



РАНХиГС
РОССИЙСКАЯ АКАДЕМИЯ НАРОДНОГО ХОЗЯЙСТВА
И ГОСУДАРСТВЕННОЙ СЛУЖБЫ
ПРИ ПРЕЗИДЕНТЕ РОССИЙСКОЙ ФЕДЕРАЦИИ

Э М И Т
И Н С Т И Т У Т
ЭКОНОМИКИ, МАТЕМАТИКИ
И ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ

РОССИЙСКАЯ АКАДЕМИЯ НАРОДНОГО ХОЗЯЙСТВА И
ГОСУДАРСТВЕННОЙ СЛУЖБЫ ПРИ ПРЕЗИДЕНТЕ РОССИЙСКОЙ ФЕДЕРАЦИИ

Оценка макроэкономических зависимостей с использованием методов снижения размерности в данных

Отчёт по научно-исследовательской работе

2018

Михаил Гареев

ЭО-15-01

[mkhlgrv@gmail.com](mailto:mkhlgry@gmail.com)

Научный руководитель: к.э.н. Полбин А.В.

- ▶ При оценке моделей из макроэкономики часто можно столкнуться с тем, что параметров относительно много, а наблюдений - мало. Иногда эту проблему решается использованием **методов снижения размерности в данных**.

Цели и задачи

Цель:

- ▶ Проверка некоторых гипотез макроэкономики при помощи методов снижения размерности в данных и создание на их основе предсказательных моделей для макроэкономических показателей.

Задачи:

1. Обзор методов снижения размерности (LASSO, Post-LASSO, Dantzig Selector и др.).
2. Применение этих методов для оценки макроэкономических зависимостей, анализ результатов, сравнение с другими методами оценивания и с результатами, полученными ранее.
3. Построение предсказательных моделей.
4. Создание процедуры мэтчинга стран на основе их макроэкономических показателей.

Методы снижения размерности

Разреженная линейная модель с высокой размерностью в данных

Модель:

$$\beta_0 + \varepsilon_i, \varepsilon_i \sim N(0, \sigma^2), \beta_0 \in \mathbb{R}^p, i = 1, \dots, n,$$

где:

- ▶ y_i — это значения объясняемой переменной,
- ▶ x_i — это значения p -размерной объясняющей переменной,
- ▶ ε_i — значения независимых случайных ошибок в каждом наблюдении i ,

при этом возможно, что $p \geq n$, но только $s < n$ компонентов вектора β_0 не равны 0.

Можно ли уменьшить размерность модели?

Методы снижения размерности

Oracle Problem

Задача (Oracle Problem):

$$\min_{\beta \in \mathbb{R}^p} \mathbb{E}_n \left[(y_i - x_i' \beta)^2 \right] + \sigma^2 \frac{\|\beta\|_0}{n}, \quad (1)$$

где $\|\beta\|_0$ — это количество ненулевых компонентов в векторе β , обобщение понятия нормы для степени 0.

Гёльдерова норма для вектора x степени p :

$$\|x\|_p = \sqrt[p]{\sum_i |x_i|^p},$$

где обычно $p \geq 1$.

Решение (1) — это баланс между ошибкой регрессии и количеством ненулевых коэффициентов из вектора β .

Методы снижения размерности оптимизируют эмпирические аналоги задачи (1).

AIC/ BIC

$$\hat{\beta} \in \arg \min_{\beta \in \mathbb{R}^p} \sum_i = 1^n \left[(y_i - x_i' \beta)^2 \right] + \frac{\lambda}{n} \|\beta\|_0,$$

где λ — параметр штрафа.

LASSO

$$\hat{\beta}^{\text{LASSO}} \in \arg \min_{\beta \in \mathbb{R}^p} \sum_i = 1^n \left[(y_i - x_i' \beta)^2 \right] + \frac{\lambda}{n} \|\beta\|_1,$$

где λ — параметр штрафа, выбирается алгоритмически.

Методы снижения размерности

Post-LASSO

1. Использовать метода LASSO, найти $\hat{\beta}$.
2. Применить МНК-регрессию, оценивая только неисключенные параметры β :

$$\tilde{\beta} \in \arg \min_{\beta \in \mathbb{R}^p} \sum_i (y_i - x_i' \beta)^2 + \frac{\lambda}{n} \|\beta\|_1, \beta_j = 0 \text{ if } \hat{\beta}_j = 0.$$

Dantzig Selector

$$\begin{aligned} & \min_{\beta \in \mathbb{R}^p} \|\beta\|_1 \\ & \text{s.t. } |x_i \beta - y_i| \leq \lambda \forall i = 1, \dots, n, \end{aligned}$$

где λ — параметр штрафа, выбирается алгоритмически.

Проверка гипотезы конвергенции с помощью методов снижения размерности

Однофакторная модель

Модель

$$g_i = \alpha + \beta \ln(G_i) + \varepsilon_i, \varepsilon_i \sim N(0, \sigma^2),$$

где:

- ▶ g_i — средний за 1980–1984 темп роста реального ВВП на душу населения,
- ▶ G_i — логарифм ВВП на душу населения в 1980 г. (в долларах) для страны $i, i = 1, \dots, 245$.

Данные:

Проверка гипотезы конвергенции с помощью методов снижения размерности.

Использование LASSO

Модель

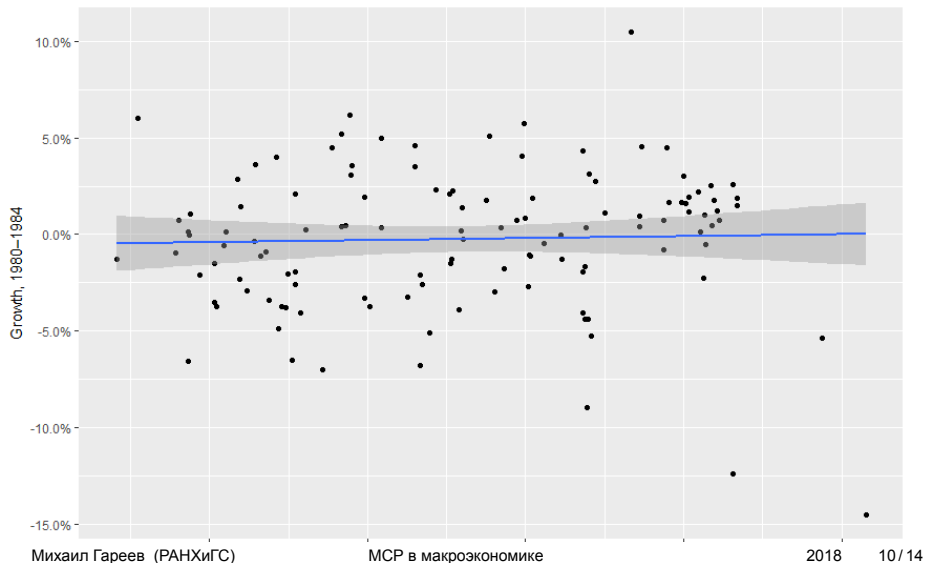
$$g_i = \alpha + \beta \ln(G_i) + \varepsilon_i, \varepsilon_i \sim N(0, \sigma^2),$$

где:

- ▶ g_i — средний за 1980–1984 темп роста реального ВВП на душу населения,
- ▶ G_i — логарифм ВВП на душу населения в 1980 г. (в долларах) для страны $i, i = 1, \dots, 245$.

Проверка гипотезы конвергенции с помощью методов снижения размерности.

Данные



Проверка гипотезы конвергенции с помощью методов снижения размерности.

Сравнение результатов

Таблица: Результаты регрессий

	g	
	(1)	(2)
G	0.001 (−0.005, 0.007)	−0.0112 (−0.022, 0.001)
C	−0.010 (−0.055, 0.034)	−0.03 (−0.032, 0.041)
Наблюдений	120	120
R ²	0.001	0.001
Adjusted R ²	−0.007	−0.007
lambda		2.7870

Примечание:

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

- ▶ Методы снижения размерности (LASSO, Post-LASSO и др.) потенциально представляют собой мощный инструмент для нахождения и проверки макроэкономических зависимостей.
- ▶ Планы на ближайшее время:
 - ⇒ Сделать осмысленные LASSO, Post-Lasso регрессии для проверки гипотезы конвергенции на основе современных данных Всемирного банка.
 - ⇒ Проверить другие макроэкономические гипотезы с помощью методов снижения размерности.

Спасибо за внимание

Оценка макроэкономических зависимостей с
использованием методов снижения размерности в
данных

Михаил Гареев

ЭО-15-01

mkhlgry@gmail.com

2018



Belloni, Alexandre and Chernozhukov, Victor
High dimensional sparse econometric models: An introduction.
Springer, 2011



Belloni, Alexandre, Victor Chernozhukov, and Christian Hansen.
Lasso methods for gaussian instrumental variables models
2011



Barro, Robert J. and Lee, Jong-Wha
Data Set for a Panel of 138 Countries
1994



Candes, Emmanuel, and Terence Tao.
The Dantzig selector: Statistical estimation when p is much larger
than n .
The Annals of Statistics 35.6 (2007): 2313-2351.



Akaike, Hirotugu.
A new look at the statistical model identification.
IEEE transactions on automatic control 19.6 (1974): 716-723.