

Some methodological issues concerning the empirical assessment of network effects

Lecture notes for the course of Economics and Policy of Networks
Prof. Fabio Sabatini
Sapienza University of Rome, Department of Economics and Law

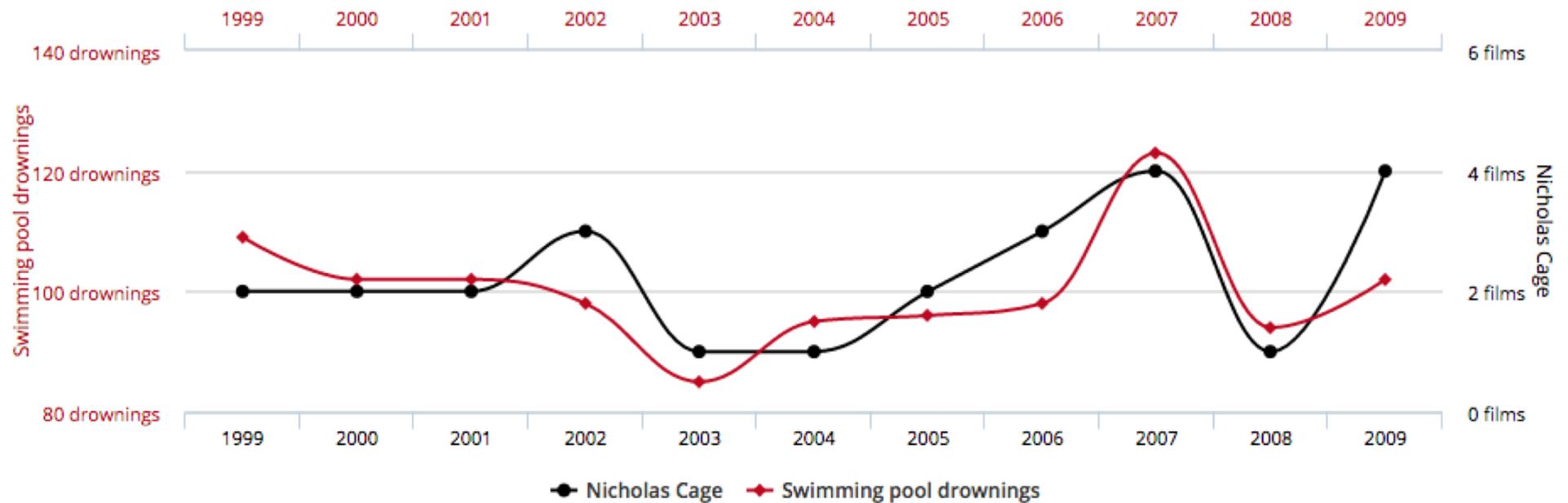


Correlation does not imply causation



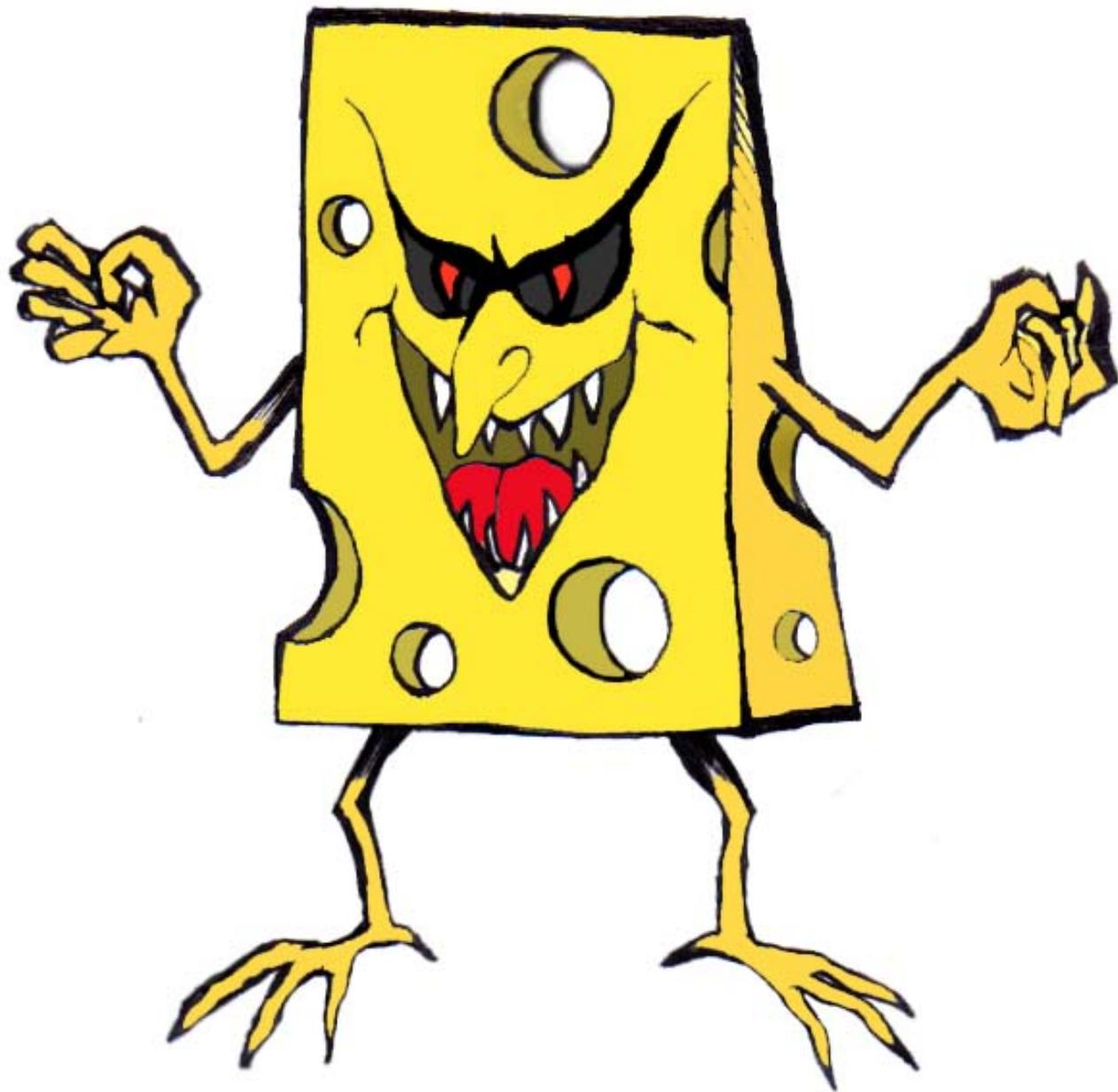
Number of people who drowned by falling into a pool correlates with Films Nicolas Cage appeared in

Correlation: 66.6% ($r=0.666004$)



Data sources: Centers for Disease Control & Prevention and Internet Movie Database

tylervigen.com

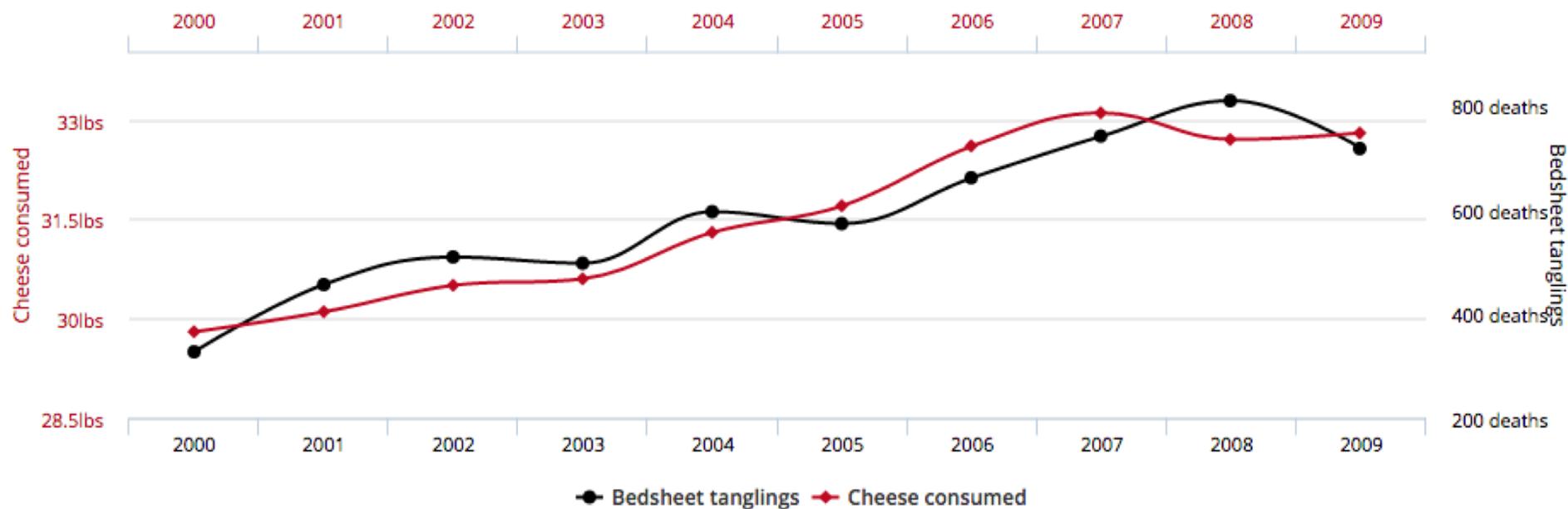


Per capita cheese consumption

correlates with

Number of people who died by becoming tangled in their bedsheets

Correlation: 94.71% ($r=0.947091$)

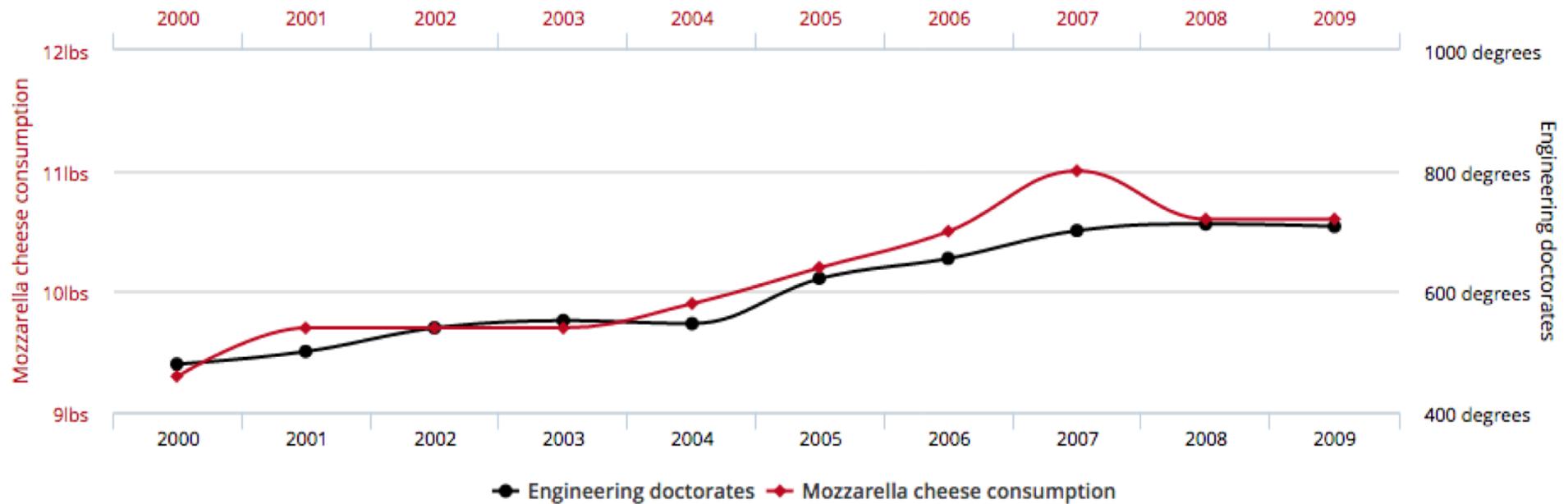


Data sources: U.S. Department of Agriculture and Centers for Disease Control & Prevention

tylervigen.com

Per capita consumption of mozzarella cheese correlates with Civil engineering doctorates awarded

Correlation: 95.86% (r=0.958648)



Data sources: U.S. Department of Agriculture and National Science Foundation

tylervigen.com

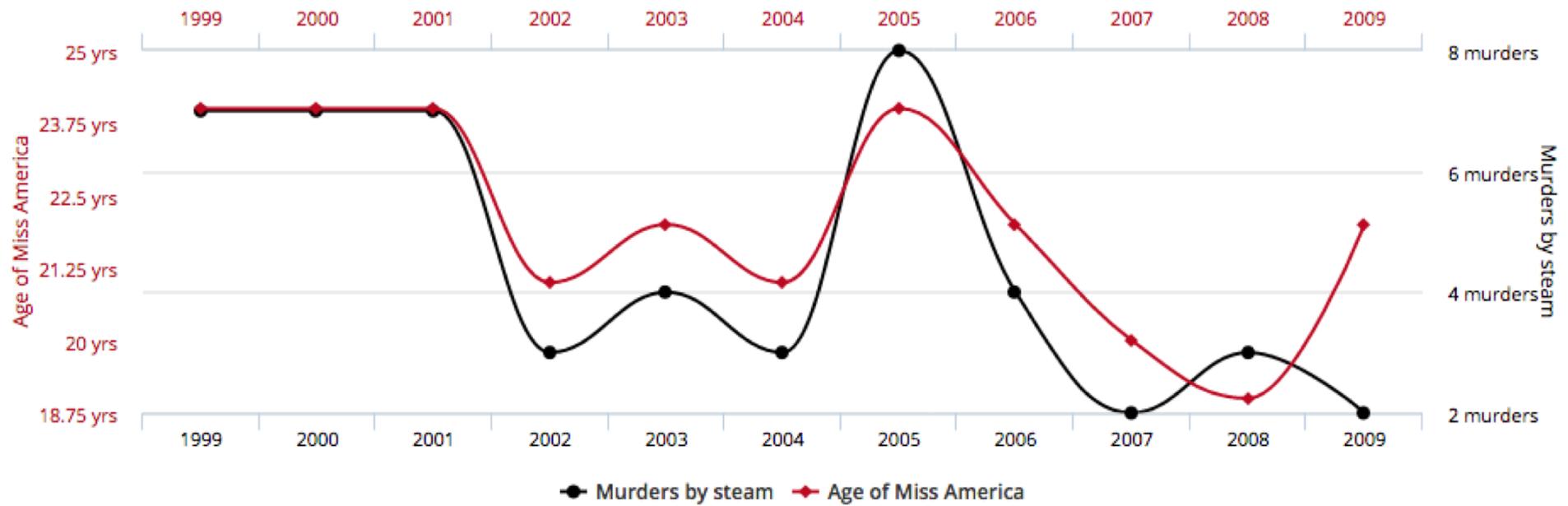


Age of Miss America

correlates with

Murders by steam, hot vapours and hot objects

Correlation: 87.01% (r=0.870127)



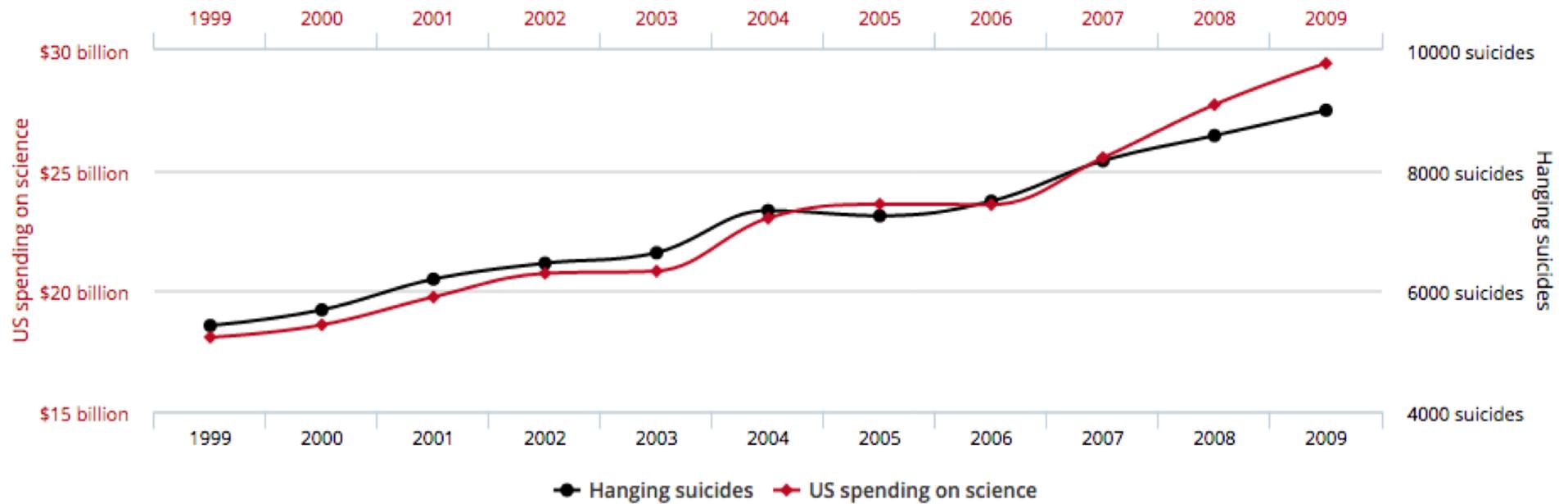
tylervigen.com

Data sources: Wikipedia and Centers for Disease Control & Prevention

US spending on science, space, and technology correlates with Suicides by hanging, strangulation and suffocation



Correlation: 99.79% (r=0.99789126)

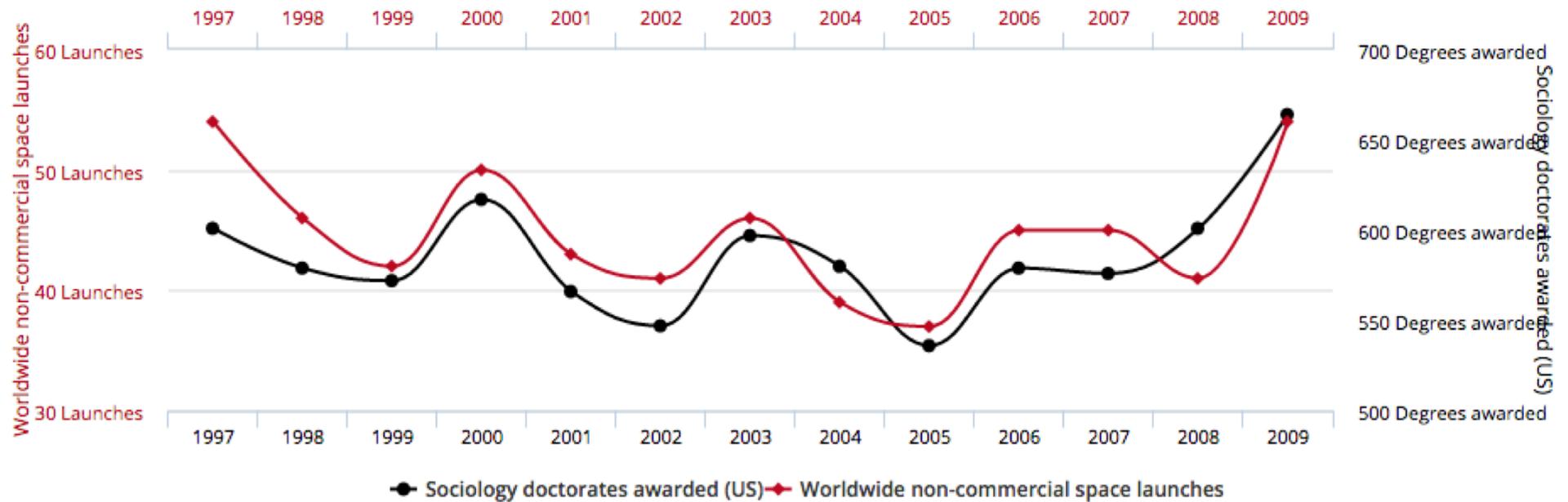


Data sources: U.S. Office of Management and Budget and Centers for Disease Control & Prevention

tylervigen.com

Worldwide non-commercial space launches correlates with Sociology doctorates awarded (US)

Correlation: 78.92% (r=0.78915)

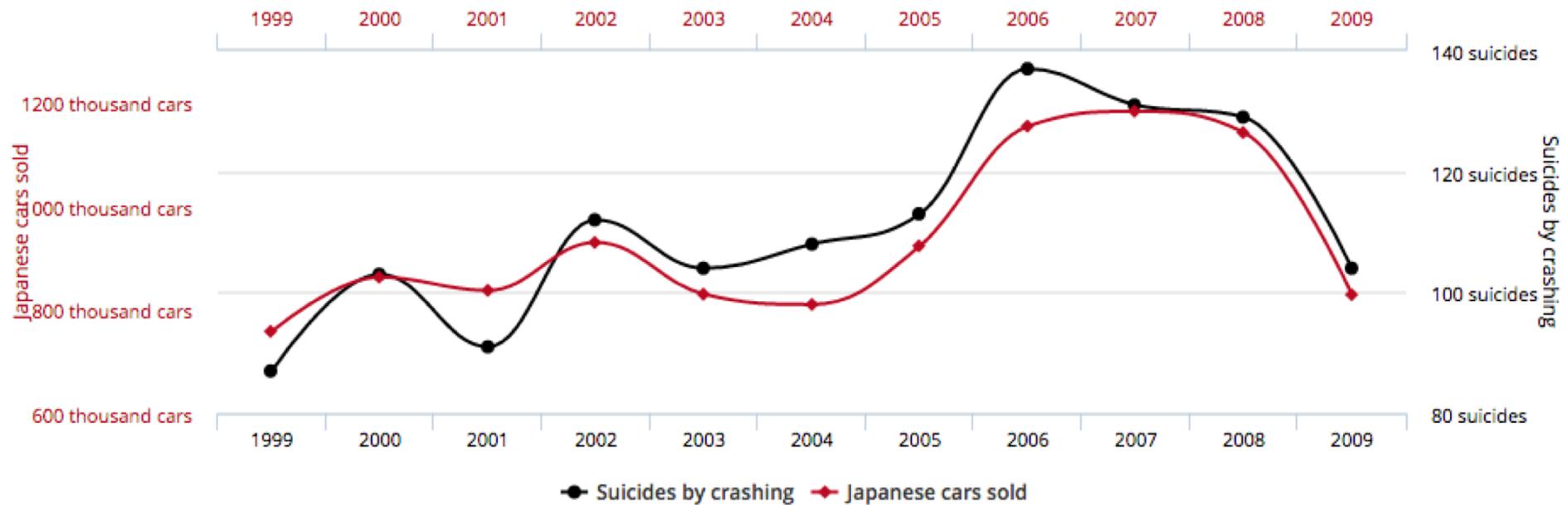


Data sources: Federal Aviation Administration and National Science Foundation

tylervigen.com

Japanese passenger cars sold in the US correlates with Suicides by crashing of motor vehicle

Correlation: 93.57% (r=0.935701)

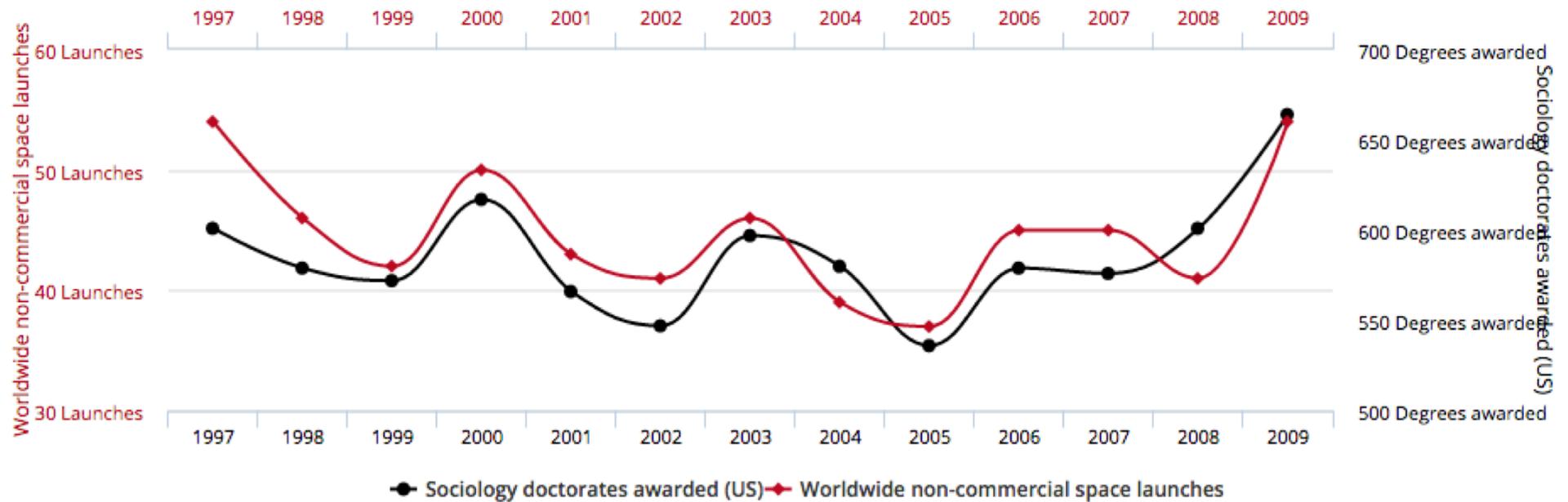


Data sources: U.S. Bureau of Transportation Statistics and Centers for Disease Control & Prevention

tylervigen.com

Worldwide non-commercial space launches correlates with Sociology doctorates awarded (US)

Correlation: 78.92% (r=0.78915)



Data sources: Federal Aviation Administration and National Science Foundation

tylervigen.com

Take 10 minutes of your time to see this very nice lecture on correlation vs. causation: <https://goo.gl/CDO1u0>

Correlation does not mean causation

- Identifying the effect of the current level of a variable (e.g. social networks, civicness, trust) on the current level of a hypothetical outcome (e.g. measures of the institutional or the economic performance) is challenging for several reasons:
 - Reverse causality
 - Common bias
 - Omitted variables
 - Measurement errors concerning the endogenous variable and the covariates.

Correlation does not mean causation

- When these factors hold, the explanatory (endogenous) variable is correlated with the error term in regressions.
- If this is the case, **the correlation between the endogenous and the dependent variable is likely to be spurious, instead of reflecting a causal relationship.**
- **As a result, estimates are biased and no general conclusion can be drawn from the empirical analysis.**

Instrumental variables: basics

- **Instrumental variables (IV) regression** is a general way to obtain a consistent estimator of the unknown coefficient of the population regression function when the regressor, X , is correlated with the error term, u .
- IV regressions are used to estimate causal relationships when controlled experiments are not feasible or when a treatment is not successfully delivered to every unit in a randomized experiment **to assess the causal relationship** between an explanatory variable X and a dependent variable y .
- **A valid instrument induces changes in the explanatory variable but has no independent effect on the dependent variable**, allowing a researcher to uncover the causal effect of the explanatory variable on the dependent variable.

Instrumental variables: basics

To understand how IV regression works, think of the variation of X (e.g. *social capital*) as having 2 parts

$$Y_i = \beta_0 + \beta_1 X_i + u_i;$$
$$i = 1, \dots, n$$

1) One part that, for whatever reason, is correlated with the disturbance term u . This is the part that causes the problems.

2) A second part that is uncorrelated with u .

If you had information allowing you to isolate the second part, then you could focus on those variation in X that are uncorrelated with u and disregard the variations in X that bias the OLS estimates.

Instrumental variables: basics

- With IV regressions, the information about the movement in X that is uncorrelated with u is obtained from one or more additional variables called “instrumental variables” or “instruments”.
- In other words, IV regressions use these variables as “instruments” to isolate the movements in X that are uncorrelated with u .
- This allows to use those movements **to consistently estimate the correlation between X and y** in the same dataset.
- The consistency of the estimates (**i.e. the fact they are not biased by the endogeneity issues summarized before**) allows to argue about a possible causal effect connecting the y to the X .

Instrumental variables: basics

- Whatever the source of the correlation between X and u , if there is a valid instrumental variable, Z , then **the effect on Y of a unit change in X can be estimated** using the instrumental variable estimator.
- Consider the regression model:

$$Y_i = \beta_0 + \beta_1 X_i + u_i;$$
$$i = 1, \dots, n$$

- IV estimation uses an additional instrumental variable, Z , to isolate that part of X that is uncorrelated with u .

Instrumental variables: basics

A valid IV must satisfy two conditions:

$$Y_i = \beta_0 + \beta_1 X_i + u_i;$$
$$i = 1, \dots, n$$

1) **Relevance:** $\text{corr}(Z_i, X_i) \neq 0$

The instrument Z_i must be correlated with the endogenous variable, and you **have a theory to justify that correlation**.

2) **Orthogonality:** $\text{corr}(Z_i, Y_i) = 0$

The instrument Z_i must be **exogenous** and uncorrelated with the dependent variable. If this is the case, **then the part of the variation in X_i captured by Z_i is in turn exogenous**, and this exogenous variation can in turn be used to estimate the coefficient β_1 .

Instrumental variables: basics

If Z_i satisfies the two conditions for a valid instrument, then the coefficient β_1 can be estimated through a **Two Stages Least Squares (TSLS)** estimator.

The first stage

$$X_i = \pi_0 + \pi_1 Z_i + v_i$$

decomposes X into 2 components:

1) A problematic component, v_i , that may be correlated with the error.

2) A problem-free component $\pi_0 + \pi_1 Z_i$:

Because Z_i is exogenous, **this component of X_i is uncorrelated with u_i , and may be used to estimate β_1 .**

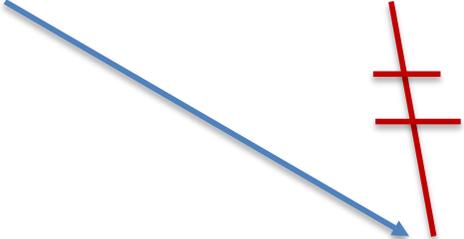
The second stage regresses Y_i on the predicted value of X_i obtained from the first stage, $\hat{X}_i = \hat{\pi}_0 + \hat{\pi}_1 Z_i$

The resulting estimators from the 2nd stage regression are the TSLS estimators,

$$\hat{\beta}_0^{TSLS}, \hat{\beta}_1^{TSLS}$$

$$X_i = \pi_0 + \pi_1 Z_i + v_i$$

First stage


$$Y_i = \beta_0 + \beta_1 X_i + u_i;$$
$$i = 1, \dots, n$$

Second stage

An example: class size → test scores

- Suppose you want to test the effect of class size on educational achievements (e.g. test scores).
- What effect do you expect to find?
- Can you imagine other factors that may influence test scores apart from class size?
- Would the estimation of class size's effect be reliable if you do not account for such other factors?

An example: class size → test scores

- Even if you have a rich dataset, there will always be unmeasured variables such as learning opportunities outside school, the family background, and the quality of teachers that you will not be able to account for in the model.
- These variables can bias:
 - The dependent variable only
 - Both the dependent and the endogenous variable at the same time.
- If data about these variables are not available, then the bias cannot be addressed by including these omitted variables in regressions.

An example: class size → test scores

- Now, consider the following hypothetical example: some schools are forced to close for repair because of a summer earthquake (a minor one, causing some damage but no injuries or psychological shocks).
- Districts close to the epicentre will have more damaged schools, so they will need to increase class size.
- Therefore, the distance from the epicentre – or other measures of the intensity with which the shakes were felt – satisfies the relevance condition, as it certainly affects class size.
- If distance from the epicentre is unrelated to any other factor that may affect students' performance, then it will be exogenous, because it is uncorrelated with the error term.
- **Thus the distance from the epicentre can be used as instrumental variable to circumvent the omitted variables bias and to estimate the effect of class size on test scores.**

Instrumental variables: basics

- When there are multiple endogenous regressors, X_{1j}, \dots, X_{kj} , the TSLS algorithm is similar, except that **each endogenous regressor requires its own first stage regression**.
- In the first stage, the dependent variable is one of the X s, and the regressors are:
 - All the instruments Z s
 - All the exogenous variables included in the structural equation (i.e. the second stage).
- In the second stage, the X s are replaced by their respective predicted values.

Instrumental variables: basics

- When there is one endogenous variable but multiple instruments, the condition for instrument relevance is that **at least one Z is useful for predicting X** , given W .
- **Instruments that explain little in the variation in X are called *weak instruments*.**
- This is a problem, because the final estimates are likely to be biased and TSLS is no longer reliable.

Instrumental variables: basics

- How relevant must the instruments be?
- We will neither deal with the technical aspects of the answer nor explain why a weak instrument is a problem but...
- Just present a simple and operational rule of thumb: you do not have to worry if the first stage F-statistics exceeds 10 (see for example Stock and Watson, 2010).

History as a source of instruments

History as a source of instruments

- Returning to Putnam's work:
- In much of the literature that studies the effect of social capital or social networks on the economic performance, separating the effect of social capital from that of formal institutions is a key problem.
- The cultural values that promote cooperation and exchange and pro-market institutions are complementary, implying that countries with strong values and high trust also choose institutions that support these values making them attractive to the population.
- Hence, in cross-country or cross-regional estimates it is hard to tell apart the effect of social capital on income per capita (or growth) from that of institutions.

History as a source of instruments

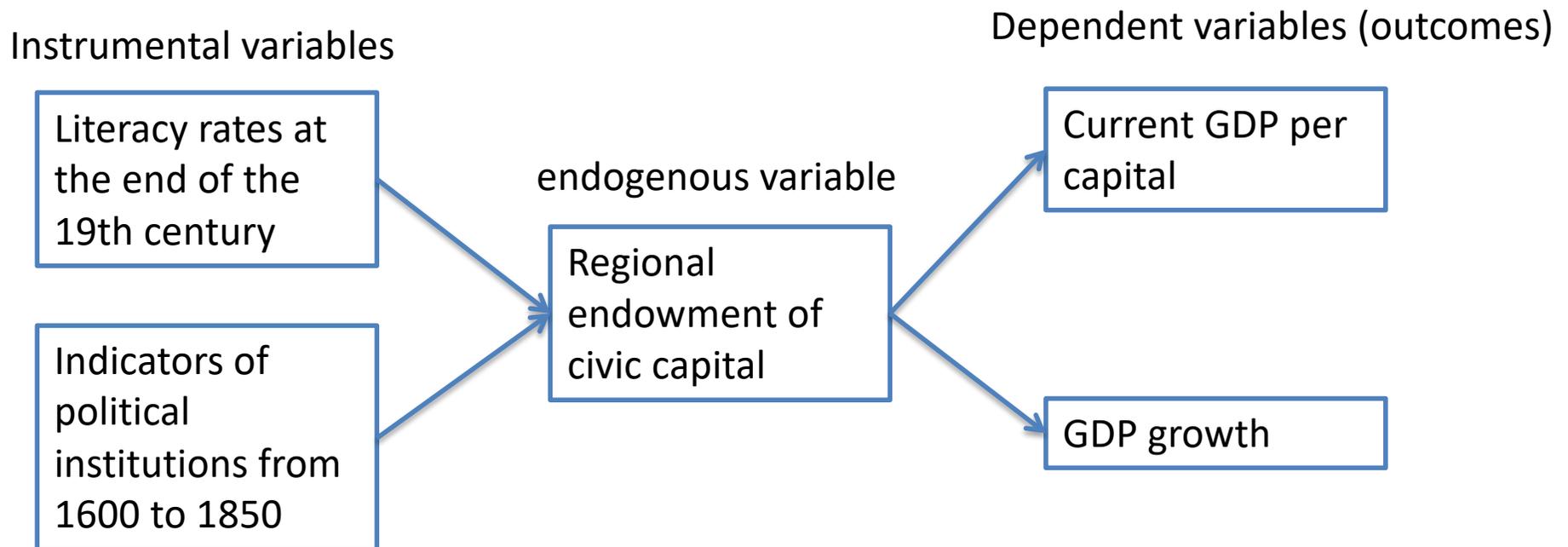
- **Long-term historical episodes can act as a random source of network formation and civic capital accumulation** and, if “properly isolated”, they can be good candidates for acting as **instruments for today norms and beliefs shared by a community**, or for today propensity to form networks such as voluntary associations or networks of enterprises.
- In fact, since culture is transmitted slowly from one generation to the next, distant but relevant historical episodes can have predictive power on today’s norms and beliefs.

History as a source of instruments

- For example, Tabellini (2010) shows that differences across regions of Europe in the current endowment of civic capital can be explained with differences in long-term history, such as:
 - The literacy rates that prevailed at the end of the 19th century;
 - Indicators of political institutions in the period from 1600 to 1850.

History as a source of instruments

- Using these measures as instruments Tabellini (2010) finds that regions with higher endowment of civic capital have higher GDP per capita today and have experienced faster GDP growth.



History as a source of instruments

- Since Tabellini (2010) uses regional variation and these regions are part of countries with common institutional design, he can exclude that civic capital captures the effect of formal institutions as they are absorbed by country fixed effects.
- Furthermore, controls for current levels of education and for the historical level of economic development suggest that civic capital is unlikely to reflect persistent differences in human capital and in productivity.
- **What is key for identification is that the historical instruments don't have a direct effect on today's output but affect the latter only because they affected the cultural traits of these populations centuries ago which are then reflected – though intergenerational transmission – in today culture.**

History as a source of instruments

- Another example: Guiso, Sapienza and Zingales (2016) rely on a similar strategy to identify the effect of civic capital on average per capita income.
- After having shown that past history of communal independence explains differences across cities in the North of Italy, they use this variation to identify the effect of civic capital on GDP per capita in year 2001, instrumenting today civic capital with past history of independence.
- They find that differences in civic capital can explain a good fraction of the differences in income per capita across towns in the North of Italy.

History as a source of instruments

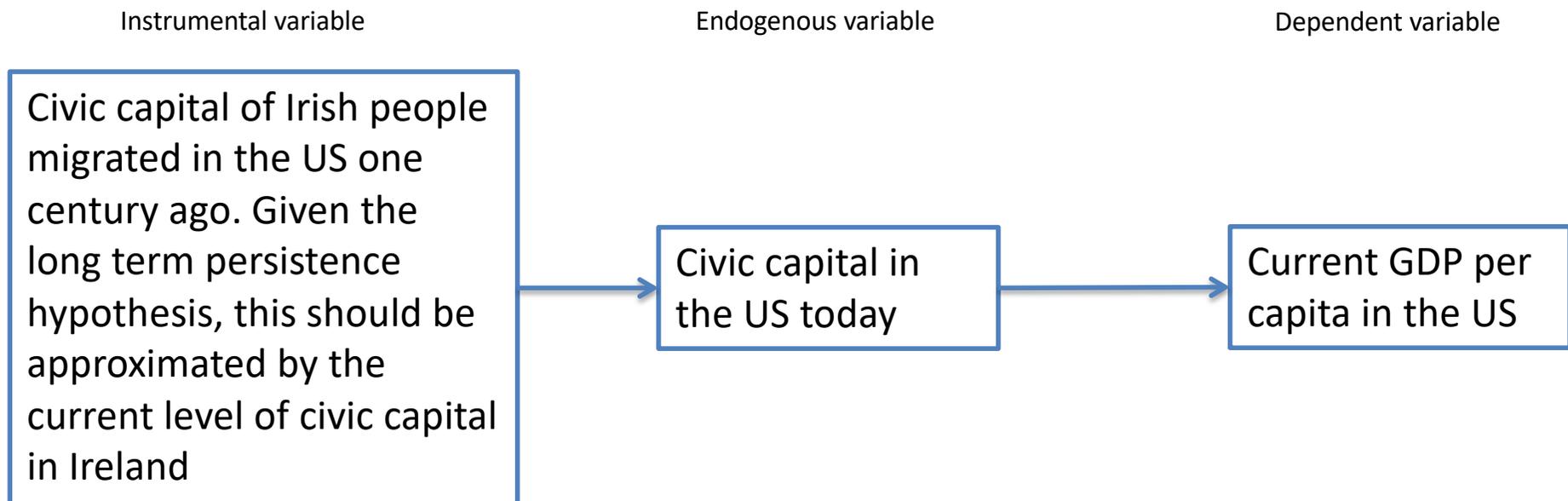
- There is however a more serious problem with this approach that invests the validity of the “exclusion restriction” for the instrument(s) for civic capital.
- **For the instruments to be valid it must be that the historical episodes that built up civic capital did not at the same time foster the accumulation of other forms of capital that have lasted up to today and still exert a direct effect on income.**
- For instance, in the GSZ (2016) context, having been a free city in the 13th century may have resulted in accumulated assets of some sort that still directly affect income today, in addition to indirectly affecting it because of its boost on civic capital.
- The general point is that historical shocks to civic capital could have also shocked other types of capital as persistent as civic capital and which may have an independent, direct effect on income.

History as a source of instruments

- **The only way to account for this is to obtain direct measures of these assets and try to control for them.**
- In addition, movers and “cultural portability” can be used as identification strategy.
- Once ingrained in the brain of individuals norms and beliefs tend to move with them and continue to affect their actions when people locate in a new environment, where different norms and beliefs prevail.
- On the other hand institutions are not portable.

History as a source of instruments

- Therefore focusing on movers' and using information on the prevailing norms and beliefs in their country of origin one can separate the effect of civic capital from that of institutions.
- This is called the “epidemiological approach”. An Example:



- But this is much more complicated and we will come back to it at a later stage of the course.

Natural experiments as a source of instruments

- Large shocks to the benefits of cooperation can induce a change in the norms and beliefs that support cooperative behavior.
- Once again, history can provide some natural experiments in this sense.
- Other possible sources of natural experiments are, for example, geography and natural disasters.

Natural experiment: the formation of communes in Italy

- One such a shock is represented by the collapse of the Holy Roman Empire at the beginning of the second millennium.
- As the opportunities for trade expanded, the North and South of Italy were subjected to two very different “**treatments**”.
- While the South was governed by an efficient and autocratic monarchy (the Norman Kings), the North was left in power vacuum.
- In some northern cities, the response to the lack of government was the formation of small groups of individuals who agreed with an informal pact to provide mutual help and collaborate to solve problems of common interest.
- Slowly, more stable institutions started to emerge from these agreements. In the mid-twelfth century a new word came into use to describe them: “commune”. The word commune is a synonym for republic (*res publica*, i.e. common property) and is used with this meaning.
- This sense of responsibility for the common good that citizens of independent towns developed and consolidated over two centuries of self-government is the “civicness” Putnam refers to and the limits to the power of the executive that Acemoglu and Johnson (2005) deem necessary for development.

Natural experiment: the formation of communes in Italy

- Guiso, Sapienza and Zingales (2016, but the work started in 2008) more in depth analyzed how past history relates to civic capital within the North.
- Not all major cities located in the North at the turn of the first millennium actually became free cities: some did not and either remained under the control of the emperor (at least for a while) or fell under the control of one of feudal lords that survived the communal experience.

The figure shows the map of Italy at around year 1167.

The red line marks the borders of the country that where the Holy Roman Empire of Germany.

All the towns marked with a **full dot** were commune.

Towns in red were commune that belonged to the Lombard League.

Towns in blue were allied to the Emperor.

The green areas mark the territories of various Principati and Feudi.

The Southern part of Italy not belonging to the Empire was under the Norman Kingdom of Sicily



A natural experiment: the formation of communes in Italy

- GSZ found that Center-Northern cities that became free cities have significantly higher levels of civic capital today.
- For example, the number of voluntary associations is 25% higher in cities that were free city-states, consistent with Putnam conjecture.
- The correlation, however, may be spurious: Past history may be a proxy for some unobservable characteristics that affect both the chances a city became independent in the middle ages and the level of civic capital today.

A natural experiment: the formation of communes in Italy

- To address causality GSZ found two instruments that, in their view, affect the cost of becoming independent at that time, but are unlikely to affect the level of civic capital today:
 - 1) Whether a city was the seat of a Bishop . This variables captures the variation in the cost of coordination, since it is documented (Tabacco, 1987) that the presence of a bishop facilitated the necessary coordination of the prominent local families to provide the public goods and made it easier to transform a city into an independent commune.
 - 2) Whether the city had been founded by the Etruscans. a proxy for how easy a city was to defend. Since the Etruscans, a pre Roman civilization, was organized as free city states, they chose to locate their cities in positions that were easy to defend.

A natural experiment: the formation of communes in Italy

- Using these two instruments GSZ are able to confirm that cities that became a commune have more civic capital today.
- Furthermore, since the affirmation of the Norman Kingdom in the South, prevented the formation of free city-states in the South they can then test the validity of their instruments, by looking at their effect in the South, where free city-states did not occur.
- That these instruments have **no effect in the South** suggests as GSZ find is evidence of the validity of the exclusion restriction, lending strong support to Putnam conjecture.

A natural experiment: slave trade

- Nunn and Wantchekon (2009) provide a historical natural experiment of how civic capital can be destroyed. They focus on slave trade to explain mistrust within Africa.
- They argue that today's level of trust among different African ethnicities is the reflection of the past exposure to the risk of being captured and sold as slave in the 18th and early 19th century.
- Because of the high payoff of selling people to slave traders, indigenous sold even people of their same ethnic group, close friends, and relatives. This created a climate of suspicion that may have resulted in an evolution of mistrust towards others and towards local leaders.

A natural experiment: slave trade

- To assess the effects of this historical experience Nunn and Wantchekon (2009) used data from the 2005 wave of Afrobarometer.
- They found that Africans whose ancestors faced a higher chance of being captured and sold as slave trust their relatives, neighbors, and local council less today.
- This conclusion is further strengthened by instrumenting the intensity of slave trade with the distance from the coast.

Geography-related natural experiments

- Another source of “natural” shocks to the benefit of cooperation is provided the geographical environments.
- The efficient exploitation of certain natural resources can only be achieved if several people, possibly a whole community, are willing to cooperate.
- For example, in mountainous areas where the main crop is slow-growing trees it is impossible to support a fragmented land ownership without a very high degree of cooperation, since farmers need to take turns in cutting their trees and pool and divide the proceeds.
- As Ostrom (1990) shows, this solution requires a considerable amount of cooperation and mutual trust. This experience of cooperation and trust, repeated over centuries, can increase the level of civic capital.
- Climate variability can also determine higher levels of trust. This correlation arises because climate variability generates a higher need for insurance, which can only be delivered if there is enough cooperation.

The limits of Putnam's work

- The methodological issues addressed so far allow us to highlight several weaknesses of Putnam's work on Italy.
 - The identification strategy does not fully account for the endogeneity of civicness. No two stages techniques are employed.
 - Confounding factors and omitted variables bias are definitely underestimated.
 - The choice of indicators is often questionable.
 - Indicators are often aggregated in a questionable way (e.g. linear combinations → loss of information, reduced reliability).

Summarizing: How Putnam et al. (1993) oriented the contemporary debate

- Despite those limitations, Putnam's work is pioneer:
 - Because, by predicting the current values of civiness through the historical values of civiness in the estimation of institutional and economic outcomes, they anticipated the approach of using TSLs where civiness is instrumented by variables picked in the remote past. This idea inspired much of the following literature on the topic, which in turn is very influential in the current political economy debate.
 - More in general, they were the first to systematically test the hypothesis that social structures – networks, norms, institutions – matter in economic outcomes.
 - They operationalized the concept of social capital offering concrete, easy-to-build, and theory-grounded measures of it, based on concepts such as norms and networks. Putnam's indicators have then been used in a gargantuan amount of later studies.
 - In doing so, they built a persistent bridge across social sciences.

Summarizing: How Putnam et al. (1993) oriented the contemporary debate

- “These unexpected, elemental links between civics and economics cast new light on the long-standing debate about the North-South development gap, not only within Italy but also globally.”
- Civic networks were robustly found to be a very good predictor of economic success and development – even though not so robustly nor globally (just in Italy): there was evident room for the intervention of economists.
- You can easily understand how the patterns discovered by Putnam et al. (1993) were mesmerizing for economists.

Suggestions for an empirical exercise

- Find indicators of networks and other related concepts we addressed in the lectures so far – trust, civicness, civic capital, social capital, etc. – in a chosen dataset.
- Examples: World Values Survey, Eurobarometer survey, British Household Panel, Survey of Household Income and Wealth, Istat Multipurpose Household Survey, Italian National Election Studies, etc. Possibly try to build latent, synthetic, indicators.
- Build local, macro or “meso” (i.e. local, regional-level or province-level) indicators.
- Empirically assess the correlation between those indicators and a range of hypothetical outcomes (e.g. GDP per capita, GDP growth, measures of well-being, health, etc.).
- In case you have cross-sections, try to find appropriate instruments and eventually perform TSLS.
- Of course your instructor will duly assist you in each stage of such an exercise, from data retrieval, to the building of indicators and performing the estimates!

References

- Guiso, L., Sapienza, P., Zingales, L. (2010). [Civic capital as the missing link](#). NBER Working paper 15845. Free download here: <https://goo.gl/IMHt8s>.
- Guiso, L., Sapienza, P., Zingales, L. (2016). [Long term persistence](#). *Journal of the European Economic Association* 14(6): 1401–1436. Free download here: <https://goo.gl/x6ChvB>.
- Nunn N, Wantchekon L. (2011). [The Slave Trade and the Origins of Mistrust in Africa](#). *American Economic Review* 101 (7): 3221-3252. Free download here: <https://goo.gl/CWXcwz>.
- Putnam, R. D., Leonardi, R., Nanetti, R. Y. (1993). *Making Democracy Work: Civic Traditions in Modern Italy*. Princeton: Princeton University Press. Available in the Department's library (a Pdf copy for personal use can be provided by the professor by request).
- Tabellini, G. (2010). [Culture and institutions: Economic Development in the Regions of Europe](#). *Journal of the European Economic Association* 8(4), 677–716. Free download here: <https://goo.gl/d0Y1tR>.

Examples of possible exam questions

- Which is, according to Putnam, the relationship between civic engagement in the 1900, socio-economic development in the 1900, the current levels of these two variables, and the current institutional performance?
- How did Putnam measure civic engagement in the late nineteenth century?

Highlights

- ...

- ...

- ...

- ...

- ...

Thank you!