced in the SEBLUP (Petrucci and Salvati, 2006; Pratesi and Salvati, 2008, 2009).

More details in the following report downloadable from s

http://www.gsars.org/wp-content/uploads/2015/09/TR-Spatial-Disaggregation-and-Small-Area-Estimation-210915.pdf

IMPROVING AGRICULTURAL & RURAL STATISTICS

Global Strategy

Technical Report Series GO-07-2015

Spatial Disaggregation and Small-Area Estimation Methods for Agricultural Surveys: Solutions and Perspectives

Design consistency

Design consistency is a general-purpose form of protection against model failures in that it guarantees that estimates make sense even if the assumed model fails completely, at least for large domains. The GREG estimator is asymptotically design-unbiased and consistent, but it can be sensible to extreme values of inclusion probabilities (Fabrizi *et al*., 2014). GREG estimators supported by LMM have turned to model-based estimation for the parameters of the model, so the efficiency of the resulting small-area estimators relies on the validity of the model assumption, and typically on the validity of the normality of residuals.

Robustness to outliers

GREG and GREG-S expressions allow for survey weighting of outlying observations, but this does not guarantee protection against the outlying observations. A robust version of GREG was proposed in Duchesne (1999).

Predictions for out-of-sample areas

Predictions for the out-of-sample areas – those with zero sample size – are based on the estimated parameters of the linear regression model and on the *X* auxiliary information:

$$\hat{\vartheta}_i^{GREG-OUT} = \frac{1}{N_i} \sum_{j \in U_i} \mathbf{x}_{ij}^T \hat{\mathbf{B}}$$