The Empirics of Agglomeration

Vera Ivanova

Chuvash State University (Cheboksary, RUSSIA)

The First CMSSE Summer School, Nizhny Novgorod, 2012

Outline

1. A general framework
2. Empirical tests evaluating different determinants of firm's location choices
3. Location choice
   - An Econometric Model of Location
   - The Determinants: Production Cost and Productivity
   - The Determinants: Market Potential
   - Discrete choice models of location
4. HME
5. Local factor prices
6. Stability of the spatial patterns
   - Historical stability of agglomerations
   - Can initial geographical advantages be turned into a disadvantage?
   - Do temporary shocks have a long-run impact?
   - Temporary shocks. Case of Japan
   - Temporary shocks. Case of Russia
   - The division of Germany as a natural experiment
7. Concluding remarks

A general framework

Profit equation (follows from the multi-regional versions of the DSK model):

$$\pi_{rs}^* = (p_r^* - m_r)\tau_{rs} q_{rs}^* = m_r \frac{\tau_{rs} q_{rs}^*}{\sigma - 1}$$

Mill pricing:

$$p_{rs}^* = \tau_{rs} p_r^* = \tau_{rs} m_r \sigma / (\sigma - 1)$$

- r, s – indices corresponding to any given region or country,
- p_r – the mill price of a variety sold by a firm located in r,
- m_r – marginal production cost,
- q_{rs} – the quantity that this firm sells on market s,
- τ_{rs} – the iceberg-type trade cost from r to s.

A general framework

Quantity:

$$q^*_{rs} = (\tau_{rs} p^*_r)^{-\sigma} \mu_s Y_s P_s^{\sigma-1},$$

- $P_s = \left(\sum_{r=1}^{R} n_r (p^*_r \tau_{rs})^{-(\sigma-1)} \right)^{-1/(\sigma-1)}$ – CES price index in region s,
- Y_s – income of region s,
- μ_s – share of good considered in the consumption of region s.

A general framework

Total profit for a firm located in r :

$$\Pi_r^* = \sum_{r=1}^{R} \pi_{rs}^* - F_r = c m_r^{-(\sigma-1)} RMP_r - F_r,$$

- F_r – firm's fixed cost,
- $RMP_r = \sum_{r=1}^{R} \phi_{rs} \mu_s Y_s P_s^{\sigma-1}$ – real market potential,
- $\phi_{rs} = \tau_{rs}^{-(\sigma-1)}$,
- c – a coefficient depending only on σ

Market potential as an indicator for the degree of accessibility to market r (Harris, 1954):

$$MP_r = \sum_{r=1}^{R} \frac{Y_s}{d_{rs}}$$

- d_{rs} – distance between r and s.

Empirical tests evaluating different determinants of firms' location choices

1. Location choice

2. Home-market effect

3. Local factor prices

4. Stability of the spatial structures

An econometric model of location

As reliable data for profits are almost always non-available, take a monotonic transformation of profits:

$$U_r \equiv \frac{-\ln c + \ln(\Pi_r^* + F)}{\sigma - 1} = \frac{1}{\sigma - 1} \ln RMP_r - \ln m_r$$

where

- RMP – real market potential,
- m_r – marginal production cost

Estimation of production costs

Assuming technology of Cobb-Douglas type, we get:

$$\ln m_r = \alpha \ln w_r + (1-\alpha)\ln v_r - \ln A_r$$

where

- w_r – wage in region r,
- α – labor share in the production process, $\alpha \in (0,1)$
- v_r – price of intermediate goods (or some production factors other than labor),
- A_r – total factor productivity in region r.

Measurement of market potentials

Why not to measure RMPs directly?

- the RMPs depend on unknown parameters, such as τ_{rs}, which we want to estimate
- the RMPs depend on CES-based price indices P_s, which are not directly observable

A widely used strategy is iterated non-linear least squares algorythms

The logit model of location

As firms maximize profits, they also maximize $\widetilde{U}_r$ given by:

$$\widetilde{U}_r = U_r + \varepsilon_r,$$

where ε_r is the error term capturing possible idiosyncrasies in firms' behavior

Assume that

$$\varepsilon_r \sim iid, \quad F(\varepsilon_r) = \exp(-\exp(-\varepsilon_r))$$

Then probability that $U_r \geq U_s \; \forall s \neq r$ is given by multinomial logit formula:

$$P_r = \frac{exp(U_r)}{\sum_{s=1}^{N} exp(U_r)}$$

Standard estimation technique is maximum likelihood

Theoretical framework

The basic HME relationship, following from the Helpman-Krugman model (1985)

$$\lambda_r^* = \frac{1}{2} + \mathscr{M}\left(\theta_r - \frac{1}{2}\right),$$

where:

- θ_r – the share of demand attributed to region r;
- λ_r^* – the share of firms located in region r;
- $\mathscr{M}$ –some measure of trade freeness:

$$\mathscr{M} \equiv \frac{1+\phi}{1-\phi}$$

Estimation strategy

Davis, Weinstein (1996, 1999, 2003) suggest an empirical counterpart (however, rather vague) of the above HME relationship:

$$y_r^k = \beta_1 share_r^k + \beta_2 idiodem_r^k + \varepsilon_r^k \ ,$$

where

- y_r^k – production of a good k in country r ,
- $share_r^k$ – production of good k in country r provided the share of sector k in this country is the same as in the rest of the world:

$$share_r^k = \frac{y_R^k}{y_R} y_r,$$

- $y_R^k = \sum_{s \neq r} y_s^k$, $y_R = \sum_k X_R^k$
- $idiodem_r^k$ – deviation of country r's expenditure in good k relative to the rest of the world's expenditure pattern:

$$idiodem_r^k = \left(\frac{E_r^k}{E_r} - \frac{E_R^k}{E_R} \right) y_r$$

Note: the above equation becomes more closely related to the theory if $\beta_1 = 1$

Estimation results: an example

Davis and Weinstein (DW) estimators in pooled regressions.

		$\text{share}_r^k\ (\hat{\beta}_1)$		$\text{idiodem}_r^k\ (\hat{\beta}_2)$	
Article/sample	factors_r^k	Result	s.d.	Result	s.d.
DW96, OECD	No	1.103	(0.002)	1.229	(0.005)
	Yes	0.259	(0.198)	0.712	(0.033)
DW99, Japan	No	1.033	(0.007)	1.416	(0.025)
	Yes	−1.744	(0.211)	0.888	(0.070)
DW03, OECD	No	0.96	(0.01)	1.67	(0.05)
	Yes	—	—	1.57	(0.10)

Local factor prices

Basic regression (Redding, Venables, 2004):

$$\ln w_r = \frac{1}{\alpha\sigma} \ln RMP_r - \frac{\gamma}{\alpha} \ln P_r - \frac{\beta}{\alpha} \ln \chi_r - \frac{1}{\sigma\alpha} \ln \frac{a}{c},$$

- w_r – wages,
- χ_r– prices of the other primary factors,
- P_r – price index of the varieties,
- RMP_r – real market potential,
- $\alpha + \beta + \gamma = 1$,
- a – constant measuring the degree of increasing return to scale, which is assuming to be the same across regions.

Stability of the spatial structures. How to measure?

Correlation between region r's share $\lambda_{r,t}$ in the total population at time t, and this share b years earlier, $\lambda_{r,t-b}$.

High correlation is to be expected for short periods.

Is the correlation markedly lower over longer periods (s.t. substantial demographic and economic developments or s.t. important shocks)**?**

Correlation between regional densities

DONALD R. DAVIS and DAVID E. WEINSTEIN. *Bones, Bombs, and Break Points: The Geography of Economic Activity.* AER, 2002.

Year	Population in thousands	Share of five largest regions	Relative variance of log population density	Zipf coefficient	Raw correlation with 1998	Rank correlation with 1998	History
1600	12,266	0.30	0.64	— 1.192 (0.068)	**0.76**	**0.83**	Reunification achieved after bloody war, extensive contact with West. Japan is a major regional trading and military power.
1721	31,290	0.21	0.43	— 1.582 (0.113)	**0.85**	**0.84**	Closure of Japan to trade with minor exceptions around Nagasaki. Capital moves to Tokyo. Political stability achieved.
1798	30,531	0.21	0.37	— 1.697 (0.120)	**0.83**	**0.81**	Population is approximately 80 percent farmers, 6 percent nobility. Population stability attributed to infanticide, birth control, and famines.
1872	33,748	0.18	0.30	— 1.877 (0.140)	**0.76**	**0.78**	Collapse of shogun's government, civil war, jump to free trade, end of feudal regime, start subsidized import of foreign technology.
1920	53,032	0.25	0.43	— 1.476 (0.043)	**0.94**	**0.93**	Industrialization and militarization in full swing, but still 50 percent of labor force is farmers. Japan is a major exporter of silk and textiles.
1998	119,49	0.41	1.00	— 0.963 (0.025)	**1.00**	**1.00**	Japan is a fully industrialized country. Tokyo, with a population of 12 million, is one of the largest cities in the world.

All time periods have 39 regions with Hokkaido and Okinawa dropped from all years. The relative variance of the log population is the variance of the log of population density in year t divided by the variance of the log of population density in 1998. The Zipf coefficient is from a regression of log rank on log population density using 1920 log density as instruments for the years prior to 725 and 1998 data for later years. Standard errors are in parentheses. The correlation columns indicate the raw and rank correlations between regional density in a given year and regional density in 1998.

Correlation between regional densities

Davis, Weinstein:

- regions' hierarchy has remained extremely stable

Correlation between populations

Steven Brakman, Harry Garretsen and Marc Schramm. *The strategic bombing of German cities during World War II and its impact on city growth.* Journal of Economic Geography, 2004.

Table 2. Rank correlation, with city size ranking for 1939 as benchmark

	W-50	**E-22**	**60**
1946	.860	.931	.875
1964	.953	.920	.946
1979	.867	.878	.856
1990	.850	.854	.847
1999	.855	.881	.841

W-50=50 largest cities in West Germany,
E-22=22 largest cities in East Germany,
60=60 largest cities in East and West Germany combined.

German urban structure displays less stability than the Japanese one.

Brakman et al.:

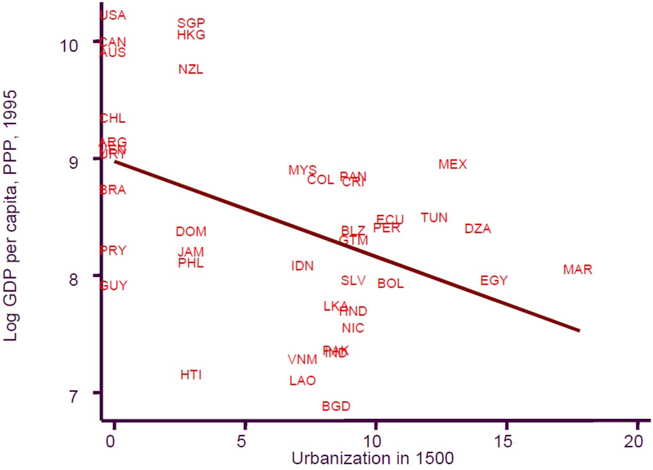
bombing had a significant but temporary impact on post-war city growth in Germany as a whole as well as in West Germany separately, but that this is not the case for city growth in East Germany.

Can initial geographical advantages be turned into a disadvatage?

Acemoglu, Daron, Simon Johnson and James A. Robinson. *Reversal Of Fortune: Geography And Institutions In The Making Of The Modern World Income Distribution.* Quarterly Journal of Economics, 2002.

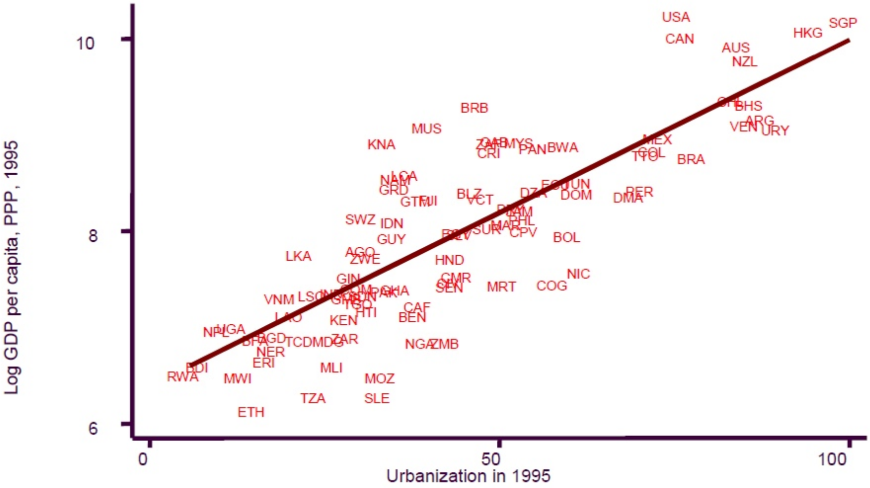
European colonization of the American, African and Oceanic continents from 1500 onward.

Urbanization in 1500 and GDP per capita in 1995



Urbanization in 1500 is percent of the population living in towns with 5,000 or more people.

Urbanization in 1995 and GDP per capita in 1995



Urbanization in 1500 and GDP per capita in 1995

Urbanization in 1500 and GDP per capita in 1995 for former European colonies.

Dependent Variable is log GDP per capita (PPP basis) in 1995	base sample	without North Africa	without the Americas	just the Americas	with continent dummies
Urbanization in 1500	-0.08 (0.03)	-0.10 (0.03)	-0.12 (0.05)	-0.05 (0.03)	-0.08 (0.03)
Asia Dummy					-1.33 (0.61)
Africa Dummy					-0.53 (0.77)
America Dummy					-0.96 (0.57)
R^2	0.19	0.22	0.26	0.13	0.32
Number of Observations	41	37	17	24	41

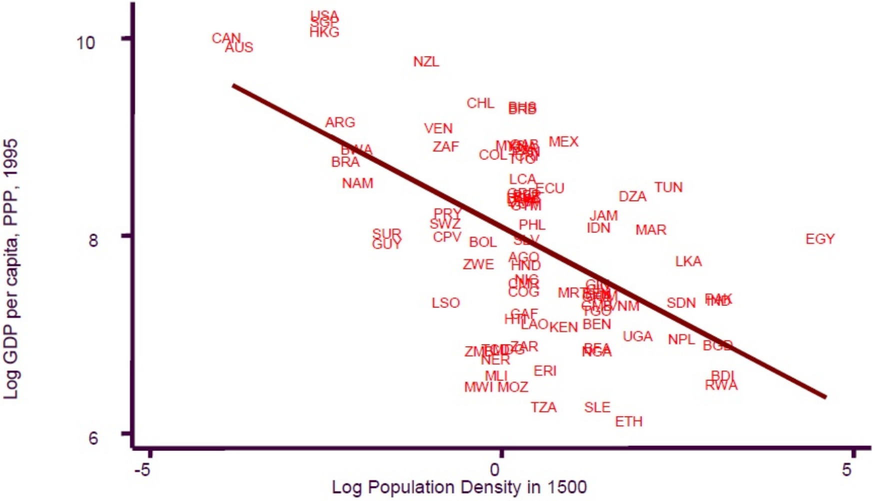
Notes. Urbanization in 1500 is percent of the population living in towns with 5,000 or more people. Depending variable – logarithm of GDP per capita, on Purchasing Power Parity Basis, in 1995.

Urbanization in 1500 and GDP per capita in 1995

Dependent Variable is log GDP per capita (PPP basis) in 1995	base sample	Without neo-Europes	controlling for latitude	controlling for other geography	controlling for colonial origin	controlling for religion
Urbanization in 1500	-0.08 (0.03)	-0.05 (0.03)	-0.07 (0.03)	-0.09 (0.04)	-0.07 (0.03)	-0.06 (0.03)
P-Value for Temperature				[0.23]		
P-Value for Humidity				[0.67]		
P-Value for Soil Quality				[0.95]		
P-Value for Natural Resources				[0.92]		
Dummy for Landlocked				-1.14(0.63)		
Latitude			1.42 (0.92)			
Former French Colony					-0.59 (0.39)	
Former Spanish Colony					0.06 (0.29)	
P-value for Religion						[0.47]
R^2	0.19	0.09	0.24	0.62	0.27	0.25
Number of Observations	41	37	41	41	41	41

Notes. The regression that includes continent dummies omits Oceania. The neo-Europes are the USA, Canada, Australia and New Zealand.
The "other geography" regression includes 5 measures of temperature, 4 measures of humidity, 7 measures of soil quality, and 5 measures of natural resources.
The regression that controls for colonial origin includes dummies for former French colony, Spanish colony, Portuguese colony, Belgian colony, Italian colony, German colony, and Dutch colony. British colonies are omitted. The religion variables are percent of the population who are Muslim, Catholic, and "other"; percent Protestant is omitted.

Population density in 1500 and GDP per capita in 1995



Notes. Population density in 1500 is total population divided by arable land area.

Population density in 1500 and GDP per capita in 1995

Dependent Variable is log GDP per capita (PPP basis) in 1995	base sample	without North Africa	without the Americas	just the Americas	with continent dummies
Log Population Density in 1500	-0.38 (0.06)	-0.40 (0.05)	-0.32 (0.07)	-0.25 (0.09)	-0.26 (0.05)
Asia Dummy					-0.91 (0.55)
Africa Dummy					-1.7 (0.52)
America Dummy					-0.69 (0.51)
R^2	0.34	0.55	0.27	0.22	0.56
Number of Observations	91	47	58	33	91

Notes. Population density in 1500 is total population divided by arable land area.

Population density in 1500 and GDP per capita in 1995

Dependent Variable is log GDP per capita (PPP basis) in 1995	base sample	Without neo-Europes	controlling for latitude	controlling for other geography	controlling for colonial origin	controlling for religion
Log Population Density in 1500	-0.38 (0.06)	-0.32 (0.06)	-0.33 (0.06)	-0.32 (0.07)	-0.32 (0.06)	-0.37 (0.07)
P-Value for Temperature				[0.30]		
P-Value for Humidity				[0.04]		
P-Value for Soil Quality				[0.32]		
P-Value for Natural Resources				[0.75]		
Dummy for Landlocked				-0.43 (0.27)		
Latitude			2.09 (0.74)			
Former French Colony					-0.48(0.20)	
Former Spanish Colony					0.25(0.22)	
P-value for Religion						[0.47]
R^2	**0.34**	**0.24**	**0.40**	**0.62**	**0.48**	**0.36**
Number of Observations	**91**	**87**	**91**	**85**	**91**	**85**

Notes. Population density in 1500 is total population divided by arable land area.

Do temporary shocks have a long-run impact?

Assume that shocks are multiplicative:

$$\ln \lambda_{r,t+a} - \ln \lambda_{r,t} = \alpha + \beta(\ln \lambda_{r,t} - \ln \lambda_{r,t-b}) + \varepsilon_{r,t}$$

- a – time that has elapsed since the end of a shock occuring at time $t - b$
- b – duration of the shock.

If $\hat{\beta} \approx 0$, the size of cities evolves randomly: temporary shocks would then have a permanent effect.
If $\hat{\beta} \approx -1$, the shocks are totally absorbed after a years.

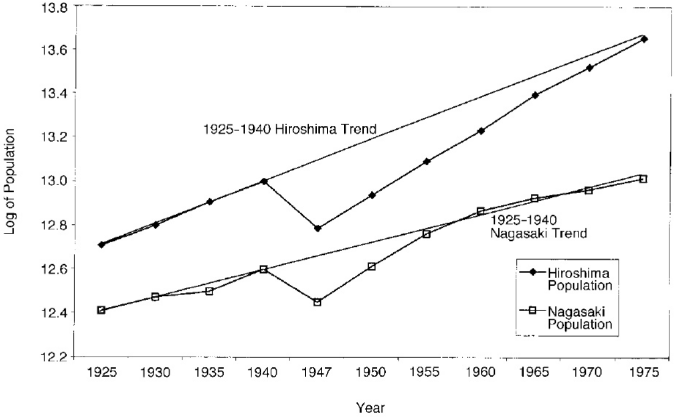
Temporary shocks. Case of Japan

DONALD R. DAVIS and DAVID E. WEINSTEIN. *Bones, Bombs, and Break Points: The Geography of Economic Activity.* AER, 2002.

Dependent variable	growth rate of population between 1947 and 1960		growth rate of population between 1947 and 1965
Independent variable			
Growth rate of population between 1940 and 1947	— 1.048 (0.097)	— 0.759 (0.094)	— 1.027 (0.163)
Government reconstruction expenses	1.024 (0.387)	0.628 (0.298)	0.392 (0.514)
Growth rate of population between 1925 and 1940		0.444 (0.054)	0.617 (0.092)
R^2	0.279	0.566	0.386
Number of observations	303	303	303

Two-stage least-squares estimates of impact of bombing on cities (instruments: deaths per capita and buildings destroyed per capita). Standard errors are in parentheses.

Hiroshima and Nagasaki



Even Nagasaki and Hiroshima (which suffered nuclear bombings that reduced their population by 8.5% and 20%, respectively) saw their population growth rates revert to those of the prewar years of 1925–1940 by 1960 for Nagasaki and by 1975 for Hiroshima, while prewar population levels had been reached long before these dates in both cities.

Temporary shocks. Case of Russia

Tatiana Mikhailova. *Looking for Multiple Equilibria in Russian Urban System*. WP No10/08E, June 15, 2011.

Dependent variable is Ln(Population$_t$)-Ln(Population$_{t-1}$)

Date$_t$ - Date$_{t-1}$	1926 - 1897 (1)	1939 - 1926 (2)	1959 - 1939 (3)	1970 - 1959 (4)	1979 - 1970 (5)	1989 - 1979 (6)	2002 - 1989 (7)
Ln population$_{t-1}$	-.18 (.064)	-.16 (.050)	-.11 (.035)	-.015 (0.11)	-.018 (.016)	-.006 (.007)	-.012 (.010)
Ln urban population$_{t-1}$ inside 20 km radius	.075 (.061)	.031 (.047)	.044 (.029)	-.019 (.01)	-.010 (.009)	-.017 (.006)	.010 (.008)
Ln urban population$_{t-1}$ inside 100 km radius	.027 (.019)	.060 (.021)	.00 (.011)	-.008 (.007)	-.005 (.008)	.002 (.003)	.001 (.004)
Growth$_{t-1}$		.097 (.070)	.14 (.027)	.089 (.023)	.26 (.032)	.32 (.028)	.19 .027)
Oblast center dummy	.43 (.068)	.50 (.088)	.29 (.051)	.25 (.028)	.13 (.031)	.07 (.021)	.043 (.021)
Geography controls	yes	yes	yes	yes	yes	yes	yes
N of obs	500	459	624	756	902	946	955
R^2	0.19	0.25	0.20	0.18	0.25	0.35	0.31

Temporary shocks. Case of Russia

Dependent variable is Ln(Population$_t$)-Ln(Population$_{t-1}$)

Date$_t$ - Date$_{t-1}$	1939 - 1926		1959 - 1939		1959 - 1926	1970 - 1926	1989 - 1926
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Ln population$_{t-1}$	-.16	-.13	-.11	-.11	-.23	-.28	-.29
	(.050)	(.047)	(.035)	(0.35)	(.062)	(.066)	(.073)
Ln urban population$_{t-1}$ inside 20 km radius	.031	-0.007	.044	-.034	-.028	.024	.022
	(.047)	(.044)	(.029)	(.030)	(.055)	(.060)	(.064)
Ln urban population$_{t-1}$ inside 100 km radius	.060	.021	.001	-.008	-.020	.007	-.005
	(.021)	(.025)	(.011)	(.013)	(.032)	(.036)	(.040)
Growth$_{t-1}$	.097	.086	.14	.13	.14	.17	.15
	(.070)	(.066)	(.027)	(.026)	(.082)	(.085)	(.090)
Oblast center dummy	.50	.45	.29	.29	.76	.86	1.10
	(.088)	(.092)	(.051)	(.052)	(.11)	(.16)	(.17)
War dummy				-.026	-.087	-.079	-.020
				(.032)	(.074)	(.086)	(.089)
GULAG camps in 20 km radius dummy		.22		.081	.33	.35	.39
		(.081)		(.041)	(.10)	(.11)	(.11)
GULAG camps in 100 km radius dummy		.11		-.001	.093	.049	.063
		(.059)		(.037)	(.075)	(.081)	(.089)
Ln (GULAG prisoners per capita in 20 km radius + 1)		.053		.024	.079	.083	.094
		(.059)		(.032)	(.072)	(.077)	(.087)
Ln (GULAG prisoners per capita in 100 km radius + 1)		-.12		-.007	-.14	-.13	-.16
		(.042)		(.037)	(.068)	(.068)	(.077)
Migration restrictions since 1959 dummy						.26	.12
						(.16)	(.17)
Expansion restrictions since 1959 dummy						.49	.51
						(.33)	(.37)
Geography controls	yes	yes	yes	yes	yes	yes	yes
N of obs	459	459	624	624	454	454	456
R^2	0.29	0.25	0.20	0.21	0.37	0.36	0.36

Temporary shocks. Case of Russia

Tatiana Mikhailova finds that **WWII does not have a statistically signifcant long-term effect on city growth, controlling for other factors, while GULAG system does**. The growth of an average city in 1960s exhibits partial mean-reversion after the shocks of 1930s-1950s.

The dynamics is consistent with multiple equilibria hypothesis: cities that received a lot of investment (as measured by the GULAG population) in the 1930s–1950s, have a higher chance not to revert to the previous trajectory, but to continue growing, while neglected cities are more likely to decline.

The division of Germany as a natural experiment

Stephen J. Redding and Daniel M. Sturm. *The Costs of Remoteness: Evidence from German Division and Reunification.* AER, 2008.

$$Popgrowth_{ct} = \beta Border_c + \gamma(Border_c \times Division_t) + d_t + \varepsilon_{ct}$$

- $Popgrowth_{ct}$ – the annualized rate of population growth over the periods 1919-25, 1925-33, 1933-39, 1950-60, 1960-70, 1970-80 and 1980-88 in West German city c at time t;
- $Border_c$ – a dummy which is equal to one when a city is a member of the treatment group of cities close to the East-West border (cities are classified as close to the East-West border if they were within 75 kilometers of this border);
- $Division_t$ – a dummy which is equal to one when Germany is divided;
- d_t – a full set of time dummies
- ε_{ct} – the error term.

The division of Germany as a natural experiment

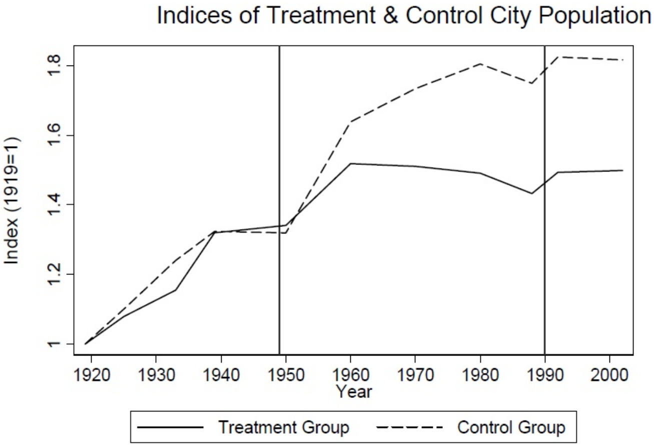
Dependent variable: Population Growth	**(1)**	**(2)**	**(4)**	**(5)**
Border × Division	-0,746*** (0,182)		-1,097*** (0,260)	-0,383 (0,252)
Border × Year 1950-60		-1,249*** (0,348)		
Border × Year 1960-70		-0,699** (0,283)		
Border × Year 1970-80		-0,640* (0,355)		
Border × Year 1980-88		-0,397*** (0,147)		
Border	0,129 (0,139)	0,129 (0,139)	0,233 (0,215)	-0,009 (0,148)
Year Effects	Yes	Yes	Yes	Yes
City Sample	All Cities	All Cities	Small Cities	Small Cities
Observations	833	833	420	413
R^2	0,21	0,21	0,23	0,30

The division of Germany as a natural experiment

Dependent variable: Population Growth	(3)
Border × Division 0-25 km	-0,702*** (0,257)
Border × Division 25-50 km	-0,783*** (0,189)
Border × Division 50-75 km	-0,620* (0,374)
Border × Division 75-100 km	0,399 (0,341)
Border 0-25км	-0,110 (0,185)
Border 25-50км	0,144 (0,170)
Border 50-75км	0,289 (0,272)
Border 75-100км	-0,299* (0,160)
Year Effects	Yes
City Sample	All Cities
Observations	833
R^2	0,21

The imposition of the East-West border

The figure graphs the evolution of total city population in the treatment group of cities along the East-West German border and the control group of other West German cities.
For each group, total population is expressed as an index relative to its 1919 value.
The two vertical lines indicate the year 1949 when the Federal Republic of Germany (West Germany) and the German Democratic Republic (East Germany) were established and the year 1990 when East and West Germany were reunified.



Concluding remarks

1. We examine different approaches to testing spatial structures stability.
2. The methods allow to get tractable results, but require historical data of high quality.
3. A weakness of approaches decribed above is that thay have no thorough theorethical foundation.

Thanks

Thank you for your attention!