GDP FLASH ESTIMATES: SOPHISTICATION THROUGH SIMPLICITY

Session number: 11
Date: 1st June, 2016

Kristina Kiriliauskaitė
Statistic Lithuania
kristina.kiriliauskaite@stat.gov.lt

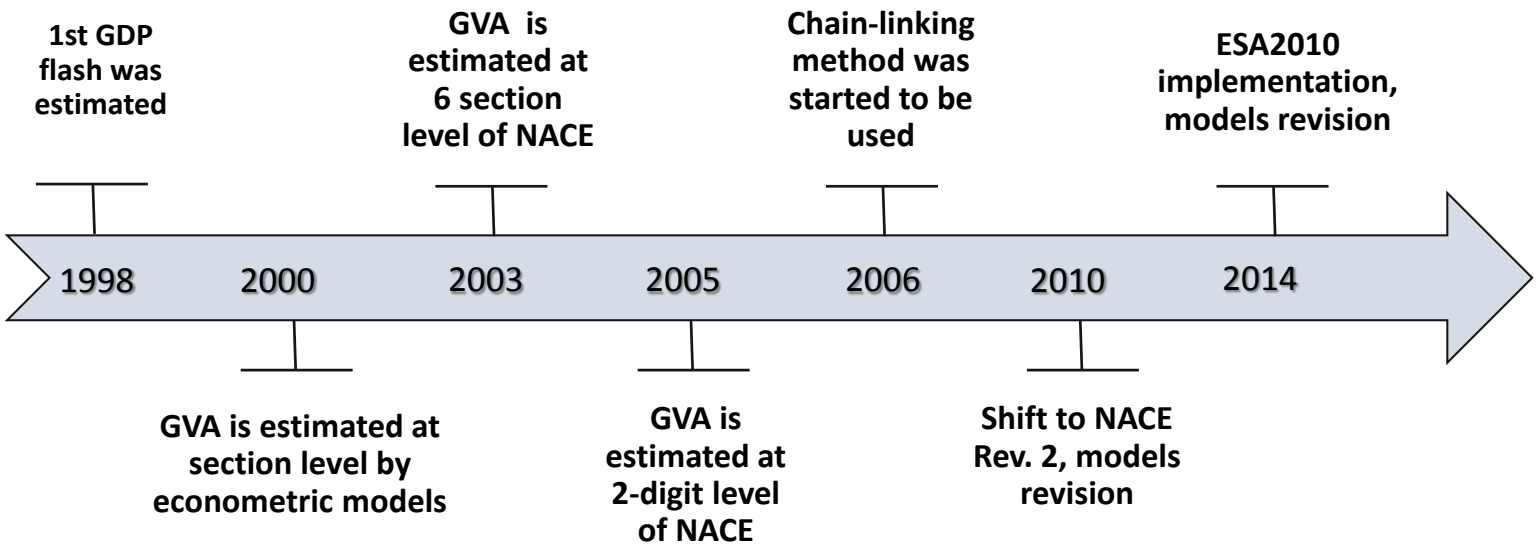
Madrid, May 31 - June 3

Outline

- Statistics Lithuania‘s (SL) GDP estimates evolution
- Methodology background
- Data sources
- Modelling techniques
- Combined forecasting
- Quality assurance
- Final remarks

SL first estimate

Historical development and methodological changes in Gross domestic product (GDP) and Gross Value Added (GVA) estimation process:



Methodology background

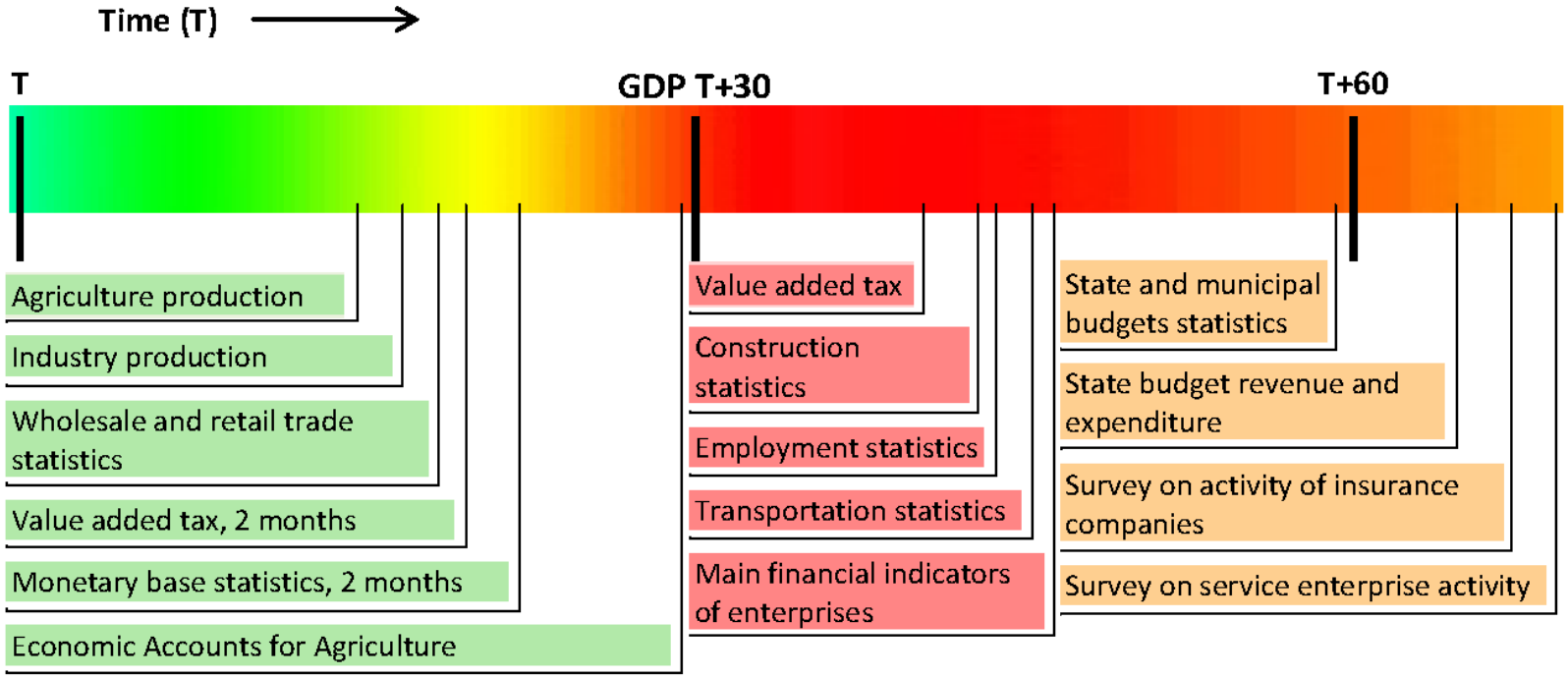
SL GDP flash estimate is based on production approach method.
GDP flash values are estimated at current and previous year's price using indirect method:

$$\widehat{GDP} = \sum_{i=1}^{n} \widehat{VA}_i + \widehat{D.21} - \widehat{D.31},$$

here $\widehat{VA}_i$ – value added at activity i, $\widehat{D.21}$ - taxes on products, $\widehat{D.31}$ - subsides on products.

The value added (VA) is estimated at the 2-digit level of NACE Rev. 2.
VA of 88 economical activities at current prices and 88 at previous prices are estimated using econometric models.
This composes the set of 176 endogenous variables to be observed.
Chain linking method is used.

Data source and availability



The necessary data is available between 20th-26th days after the end of the corresponding quarter. Some monthly data are received only for two months of quarter.

40% of all necessary data for GDP t+30 preparation is accessible and remaining 60 % of missing data must be extrapolated or nowcasted using mathematical, econometrical techniques.

Modelling techniques (1)

Econometric methods used for SL GDP t+30 modelling:

- *multivariate regression models with trend and seasonal dummies, outliers:*

$$y_{i,t} = \mu_{i,t} + \alpha_1 x_{1,t-l} + \alpha_2 x_{2,t-l} + \cdots + \alpha_n x_{n,t-l} + \beta_k S_{k,t} + \tau_i O_{i,t} + \varepsilon_{i,t}$$

$y_{i,t}$ – dependent variable of the VA of i economical activity at time – t, $t = (1, \cdots, T)$,
T – time series length,
$\mu_{i,t}$ – trend of the value added i,
$x_{j,t-l}$ – regressor j, $j = \overline{1,n}$, with a time lag $l = (0, \cdots, 4)$,
$S_{k,t}$ – seasonal variables, which are equal to 1 if quarter depends to first, second, third or fourth season, otherwise – 0, $k = (1, \cdots 4)$,
$O_i(t)$ represents outliers: additive outlier, level shift or transitory change;
α_j, β_k, τ_i – model parameters, coefficients of regressors,
$\varepsilon_{i,t} \sim N(0, \sigma^2)$ is random disturbance.

Model's parameters are estimated by Least Squares method.

Modelling techniques (2)

- *ARIMA family models (AR(p), MA(q), ARMA(p,q), ARIMA(p,d,q), ARIMAX(p,d,q)):*

$$\phi(L)(1-L)^d Y_t = \Theta(L)X_t + \theta(L)\varepsilon_t,$$

$\phi(L)$ is an autoregressive polynomial,
$\theta(L)$ – moving average polynomial,
$\Theta(L)$ – explanatory variables polynomial,
L – a lag operator,
d– d^{th} difference operator.

This model is extended ARIMA model with the addition regressors $X_t$included into. It includes more realistic dynamics and additional information from the explanatory variables.

ARIMA(X) model selection process is done by The Box-Jenkins procedure.

Example (1)

Wholesale and retail trade (G) value added is estimated using indirect method and extended multivariate regression models:

$$y_{G,t} = \sum_{i=45}^{47} y_{i,t}$$

i is value added of G section activities (45, 46, 47) at 2-digit level of NACE Rev. 2.

Example (2)

- *Wholesale and retail trade and repair of motor vehicles and motorcycles*:

$$y_{45,t} = 86 + 0.0002 \cdot x_{1,t} - 0.0001 \cdot x_{2,t} - 30.9 \cdot S_{1,t} - 5.7 \cdot S_{4,t} + 10 \cdot O_t + \varepsilon_{45,t},$$

$x_{1,t}$ – turnover of G45,
$x_{2,t}$ – value add tax of G45,
$S_{1,t}, S_{4,t}$ – 1st and 4th seasonal dummies,
O_t – level shift outlier.

Statistical characteristics	μ	α_1	α_2	β_1	β_2	τ_1	$\boldsymbol{R^2}$
Significance/ p-value	0.000	0.000	0.004	0.000	0.049	0.049	**0.89**

Example (3)

- *Wholesale trade except of motor vehicles and motorcycles*:

$$y_{46,t} = 163 - 0.0001 \cdot x_{1,t} + 0.0001 \cdot x_{2,t} - 29.5 \cdot S_{1,t} - 32.6 \cdot S_{4,t} + 6.7 \cdot O_t + \varepsilon_{i,t},$$

$x_{1,t}$ – value add tax of G46,
$x_{2,t}$ – wholesales trade results,
$S_{1,t}, S_{4,t}$ – 1st and 4th seasonal dummies,
O_t – level shift outlier.

Statistical characteristics	μ	α_1	α_2	β_1	β_2	τ_1	$\boldsymbol{R^2}$
Significance/ p-value	0.024	0.049	0.006	0.049	0.049	0.049	**0.92**

Example (4)

- *Retail trade, except of motor vehicles and motorcycles*:

$$y_{47,t} = 81 + 0.0002 \cdot x_{1,t} + 0.0003 \cdot x_{2,t} - 35.7 \cdot S_{1,t} - 4.4 \cdot S_{4,t} + \varepsilon_{i,t}$$

$x_{1,t}$ – retail trade results,
$x_{2,t}$ – value add tax of G47,
$S_{1,t}, S_{4,t}$ – 1st and 4th seasonal dummies.

Statistical characteristics	μ	α_1	α_2	β_1	β_2	$\boldsymbol{R^2}$
Significance/ p-value	0.049	0.000	0.049	0.004	0.049	**0.93**

Example (5)

ARIMAX model is used for telecommunications (J61) value added estimation:

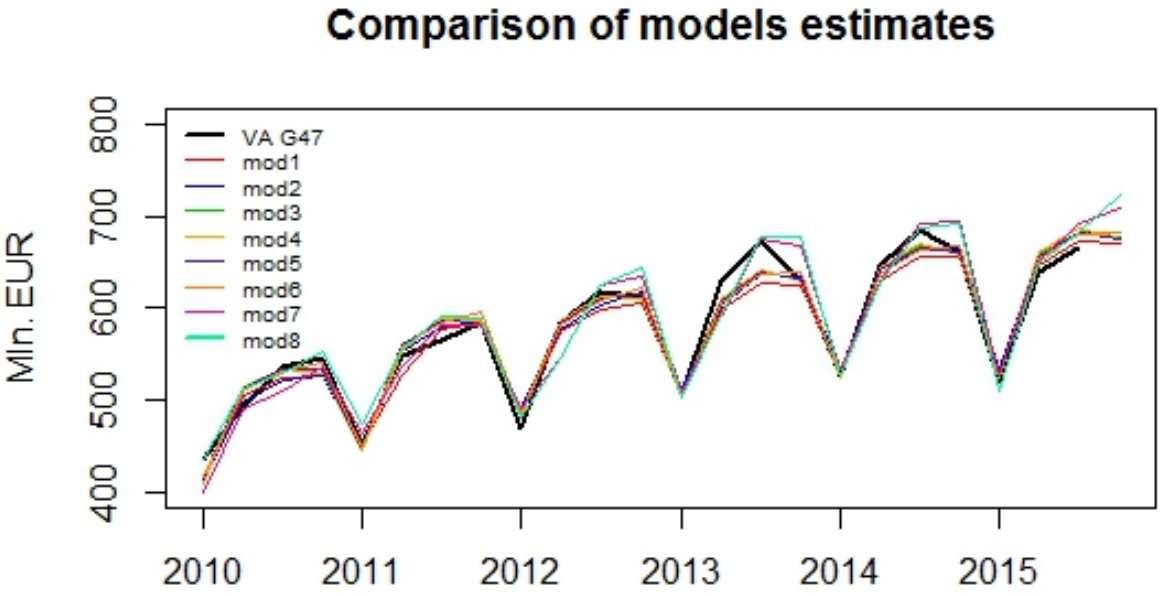
$$y_{61,t} = 8.1 + 0.0015 \cdot y_{61,t-1} + 0.0001 \cdot x_{1,t} + 0{,}52 \cdot x_{2,t} - 2{,}34.4 \cdot S_{4,t} + \varepsilon_{i,t},$$

$x_{1,t}$ – value add tax of J61,
$x_{2,t}$ – J61 services incomes.

Statistical characteristics	μ	ϕ_1	θ_2	θ_1	β_1	AIC
Significance/ p-value	0.049	0.04	0.049	0.001	0.049	207

Combined nowcasting

G47 estimates for 2015 Q4



Model no.	MAPE	$\Delta t-4$, % 2015 Q4
mod1	2.29	1.61
mod2	2.26	3.23
mod3	1.88	2.45
mod4	1.91	2.19
mod5	1.95	2.38
mod6	2.11	3.37
mod7	3.05	7.44
mod8	2.76	9.47

$$\min_i \hat{f}_i = 1.61, \quad \max_i \hat{f}_i = 9.47,$$

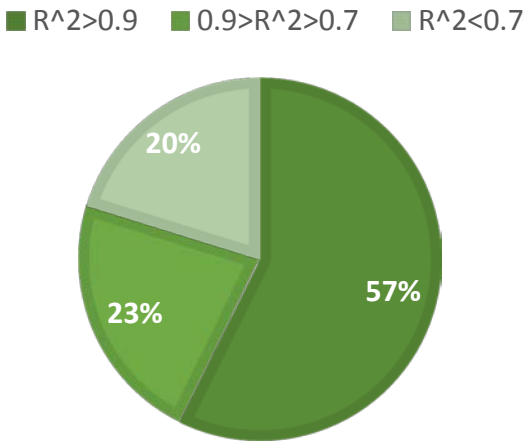
$$\text{average: } \hat{f} = \frac{1}{8}\sum_{i=1}^{8} \hat{f}_i = 4.02,$$

$$\text{weighted average: } \hat{f} = 0.05 \cdot 1.16 + 0.16 \cdot 3.23 + 0.16 \cdot 2.45 + 0.16 \cdot 2.19 + 0.16 \cdot 2.38 + 0.16 \cdot 3.37 + 0.05 \cdot 7.44 + 0.05 \cdot 9.47 = 3.29.$$

Quality assurance

Quality criteria's:

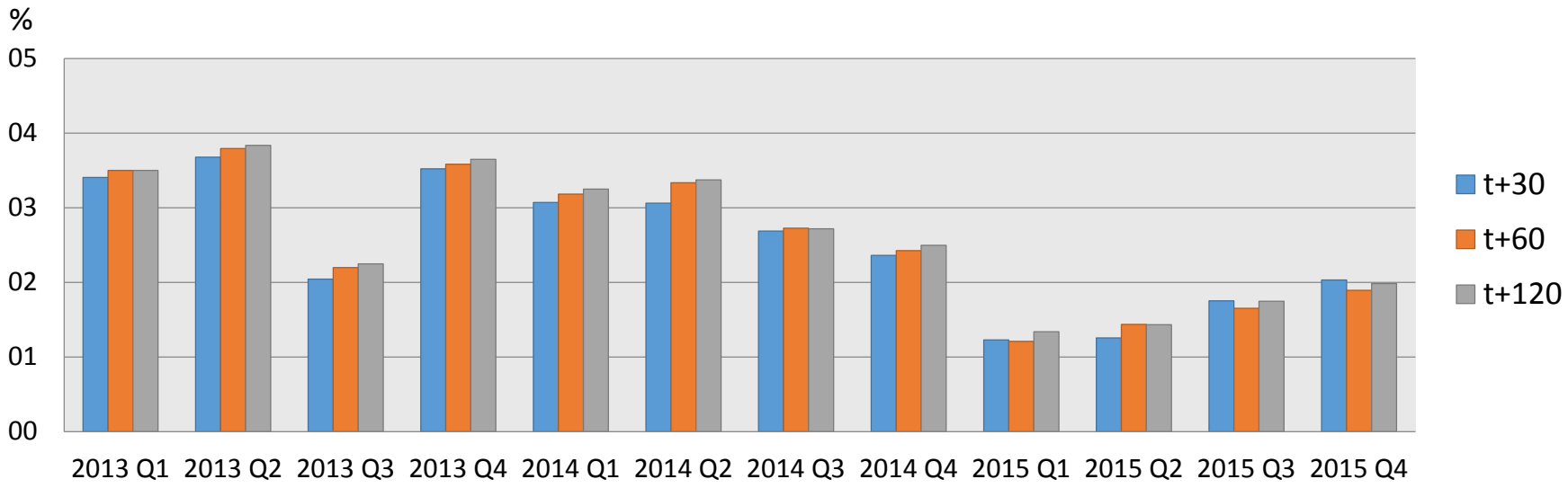
- Selected model is adequative: high R^2, low *AIC*
  - *80 % of models* $R^2 \geq 0.7$
- Model's parameters are significant
- Errors fulfils models acceptance criteria's
  - $(\varepsilon_{i,t} \sim N(0, \sigma^2), cov(\varepsilon_{i,t} = const)$
- $\mathrm{APE(t)} < 5\ \%, MAPE < 5\ \%$



Activities/ MAPEs at constants price	A	B-E	C	F	G-I	J	K	L	M-N	O-Q	R-T	GDP
Before ESA2010	5.6	1.9	1.9	5.5	3.5	3.0	5.5	38	3.4	3.2	5.5	**2.1**
After ESA2010	4.6	1.6	1.6	5.4	3.0	2.9	4.8	3.4	2.9	2.7	5.0	**0.3**

GDP estimate‘s growth rate‘s $\Delta t - 4$ comparative analysis

GDP t+30 and t+60 estimates are very close to t+120 estimates:



GDT t+30 estimated growth rates $\Delta t - 4$ do not differ from later estimation more than ∓ 1 percentage point.

Final remarks

- Estimation process is very intensive (6-10 days).
- SL GDP t+30 estimation methodology is a complex system with many classical extended econometrics' equations, mathematical tools.
- Extended multiple regressions and ARIMA family models compose stable system to obtain GDP t+30 estimates. Just model parameters are re-estimated every quarter.
- Estimation at 2 digits level of NACE rev. 2 derives better results.
- GDP t+30 estimates MAPEs do not reach 1 %.
- Global methodological changes (like ESA 2010) allowed retaining the same high quality of estimates as before or even increasing it.
- Constant models and quality monitoring is necessary.

Q 2016
European Conference on
Quality in Official Statistics

Thank you