Alexander Goryunov

Data science, data analysis, econometrics, causal inference

Mobile: +972-53-571-0490
Email: alex.goryunov@gmail.com
Repo: https://github.com/agoryuno
LinkedIn: Alexander Goryunov
Blog: https://agoryuno.github.io

Employment history

2019-2022: Federal comptroller's office of Russia, Moscow - *senior data analyst*

- quantitative evaluation of government policy
- evangelizing modern approaches to using data for policy decision making with in-house auditors and general government
- code and audit reports review
- overseeing outsourced data engineering projects to ensure quality and proper interfacing with the in-house data pipeline
- assigning junior analysts to projects and supervising their progress

2012-2019: Economic research institute, Russia, Khabarovsk - *deputy director for science*

- overseeing the institute's academic program
- overseeing departmental planning and reporting
- designing and reading courses on economic theory and data analysis
- advising research staff on technical solutions to research problems
- designing and implementing a data warehouse to support research activities

2008-2012: Economic research institute, Russia Khabarovsk - *head of economic theory department*

- conducting research in regional development and international economics
- organizing regular methodological seminars
- planning the department's research activities
- editing the department's research reports

Tech stack

Analysis

- Git
- Cython
- Numpy
- Pandas
- Scikit Learn
- PyTorch
- Statsmodel
- PyMC3
- R

Engineering

- SQLAlchemy
- SQL
- Apache Beam
- Google Cloud Platform

Visualization

- Matplotlib
- Seaborn

Web interfaces

- Starlette
- Flask
- React.js

Alexander Goryunov

Data science, data analysis, econometrics, causal inference

Mobile: +972-53-571-0490
Email: alex.goryunov@gmail.com
Repo: https://github.com/agoryuno
LinkedIn: Alexander Goryunov
Blog: https://agoryuno.github.io

Recent projects

A database of real estate prices in major Russian cities from one of the two main open sources of classifieds. XHR calls to the website's internal API via Selenium used to scrape data, cloud-hosted Postgres with SQLAlchemy for data modeling used to store data. A simple binary tree-based method used to partition API queries used to improve reliability of the scraping service: https://github.com/agoryuno/binrange

Estimation of quality of high schools in 20 Russian regions based on the students' academic outcomes. Data was received from a Russian official agency and covered several years of anonymized standardized exam results for every student in the regional sample. A mixed effects model was chosen to extract the school's effect on individual outcomes with controls for age, gender and a proxy for extra-curricular activities of students. A quarter of an average student's result was attributable to their school with extreme regional inequality in access to education. The estimation method was proposed to the government agency as an alternative way of analyzing access to education.

A **microsimulation model** for long term projections of the Russian poverty level. The model was built in Python, based on data from the national household incomes survey. The model includes modules responsible for different scenarios of demographic change, and income and labour participation effects of the ongoing pension reform. It allows for scenario-based estimation of social transfers' income and poverty reduction effects.

Estimation of direct fiscal cost of additional subsidies for poor families with children. In the absence of data on family incomes, as opposed to household incomes, as well as a discrepancy in the definitions of monetary poverty used by the statistics and social support agencies, meant that neither total number of families with children, nor the number of poor families with children were directly known from official statistics. This gap in data was bridged by an MCMC estimation (PyMC3) utilizing partial information on existing family subsidies from the responsible agency. The total expenditure calculation was supplemented with sensitivity analysis for the possible margin of error, as well as estimates of the labor substitution effect of the new subsidy. The results were accepted by the government and the new subsidy was instituted about 6 months later.

Finding irregularities in the voting patterns of a suspected rigged gubernatorial election in a Russian region in 2018. Data scraping of official websites of polling stations for voting results and station coordinates (mechanize + BeautifulSoup). PCA, spatial autoregression, and visual analysis in R showed extreme spikes of votes for the winning candidate in two large cities, which were also tightly spatially correlated within the cities, suggesting that "networks" of administrators on the ground were involved in the doctoring of results as opposed to centralized vote counting fraud.