STATISTICS LITHUANIA

INCOMPLETE ADMINISTRATIVE DATA FOR PRODUCTION OF OFFICIAL STATISTICS

Milda Šličkutė-Šeštokienė

Statistics Lithuania

STATISTICS LITHUANIA

ESSnet - Use of Administrative Data in Business Statistics

WP4 - Timeliness of Administrative Sources

Joint work of 7 NSI:

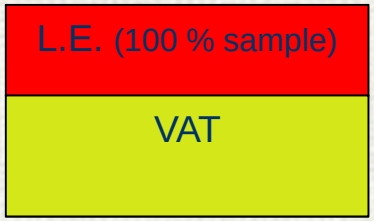
DESTATIS (Germany),
Statistics Estonia,
ISTAT (Italia),
Statistics Finland,
Statistics Lithuania,
Statistics Netherlands,
ONS (UK)

Aim of the project: **to create recomendations for all countries for dealing with incomplete administrative data**

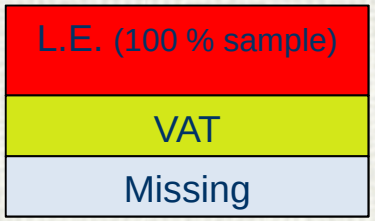
STATISTICS LITHUANIA

VAT (Value Added Tax) for STS (Short Term Statistics)

Final situation: (after year)
- all admin data are available for NSIs
- data cover the population



Monthly and quarterly estimates:
Part of admin data are 'missing'



Questions to be answered:

1. which of the identified and improved estimation methods produce good quality estimates for STS;
2. how estimates should be revised when more complete administrative data become available.

STATISTICS LITHUANIA

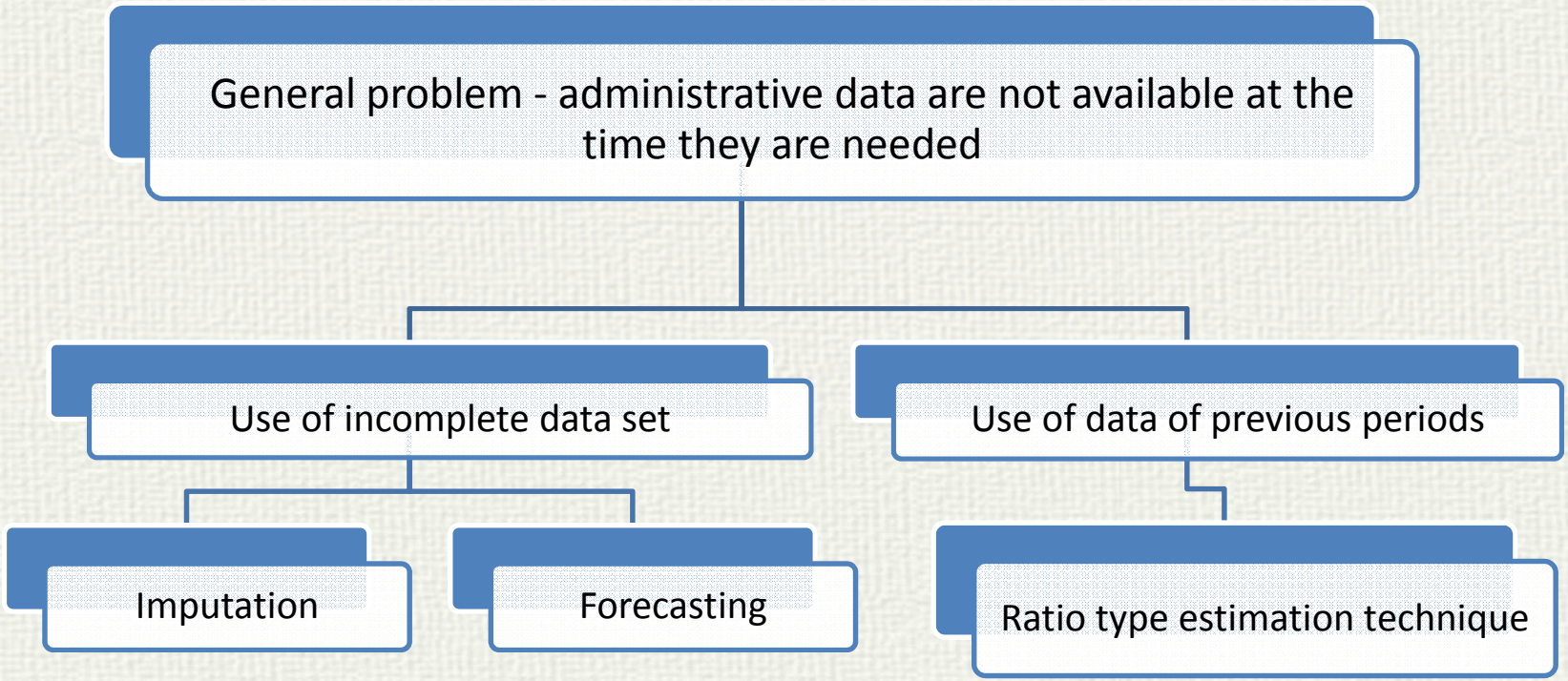
Two situations:

a. **VAT fairly complete and representative**

short summary of established techniques

b. **VAT not complete and/or not-representative**

regression techniques should be applied


General problem - administrative data are not available at the time they are needed
Use of incomplete data set
Use of data of previous periods
Imputation
Forecasting
Ratio type estimation technique

STATISTICS LITHUANIA

VAT not complete and/or not-representative

Horvitz–Thompson: $$\hat{Y}^{HT} = \sum_{k \in s} \frac{1}{\pi_k} y_k$$

Ratio: $$\hat{Y}^{Ratio} = X_1 \frac{\hat{Y}^{HT}}{\hat{X}_1^{HT}} = \hat{Y}^{HT} \frac{X_1}{\hat{X}_1^{HT}}$$

Generalized regression (GREG) $$\hat{Y}^{GREG} = \hat{Y}^{HT} + \sum_{j=1}^{J} \hat{B}_j (X_j - \hat{X}_j^{HT})$$

where $$X_j = \sum_{k \in U} x_{kj}$$ – the total of the auxiliary variable x_j

$$\hat{\mathbf{B}} = (\hat{B}_1, ..., \hat{B}_J) = (\sum_{k \in s} \mathbf{x_k} \mathbf{x_k}' / \sigma_k^2 \pi_k)^{-1} \sum_{k \in s} \mathbf{x_k} y_k / \sigma_k^2 \pi_k$$

the vector of regression coefficients derived applying a linear regression model

STATISTICS LITHUANIA

Lithuanian VAT data

- No VAT data of reference period are available for data estimation;
- VAT data for (t-1) period are used as auxiliary information;
- Ratio and GREG estimators were analyzed.

STATISTICS LITHUANIA

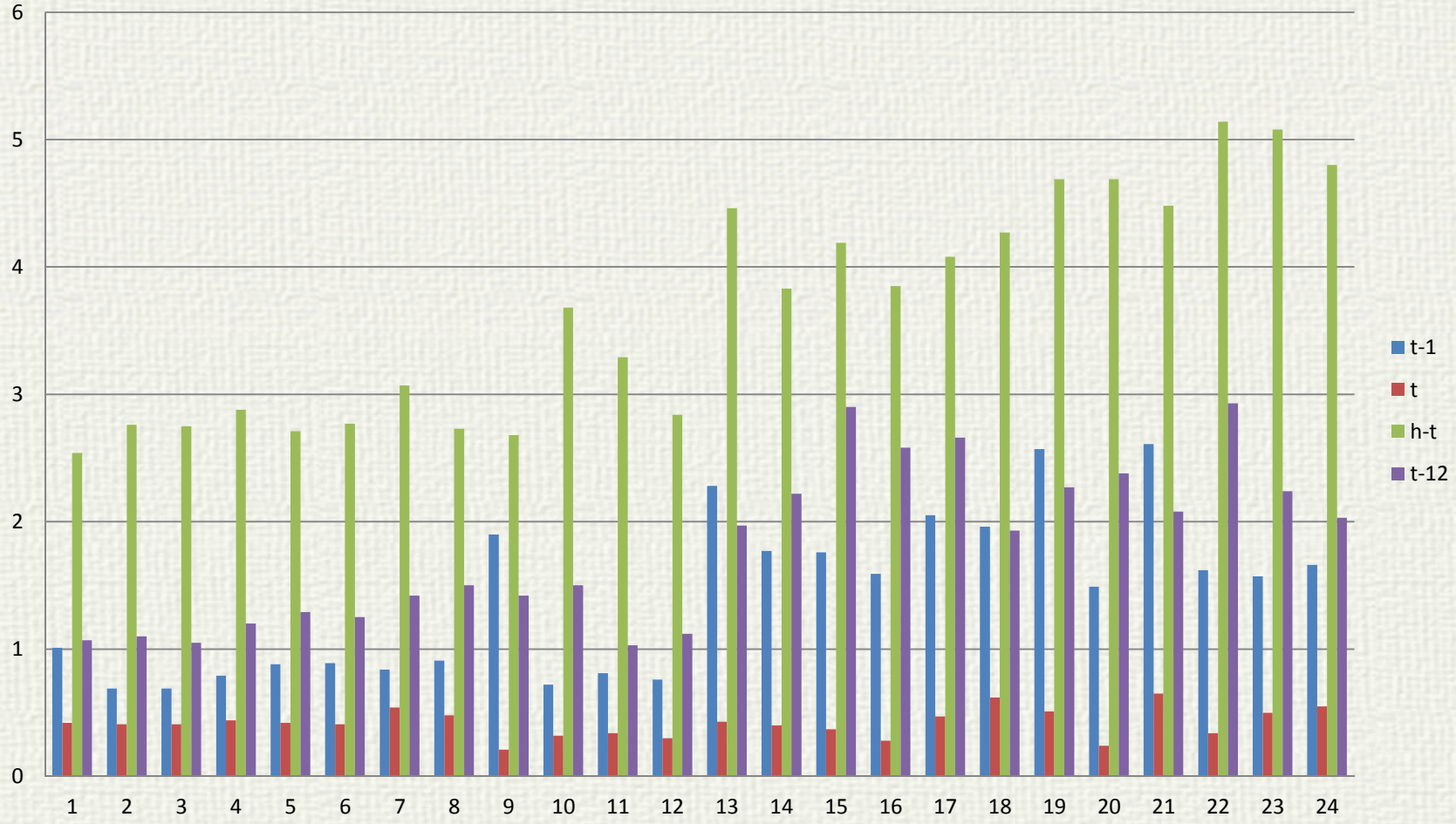
Coefficients of correlation for Income (survey) / (VAT)

	Month	1	2	3	4	5	6	7	8	9	10	11	12
2010	Inc.(t)/VAT(t-1)	0,996	0,999	0,973	0,996	0,999	0,999	0,999	0,998	0,997	0,998	0,997	0,997
	Inc.(t)/VAT(t)	1,000	0,974	1,000	1,000	1,000	1,000	1,000	1,000	0,999	0,999	0,998	0,983
2011	Inc.(t)/VAT(t-1)	0,982	0,996	0,998	0,998	0,998	0,999	0,999	0,999	0,999	1,000	1,000	0,982
	Inc.(t)/VAT(t)	0,999	0,999	0,999	0,999	0,999	1,000	0,998	0,997	0,998	0,991	0,993	0,999

STATISTICS LITHUANIA

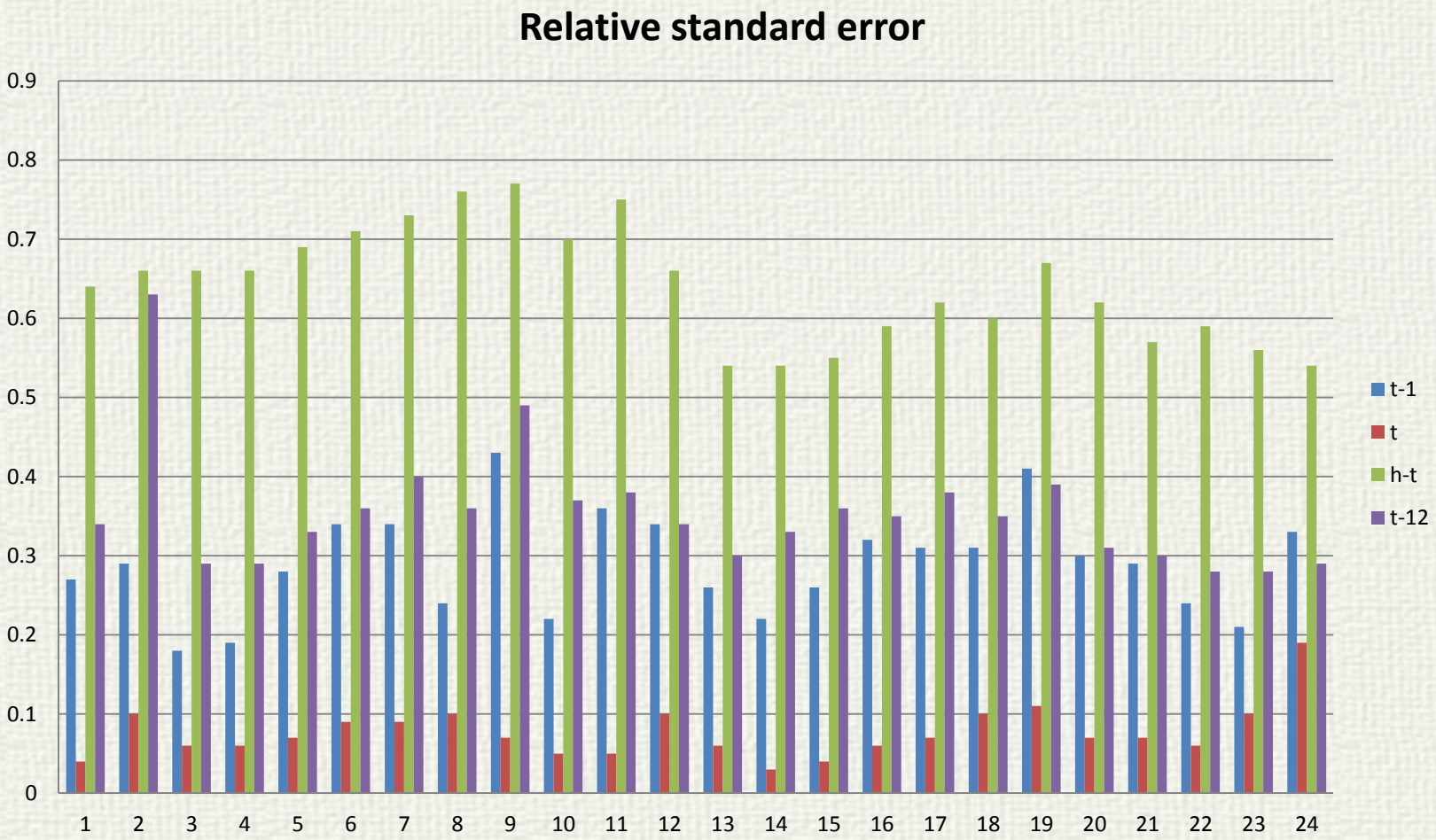
I56 Food and beverage service activities

Relative standard error



STATISTICS LITHUANIA

G47 Retail trade, except of motor vehicles, motorcycles and fuel



STATISTICS LITHUANIA

Future plans
Italian solution for timeliness (used till 2008 for employment data)

- Survey for large enterprises (>500 employees);
- Only incomplete administrative data for small and medium enterprises – **referred as sample;**
- GREG estimator was applied.

Problems that have to be solved

- Non randomness of the sample;
- Small areas.

STATISTICS LITHUANIA

Thank you for your attention ☺