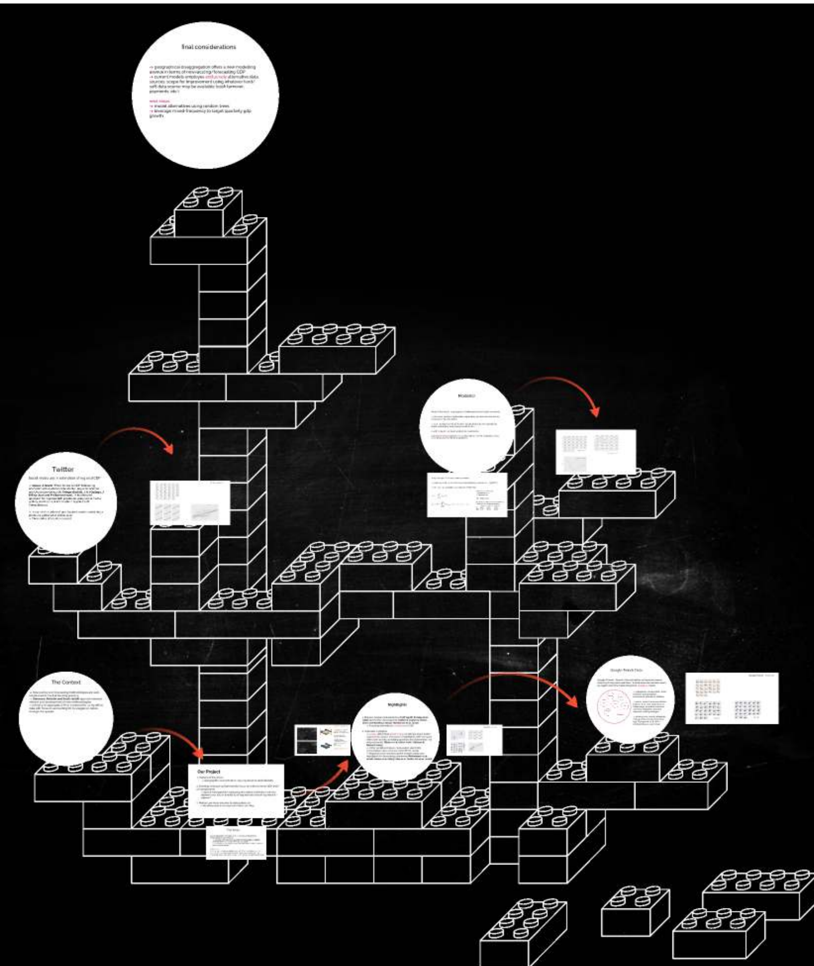
Warcasting-Nowcasting GDP without economic data

Constantinescu, Kappner, Szumilo



Mihnea Constantinescu
National Bank of Ukraine and
Kyiv School of Economics

BCC Conference
September 22-23, 2022

The Context

>> Nowcasting and forecasting methodologies are well-established in Central Banking practice
-> **Giannone, Reichlin and Small (2008)** spurred renewed interest and development of new methodologies
-> primarily of aggregate GDP or components, using official data with focus on accounting for its staggered nature through the quarter

The Context

Covid19

>> Nowcasting and forecasting methodologies are well-established in Central Banking practice
-> **Giannone, Reichlin and Small (2008)** spurred renewed interest and development of new methodologies
-> primarily of aggregate GDP or components, using official data with focus on accounting for its staggered nature through the quarter

>> Recent Covid experience highlighted the need for alternative data sources
-> VoxEU of **Diebold (2020)**; **Woloszko (2020)** on use of Google Trends; **Blanchflower and Bryson (2021)** on qualitative surveys or the "economics of walking about"

The Context

Covid19

>> Nowcasting and forecasting methodologies are well-established in Central Banking practice
-> **Giannone, Reichlin and Small (2008)** spurred renewed interest and development of new methodologies
-> primarily of aggregate GDP or components, using official data with focus on accounting for its staggered nature through the quarter

>> Recent Covid experience highlighted the need for alternative data sources
-> VoxEU of **Diebold (2020)**; **Woloszko (2020)** on use of Google Trends; **Blanchflower and Bryson (2021)** on qualitative surveys or the "economics of walking about"

>> Then the russian invasion starts...

The Context

>> Nowcasting and forecasting methodologies are well-established in Central Banking practice
-> **Giannone, Reichlin and Small (2008)** spurred renewed interest and development of new methodologies
-> primarily of aggregate GDP or components, using official data with focus on accounting for its staggered nature through the quarter

>> Recent Covid experience highlighted the need for alternative data sources
-> VoxEU of **Diebold (2020)**; **Woloszko (2020)** on use of Google Trends; **Blanchflower and Bryson (2021)** on qualitative surveys or the "economics of walking about"

>> Then the russian invasion starts...

Covid19

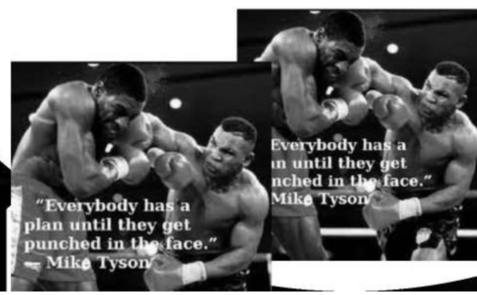
The Context

>> Nowcasting and forecasting methodologies are well-established in Central Banking practice
-> **Giannone, Reichlin and Small (2008)** spurred renewed interest and development of new methodologies
-> primarily of aggregate GDP or components, using official data with focus on accounting for its staggered nature through the quarter

>> Recent Covid experience highlighted the need for alternative data sources
-> VoxEU of **Diebold (2020); Woloszko (2020)** on use of Google Trends; **Blanchflower and Bryson (2021)** on qualitative surveys or the "economics of walking about"

>> Then the russian invasion starts...

Covid19



The Context

>> Nowcasting and forecasting methodologies are well-established in Central Banking practice
-> **Giannone, Reichlin and Small (2008)** spurred renewed interest and development of new methodologies
-> primarily of aggregate GDP or components, using official data with focus on accounting for its staggered nature through the quarter

>> Recent Covid experience highlighted the need for alternative data sources
-> VoxEU of **Diebold (2020); Woloszko (2020)** on use of Google Trends; **Blanchflower and Bryson (2021)** on qualitative surveys or the "economics of walking about"

Covid19

>> Then the russian invasion starts...

Our Project

1. Nature of the shock
   -> Geographic concentration, varying duration and intensity

2. Existing nowcasting frameworks focus on national level GDP and/or components
   -> Spatial heterogeneity: replicating the national estimation exercise depends crucially on availability of regional data (assuming national = regional)

3. Martial Law froze any and all data collection
   -> No official stats or surveys from March until May

First Steps

Rely on alternative indicators and build bayesian hierarchical models of key macro variables

-> using payments data as in **Galbraith and Tkacz (2018)**, **Chapman and Desai (2021)** : unavailable timespan and depth

-> using electricity and google/apple mobility data: no longer available due to security reasons;

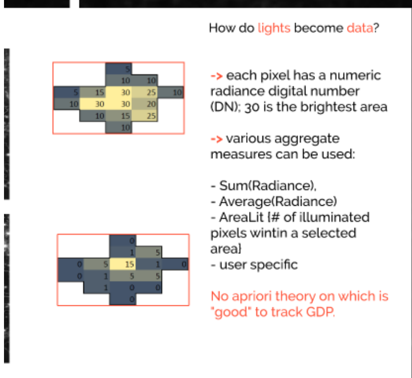
Bottom-line:

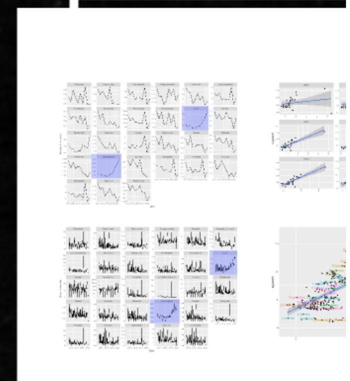
1. Some regional data available (yearly GDP at oblast level) but no regional variables as needed in established nowcasting exercise
2. Available data: nightlights, social media activity, Google Search Data

Nightlights

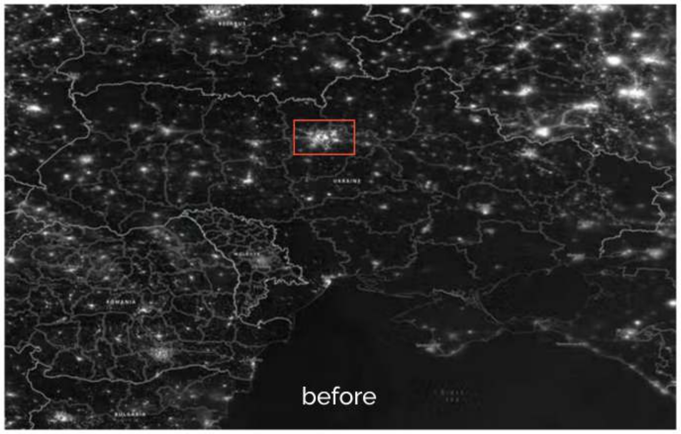
1. Mature literature kickstarted by **Croft (1978), Elvidge et al. (1997)** and further developed by **Sutton & Costanza (2002), Chen and Nordhaus (2011), Henderson et al. (2012)**
   -> Focusing primarily on country level GDP

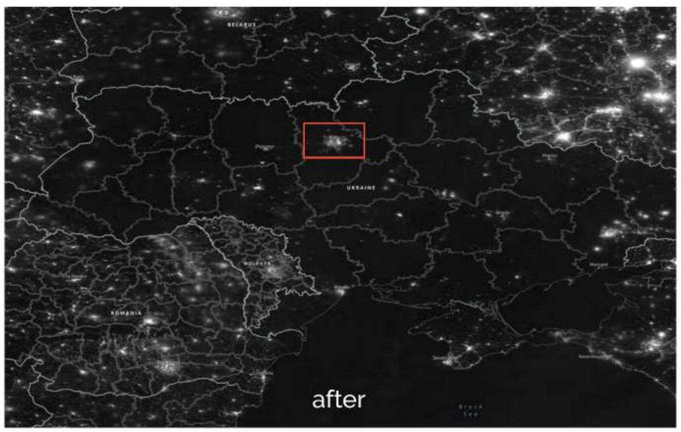
2. Important highlights
   -> Levels rather than growth rates of GDP are much better captured by various measures of nightlights (GDP can grow with urban activity spreading upwards and downwards, not only outwards); **Gibson et al. (2020, 2021), Addison & Stewart (2015)**
   -> Other variables in levels (population, electricity consumption) also correlate well with NL levels
   -> Regional cross-sectional panel models using only nightlights for developing economies **Bickenback et al. (2016), Keola et al. (2015), Xiao et al. (2021), Lin et al. (2022)**





How do lights become data?





-> each pixel has a numeric radiance digital number (DN); 30 is the brightest area

-> various aggregate measures can be used:

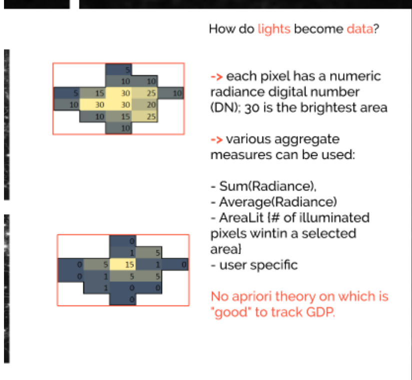
- Sum(Radiance),
- Average(Radiance)
- AreaLit {# of illuminated pixels wintin a selected area}
- user specific

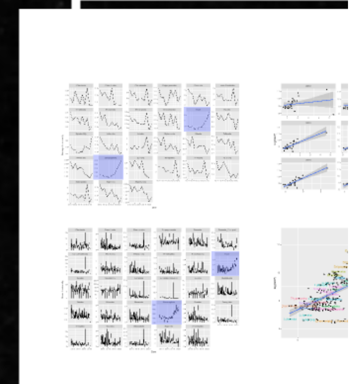
No apriori theory on which is "good" to track GDP.

Nightlights

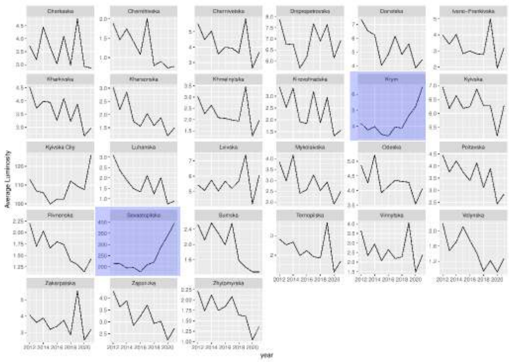
1. Mature literature kickstarted by **Croft (1978), Elvidge et al. (1997)** and further developed by **Sutton & Costanza (2002), Chen and Nordhaus (2011), Henderson et al. (2012)**
   -> Focusing primarily on country level GDP

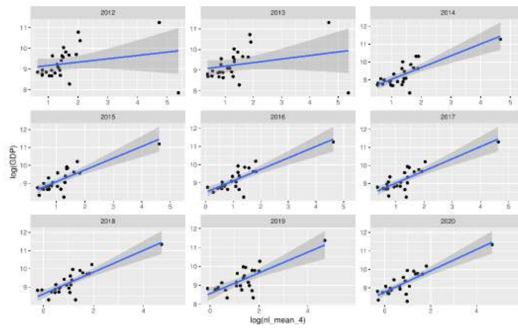
2. Important highlights
   -> Levels rather than growth rates of GDP are much better captured by various measures of nightlights (GDP can grow with urban activity spreading upwards and downwards, not only outwards); **Gibson et al. (2020, 2021), Addison & Stewart (2015)**
   -> Other variables in levels (population, electricity consumption) also correlate well with NL levels
   -> Regional cross-sectional panel models using only nightlights for developing economies **Bickenback et al. (2016), Keola et al. (2015), Xiao et al. (2021), Lin et al. (2022)**

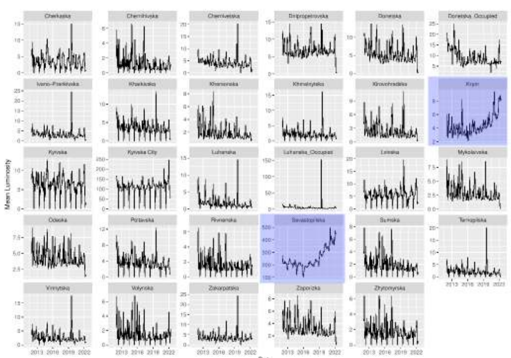


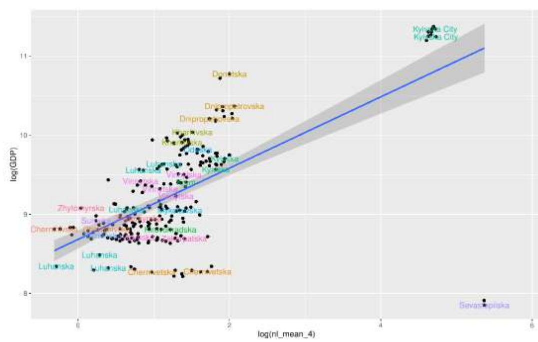


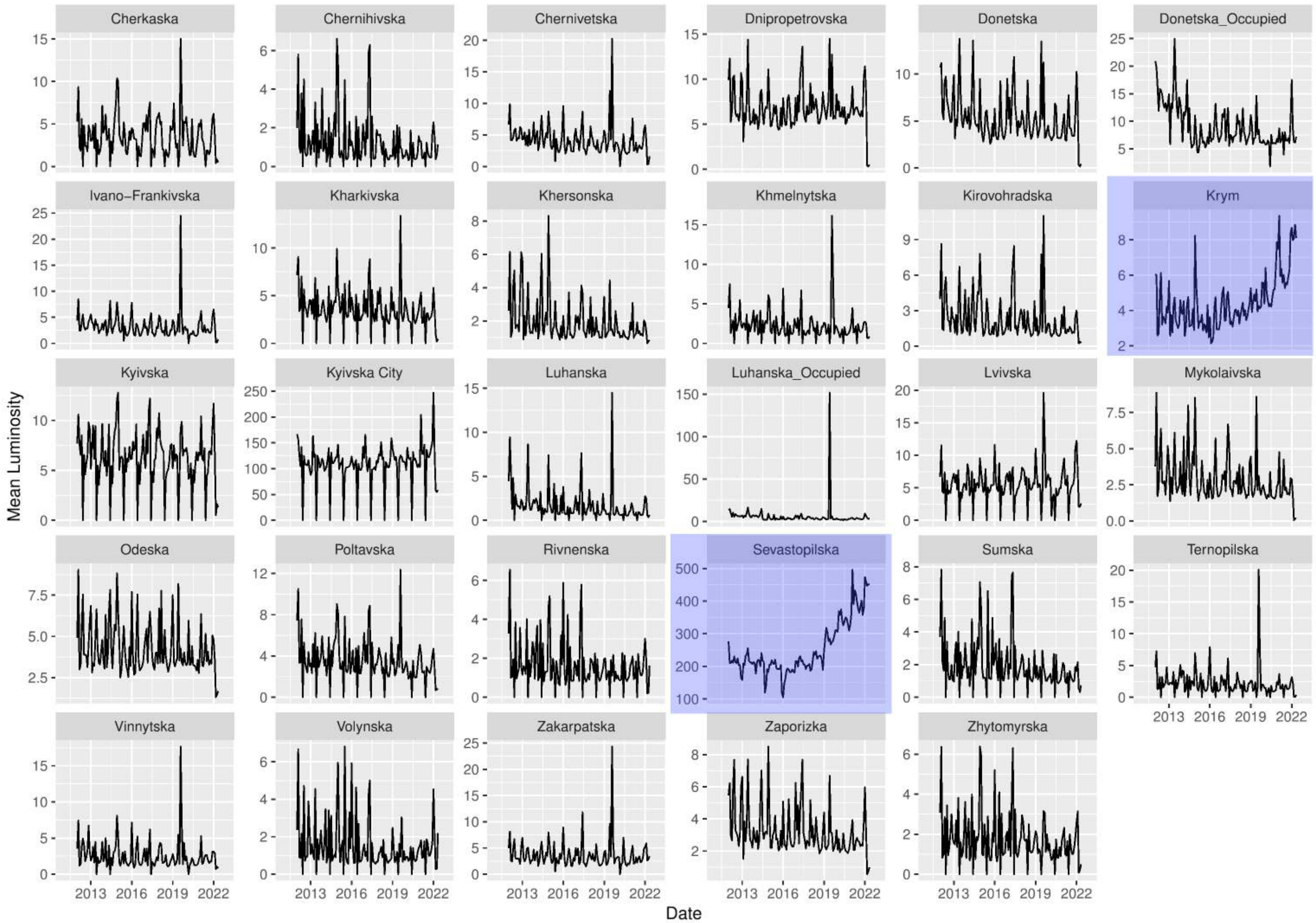
Nighlights - empirics



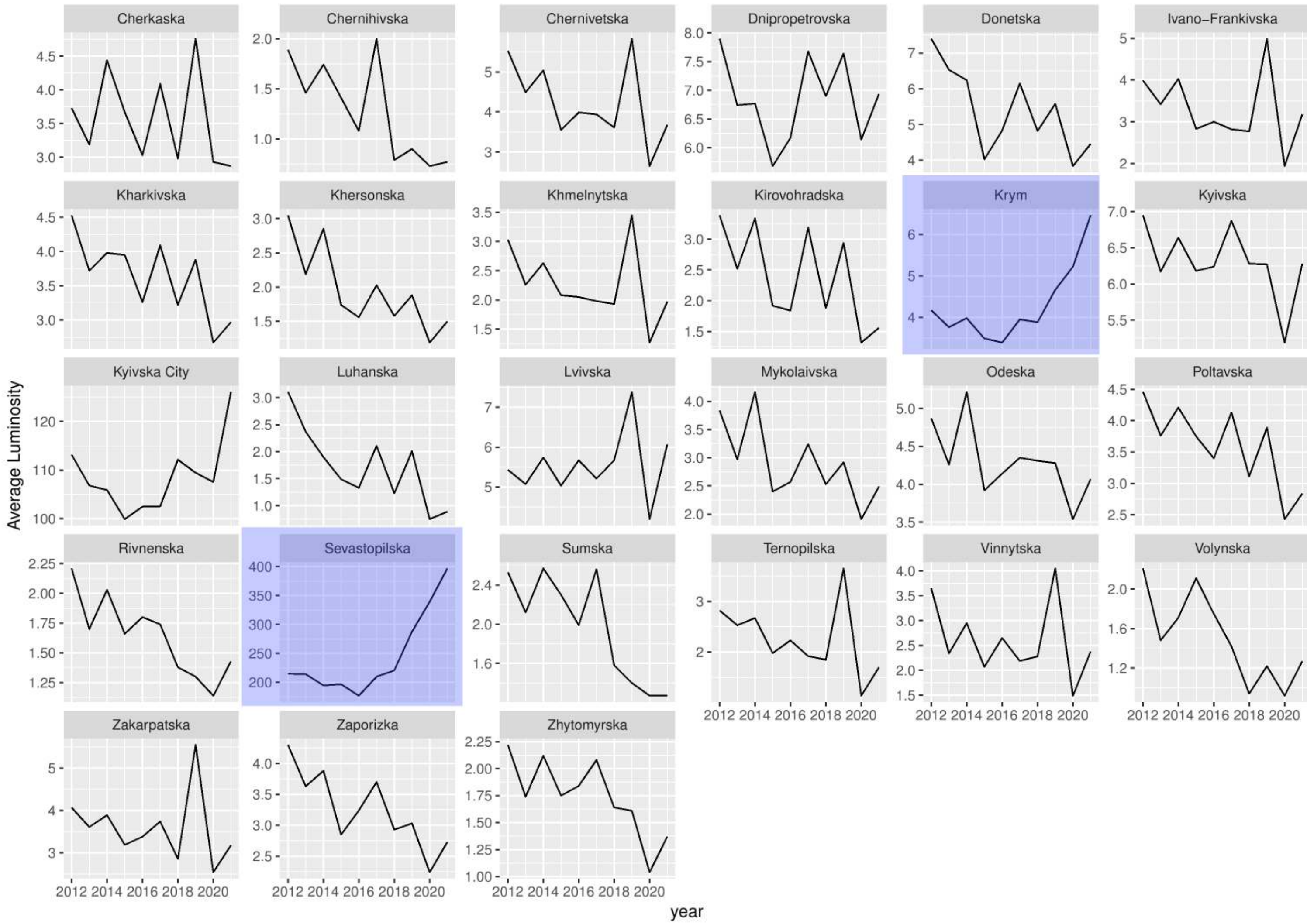


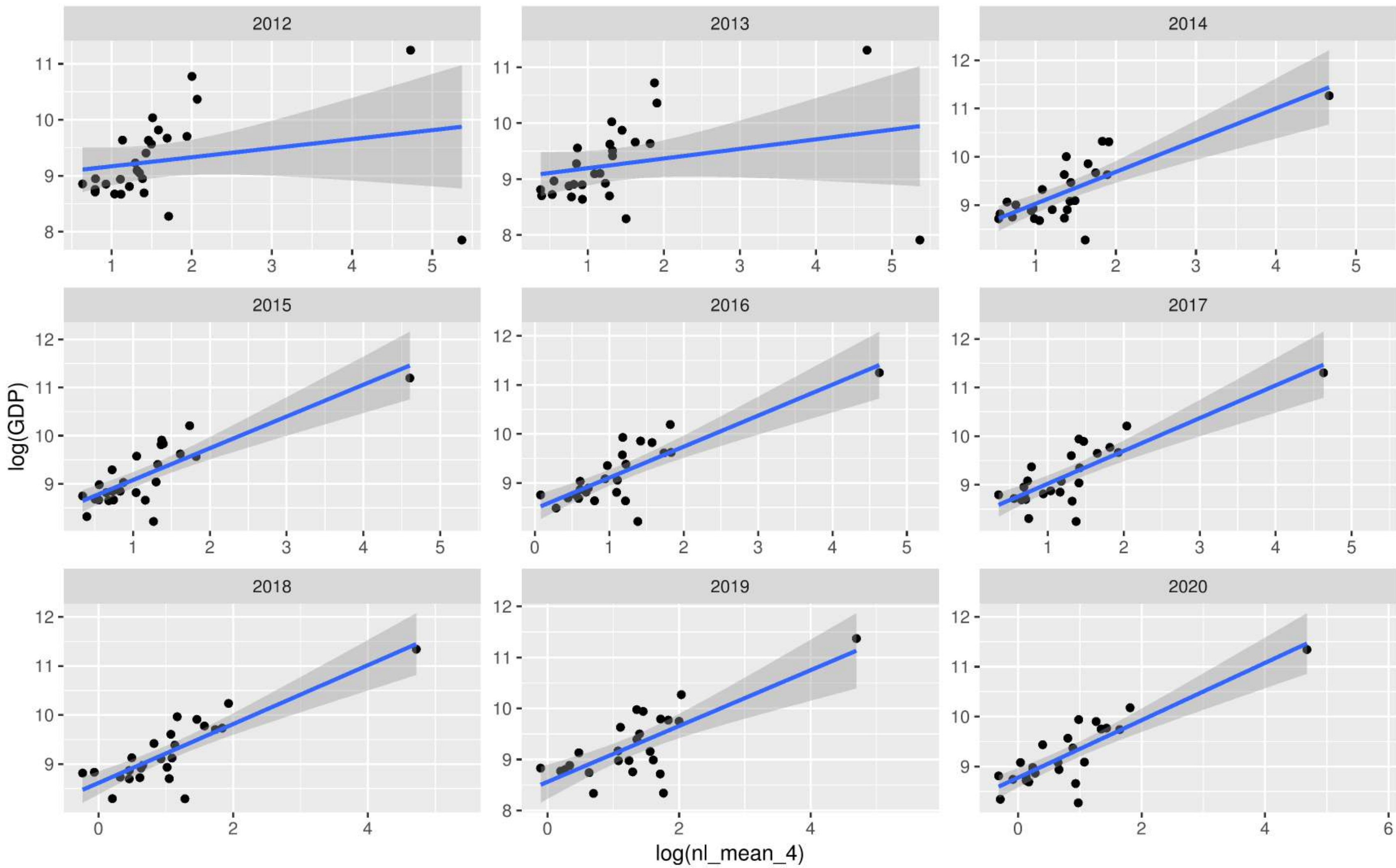


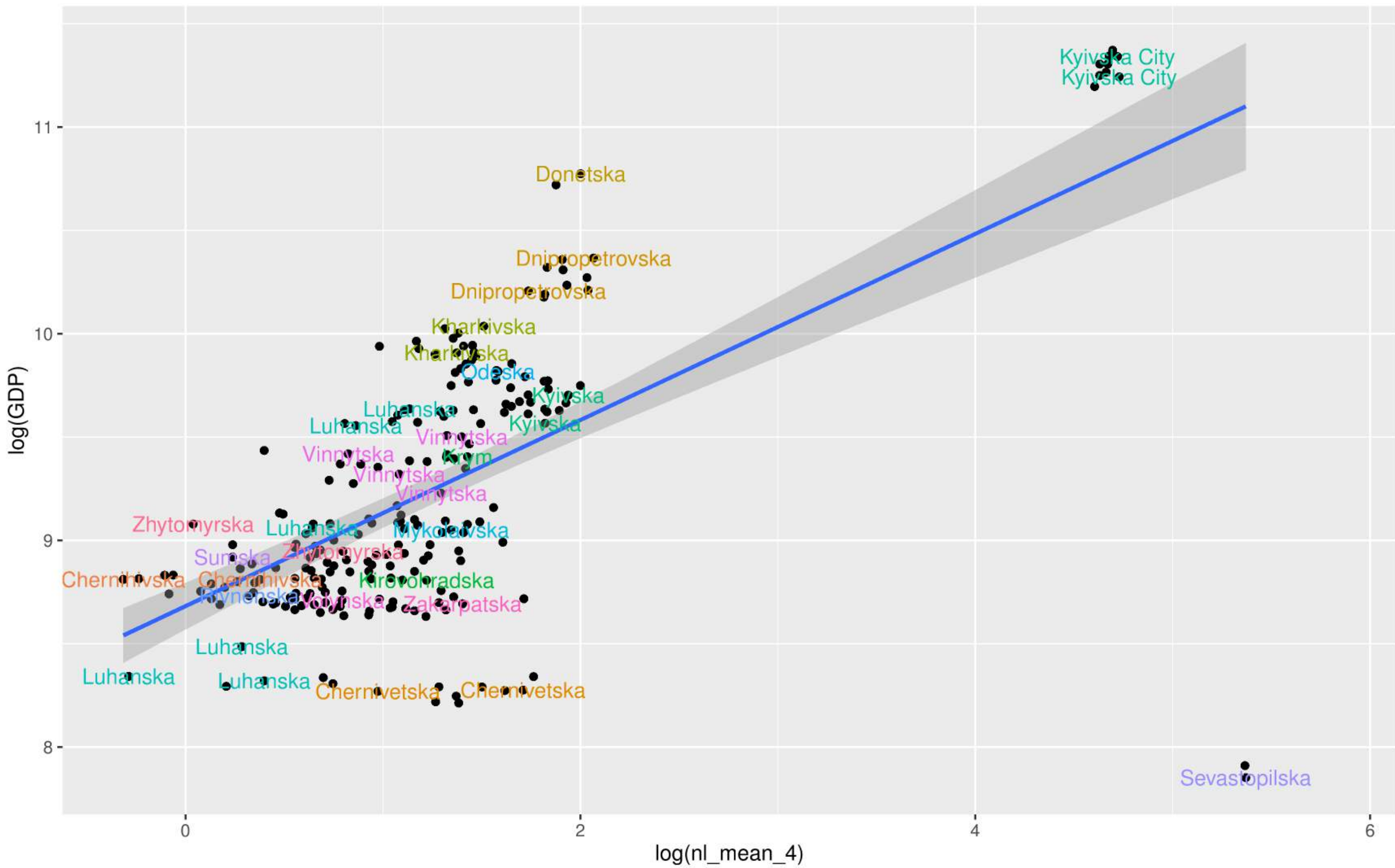


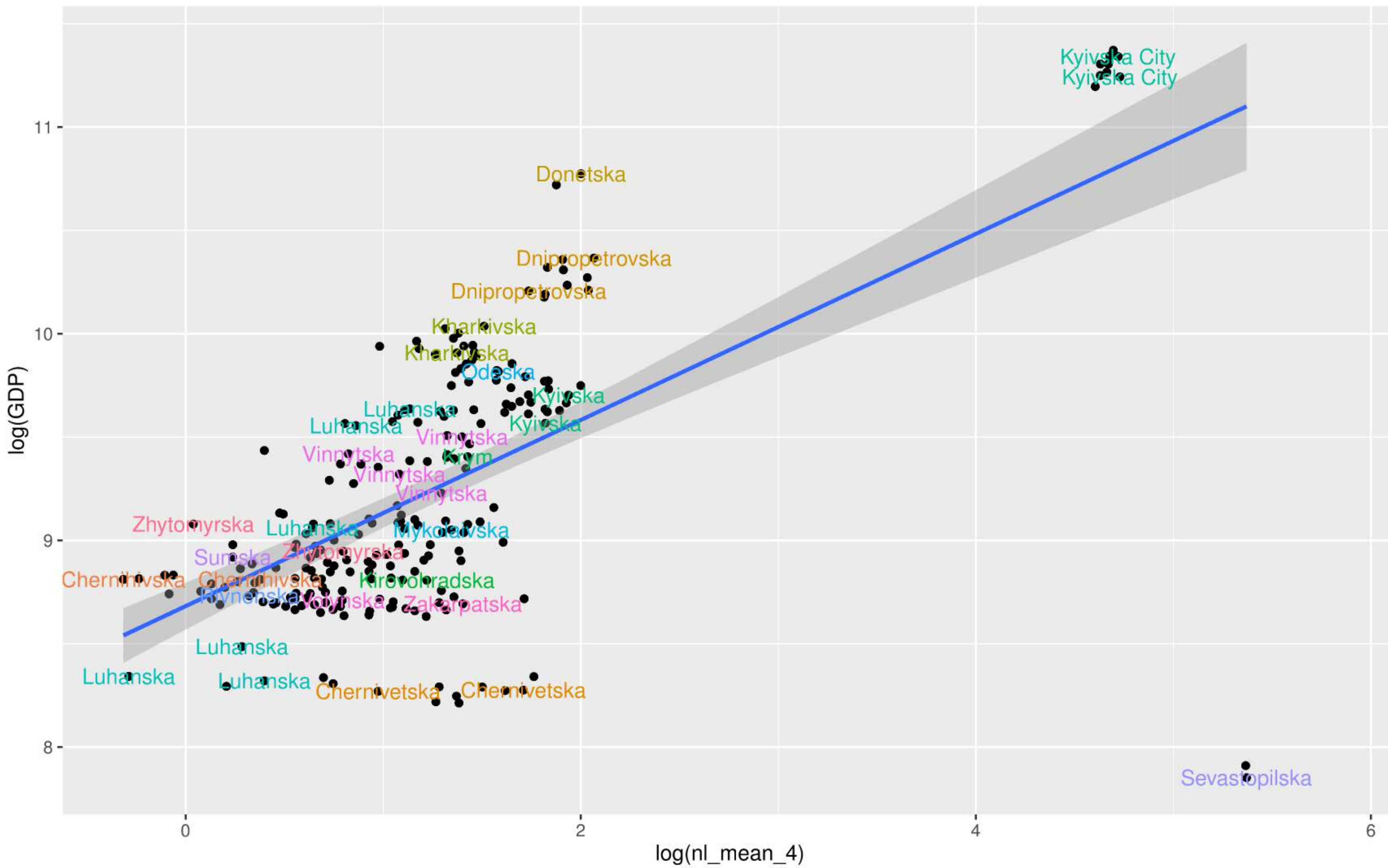


Cherkaska
Chernihivska
Chernivetska
Dnipropetrovska
Donetska
Donetska_Occupied
Ivano−Frankivska
Kharkivska
Khersonska
Khmelnytska
Kirovohradska
Krym
Kyivska
Kyivska City
Luhanska
Luhanska_Occupied
Lvivska
Mykolaivska
Odeska
Poltavska
Rivnenska
Sevastopilska
Sumska
Ternopilska
Vinnytska
Volynska
Zakarpatska
Zaporizka
Zhytomyrska
Mean Luminosity
Date
2013
2016
2019
2022


Cherkaska
Chernihivska
Chernivetska
Dnipropetrovska
Donetska
Ivano-Frankivska
Kharkivska
Khersonska
Khmelnytska
Kirovohradska
Krym
Kyivska
Kyivska City
Luhanska
Lvivska
Mykolaivska
Odeska
Poltavska
Rivnenska
Sevastopilska
Sumska
Ternopilska
Vinnytska
Volynska
Zakarpatska
Zaporizka
Zhytomyrska
Average Luminosity
year
2012 2014 2016 2018 2020

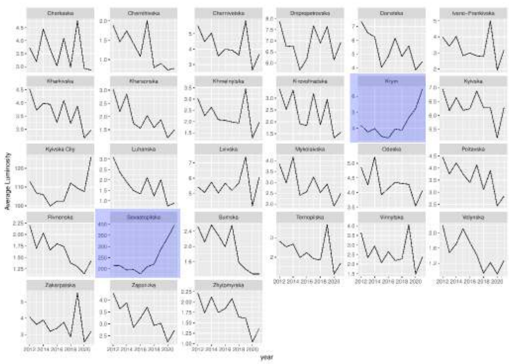

2012
2013
2014
2015
2016
2017
2018
2019
2020
log(GDP)
log(nl_mean_4)

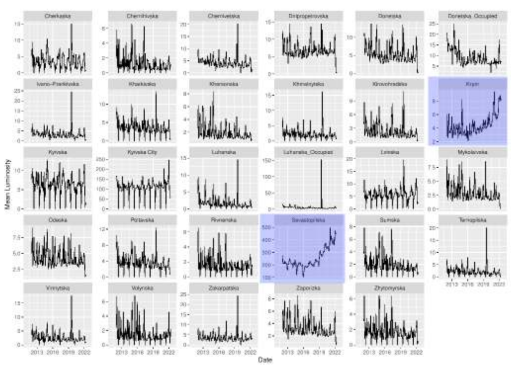

log(GDP)
log(nl_mean_4)
11
10
9
8
0
2
4
6
Kyivska City
Kyivska City
Donetska
Dnipropetrovska
Dnipropetrovska
Kharkivska
Kharkivska
Odeska
Kyivska
Kyivska
Luhanska
Luhanska
Vinnytska
Vinnytska
Vinnytska
Vinnytska
Krym
Zhytomyrska
Zhytomyrska
Luhanska
Mykolaivska
Sumska
Chernihivska
Chernihivska
Rivnenska
Volynska
Kirovohradska
Zakarpatska
Luhanska
Luhanska
Luhanska
Chernivetska
Chernivetska
Sevastopilska

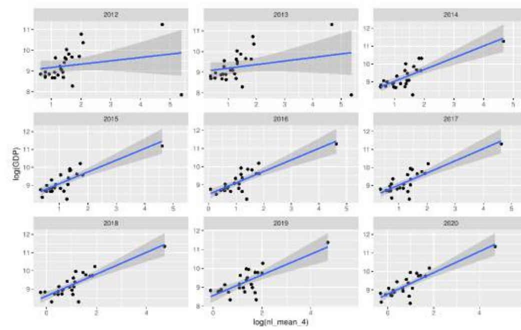


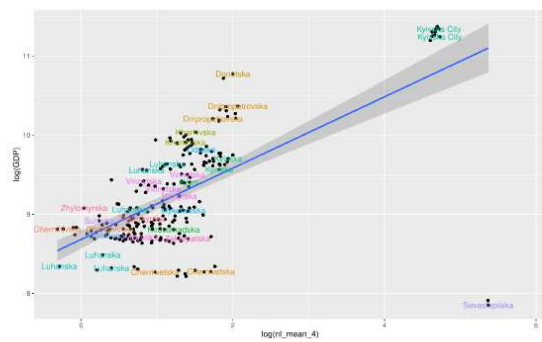
Kyivska City
Kyivska City
Donetska
Dnipropetrovska
Dnipropetrovska
Kharkivska
Kharkivska
Odeska
Kyivska
Kyivska
Luhanska
Luhanska
Vinnytska
Vinnytska
Vinnytska
Vinnytska
Krym
Zhytomyrska
Luhanska
Mykolaivska
Sumska
Zhytomyrska
Chernihivska
Chernihivska
Kirovohradska
Rivnenska
Volynska
Zakarpatska
Luhanska
Luhanska
Luhanska
Chernivetska
Chernivetska
Sevastopilska
log(GDP)
log(nl_mean_4)
11
10
9
8
0
2
4
6

Nighlights - empirics









Key Takeaways

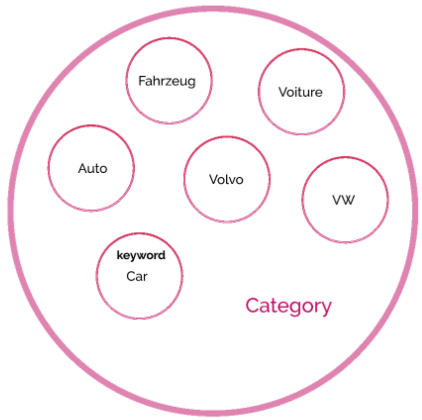
-> strong support for relationship in the cross-section in levels (log-log); stable over time

-> regions have different levels of nightlights

-> within region, no clear longitudinal relationship

Google Trends Data

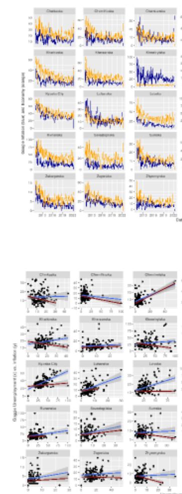
Google Trends = Search Volume indices of keyword search intensity (# keyword searches/ # total searches) broken down by region and time frame {Keyword, Category, Topic}



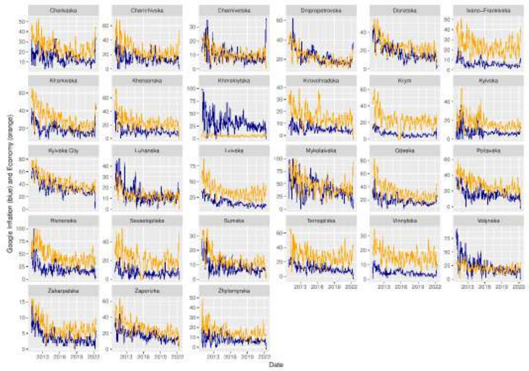
-> categories: consumption, labor markets, transportation, investment, education, inflation

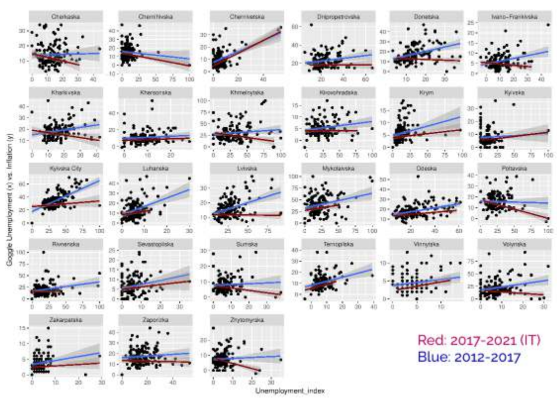
-> some issues: these are relative indices (w.r.t. total searches); as total usage increases over time and new categories become relevant, ranking changes.

-> during a war, some categories may give the wrong impression (eg. Transportation & GDP in normal times vs. war times)

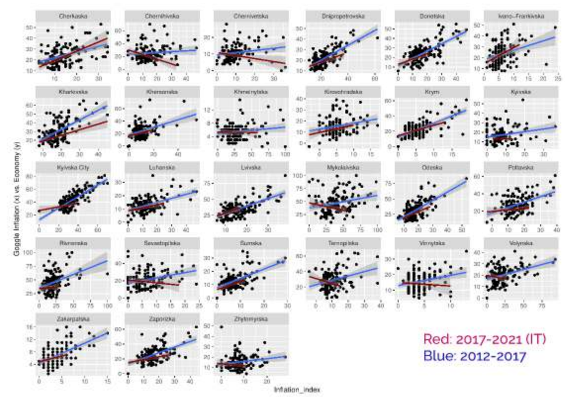


Google Trends - empirics

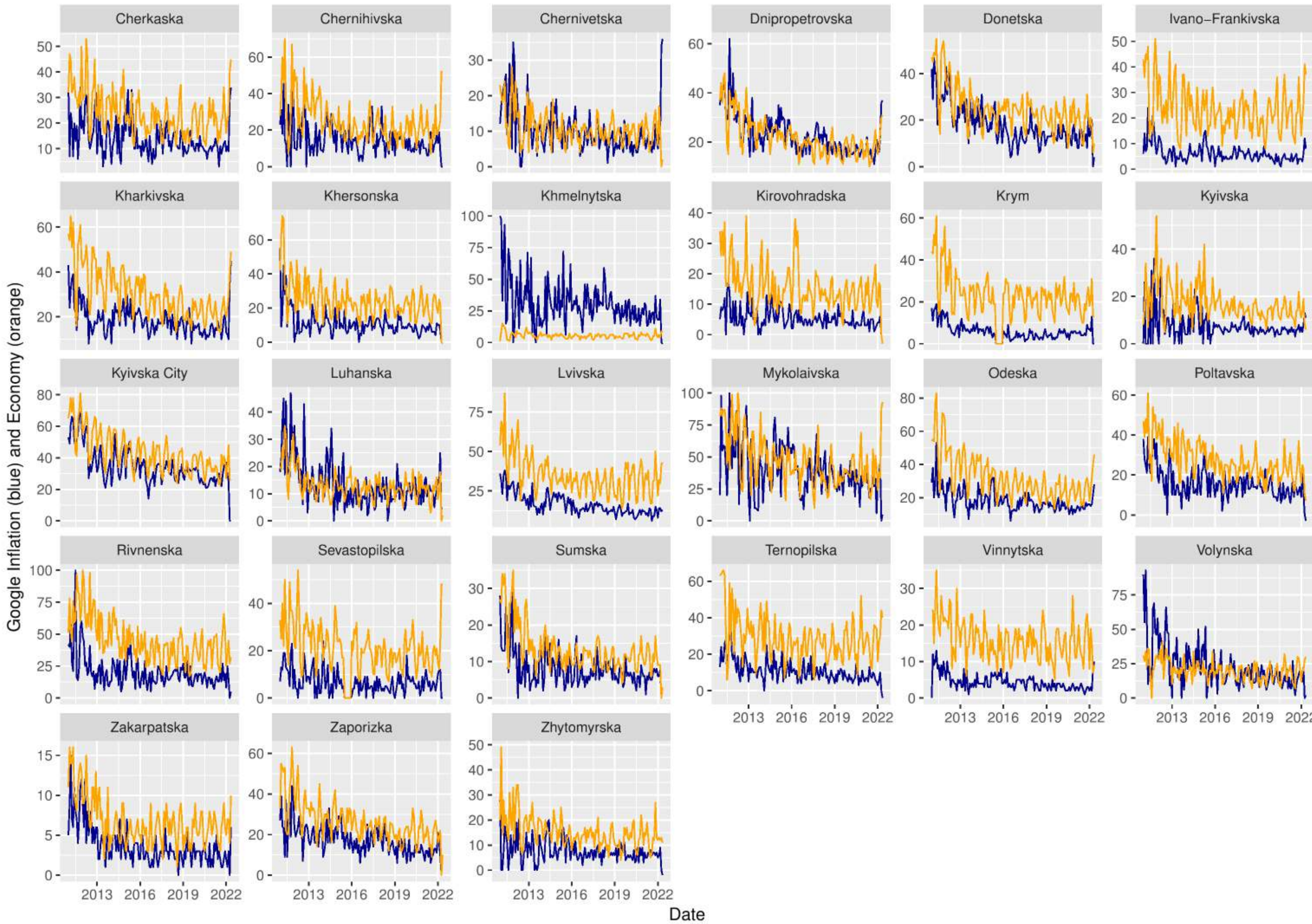


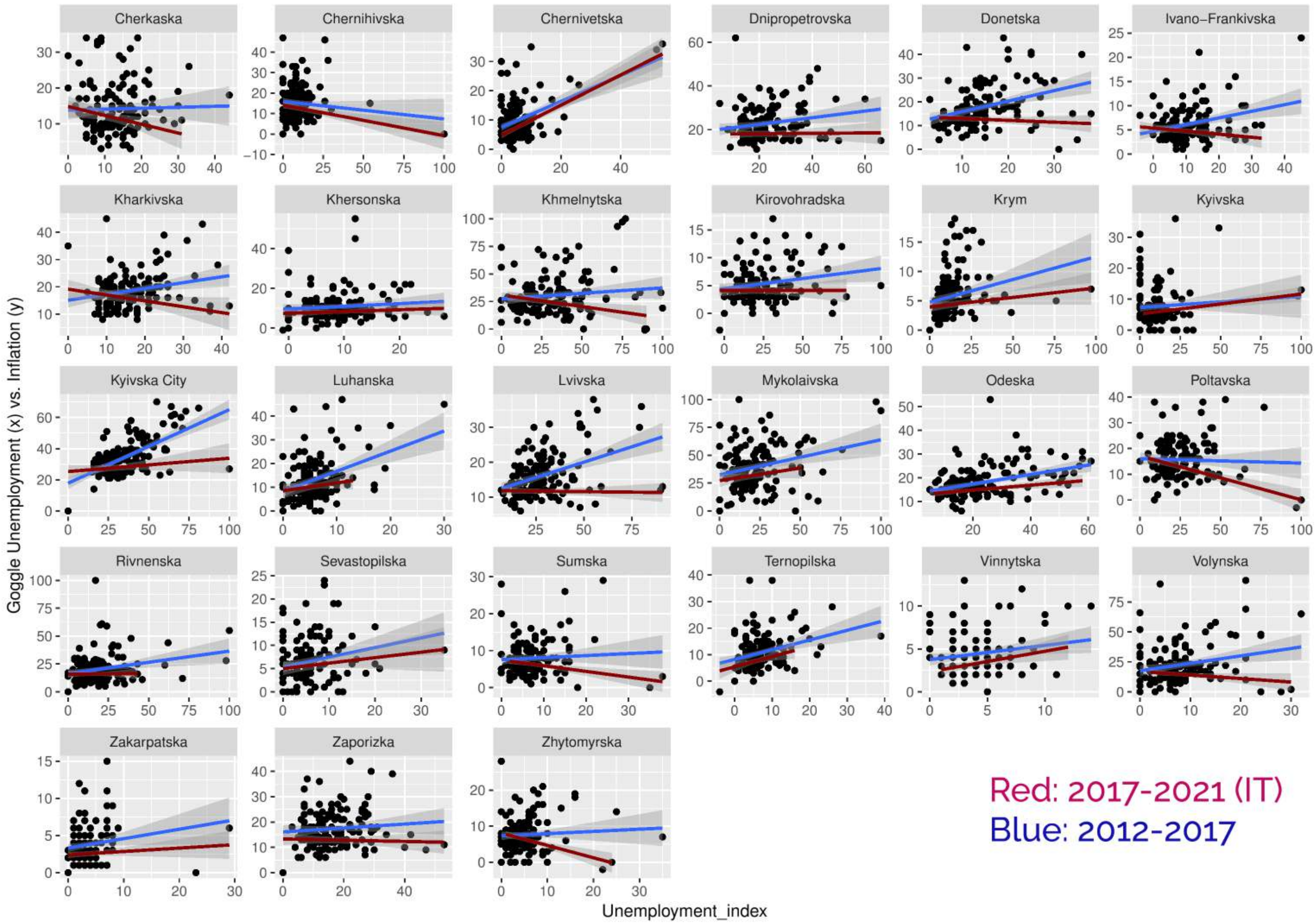


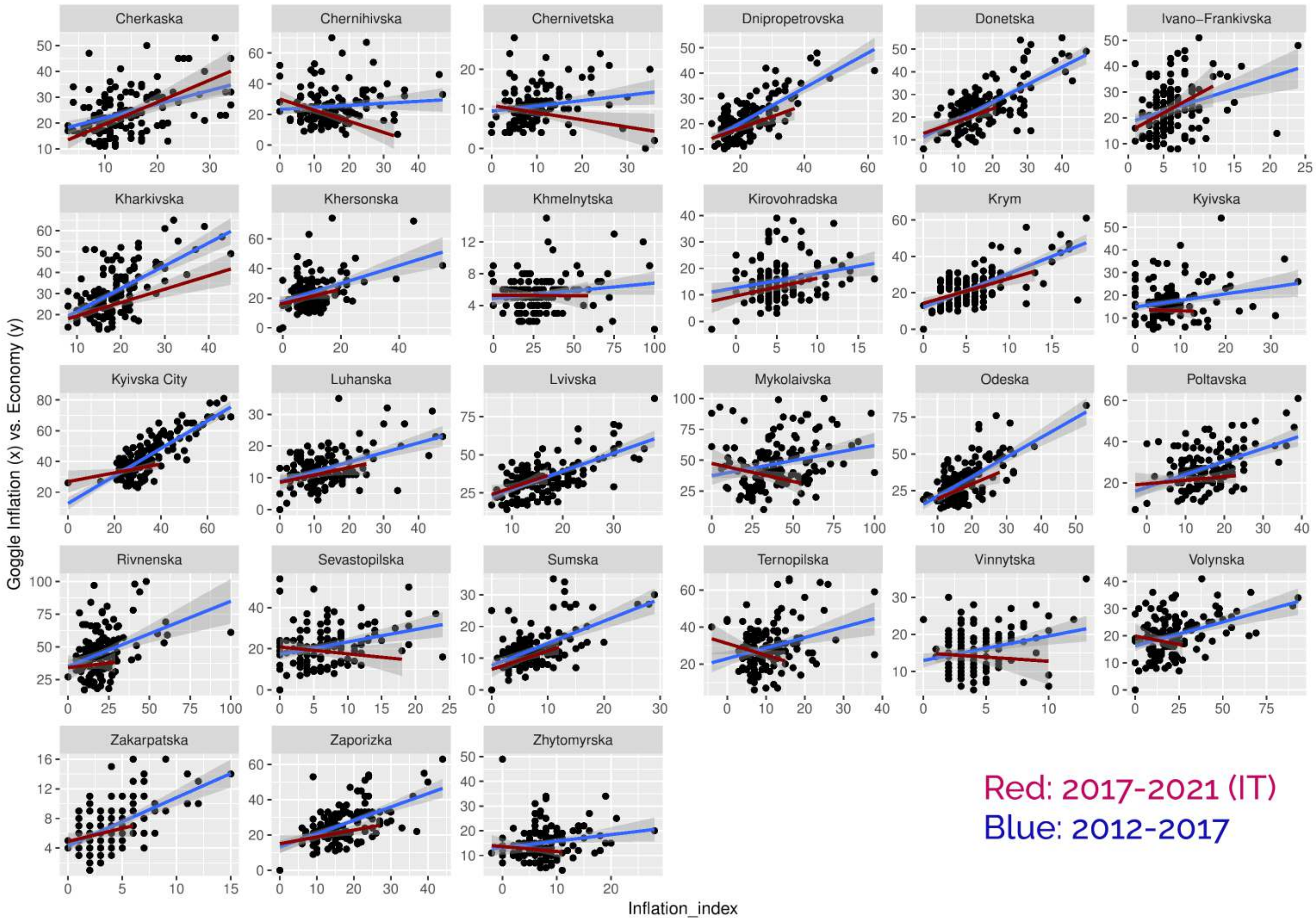
Red: 2017-2021 (IT)
Blue: 2012-2017

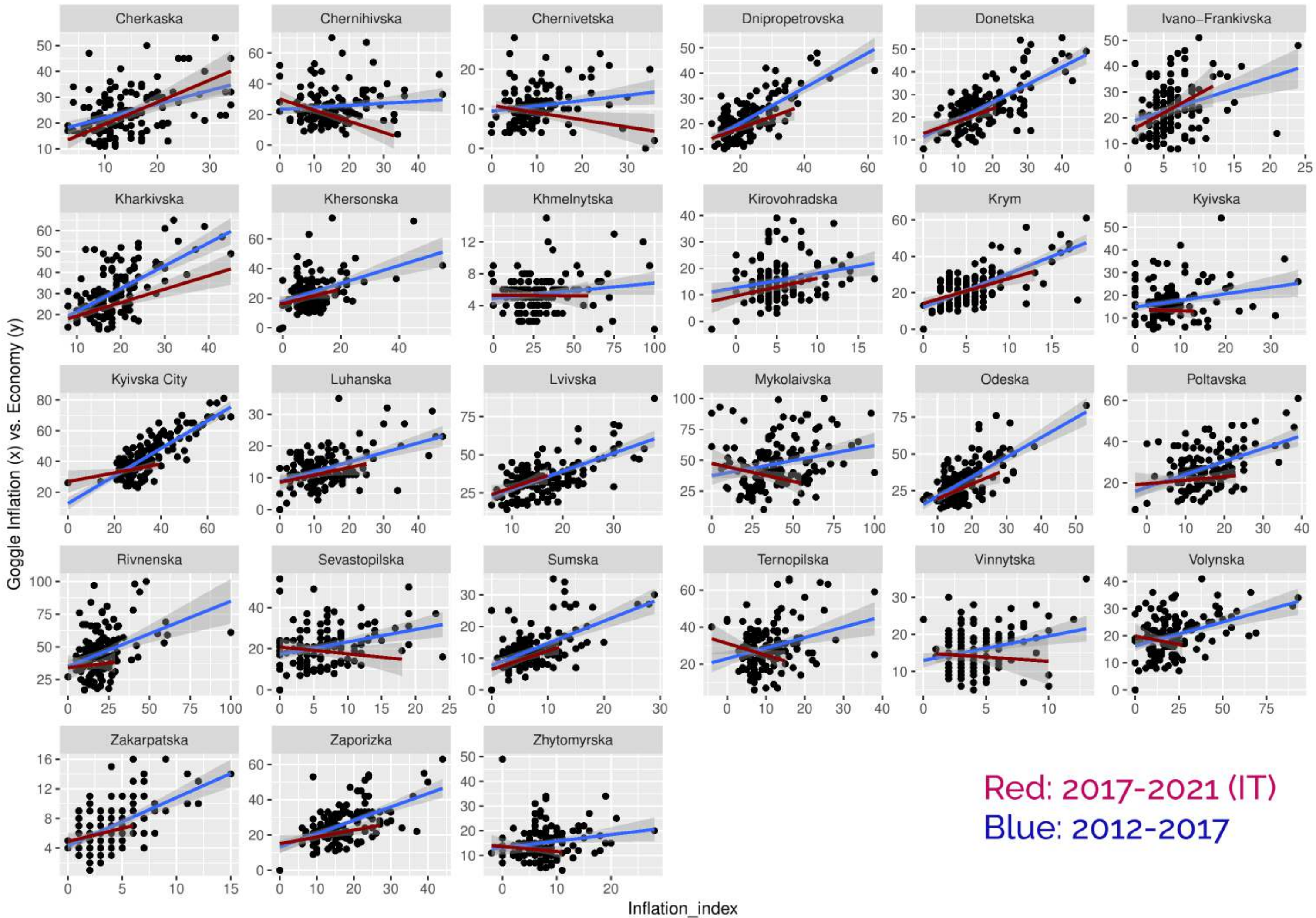


Red: 2017-2021 (IT)
Blue: 2012-2017

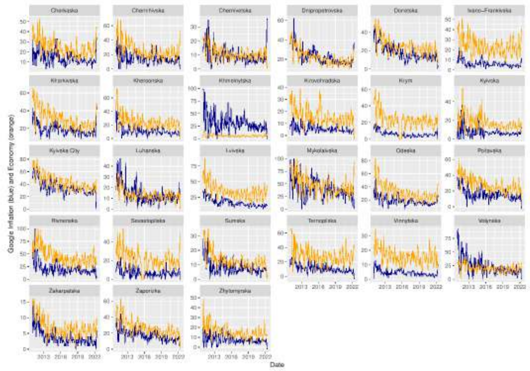

Cherkaska
Chernihivska
Chernivetska
Dnipropetrovska
Donetska
Ivano−Frankivska
Kharkivska
Khersonska
Khmelnytska
Kirovohradska
Krym
Kyivska
Kyivska City
Luhanska
Lvivska
Mykolaivska
Odeska
Poltavska
Rivnenska
Sevastopilska
Sumska
Ternopilska
Vinnytska
Volynska
Zakarpatska
Zaporizka
Zhytomyrska
Google Inflation (blue) and Economy (orange)
Date
2013 2016 2019 2022


Cherkaska
Chernihivska
Chernivetska
Dnipropetrovska
Donetska
Ivano-Frankivska
Kharkivska
Khersonska
Khmelnytska
Kirovohradska
Krym
Kyivska
Kyivska City
Luhanska
Lvivska
Mykolaivska
Odeska
Poltavska
Rivnenska
Sevastopilska
Sumska
Ternopilska
Vinnytska
Volynska
Zakarpatska
Zaporizka
Zhytomyrska
Goggle Unemployment (x) vs. Inflation (y)
Unemployment_index
Red: 2017-2021 (IT)
Blue: 2012-2017


Cherkaska
Chernihivska
Chernivetska
Dnipropetrovska
Donetska
Ivano−Frankivska
Kharkivska
Khersonska
Khmelnytska
Kirovohradska
Krym
Kyivska
Kyivska City
Luhanska
Lvivska
Mykolaivska
Odeska
Poltavska
Rivnenska
Sevastopilska
Sumska
Ternopilska
Vinnytska
Volynska
Zakarpatska
Zaporizka
Zhytomyrska
Goggle Inflation (x) vs. Economy (y)
Inflation_index
Red: 2017-2021 (IT)
Blue: 2012-2017


Cherkaska
Chernihivska
Chernivetska
Dnipropetrovska
Donetska
Ivano-Frankivska
Kharkivska
Khersonska
Khmelnytska
Kirovohradska
Krym
Kyivska
Kyivska City
Luhanska
Lvivska
Mykolaivska
Odeska
Poltavska
Rivnenska
Sevastopilska
Sumska
Ternopilska
Vinnytska
Volynska
Zakarpatska
Zaporizka
Zhytomyrska
Goggle Inflation (x) vs. Economy (y)
Inflation_index
Red: 2017-2021 (IT)
Blue: 2012-2017

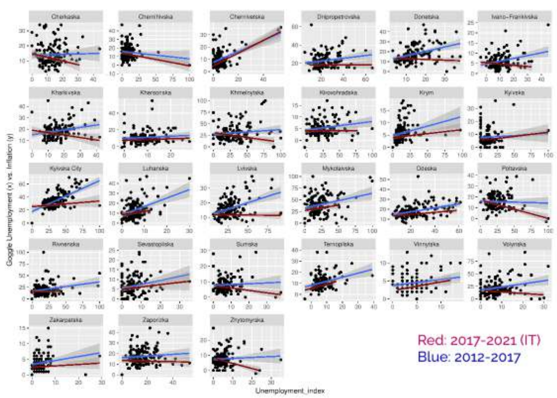
Google Trends - empirics

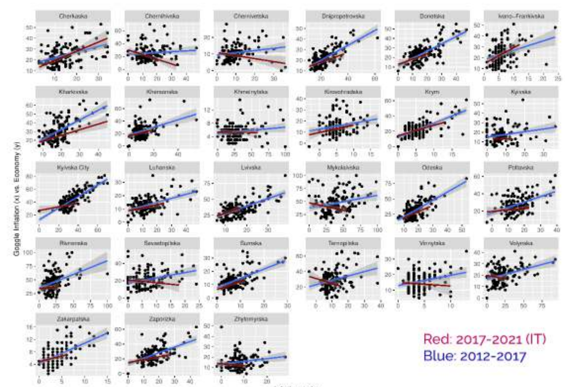


-> About 30 categories are used

-> Monthly time series contain plenty of variation (too much?)

-> Cross-index correlations appear reasonable at first sight but large number implies some shrinkage/ dimensionality reduction is needed





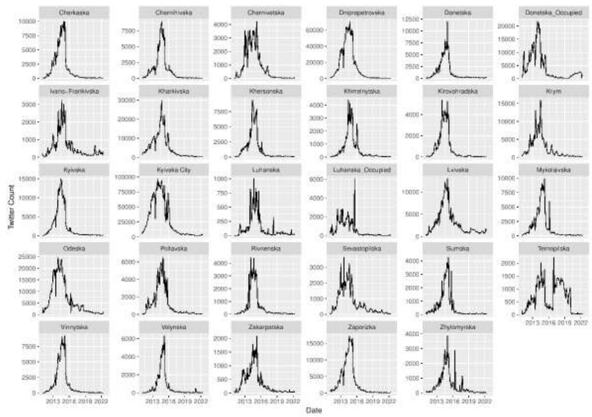
Twitter

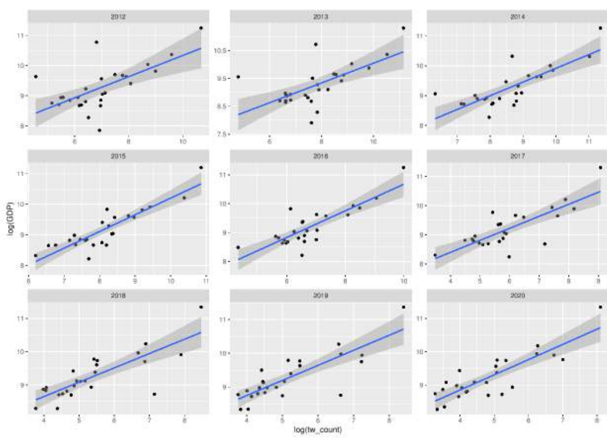
Social media use in estimation of regional GDP

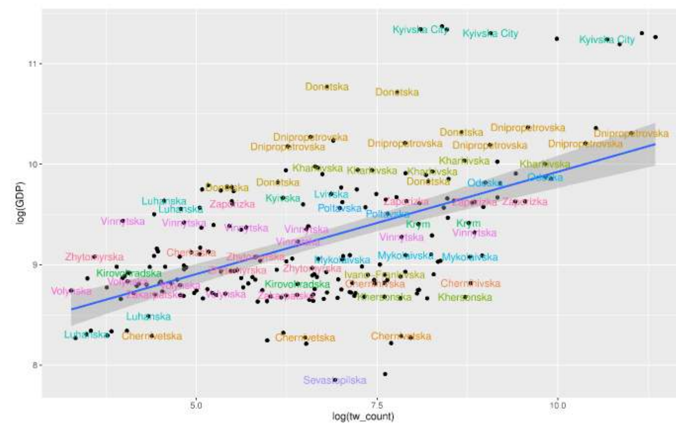
-> **Indaco, A (2020)**, "From twitter to GDP: Estimating economic activity from social media", Regional Science and Urban Economics 85; **Ortega-Bastida, J, A J Gallego, J R Rico-Juan and P Albarran (2021)**, "A Multimodal approach for regional GDP prediction using social media activity and historical information", Applied Soft Computing 111

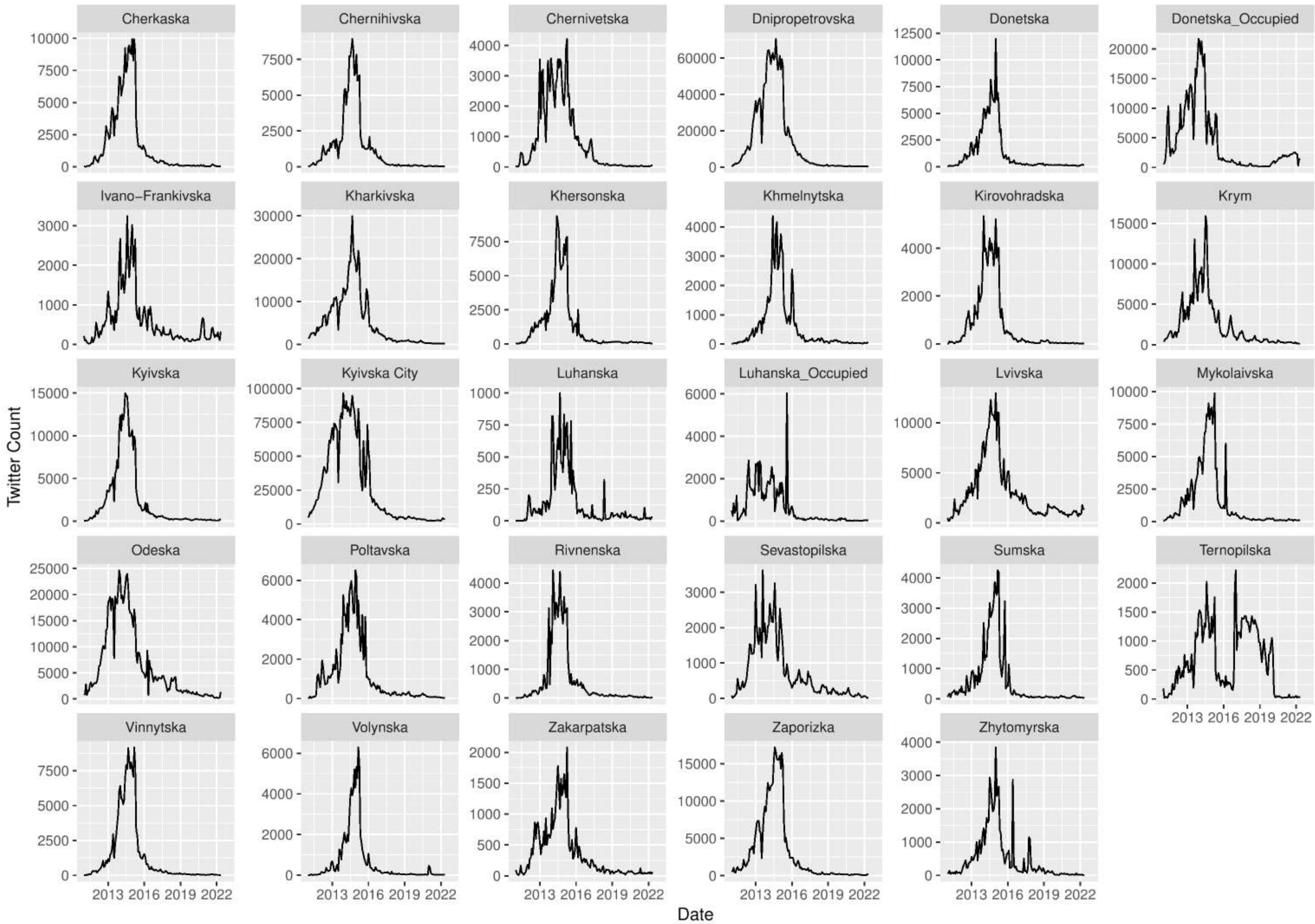
-> In our case, number of geo-located tweets containing a photo are gathered at oblast level
-> Time series of counts are used

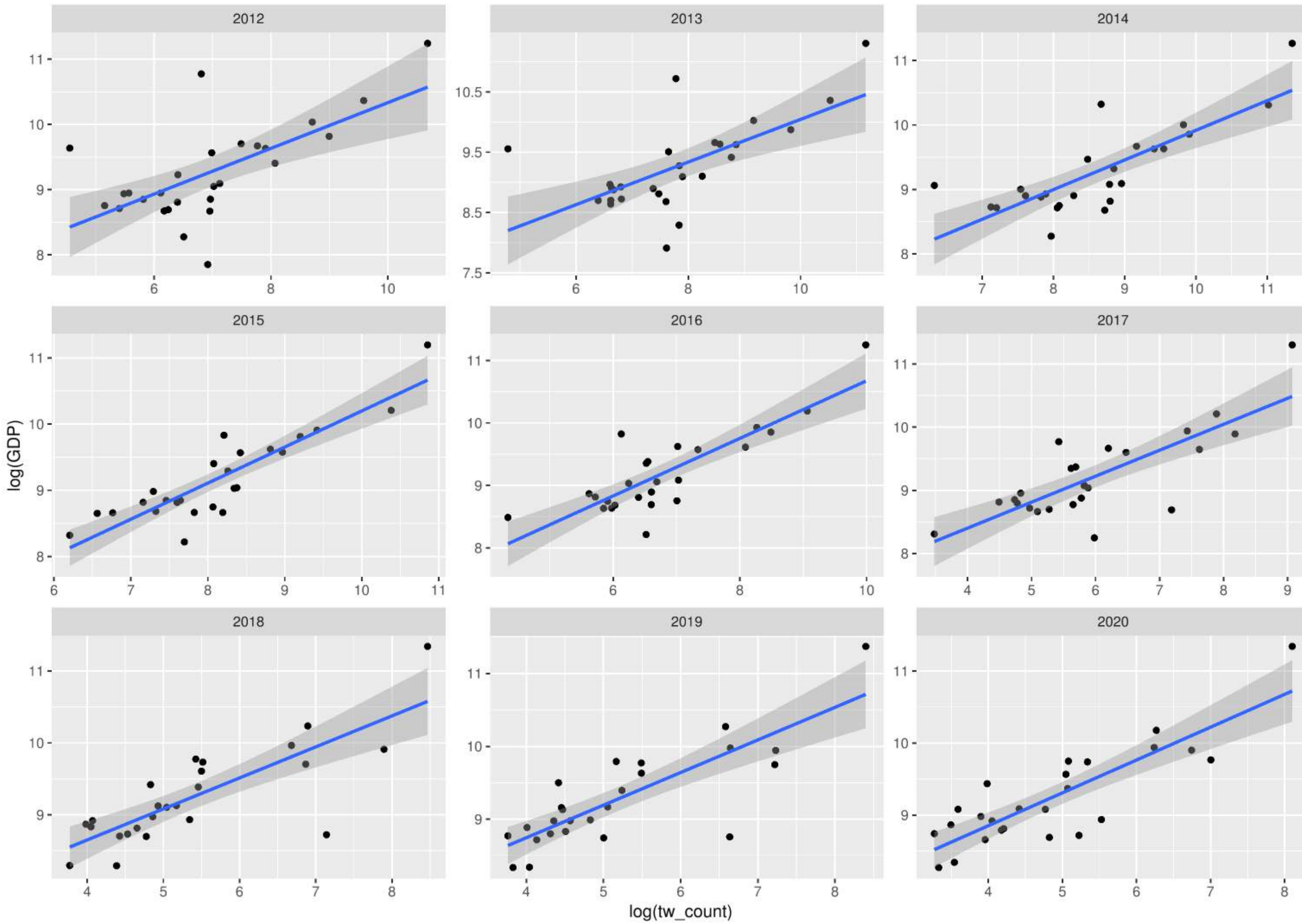
Twitter - empirics

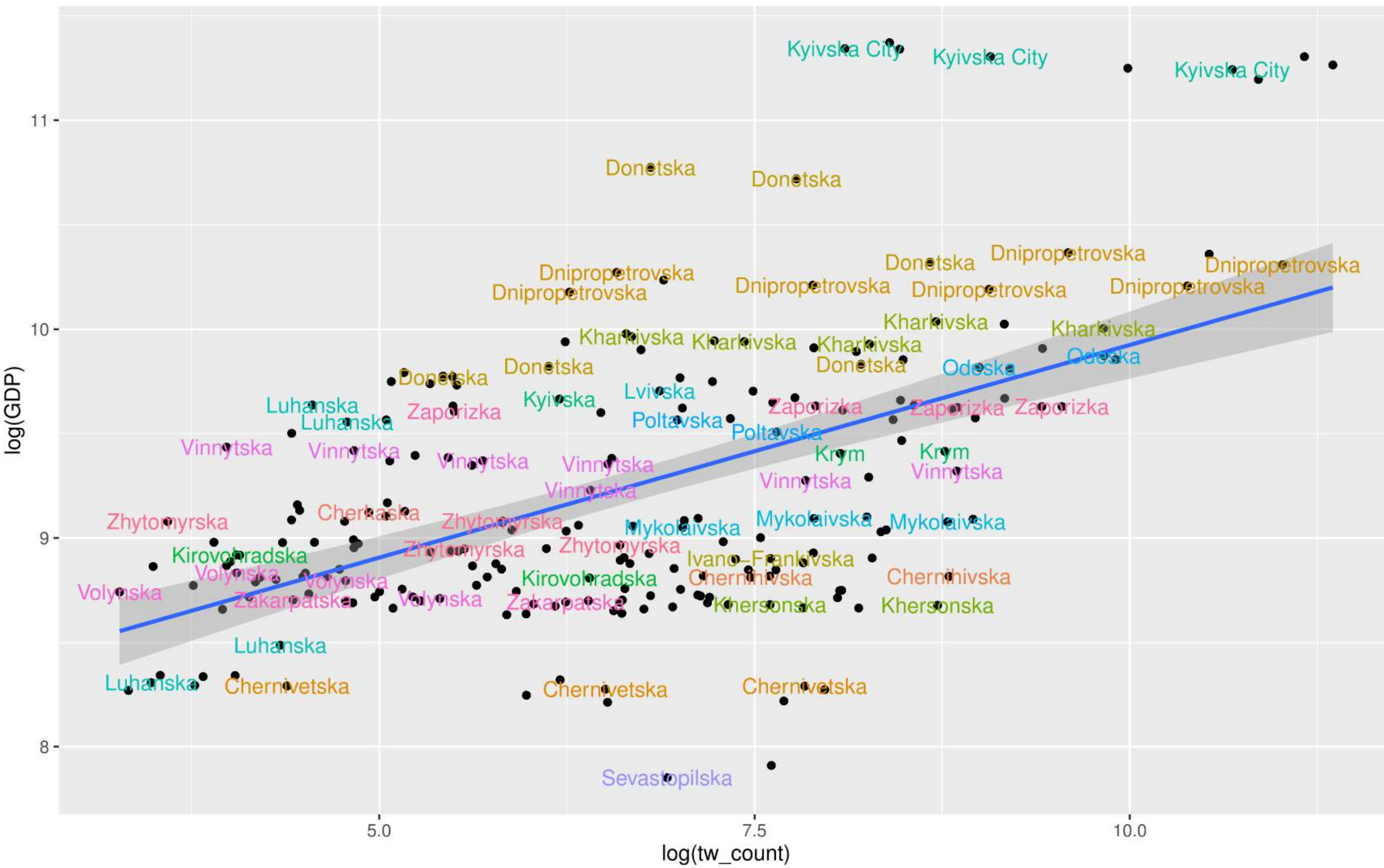




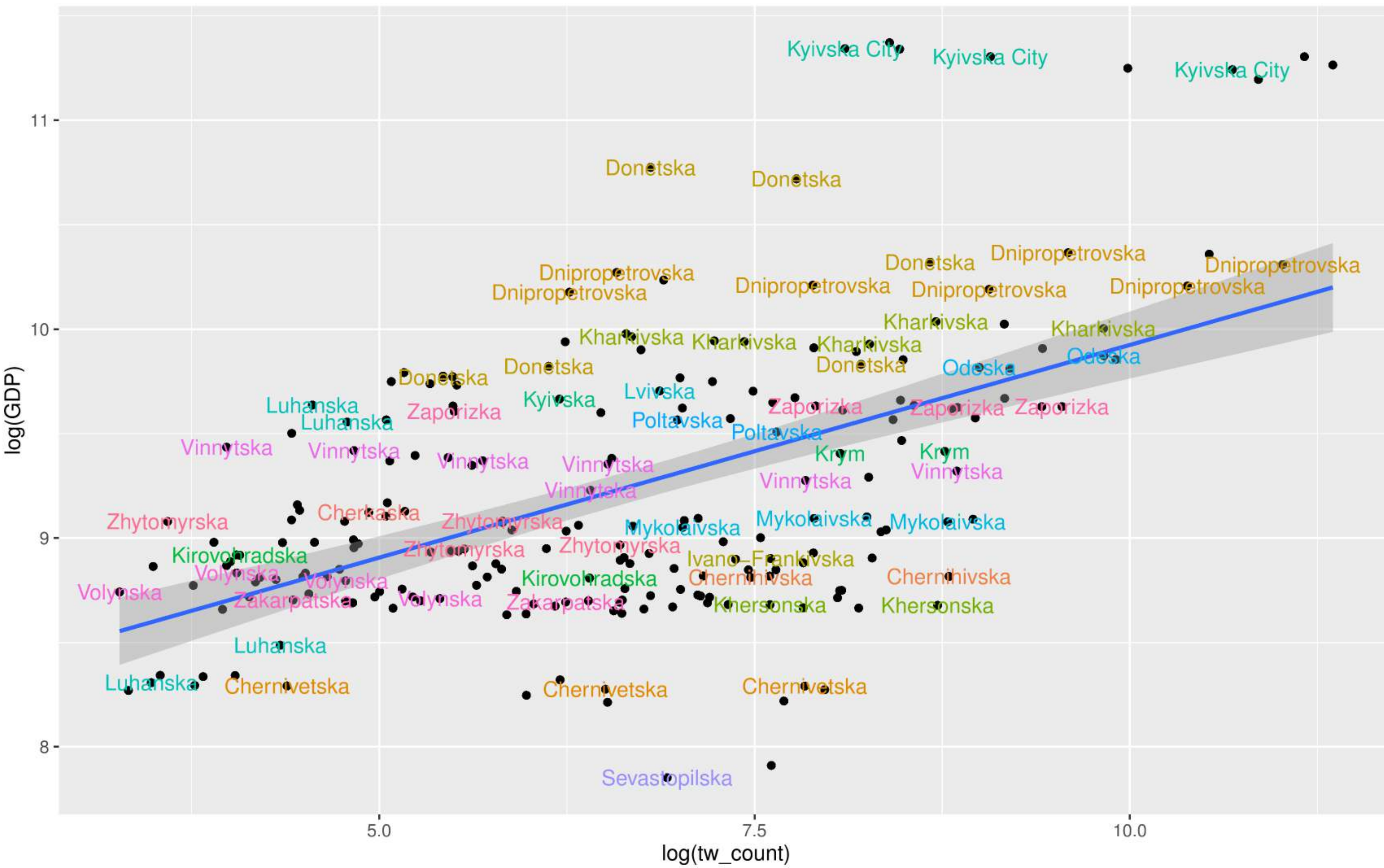



Cherkaska
Chernihivska
Chernivetska
Dnipropetrovska
Donetska
Donetska_Occupied
Ivano−Frankivska
Kharkivska
Khersonska
Khmelnytska
Kirovohradska
Krym
Kyivska
Kyivska City
Luhanska
Luhanska_Occupied
Lvivska
Mykolaivska
Odeska
Poltavska
Rivnenska
Sevastopilska
Sumska
Ternopilska
Vinnytska
Volynska
Zakarpatska
Zaporizka
Zhytomyrska
Twitter Count
Date
2013 2016 2019 2022


2012
2013
2014
2015
2016
2017
2018
2019
2020
log(GDP)
log(tw_count)

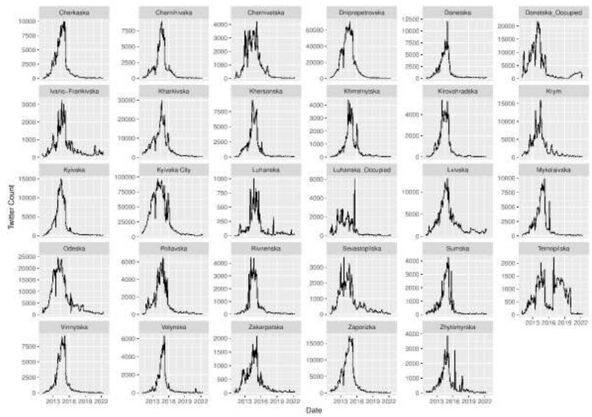


log(GDP)
log(tw_count)
11
10
9
8
5.0
7.5
10.0
Kyivska City
Donetska
Dnipropetrovska
Kharkivska
Odeska
Lvivska
Kyivska
Luhanska
Zaporizka
Poltavska
Vinnytska
Krym
Zhytomyrska
Cherkaska
Mykolaivska
Kirovohradska
Volynska
Ivano-Frankivska
Chernihivska
Zakarpatska
Khersonska
Chernivetska
Sevastopilska



log(GDP)
11
10
9
8
5.0
7.5
10.0
log(tw_count)

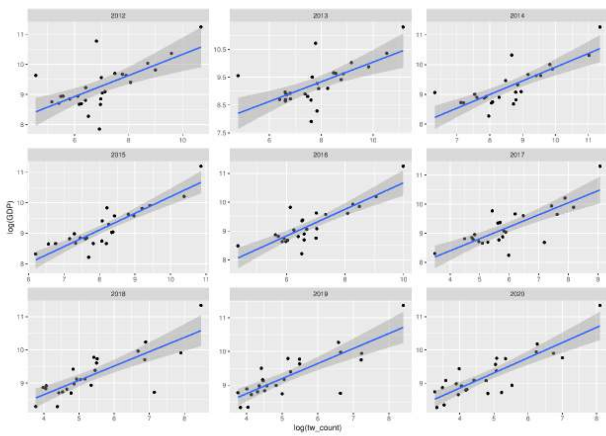
Twitter - empirics

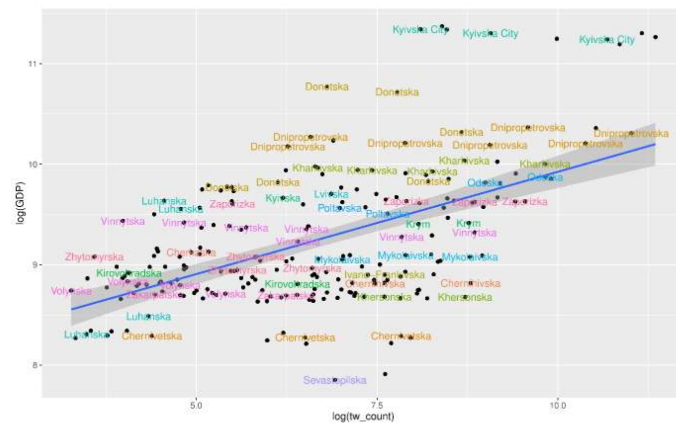


Key Takeaways
-> strong support for relationship in the cross-section in levels (log-log); failry stable over time

-> regions have markendly different levels of twitter activity

-> within region, no clear longitudinal relationship





Model(s)

Model is the result in equal parts of deliberate choice & data constraints

-> rich cross-section of potentially explanatory variables but too rich for a t=9 years, i=25 units panel

-> prior literature on (NL & Twitter) indicate levels are the variables to target; exploratory data analysis confirms this

-> within region, no clear longitudinal relationship

Considered shrinkage (added benefit is interpretability of the ML model) but unstable unless all Oblasts share the same parameters.

Model is the result in equal parts of deliberate choice & data constraints

-> rich cross-section of potentially explanatory variables but too rich for a t=9 years, i=25 units panel

-> prior literature on (NL & Twitter) indicate levels are the variables to target; exploratory data analysis confirms this

-> within region, no clear longitudinal relationship

Considered shrinkage (added benefit is interpretability of the ML model) but unstable unless all Oblasts share the same parameters.

A regional-specific factor model is estimated

-> each region has its own 3-factor model explaining observed y_i = log(GDP_i)

-> m = 3, p = 31 explanatory variables {GT, Twitter, NL}

$$Z_m = \sum_{j=1}^{p} \phi_{jm} X_j$$

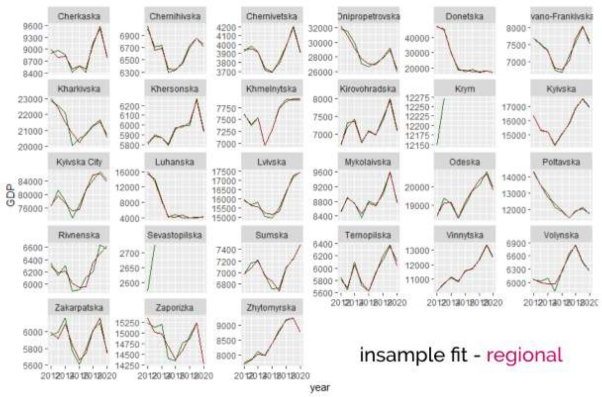
$$y_i = \theta_0 + \sum_{m=1}^{M} \theta_m z_{im} + \epsilon_i, \quad i = 1, \ldots, n,$$

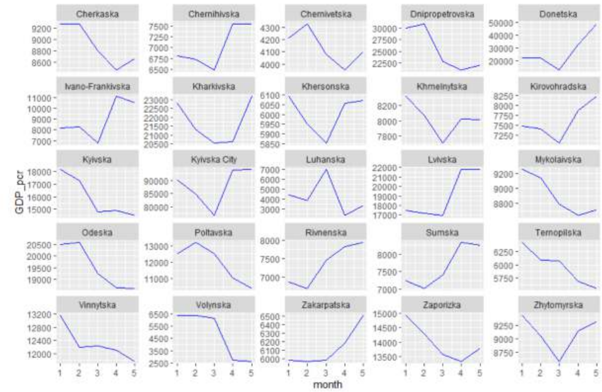
Data (Ivano-Frankivska):
X dimension: 9 31
Y dimension: 9 1

Fit method: svdpc

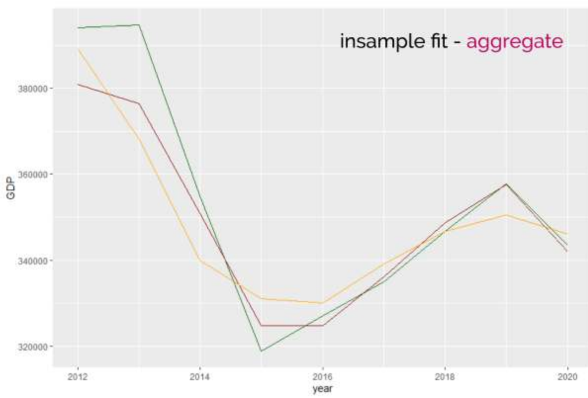
Number of components considered: 3
TRAINING: % variance explained

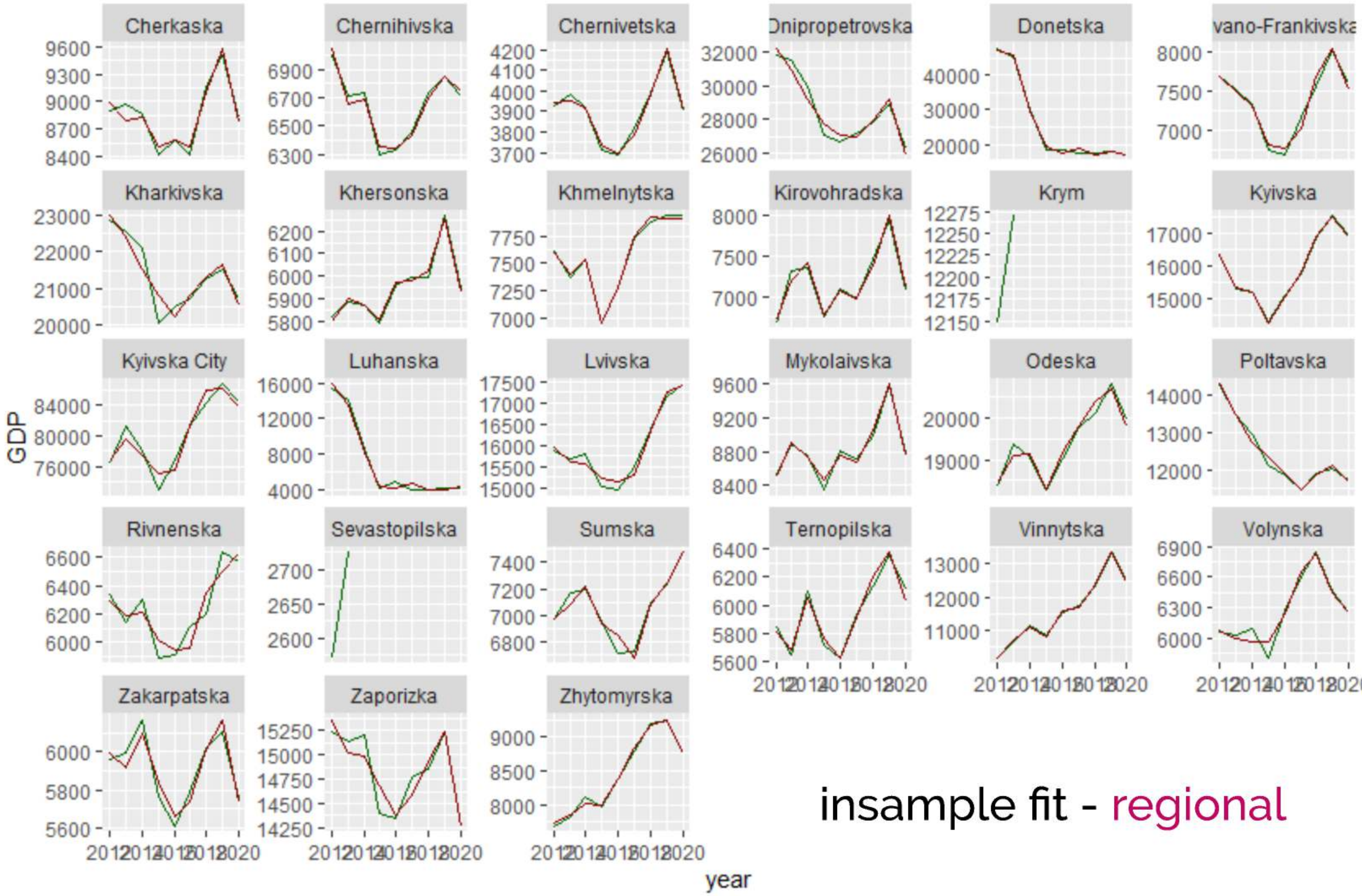
	1 comps	2 comps	3 comps
X	54.32	69.31	79.08
lng	15.17	68.76	81.61





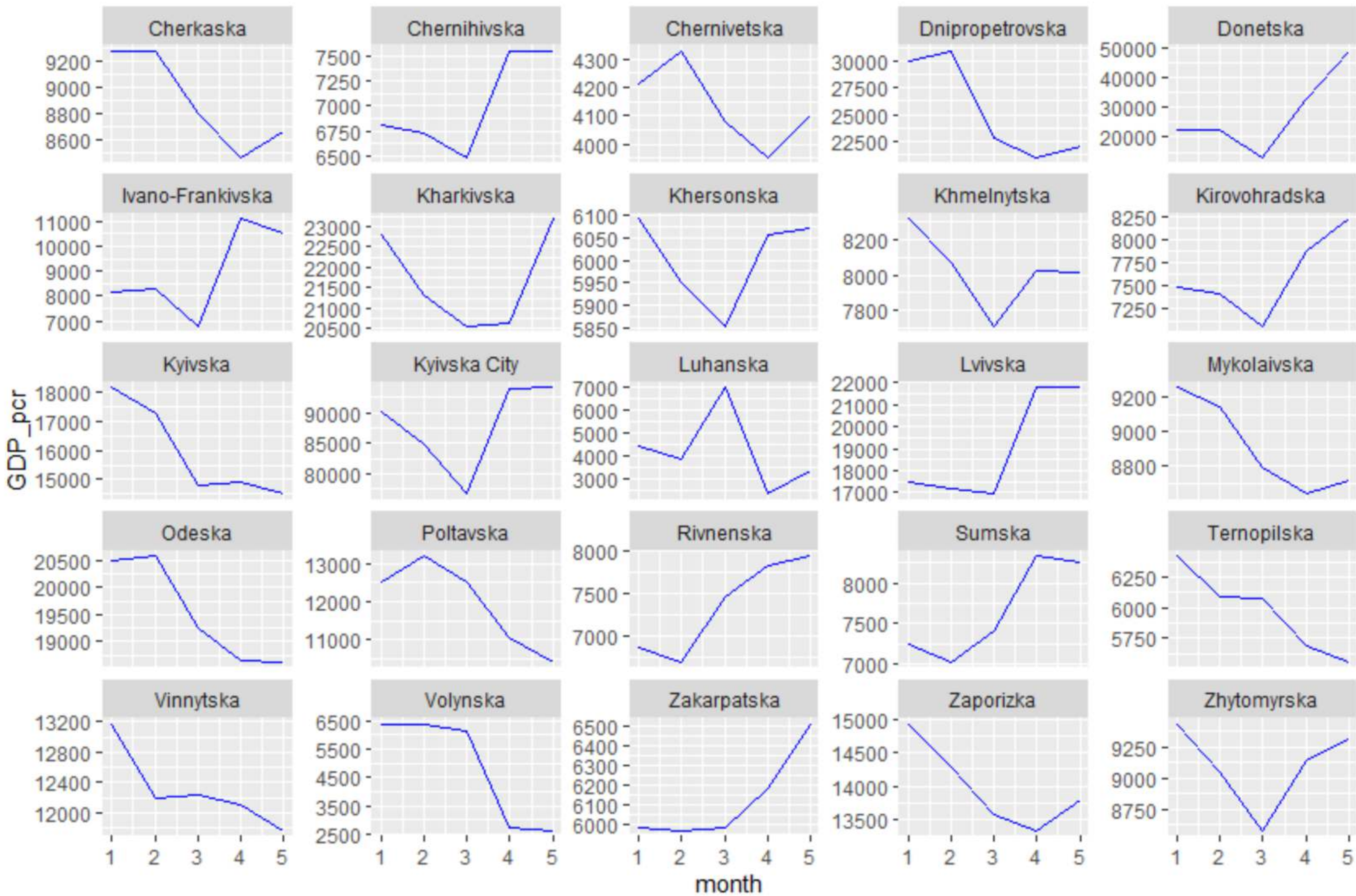
out-of-sample - regional





insample fit - regional


insample fit - aggregate
380000
360000
340000
320000
GDP
2012
2014
2016
2018
2020
year


GDP_pcr
month
Cherkaska
Chernihivska
Chernivetska
Dnipropetrovska
Donetska
Ivano-Frankivska
Kharkivska
Khersonska
Khmelnytska
Kirovohradska
Kyivska
Kyivska City
Luhanska
Lvivska
Mykolaivska
Odeska
Poltavska
Rivnenska
Sumska
Ternopilska
Vinnytska
Volynska
Zakarpatska
Zaporizka
Zhytomyrska

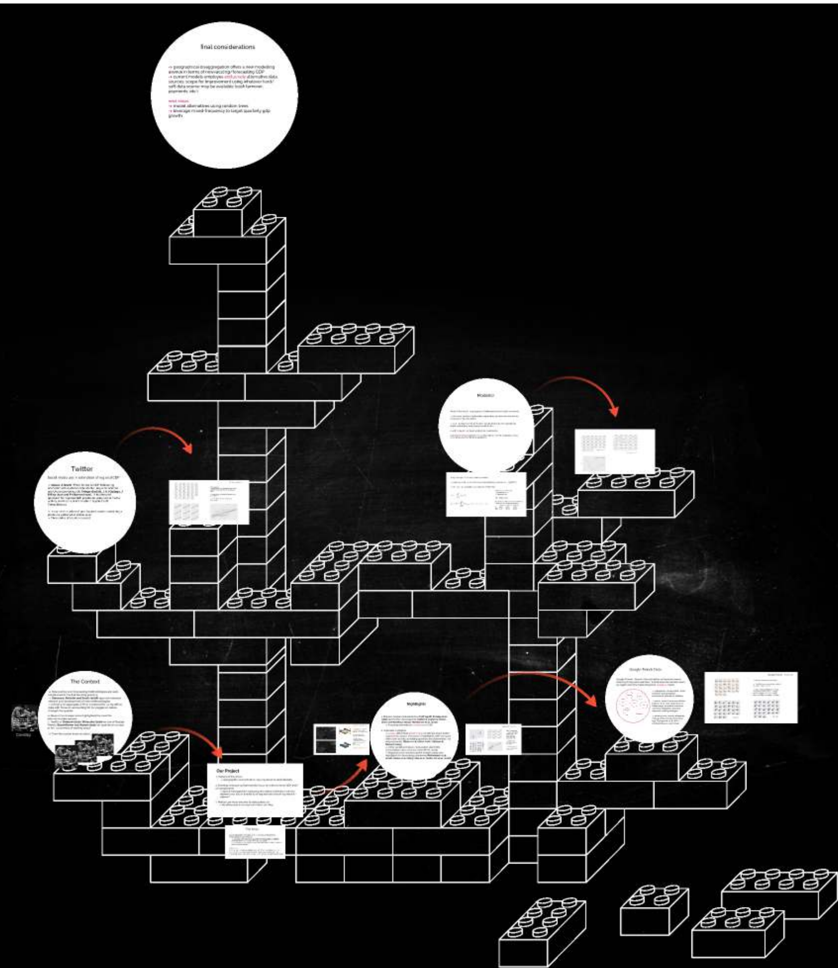
final considerations

-> geographical disaggregation offers a new modelling avenue in terms of nowvacsting/forecasting GDP
-> current models employes exclusively alternative data sources; scope for improvement using whatever hard/soft data source may be available (cash turnover, payments, etc.)

next steps:
-> model alternatives using random trees
-> leverage mixed-frequency to target quarterly gdp growth

Warcasting-Nowcasting GDP without economic data

Constantinescu, Kappner, Szumilo



Mihnea Constantinescu
National Bank of Ukraine and
Kyiv School of Economics

BCC Conference
September 22-23, 2022