Use of Structure Preserving Estimator (SPREE) to Estimate the Number of Unemployed

Tomasz Józefowski[1] Tomasz Klimanek[2]
Marcin Szymkowiak[3,1]

[1]Centre for Small Area Estimation
Statistical Office in Poznan

[2]Statistical Office in Poznan

[3]Department of Statistics
Poznan University of Economics and Business

Outline

The phenomenon of unemployment

- Unemployment is one of the most serious socio-economic phenomena and remains a challenge, which is particularly difficult to solve.
- It can lead to poverty, social pathology and exclusion of entire social groups and areas which it affects.This in turn can cause social tensions.
- In order to make appropriate decisions and counteract the negative effects of unemployment, national and local authorities require reliable and detailed information about the number of unemployed.
- Since unemployment undergoes seasonal fluctuations, it is necessary to have access to frequently updated information at the lowest possible levels of spatial aggregation.

Data sources on unemployment

There are two data sources which can provide information about the number of unemployed persons often then ones a year:

- **LFS** (Labour Force Survey) - quartely information,
- Registered unemployment - monthly information.

Labour Force Survey

- The Labour Force Survey (LFS) is the basic source of information published by CSO (the Central Statistical Office of Poland) about the situation regarding the economic activity of the population, i.e. the fact of being employed, unemployed, or economically inactive, both for the country as a whole and at the regional level (NUTS 2 – province).
- It means that province (NUTS 2) is the lowest level of administrative division in Poland at which reliable LFS data are available (some basic subdivisions are also presented e.g. NUTS2 x sex, NUTS2 x place of residence). This is mainly due to the LFS sample size.
- Data for lower levels of territorial division (NUTS 3 or NUTS 4) are biased: the sampling error is too high; the same problem exists for more detailed domains obtained by cross-classifying province with age, sex or place of residence (NUTS2 x sex x place of residence x age groups).

Registered unemployment

- Information based on data delivered by district labour offices.
- People are registered as unemployed when

  - they are not employed or not performing any other kind of paid work,
  - are available for full-time work,
  - are not attending any full-time schools (with the exception of schools for adults or to take extra-curricular exams),
  - are registered in their respective district labour offices,
  - seek employment or any other paid work.
- The information on the unemployed is broken down by
  - sex,
  - age,
  - level of education,
  - duration of unemployment,
  - region, province, subregion, district and communes.

Registered unemployment vs LFS unemployment

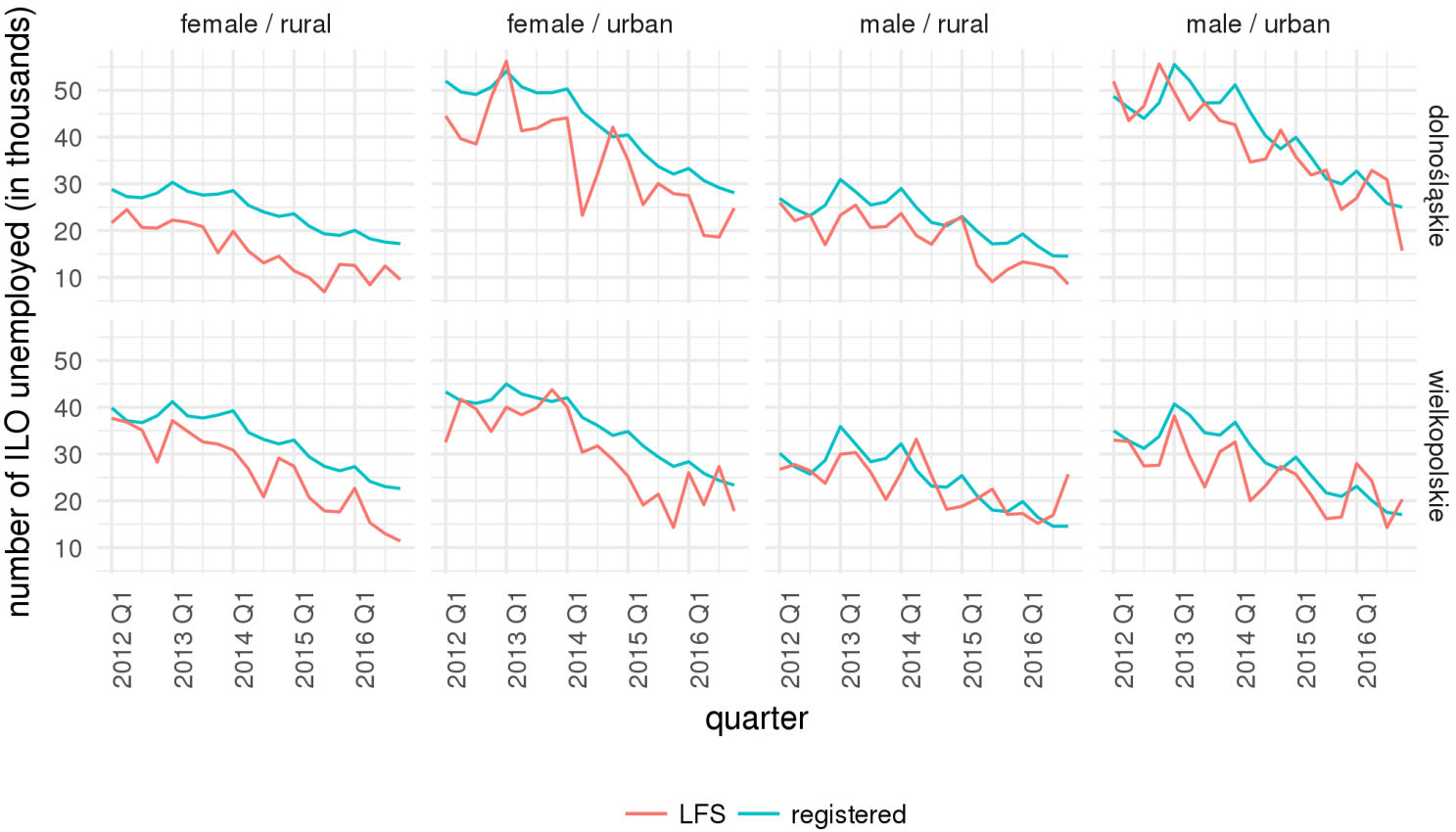

Fig.1: Number of unemployed in Poland, 2012-2016

Registered unemployment vs LFS unemployment

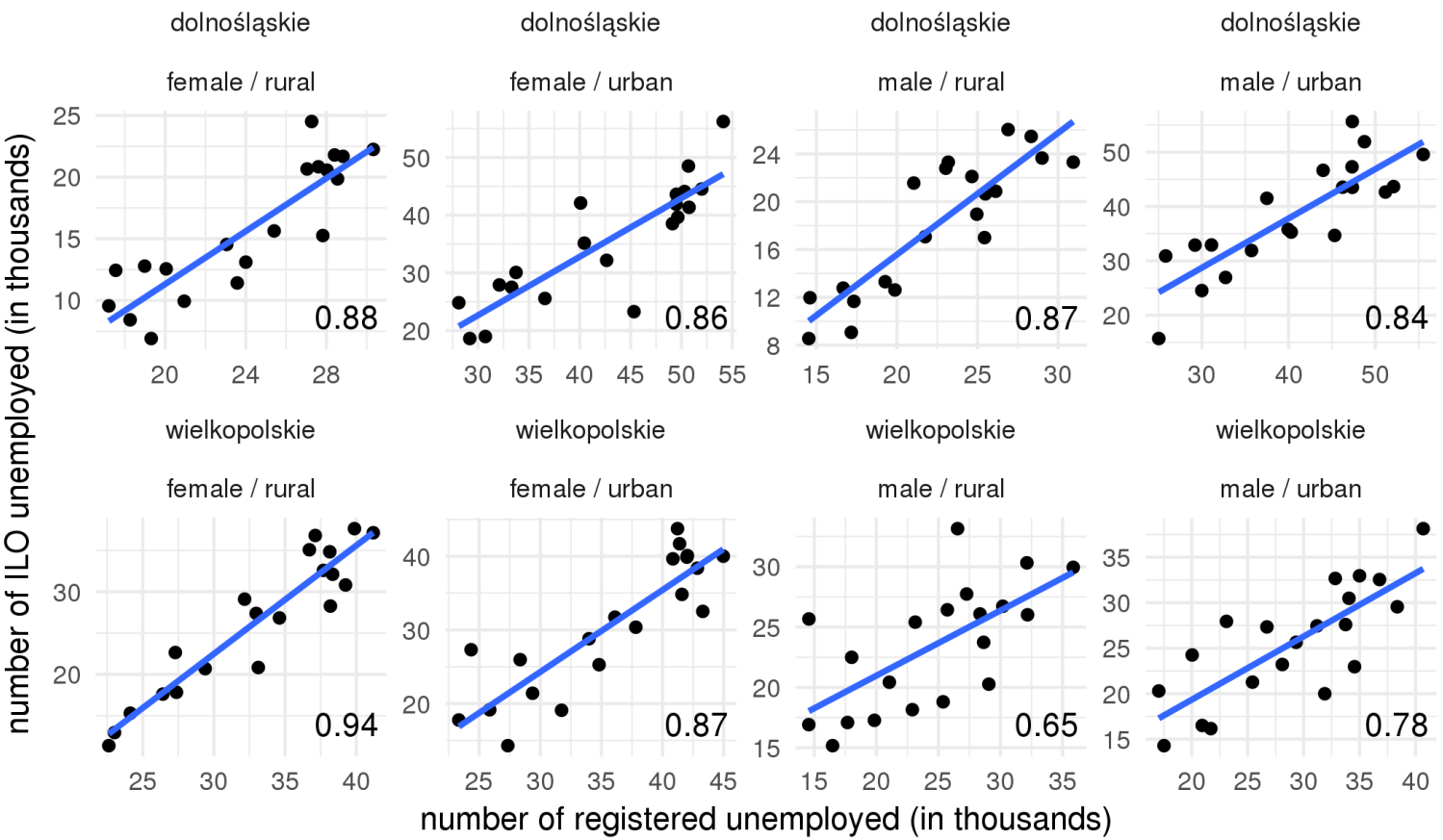

Fig.2: Correlation between registered and LFS unemployment, 2012-2016

Precision of direct estimation of unemployed number

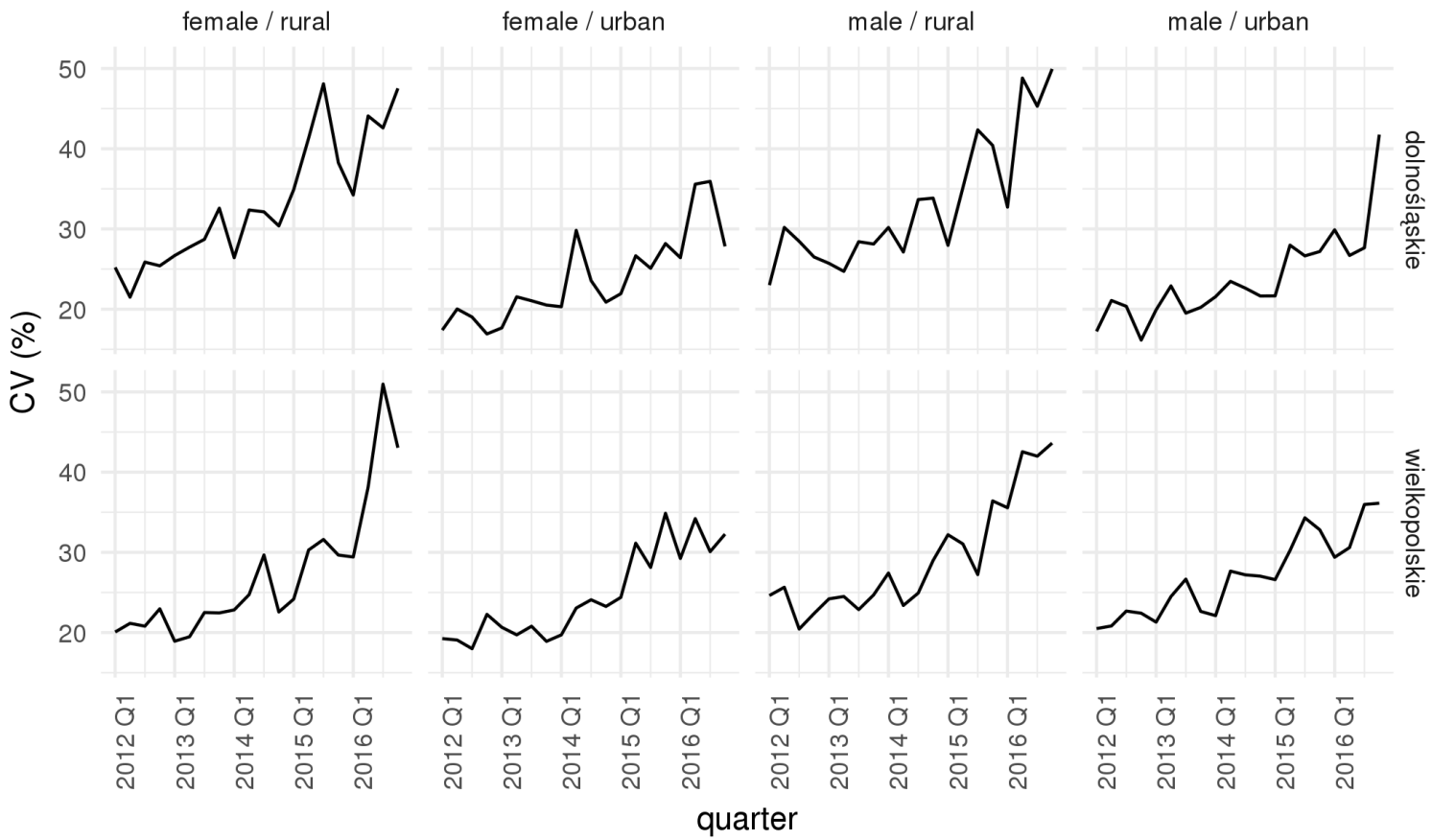

Fig.3: Precision of HT estimates of unemployed number, 2012-2016

Purpose of the study

- The main goal of the study is to obtain quarterly estimates of the number of unemployed people for domains that have not been published so far, i.e. the number of unemployed cross-classified by province, sex and place of residence.

The study

- Data: registered unemployment in 2012–2016
- LFS in 2012–2016
- Domain – province (16) x sex (2) x place of residence (2)
- Estimated parameter – number of unemployed (as defined in LFS)
- Association structure – registered unemployment (proxy)
- Allocation structure – LFS-based marginal estimates for the target contingency table
- Methods: SPREE, GLMM, GLSMM

The study

Study – the contingency table used for the estimation of the number of unemployed

Province	Sex/Place of residence				In total
	male/ urban	male/ rural	female/ urban	female/ rural	
Dolnośląskie					
...					
Zachodniopomorskie					
In total					

association structure from registered unemployment

allocation structure based on the LFS

The main idea of the SPREE estimator

- Structure Preserving Estimation is a generalised class of synthetic estimators in the sense that they fully exploit information provided by direct estimates.
- The method involves updating cell counts in a multi-way contingency table so that adjusted values add up to known marginal totals.
- Input counts inside contingency cells can come from a census or administrative register, while marginal totals represent reliable direct estimates obtained from a sample survey (for instance from LFS).
- SPREE estimators can be used to produce estimates of totals for small domains in intercensal periods and for more detailed defined cross-sections.
- SPREE estimators apply reweighting in contingency tables, by means of so-called iterative proportional fitting (IPF).

SPREE, GLSM and GLSMM estimators

- Y_{aj}, X_{aj} – denote the target variable and the proxy count for the domain of interest in the cell of a two-way contingency table, where a denotes a small area / domain identifier, while j denotes levels of a classification variable. Let us also assume that $logY_{aj}$, $logX_{aj}$ can be expressed in the form of a saturated log-linear model as:

$$logY_{aj} = \alpha_0^Y + \alpha_a^Y + \alpha_j^Y + \alpha_{aj}^Y := \xi_{aj}^Y, \tag{1}$$

$$logX_{aj} = \alpha_0^X + \alpha_a^X + \alpha_j^X + \alpha_{aj}^X := \xi_{aj}^X, \tag{2}$$

  where
  - $\alpha_0^Y = \overline{\xi_{..}^Y}$ - „overall" effect
  - $\alpha_a^Y = \overline{\xi_{a.}^Y} - \alpha_0^Y$ - effect of domain a
  - $\alpha_j^Y = \overline{\xi_{.j}^Y} - \alpha_0^Y$ - effect of j level of classification variable
  - $\alpha_{aj}^Y = \xi_{aj}^Y - \alpha_0^Y - \alpha_a^Y - \alpha_j^Y$ - effect of interaction

  for $a = 1, \ldots, A$ and $j = 1, \ldots, J$.
- The SPREE estimator is based on the assumption that:

$$\alpha_{aj}^Y = \alpha_{aj}^X. \tag{3}$$

SPREE, GLSM and GLSMM estimators

- Let's assume that:

$$\theta_{aj}^{Y} = Y_{aj}/Y_{a.} \quad \tau_{aj}^{Y} = log\theta_{aj}^{Y} - \frac{1}{J}\sum_{j=1}^{J} log\theta_{aj}^{Y} \tag{4}$$

$$\theta_{aj}^{X} = X_{aj}/X_{a.} \quad \tau_{aj}^{X} = log\theta_{aj}^{X} - \frac{1}{J}\sum_{j=1}^{J} log\theta_{aj}^{X} \tag{5}$$

- The assumption (3) can be relaxed by modeling τ_{aj}^{Y} using GLSM. Then we obtain Generalised SPREE (GSPREE) estimators, which assume proportionality between the association structures of the target and proxy compositions:

$$\tau_{aj}^{Y} = \theta_j + \beta\tau_{aj}^{X} \quad \longrightarrow \quad \alpha_{aj}^{Y} = \beta\alpha_{aj}^{X}. \tag{6}$$

- If model (6) is extended to a GLSMM model, which contains random effects μ_{aj} for different cells of the contingency table in order to decrease bias, we obtain the so-called GSPREE estimator with random effects:

$$\tau_{aj}^{Y} = \theta_j + \beta\tau_{aj}^{X} + \mu_{aj} \quad \longrightarrow \quad \alpha_{aj}^{Y} = \beta\alpha_{aj}^{X} + \mu_{aj}. \tag{7}$$

- Parameters of GLSM and GLSMM can be estimated by the Iteratively Weighted Least Squares Algorithm (IWLS).

Selected estimates

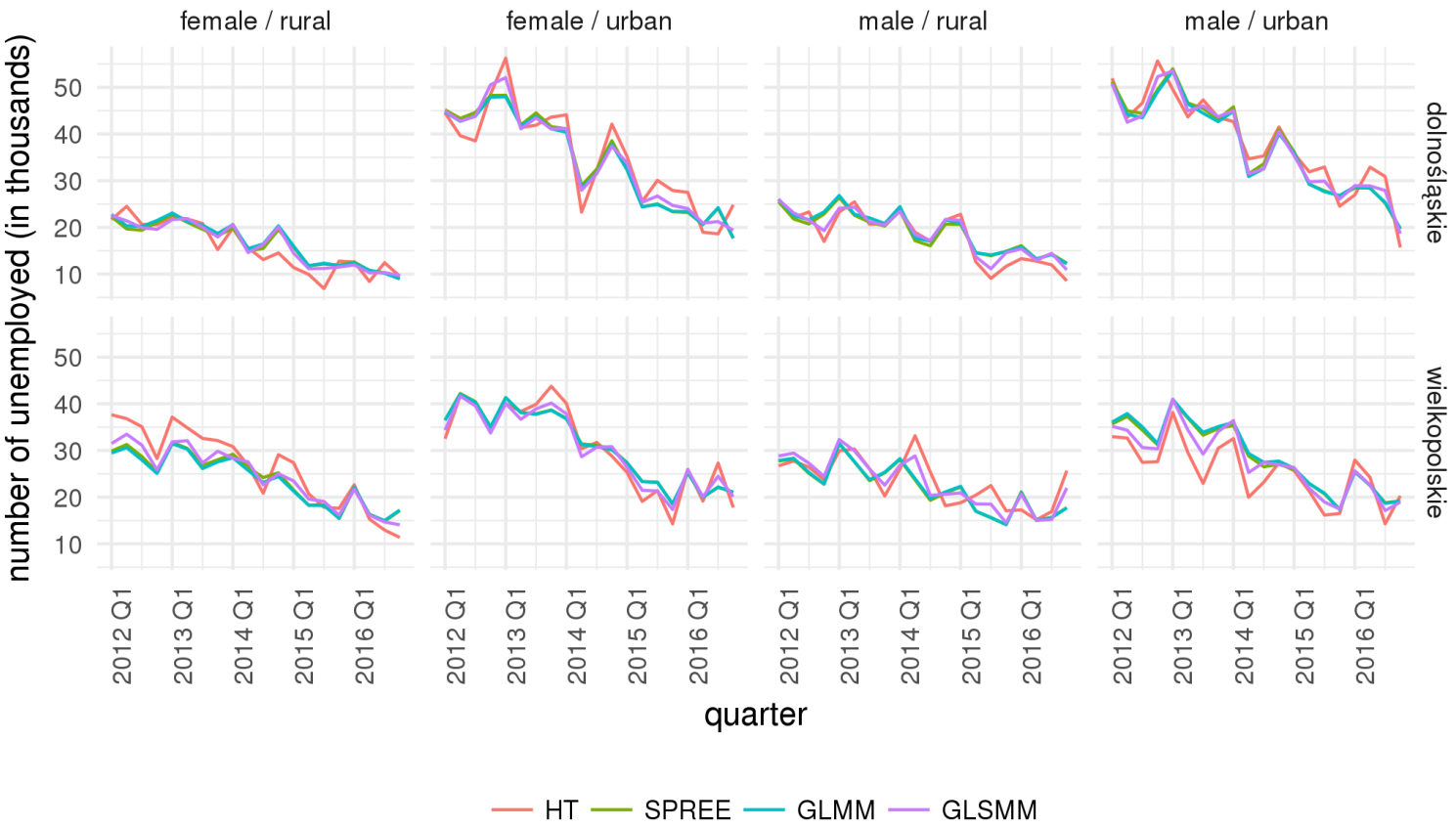

Fig.4: Estimates of the number of unemployed, 2012-2016

Selected estimates

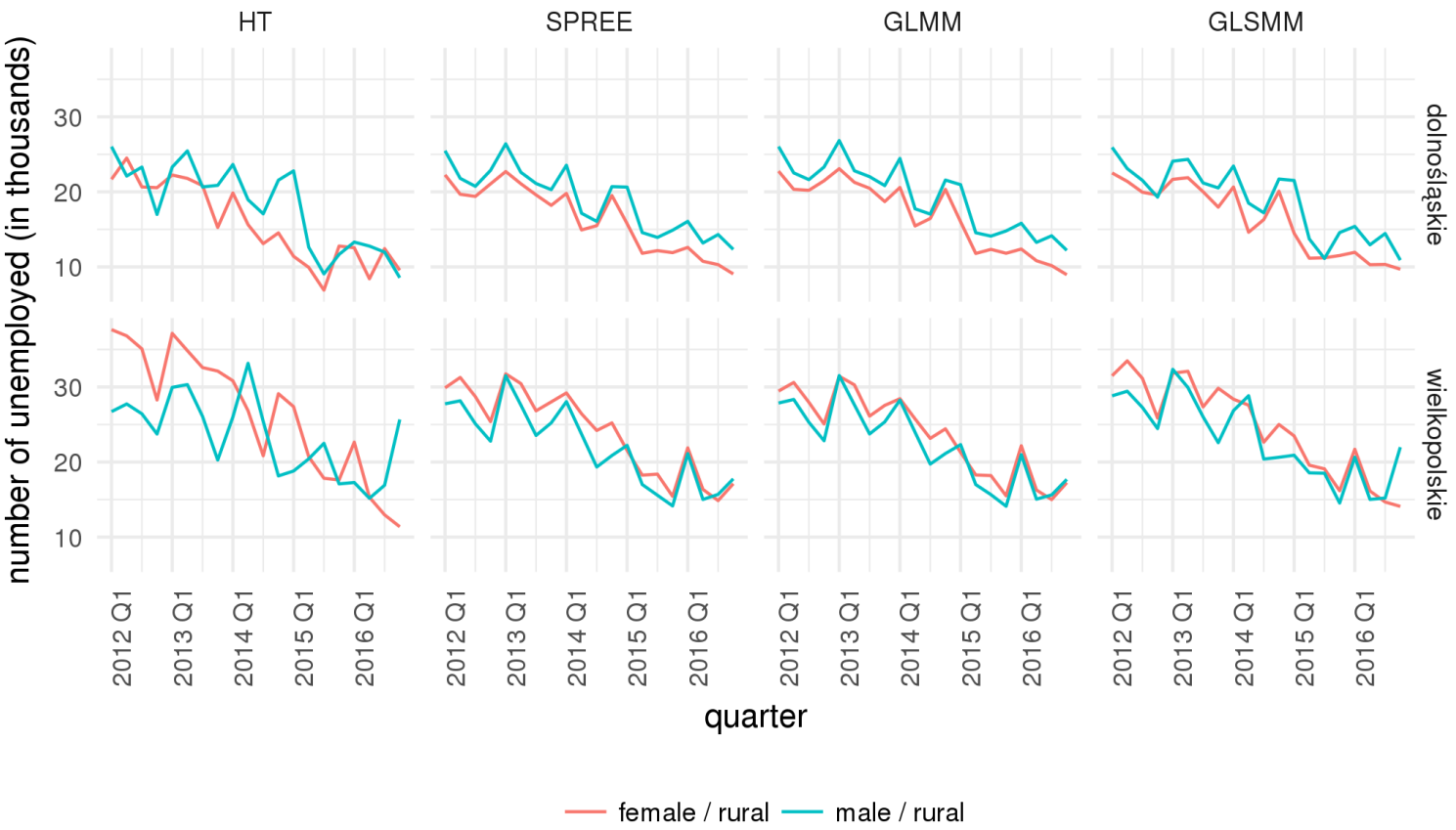

Fig.5: Estimates of the number of unemployed in rural areas, 2012-2016

Coefficient of variation

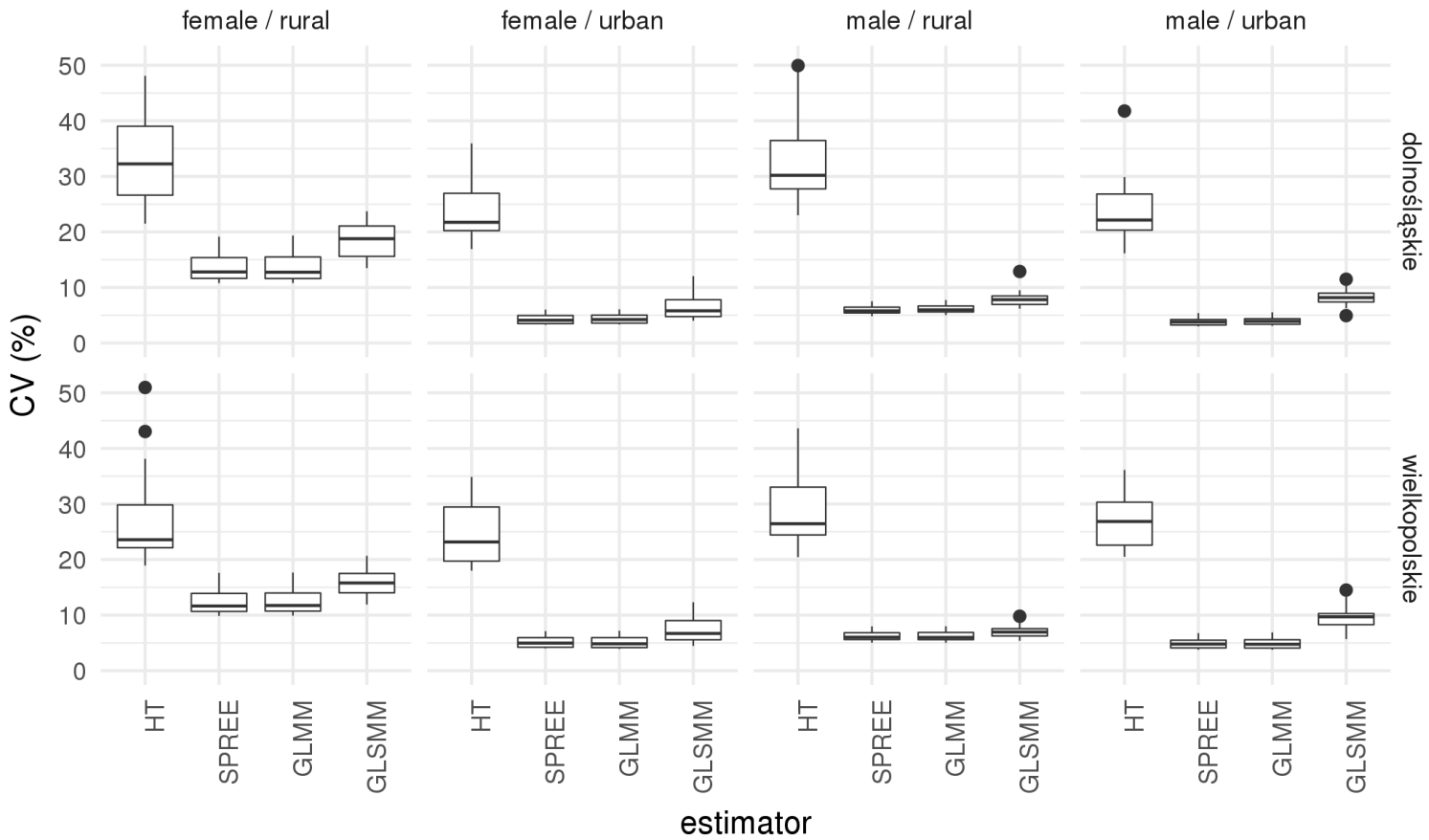

Fig.7: Coefficient of variation of the estimated number of unemployed, 2012-2016

Comparison of bias

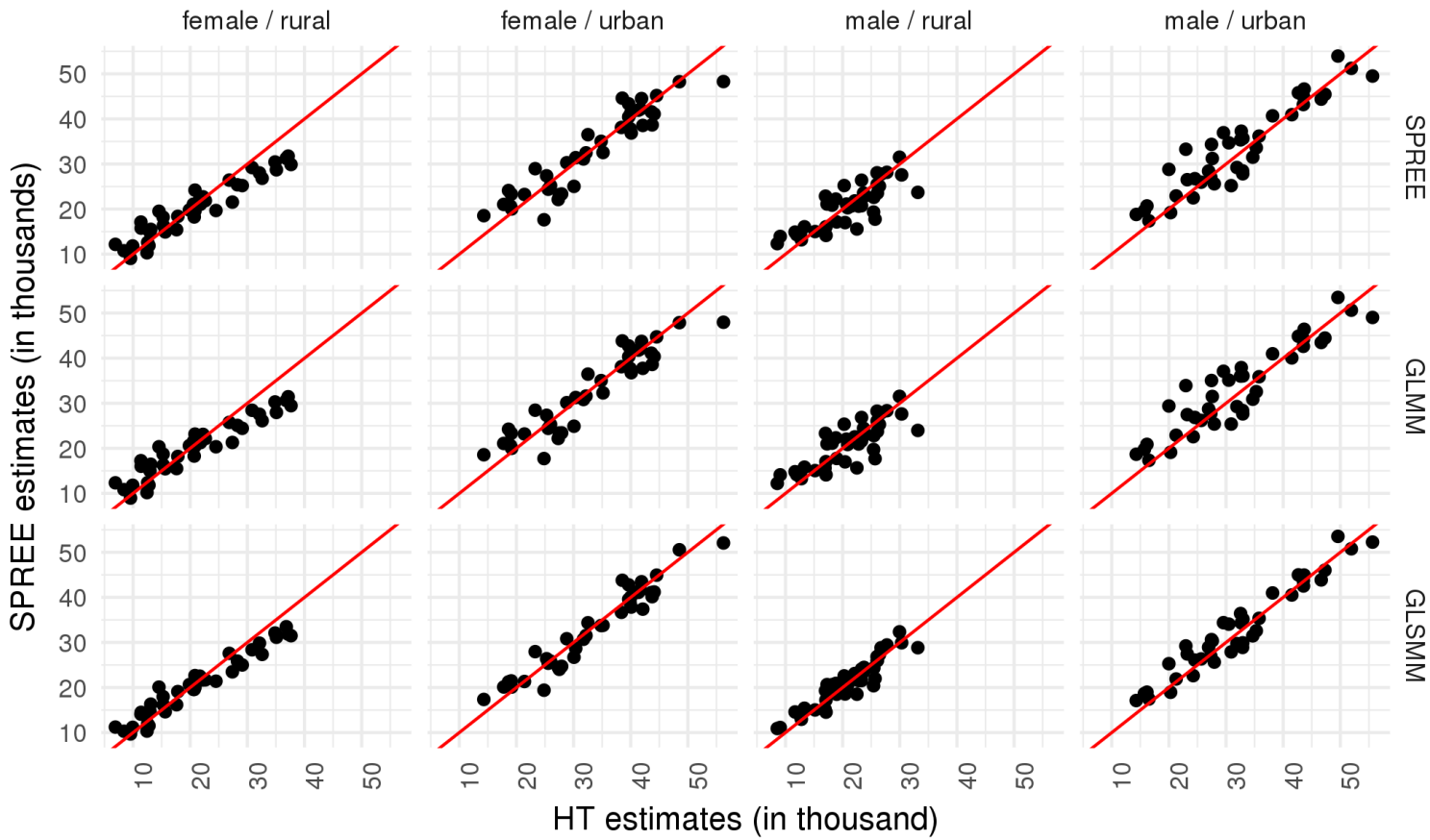

Fig.8: Bias diagnostic of the estimated number of unemployed, 2012-2016q

Summary

- SPREE-based estimates are consistent with direct estimates at a higher level spatial aggregation (benchmarking).
- SPREE estimators are more stable over time and retain the structure of estimates with respect to sex.
- Compared to the direct estimator, SPREE estimators are characterised by higher precision.
- The use of SPREE estimators makes it possible to obtain more relable estimates of the number of unemployed people for detailed domains.

References

- Eurarea Project (2004), *Enhancing Small Area Estimation Techniques to Meet European Needs*, Project documentation.
- Rao J.N.K., Molina I. (2015), *Small Area Estimation*, 2nd Edition, Wiley.
- Zhang Li-Chun, Chambers R. (2004), *Small area estimates for cross-classifications*, J. R. Statist. Soc. B, 66, Part 2.
- Zhang Li-Chun. (2009), *Estimates for small area compositions subjected to informative missing data*, Survey Methodology, Vol. 35, No. 2.

Thank you for your attention