

Влияние железных дорог на экономику: На примере Российской Империи конца XIX в.

Кирилл Сафонов

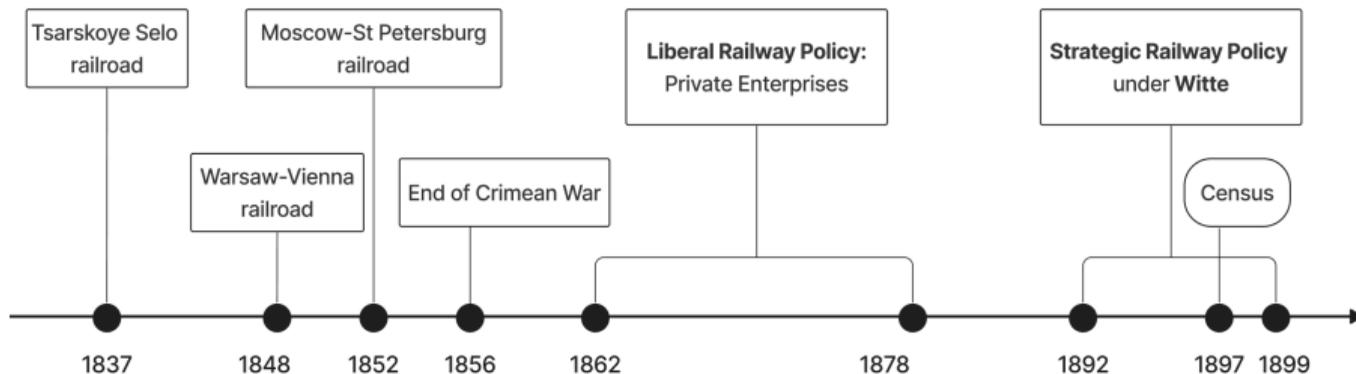


Мотивация

Дискуссия среди экономистов: оказывают ли крупные инфраструктурные проекты, такие как железные дороги, значительное влияние на экономический рост?¹

Контекст Российской Империи:

- 1 Россия была одним из крупнейших развивающихся рынков в конце XIX века.
- 2 Строительство железных дорог развивалось чрезвычайно быстро.
- 3 Железные дороги строились как по военным, так и по экономическим причинам.



¹ См. Fogel, 1964, Nerlove, 1966, Metzer, 1973, White, 1976, Allen & Arkolakis, 2014, Donaldson & Hornbeck, 2016, Donaldson, 2018

Структура

- 1 КОНТЕКСТ: РОССИЙСКАЯ ИМПЕРИЯ В КОНЦЕ XIX ВЕКА
- 2 ЛИТЕРАТУРА ПО ЭКОНОМИЧЕСКОМУ ЭФФЕКТУ ТРАНСПОРТНОЙ ИНФРАСТРУКТУРЫ
- 3 ДАННЫЕ И РЕКОНСТРУКЦИЯ ГРАФА ДОРОГ
- 4 МОДЕЛЬ “OPTIMAL TRANSPORT NETWORK FRAMEWORK”
- 5 ИДЕНТИФИКАЦИОННАЯ СТРАТЕГИЯ И РЕЗУЛЬТАТЫ

Первые железные дороги

- 1837 год – Царскосельская железная дорога
- 1848 год – Варшаво-Венская железная дорога
- 1852 год – Николаевская железная дорога

Российская Империя была пионером в освоении нового типа инфраструктуры, но к середине 19 века темпы строительства сильно замедлились.

С 1866 по 1899 Россия переживала бум строительства железных дорог. Два основных периода:

- "Либеральный"
- Стратегический

Две системы

Частные дороги, 1860-1870

- Незавершенные проекты и проблема с кредитованием
- Часто несогласованные между собой и не соответствующие военным целям

Государственные дороги, 1890-е:

Длина железнодорожной сети увеличилась почти на 30% с 1892 по 1896 год, а затем еще на 40% за следующие 4 года

Литература. Экономическое влияние строительства железных дорог

- [Fogel \(1964\)](#) предлагает первый *количественный* подход, модель “социальных сбережений”, и оценивает влияние железных дорог в США XIX века.
- Работы с моделями Difference-in-difference и Instrumental Variable (американский Средний Запад — [Atack и др., 2010](#), Швеция – [Berger & Enflo, 2017](#)).
- Подход на основе модели общего равновесия с моделированием транспортной технологии (США XIX века – [Williamson, 1974](#), [Herrendorf и др., 2009](#), [Perez-Cervantes, 2013](#)).

Литература

Рассмотрим эффект на уровне *уезда*. Основная проблема — эндогенность.

Решения в литературе:

- Использовать зависимую переменную, которая не связана со строительством железных дорог: "market access" в США XIX века ([Donaldson & Hornbeck, 2016](#)).
- Строительство железных дорог исключительно по военным причинам в Индии ([Donaldson, 2018](#)).
- DiD модели и специфические инструментальные переменные на Среднем Западе ([Atack и др., 2010](#)) и Швеции ([Berger & Enflo, 2017](#)).

Эта работа

- Использует синтетически сконструированную оптимальную транспортную сеть в качестве инструментальной переменной;
- Оценивает влияние строительства железных дорог на *урбанизацию* в Российской империи;
- Оценивает убытки от экономически не оптимального строительства железных дорог в Российской империи.

Данные

Основные переменные в анализе:

- Население (Перепись населения Российской империи, 1897);
- Урбанизация (Ibid.)²;
- Добавленная стоимость. В качестве прокси-переменной используются:
 - индустриальное производство в 1900 г. (Gregg, 2020);
 - урожайность зимней и летней пшеницы и ржи в 1913 г. (Dower & Markevich, 2018);
- Граф *фактической* железнодорожной сети на 1897 год — нужна **реконструкции**.

Текущий анализ ограничен европейской частью Российской империи, так как это наиболее густонаселённая и экономически активная часть.

² Данные переписи доступны здесь: Demoscope Weekly, www.demoscope.ru/weekly/ssp/rusgub97.php. ↻ 🔍 🔄

Распределение сельскохозяйственного и промышленного производства

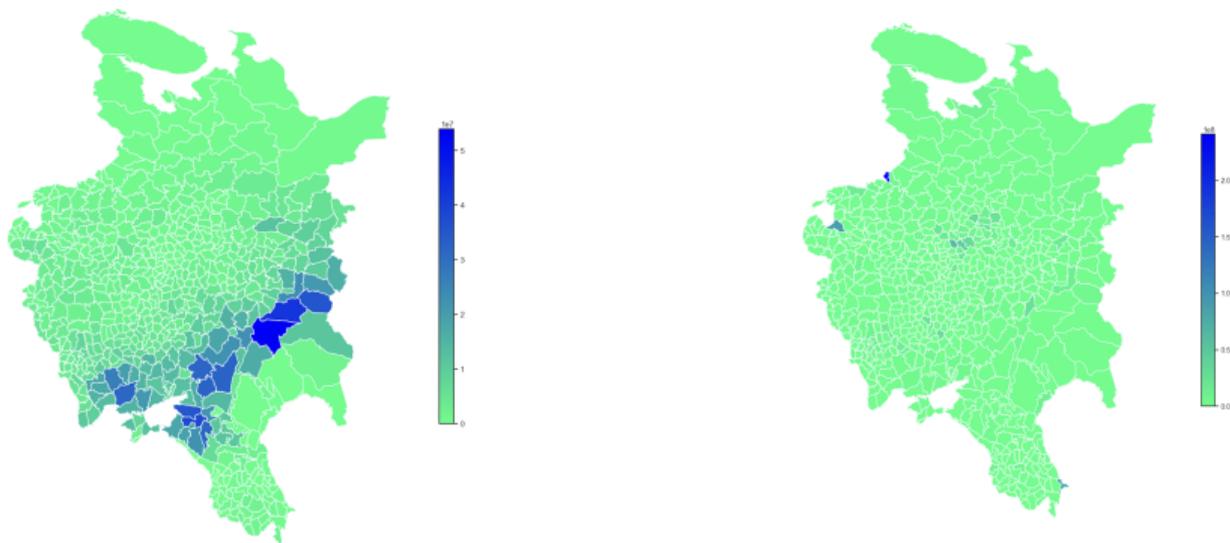


Рис. 1: Распределение сельскохозяйственного (слева) и промышленного (справа) производства в Царской России

Реконструкция современной железнодорожной сети

Данные для строительства текущей железнодорожной сети в 15 странах на постсоветском пространстве:

- Идентификаторы станций и железнодорожных линий из справочника (Тарифный справочник №4) — 26,834 наблюдений;
- Данные с *wikidata*, *openstreetmap* и *alta*, чтобы получить координаты каждой станции;
- Исправление ошибок: явно неверные координаты, пропущенные значения, неправильный порядок в пределах ж/д линии. Работа с ошибками частично автоматизирована: написанный мной алгоритм устранил около 70% ошибок. Итоговый граф требует ручной проверки.

Восстановленная современная железнодорожная система содержит 1,982 секторов и 21,634 станций.

Работа с данными

Как я ускорил процесс предварительной подготовки данных:

- Использовал автоматический сбор структурированной информации (в моем случае – web scraping)
- Нашел закономерности в ошибках и исправлял их алгоритмически. Пример алгоритма:
 - 1 Вычислить расстояние по формуле гаверсина (haversine) между координатами и сравнить его с расстоянием в источнике;
 - 2 Перестроить порядок станций в пределах одной линии, если выявлены несоответствия;
 - 3 Вернуть данные для дальнейшей ручной проверки, если ошибку не удаётся устранить автоматически.

Современная сеть железных дорог



Рис. 2: Современная сеть железнодорожного сообщения в странах постсоветского пространства

Реконструкция исторического графа

Я использовал справочник для определения железнодорожных линий, построенных к 1897 году: 445 линий и 7166 станций³.

В 1871 г. введены в эксплуатацию участки			
Время открытия для правильного движения	Наименование железнодорожных участков	Строительная длина	
		верста	сажень
1871 г.			
январь	Умет — Аткарск	161	354
февраль	Иваново — Кинешма	87	121
апрель	Верховье — Ливны	58	236
май	Чудово — Новгород	68	166
июнь	Харьков — Полтава	131	365
июль	Филоново — Царицын	264	73
июль	Аткарск — Саратов II	87	37
июль	Алексиково — Урюпино	32	378
август	Поти — Квирилы	118	163
август	Тирасполь — Кишинев	66	433
сентябрь	Жмеринка — Волочиск	155	278
сентябрь	Кошечары — Либава	296	222
сентябрь	Брест — соединительная ветвь с Московско-Брестской ж. д.	3	484
ноябрь	Смоленск — Брест	627	431
ноябрь	Отрожка — Шахтная	517	65
	Итого за 1871 г.	2677	306

Рис. 3: Пример одного года из справочника

³Информация о российских железных дорогах с 1838 по 1917 год: <https://istmat.org/node/42966>

Реконструкция графа. Валидация I

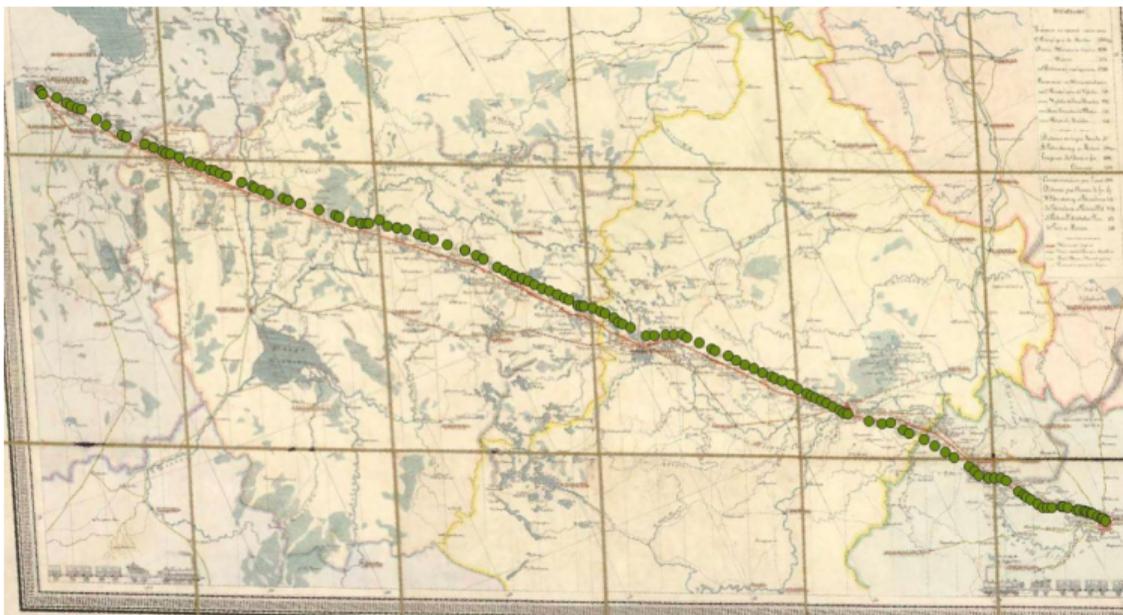
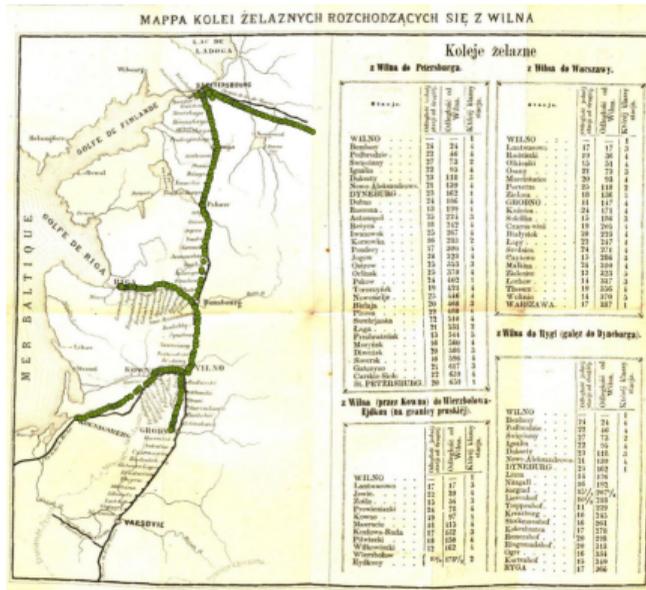


Рис. 4: Железнодорожная линия Москва–Санкт-Петербург (Николаевская). Историческая карта (красная линия) и реконструированная дорога (зеленые точки)

Реконструкция графа. Валидация II



Реконструкция графа и Дискретизация

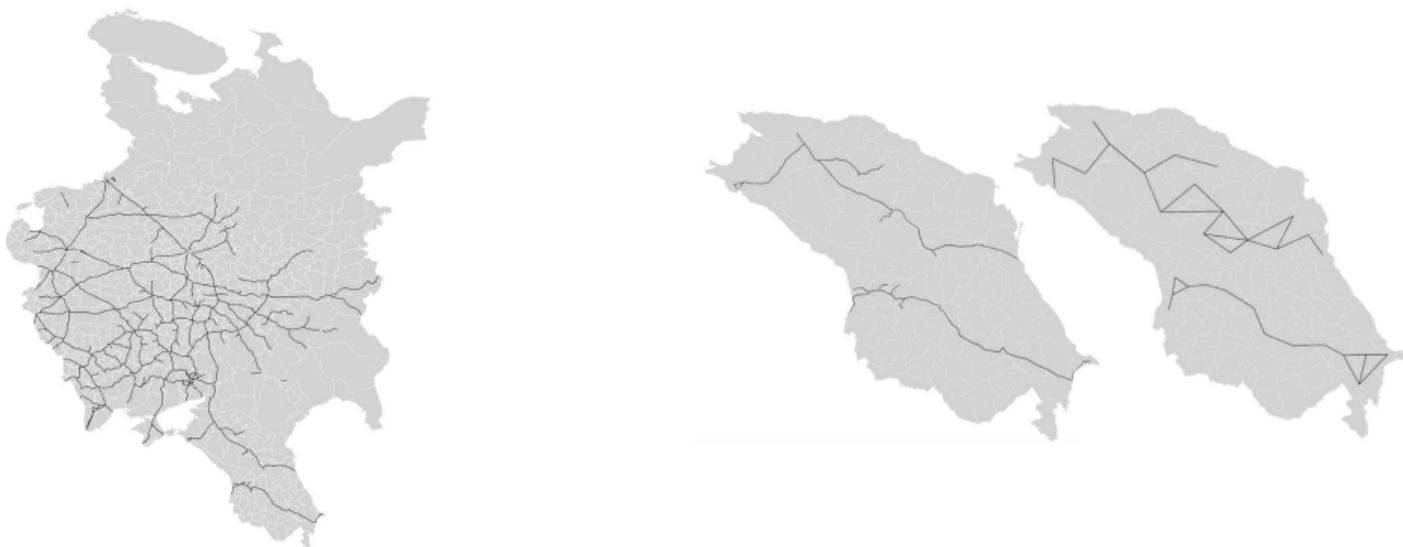


Рис. 6: Реконструированная историческая железнодорожная сеть до 1897 года (слева), результат дискретизации (справа).

Optimal Transport Networks Framework

The Social Planner optimizes:

$$W = \max_{l_{jk}} \max_{Q_{jk}^n} \max_{\{c_j, h_j, D_j^n, L_j^n, z_j^n\}} \sum_j L_j U(c_j, h_j)^4.$$

- L_j is the population in every province j in year 1897;
- $U(c_j, h_j)$ is the utility of an individual worker who consumes c units of the traded goods bundle and h units of the non-traded good in location j ;
- l_{jk} is a binary indicator of counties j and k connectivity with each other (common border);
- Q_{jk}^n is a trade flow from j to k of a sector n good;
- $\{c_j, h_j, D_j^n, L_j^n, z_j^n\}$ is a set of county-level variables.

⁴Fajgelbaum & Schaal (2020).

Optimization Space

The general equilibrium model operates in the space of potential solutions. The space is a mathematical representation of the topology of the country. The graph is based on adjacency matrix of all counties. The nodes are centroids of the counties.

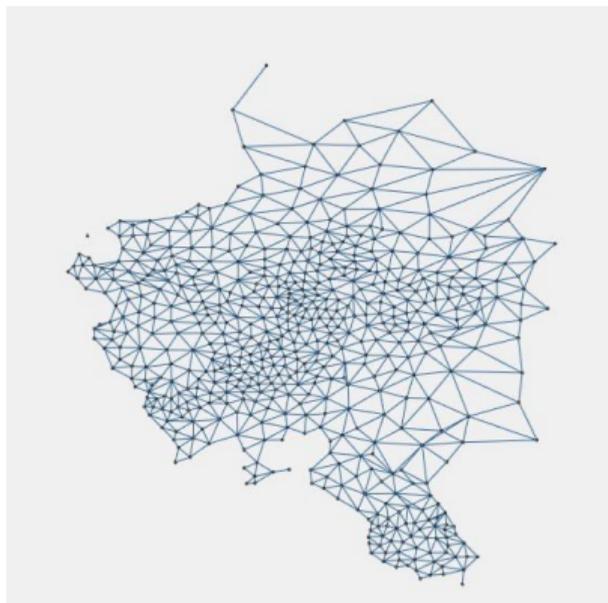


Figure. 7: Space of potential solutions for infrastructure network in the general equilibrium model

Innermost Problem

$$\max_{\{c_j, h_j, D_j^n, L_j^n, z_j^n\}} \sum_j L_j U(c_j, h_j).$$

- $U(c_j, h_j)$ is the utility of an individual worker who consumes c units of the traded goods bundle and h units of the non-traded good in location j . I use Cobb-Douglas utility function: $U(c, h) = c^\alpha h^{1-\alpha}$, where $\alpha = 0.4^a$.
- L_j is the population in every province j in 1897.
- $\{c_j, h_j, D_j^n, L_j^n, z_j^n\}$ – county-level variables.
- The number of tradable sectors is $N + 1$.
- Tradable sectors are combined with the CES aggregator into one good C_j demanded in province j :
$$C_j = \left(\sum_{n=1}^{N+1} (D_j^n)^{\frac{\sigma-1}{\sigma}} \right)^{\frac{\sigma}{\sigma-1}}$$
, where:
 - D_j^n is the demand in location j for the output in sector n ;
 - each worker consumes $c_j = C_j/L_j$;
 - $\sigma = 5$ in the CES aggregator (Head & Mayer, 2014).

^aFajgelbaum & Schaal (2020).

Second Nested Problem

The optimal flow problem: $\max Q_{jk}^n$:

- Q_{jk}^n is a trade flow from county j to k of a sector n good.

Transportation from j to k is described as *iceberg* transportation costs τ_{jk} . I use the log-linear parametrization of transport costs:

$$\tau_{jk} = \delta_{jk} \frac{Q^{\beta}}{I^{\gamma}},$$

- δ_{jk} are geographic frictions;
- Q is the quantity, i.e. trade flow;
- I is the infrastructure;
- $\beta = 0.13$: value calibrated by [Fajgelbaum & Schaal, 2020](#);
- $\gamma = 0.10$: value calibrated by [Fajgelbaum & Schaal, 2020](#).

Transport costs multiplier is a linear function of the Euclidean distance: $\delta_{jk} = \delta_0 \cdot dist_{jk}$

Outermost Problem

The outermost nested problem, $\max_{I_{jk}}$, refers to choosing the optimal infrastructure network. I parameterize the building costs:

$$\log\left(\frac{\delta_{jk}^I}{dist_{jk}}\right) = \log(\delta_0^I) - 0.11 \cdot \mathbb{1}(dist_{jk} > 50 \text{ km}),$$

where:

- δ_{jk}^I is the building costs;
- δ_0^I is the calibrated parameter for infrastructure;
- $dist_{jk}$ is the Euclidean distance between province j and k .

Применение

Этап 1. Я использую наблюдаемые производственные показатели для определения уровней продуктивности z_j в существующей сети I_{jk} . Эта калибровка приводит к набору оптимальных фундаментальных значений (z_j, H_j) .

$$(z, H) = \arg \min_{\{z, H\}} \sum_j (\hat{y}_j - y_j)^2$$

Этап 2. Я использую уже откалиброванные значения (z_j, H_j) , чтобы ещё раз решить задачу оптимизации сети. Наложённое ограничение на максимизацию состоит в том, что оптимизированная сеть использует такое же количество строительных ресурсов, например, железа, как и реальная историческая сеть.

Идентификационная стратегия

Основная модель:

$$Y_j = \alpha + \beta Z_j + \gamma X_j + \varepsilon_j,$$

- $j \in \mathcal{J} = 1, \dots, J$ — это набор локаций, т.е. уездов;
- Z_j — скаляр, идентифицирующий наличие железной дороги в уезде j до 1897 года;
- Y_j — скаляр, представляющий экономическую переменную (урбанизацию) в уезде j в 1897 году;
- X_j — вектор контрольных переменных для уезда j ;
- ε_j — ошибка для графства j .

В качестве инструмента используется синтетическая оптимальная железнодорожная сеть во всех локациях \mathcal{J} :

$$Z_j = \theta + \delta Z'_j + \pi X_j + \varphi_j,$$

- Z'_j — скаляр, идентифицирующий наличие *оптимальной* железной дороги в уезде j

Результаты. Реализованный граф VS синтетический

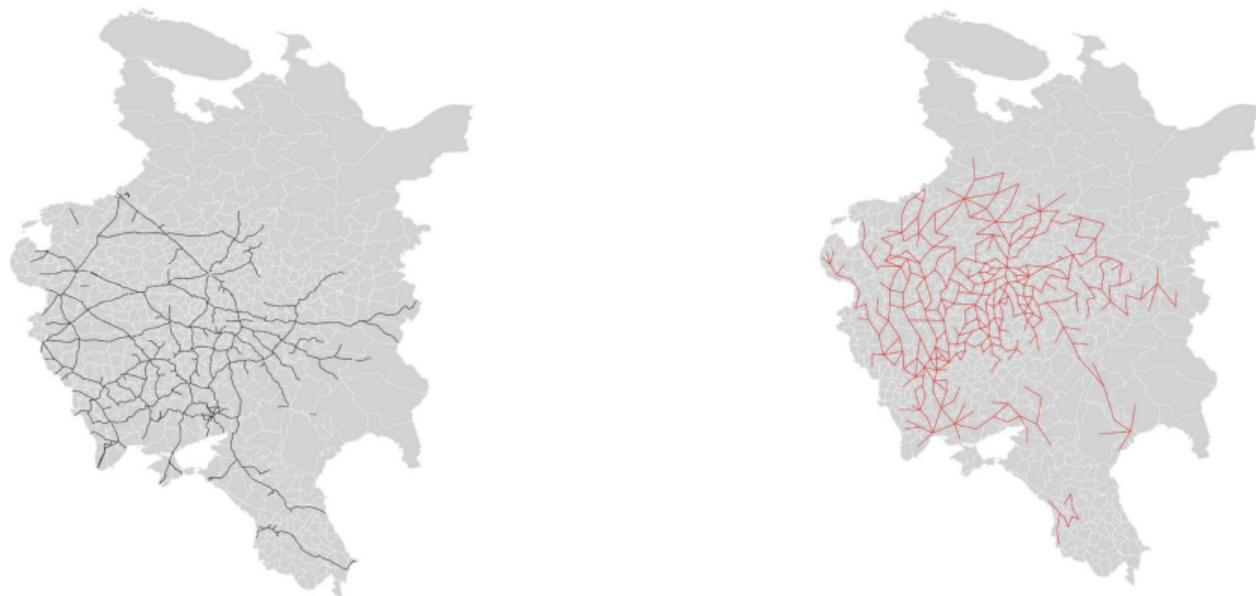


Рис. 8: Фактическая железнодорожная сеть (слева) и оптимальная железнодорожная сеть на основе модели (справа)

Потери общественного благосостояния

Оценка потерь общественного благосостояния:
разница между социальным благосостоянием при фактической (исторической) и синтетически сконструированной ж/д сетями ⁵.

В результате, оценка модели: произошло **неэффективное распределение ресурсов** и снижение благосостояния на около 2.3%.

⁵Социальный планировщик может строить железные дороги в любом месте и тратить то же количество ресурсов, которые были потрачены на реальную сеть

Влияние на урбанизацию

СПЕЦИФИКАЦИИ ПЕРВОГО ЭТАПА (FIRST STAGE):

- (1) OLS-регрессия фактической исторической сети на оптимальную, $F\text{-test} = 15.4$.
- (2) OLS с географическими контролями, $F\text{-test} = 5.4$.
- (3) OLS с географическими контролями, населением и квадратом населения, $F\text{-test} = 7.4$.
- (4) OLS с географическими контролями, населением, квадратом населения и близостью к Крыму, $F\text{-test} = 10.2$. **Итоговая модель.**

ВТОРОЙ ЭТАП (SECOND STAGE):

- Строительство железных дорог не приводит к статистически значимому экономическому эффекту на урбанизацию.

Таблица 1(а): Влияние железнодорожной инфраструктуры на урбанизацию. 2SLS

	<i>Dependent variable:</i>				
	Actual historical network				Urbanization
	OLS				2SLS
	(1)	(2)	(3)	(4)	(5)
Actual historical graph					-0.001 (0.089)
Optimal network	0.163*** (0.042)	0.103** (0.045)	0.116*** (0.043)	0.138*** (0.043)	
Border				0.120 (0.080)	0.075* (0.035)
Population			3.201*** (0.483)	2.473*** (0.488)	1.036*** (0.274)
Population ²			-1.712*** (0.465)	-1.202*** (0.462)	0.408** (0.134)
1 (Crimea distance <1000 km)				0.241*** (0.045)	0.013 (0.024)
Geographical Region Control	X	✓	✓	✓	✓
Constant	0.527*** (0.033)	0.333 (0.270)	0.051 (0.261)	0.119 (0.255)	-0.050 (0.072)
Observations	564	564	564	564	564
R ²	0.026	0.074	0.155	0.201	0.163
Adjusted R ²	0.025	0.064	0.143	0.186	0.147
F-test	15.390	5.377	7.398	10.24	

Note: * p<0.1; ** p<0.05; *** p<0.01. F-test checks the weakness of the optimal network as an instrumental variable. The standard errors are clustered at the geographical region level.

Таблица 1(6): Влияние железнодорожной инфраструктуры на урбанизацию. 2SLS

	<i>Dependent variable:</i>				
	Actual historical network				Urbanization
	<i>OLS</i>				<i>2SLS</i>
	(1)	(2)	(3)	(4)	(5)
Actual historical graph					-0.001 (0.089)
Optimal network	0.163*** (0.042)	0.103** (0.045)	0.116*** (0.043)	0.138*** (0.043)	
Border				0.120 (0.080)	0.075* (0.035)
Population			3.201*** (0.483)	2.473*** (0.488)	1.036*** (0.274)
Population ²			-1.712*** (0.465)	-1.202*** (0.462)	0.408** (0.134)
1(Crimea distance <1000 km)				0.241*** (0.045)	0.013 (0.024)
Geographical Region Control	X	✓	✓	✓	✓
Constant	0.527*** (0.033)	0.333 (0.270)	0.051 (0.261)	0.119 (0.255)	-0.050 (0.072)
Observations	564	564	564	564	564
R ²	0.026	0.074	0.155	0.201	0.163
Adjusted R ²	0.025	0.064	0.143	0.186	0.147
F-test	15.390	5.377	7.398	10.24	

Note: * p<0.1; ** p<0.05; *** p<0.01. F-test checks the weakness of the optimal network as an instrumental variable. The standard errors are clustered at the geographical region level.

Заклучение

Эта работа:

- Методологически, предлагает использовать оптимальную транспортную сеть на основе модели всеобщего равновесия в качестве инструментальной переменной для оценки экономического эффектов реальной транспортной сети;
- Не находит статистически значимого влияния строительства железных дорог на урбанизацию;
- Оценивает потери общественного благосостояния в Российской империи из-за неоптимального строительства железнодорожной сети в 2.3%.

Для дальнейшего исследования

- Учесть международную торговлю или уникальные территориальные характеристики (возвышенности, низины, реки и пр.);
- Рассмотреть влияние железнодорожного строительства к 1910-1920-ым годам, чтобы оценить влияние строительного бума 1890-х годов;
- Калибровка важных параметров модели в контексте Российской империи: выявить численные значения доли расходов, которую потребители отводили на торгуемые товары в потребительских предпочтениях (α) и эластичности замещения между разновидностями товара (σ).
- Учесть эффект крупных инфраструктурные проекты (Транссибирская магистраль, Байкало-Амурская магистраль и т.д.).

Спасибо!

References I

- Allen, T., & Arkolakis, C. (2014). Trade and the topography of the spatial Economy. *The Quarterly Journal of Economics*, 129(3), 1085–1140.
- Atack, J., Bateman, F., Haines, M., & Margo, R. A. (2010). Did Railroads Induce or Follow Economic Growth?: Urbanization and Population Growth in the American Midwest, 1850–1860. *Social Science History*, 34(2), 171–197.
- Berger, T., & Enflo, K. (2017). Locomotives of local growth: The short- and long-term impact of railroads in Sweden. *Journal of Urban Economics*, 98, 124–138.
- Donaldson, D. (2018). Railroads of the Raj: Estimating the impact of transportation infrastructure. *The American Economic Review*, 108(4–5), 899–934.
- Donaldson, D., & Hornbeck, R. (2016). Railroads and American Economic Growth: A “Market Access” approach. *The Quarterly Journal of Economics*, 131(2), 799–858.
- Dower, P., & Markevich, A. (2018). Labor Misallocation and Mass Mobilization: Russian Agriculture during the Great War. *The Review of Economics and Statistics*, 100 (2), 245–259.

References II

- Fajgelbaum, P., & Schaal, E. (2020). Optimal transport networks in spatial equilibrium. *Econometrica*, 88(4), 1411–1452.
- Fogel, R. W. (1964). Railroads and American economic growth: Essays in Econometric History. *Baltimore, Johns Hopkins Press*.
- Gregg, A. (2020). Factory Productivity and the Concession System of Incorporation in Late Imperial Russia, 1894–1908. *American Economic Review*, 110 (2), 401–427.
- Head, K., & Mayer, T. (2014). Chapter 3—Gravity Equations: Workhorse, Toolkit, and Cookbook. In G. Gopinath, E. Helpman, K. Rogoff (Eds.), *Handbook of International Economics*. Elsevier, 4, 131–195.
- Herrendorf, B., Schmitz, J., & Teixeira, A. (2009). Transportation and Development: Insights from the U.S., 1840–1860. *Federal Reserve Bank of Minneapolis Research Department Staff Report 425*.
- Metzer, J. (1973). Some Economic Aspects of Railroad Development in Tsarist Russia. *The Journal of Economic History*, 1, 314–316.

- Nerlove, M. (1966). Railroads and American Economic Growth. *The Journal of Economic History*. *The Journal of Economic History*.
- Perez-Cervantes, F. (2013). Railroads and Economic Growth: A Trade Policy Approach. *University of Chicago Mimeo*.
- White, C. M. (1976). The concept of social saving in theory and practice. *The Economic History Review*, 29 (1), 82—100.
- Williamson, J. G. (1974). Late Nineteenth-Century American Development: a General Equilibrium History. *New York: Cambridge Univeristy Press*.

Appendix - Descriptive Statistics

Table 2: Descriptive Statistics

Statistic	Mean	St. Dev.	Q1	Median	Q3	Max
Population (ppl)	172	113	105	153	220	1,318
Urban Population (ppl)	23	75	4.6	8.9	18	1,265
Agricultural Output (tonnes)	4,601	6,327	1,171	2,420	5,028	53,967
Industrial Output (rubles)	2,643	12,228	48	315	1,422	241,162
Urbanization (%)	10.5%	13.1%	3.6%	6.0%	11.8 %	96.0 %

Note: The urbanization rate is in percentage points. All other measurements are in thousands. It contains data of present population which was located in the area at the time of measurement.

[Return to the presentation](#)

Appendix – Contemporary Railway Network

The map of the constructed contemporary railway network for 15 countries: Armenia, Azerbaijan, Belarus, Estonia, Georgia, Kazakhstan, Kyrgyzstan, Latvia, Lithuania, Moldova, Russia, Tajikistan, Turkmenistan, Ukraine, Uzbekistan.



Figure 9: Contemporary network of the railway network in the post-Soviet countries

Appendix – Network Discretization

This is a subsample of the network discretization algorithm.

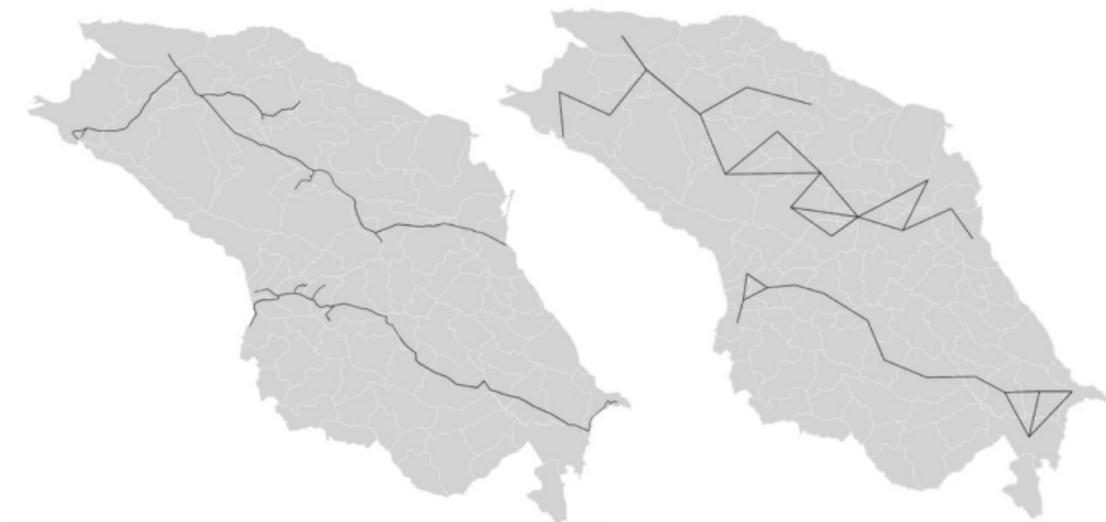


Figure 10: Example of network discretization in Tsarist Russia

Appendix – List of Geographical Regions

- (1) Northwestern Region: Grodno governorate, Kovno governorate, Minsk governorate, Mogilev governorate, Vilna governorate, Vitebsk governorate
- (2) Southwestern Region: Bessarabia governorate, Kiev governorate, Podolsk governorate, Volhynian governorate
- (3) Caucasus Region: Baku governorate, Black Sea governorate, Dagestan region, Elisabethpol governorate, Erevan governorate, Kars governorate, Kuban region, Kutaisi governorate, Stavropol governorate, Terek region, Tiflis governorate
- (4) Ostsee (Baltic) Region: Estonia governorate, Kurland governorate, Livonia governorate
- (5) Central Region: Arkhangelsk governorate, Astrakhan governorate, Chernigov governorate, Don Cossack host lands, Ekaterinoslav governorate, Kaluga governorate, Kazan governorate, Kharkov governorate, Kherson governorate, Kostroma governorate, Kursk governorate, Moscow governorate, Nizhny Novgorod governorate, Novgorod governorate, Olonets governorate, Orel governorate, Penza governorate, Poltava governorate, Pskov governorate, Ryazan governorate, Samara governorate, Saratov governorate, Simbirsk governorate, Smolensk governorate, St. Petersburg governorate, Tambov governorate, Taurida governorate, Tula governorate, Tver governorate, Ural region, Viatka governorate, Vladimir governorate, Vologda governorate.