ABSTRACT

Title of Dissertation:	APPLIED ECONOMICS ESSAYS
	Hao Bo, Doctor of Philosophy, 2020
Dissertation directed by:	Professor Sebastian Galiani, Department of Economics

This dissertation uses mathematical economic models and advanced statistical and econometric tools to study knowledge markets, external validity, and institutional changes. In Chapter 1, I focus on knowledge markets exploring how network relationships between knowledge consumers impact the equilibrium number of opinion leaders. Both a theoretical model and empirical analysis show that there'll be more opinion leaders in a knowledge market if the most active knowledge consumers occupy more central positions in a social network connecting consumers. This is the first work to formally quantify opinion leaders, knowledge markets, and consumer attention. While the existing literature emphasizes the role of opinion providers' network positions on the making of opinion leaders, this work shows that the network positions of active consumers also matter because active consumers serve as a propaganda machine. In Chapter 2, Professor Sebastian Galiani and I provide a formal, general exploration of the question of external validity and propose a simple and generally applicable method for evaluating the external validity of randomized

controlled trials. This is important. Once researchers have conducted an internally valid analysis, that analysis yields an established set of findings for the specific case in question. As for the future usefulness of that result, what matters is its degree of external validity. In Chapter 3, I theoretically argue that people weigh specialization gains against trade costs when they decide whether to specialize and trade or self-produce all goods by themselves, and thus more people participate in trade under better institutions. I show that the better the institution of an economy's trade partner, the more prosperous the economy is, thanks to expanded trade. Moreover, when more people trade, more people would like to fight for a better institution and may induce institutional improvement. Better initial institutions or lower trade costs facilitate institutional improvements; but with very high initial institutional quality, people may lose their incentive to protest. I also provide historical evidence consistent with the theory.

APPLIED ECONOMICS ESSAYS

Hao Bo

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2020

Advisory Committee:
Professor Sebastian Galiani, Chair
Professor Peter Murrell
Professor John Wallis
Professor Ingmar Prucha
Professor Susane Parker

Table of Contents

Chapter 1: Socially Embedded Knowledge Networks and the Making of Opinion Leaders: Evidence from Twitter

1. Introduction and related literature

The importance of knowledge markets is acknowledged in economics (Coase and Wang, 2012), but research on knowledge markets is scarce compared with work on commodity, service, or factor markets. Probably, the most important economic insight on knowledge markets is still from Hayek's 1945 paper "The Use of Knowledge in Society." Hayek (1945) argues that the decentralized use of knowledge in a knowledge market is essential to the workings of a society and no centralized mechanism can substitute. In that paper, the underlying network structure of the knowledge market is assumed symmetric and regular: each agent is both a knowledge supplier and a knowledge consumer, and their interactions are homogenous, that is, well-stirred. This assumption helps Hayek and generations of his readers to focus on his specific insights, but in the social networks underlying real knowledge markets, it is often the case that there are "a handful of very influential celebrities on one side and many millions of people with just a couple of followers on the other" (Bruner, 2013).

A common theme in knowledge markets is self-organized concentration of to whom knowledge consumers pay attention, which arises from a decentralization process where social influence between consumers plays an important role (Salganik, Dodd, and Watts, 2006). To whom people pay attention is an extensively studied topic (Jackson, 2019; Katz and Lazarsfeld, 1955). But to the best of our knowledge, there has been no formal treatment of the question of how the number of opinion leaders in a topic (in a knowledge market) is determined, apart from the influence of the total amount of attention people are willing to pay to the opinion leaders. In this paper we show that the structure of knowledge-consumer interactions has a significant impact on the number of opinion producers.

This is an important step in understanding the development of knowledge markets for two reasons. First, more opinion leaders means more competition in a knowledge market, and more competition usually implies higher information quality and more responsiveness to consumer welfare. Second, more opinion leaders can also indicate more diversity among widely heard opinions, and such diversity keeps non-mainstream opinions alive, enhancing a society's adaptive efficiency (a concept emphasized by North (1990) and North, Wallis, and Weingast (2009)).

In our paper we emphasize the role played by specific social embeddedness of economic agents, since another motivation of this paper is to understand the implications for knowledge market structure of the network effects via pre-existing relationships between consumers.

Network effects, especially for information products and technologies, are well documented in the field of industrial organization (Belleflamme and Peitz, 2015). For our purpose, we define network effects as existing when a consumer of a product is connected with other consumers in a network and the consumption activities of his connected consumers increase his consumption activity. This is a kind of social influence. Note our definition is slightly different from the usual definition of network effects, as given by Belleflamme and Peitz (2015): "a product is said to exhibit network effects if each user's utility is increasing in the number of other users of that product or of products compatible with it." We think it's usually the case that not all other users' consumption would have an impact on a user but rather his connected consumers would. Besides, the number of other users is only one part of other users' consumption activities and the frequency or amount of others' consumption should also matter. Consistent with Belleflamme and Peitz's (2015) definition, the current literature on network effects mainly focuses on well-stirred interactions: it is usually implicitly assumed that the probability of certain consumers to interact and influence each other depends only on the proportion of each type of consumer. To put it in another way, the existing literature mainly focuses on complete networks: everybody is interacting or adjacent to everybody. Such an assumption of symmetric social distribution of economic agents usually makes things tractable and thus facilitates many models yielding important and interesting economic insights. (See Belleflamme and Peitz (2015) for various examples.) In contrast, in this paper we obtain insights from networked social influence, i.e., consumers influence each other based on where they are embedded in a network connecting consumers, rather than in a homogenous way; this is why we have "socially embedded knowledge network" in the title. As will be seen, pre-existing relationships between consumers have an impact on the number of suppliers. Thus, in our paper we emphasize the role played by specific social embeddedness of economic agents, the importance of which is emphasized by Grannovetter (2017) and Thurner, Hanel, and Klimek, (2018).

In emphasizing social embeddedness, we put network positions foremost in our story. There is a large amount of literature on various implications of agents' network positions as surveyed by Easley and Kleinberg (2010), Jackson (2019), and Newman (2018). Major studies related to ours are on the relationship between the social network position of a knowledge supplier and the supplier's influence (Grannovetter, 1973) and on the concentration of attention to a few opinion leaders (Barabási and Albert, 1999). Variations under these two themes are surveyed by Easley and Kleinberg (2010), Page (2018), and Newman (2018). The importance of social networks or social embeddedness (Grannovetter, 1985 and 2017) is a common theme. However, formal studies that aggregate individual network positions to understand macro-level implications are few. Engle, Macy, and Claxton (2010) provide one example: so is our paper, given its third motivation to connect micro-level structure to macro-level performance (Schelling, 2006).

Having talked about motivation and existing literature, let's now intuitively summarize the theoretical insights and then talk about what we do and find in the empirical analysis.

Heterogeneous consumers are connected with each other in a social network. If the most active knowledge consumers, who have high ability to impact others through network effects due to their intensive consumption behavior, occupy more central positions in the network, there'll be more concentration of consumer attention to some opinion producers. This is because when more enthusiastic consumers are located in the central positions in the network, they'll have more influence on which opinion producers other consumers would pay attention to.[1] Opinion producers are assumed uniformly distributed everywhere and can become opinion leaders with enough followers. Thus, there is a concentration process that leads to more opinion leaders, in contrast to uniformly distributed attention that is not sufficient anywhere to give rise to opinion leaders. Another theoretical implication of this concentration process is that opinion leaders' similarity in terms of who follow them increases when enthusiastic consumers occupy more central positions in the social network. Such a concentration phenomenon also appears in Salganik, Dodd, and Watts (2006) where consumers, under social influence, tend to concentrate their song downloads on a few music providers.

To make matters clearer, let's consider an extreme case. If consumers have no impact on adjacent consumers, consumers follow unrelated sets of opinion producers. So, let's assume each consumer follows a unique opinion producer. Then no opinion producer would be an opinion leader, each having only a very small number of followers. In contrast, when there is social influence and therefore people concentrate all their attention on a few opinion producers, some opinion producers become opinion leaders.

Our formal theory inspires and is supported by our empirical analysis using a Twitter network and related tweets. We view a single knowledge market as consisting of the consumers and opinion producers within a community (a sub-network relatively

[1] The central position is central because it is contagious and thus generates influence to all other positions in all directions. In contrast, periphery positions are only contagious to a few other positions in a few directions.

isolated from other parts of the Twitter network) who focus on a specific topic. To identify topics, we estimate a topic model and find 45 topics in tweets of intensely followed Twitter users. To identify separate communities, we use network community detection and find four nearly isolated Twitter communities. Each combination of topic and community constitutes a knowledge market for which we formally define variables including the number of opinion leaders (the main dependent variable) and the network centrality of the most active knowledge consumers (the main explanatory variable). Each knowledge market is a sample point for regression analysis. Controlling for topic fixed effects, community fixed effects, and the total amount of consumer attention to a topic in a community, we find that when enthusiastic followers occupy more central network positions, the number of opinion leaders increases and opinion leaders' similarity in terms of who follow them increases. A variety of robustness tests supports this finding.

We use data from Hodas and Lerman (2014) a user network in Twitter and the URLs that appear in each user's tweets. We use a state-of-art topic model for text analysis over tweets, the correlated topic model (Blei and Lafferty, 2006; Roberts, Stewart, and Tingley, 2014). Gentzkow, Kelly, and Taddy (2017) provide a good survey of the economic analysis using text as data. In order to yield variables used for regression analysis, these topic modeling outcomes are combined with the network analysis explained in what follows. A sub-field of network science, community detection, focuses on dividing a network into communities within which nodes are densely connected and between which nodes are sparsely connected (Thurner, Hanel, and Klimek, 2018; Newman, 2018). We use a state-of-the-art network community detection algorithm for very large networks (Clauset, Newman, and Moore, 2004; Csardi, 2019), and divide the whole Twitter network into communities. Using the network communities that are detected and the topic modeling outcomes, we define variables in our regression analysis that are topic-community specific; an observation or a sample point in our regression analysis corresponds to a topic in a community.

To address endogeneity concerns, we use a friendship network as an instrument for the whole network. Our main explanatory variables, the network position centrality of enthusiastic consumers, are constructed based on the whole Twitter network. For a topic, this centrality measure is a function of network structure, which might be affected by some topic-community specific characteristic that also affects the number of and the similarity between opinion leaders. To address this endogeneity concern, we construct the network position centrality of enthusiastic consumers based only on a friendship network.[2] Under the assumption that the friendship network is subject to endogeneity concerns to a lesser degree than the original network, the outcome based on the friendship network would be significantly different from that based on the whole network only if endogeneity were really a serious concern. The contrapositive of this logic is this: under the assumption that the friendship network is subject to endogeneity concerns to a lesser degree than the original network,[3] if the outcome based on friendship network is not significantly different from that based on the whole network, endogeneity is not a concern. Wald tests fail to reject the equivalence of estimates based on both networks. Endogeniety is not a problem.

This paper makes the following contributions. Compared with most of the literature on network effects, the network effects explored in this paper are non-homogenously networked rather than well-stirred. This is a feature of both our theoretical and empirical work. Moreover, understanding knowledge markets, especially formally understanding the sources of opinion diversity, is a less explored field. The key variables both in our theory and in our empirical part are dealt with intuitively in the existing literature. But to the best of our knowledge, ours is the first paper to formally define opinion leaders, knowledge consumers, and amount of attention. Last but not least, as summarized in section 5, the paper makes some methodological contributions by constructing a new type of IV based on a multi-layer network

[2] We get the friendship network by dropping all the directed links in the whole network that don't have a reciprocal (this turns out to be a large change of the network, see section 3 for detail).

[3] The intuition behind this assumption: If I follow you and you don't follow me, it's probable that I treat you as an knowledge supplier; if I follow you and you also follow me it's possible but less probable that I treat you as a knowledge supplier because the mutual following may simply due to our friendship.

perspective, using network community detection techniques to construct a panel data structure, emphasizing the embeddedness perspective (Grannovetter, 1985 and 2017).

The presentation proceeds as follows: In section 2, we derive insights via a mathematical model. In section 3, we introduce the data, explain and present the text analysis outcomes, and define key variables used in the regression analysis. Section 4 explains the major empirical findings. In section 5 we talk about methodology contributions of this paper and in section 6 we conclude.

2 Theory

Having intuitively summarized the theoretical insights, we now provide the formal model.

2.1 Environment setup and opinion producers

This is a static model. We model a knowledge market for a topic with supply and demand sides and model how economic agents connect and influence each other. Before detailing behavioral assumptions for opinion producers and knowledge consumers, let's first introduce the environmental setup, especially the social embeddedness assumption. (Grannovetter (2017) emphasizes the importance of being clear about the social embeddedness of economic agents in understanding many social phenomena.)

There is a line of length 3 (see Figure 1 below on Page 9) and agents (knowledge consumers and opinion producers) are distributed along the line in a way that will be specified soon. We regard the line 0-3 as a social space. Distance in a social space measures how closely people are connected in a social network. At this moment it suffices to know that knowledge consumers near each other tend to influence each other and that knowledge consumers tend to follow (pay attention to) opinion producers near to them in the social space. For example, as a student in an economics department in US, my social network position makes me familiar with many US economists but it's harder for me to be familiar with economists in Japan.

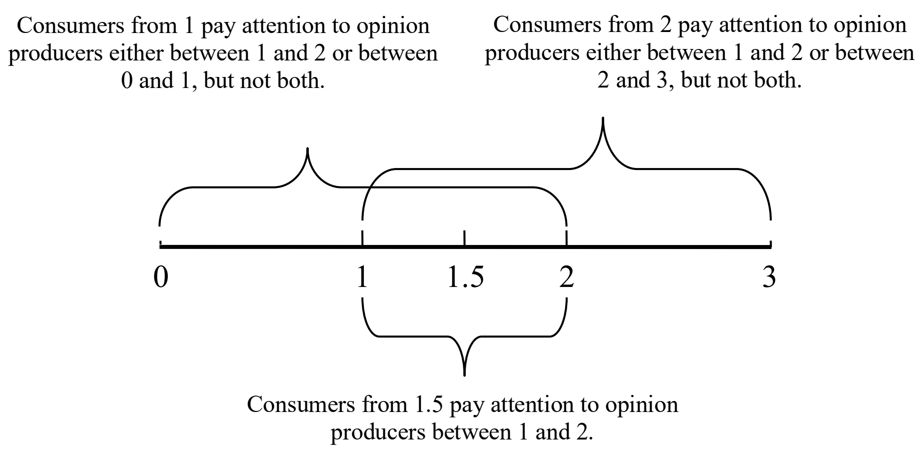
An opinion producer rests at every point on the line. Thus, opinion producers are uniformly distributed along the line. Opinion producers are made into opinion leaders by the process of consumers choosing to follow them. Specifically, each opinion producer has a probability of being an opinion leader that increases in the number of consumers following them. Let q be the total number of knowledge consumers paying attention to (following) him. We assume a continuum of consumers of which the measure is one, so $q \in [0,1]$. For concreteness, we assume the probability for an opinion producer to be an opinion leader is $\frac{2}{1+e^{\delta(1-q)}}$, where δ governs the slope of the increasing curve. We model the behavior pattern of opinion producers in this simplified form since the pivot of our story rests on the demand side at the interaction of social influence with social network structure; the simplification makes the model tractable and loses no economic insights.

Actually, any convex curve instead of $\frac{2}{1+e^{\delta(1-q)}}$ will also do. We need convexity because it helps us to capture an important intuition in the previous section which we repeat below:

> If consumers have no impact on their adjacent consumers, consumers can follow independent sets of opinion producers. So let's assume each consumer follows a unique opinion producer. Then no opinion producer would be an opinion leader due to this very small number of followers. However, when there is social influence and people concentrate all their attention to a few opinion producers, then some opinion producers will become opinion leaders.

To have this concentration effect after combining the demand and supply sides, we need the following feature: removing a marginal amount of attention from low attention areas decreases the number of opinion leaders by less than the number of opinion leaders increased from adding the same amount of attention to high attention areas.

Figure 1



__ Consumers are only from points 1, 2, and 1.5.
__ At each point between 0 and 3 rest a opinion producer.
— $\frac{\pi}{2}$ H-consumers and $\frac{1-\pi}{2}$ L-consumers rest at point 1.5; $\frac{1-\pi}{4}$ H-consumers and $\frac{\pi}{4}$ L-consumers rest at point 1 and point 2 respectively. So half are at center and half are at periphery.

Knowledge consumers rest only at points 1, 2, and 1.5 (no consumers are from other points). There is a continuum of consumers of which the measure is 1. A half of consumers are at 1.5 and the rest are evenly divided between 1 and 2. Besides, a half of consumers are enthusiastic and willing to pay H units of attention on the topic in this knowledge market (i.e., each of them spreads the H units of attention over a number of chosen opinion producers), and the other half are unenthusiastic and willing to pay L units of attention. H>L>0 and we assume (H+L)/2 =1 to streamline notation without loss of generality. We assume $\frac{\pi}{2}$ H-consumers and $\frac{1-\pi}{2}$ L-consumers (and thus 1/2 consumers in total) are at point 1.5, i.e., at the center; $\frac{1-\pi}{4}$ H-consumers and $\frac{\pi}{4}$ L-consumers are at point 1 and point 2 respectively, i.e., at the periphery.

$\pi \in (0, 1)$. When π increases, there'll be more enthusiastic consumers at the center of the social space (or of a social network) with the total attention of all consumers kept constant. So π measures the degree to which enthusiastic knowledge consumers occupy the central network position (point 1.5). π corresponds to the key explanatory variables in our regression analysis. Occupying the central position generates influence in two directions, while occupying peripheral positions only generates influence in one direction. We make a distinction between being at the center and being at peripheries and use π as the key parameter in our model because the central idea of the paper is that enthusiastic consumers in a knowledge market have more influence on other consumers through network effects when they occupy more central positions in a network connecting consumers. This greater influence leads to more opinion leaders, as will be seen when we complete the model.

We make the following assumptions to capture the importance of social space and positions therein and to model social influence as a function of how people are close to each other in a social space. Consumers from point 1.5 are only able to pay attention to opinion producers distributed between 1 and 2. The decision of paying how much attention to each opinion producers follows below. Consumers from point 1 are able to pay attention to opinion producers distributed either between 1 and 2 or between 0 and 1, but not both. Specifically, each of the consumers from point 1 chooses whether to pay his attention to the opinion producers on his left (from point 0 to point 1) or to those on his right (from point 1 to point 2); in the meanwhile, he also chooses how much attention to pay to each opinion producer on the chosen side. Consumers from point 2 are able to pay attention to opinion producers distributed either between 1 and 2 or between 2 and 3, but not both. Specifically, each of the consumers from point 2 first chooses whether to pay his attention to the opinion producers on his left (from point 1 to point 2) or to those on his right (from point 2 to point 3); in the meanwhile, he also chooses how much attention to pay to each opinion producer on the chosen side. Thus consumers from point 1.5 are more central in the sense that both halves of the area covered by their attention overlap with what're covered by consumers from point 1 and 2. As will be seen soon, this overlap in two directions means consumers at point 1.5 can generate influence in two directions

2.2 Behavioral and interaction assumptions for knowledge consumers

Every knowledge consumer distributes the attention he's willing to pay among opinion producers he is able to reach. Let a_{ik} be the amount of attention paid by consumer i to opinion producer k (who rests at position k). So for enthusiastic consumers, the attention budget constraints are $a_{ik} \geq 0$ and $\int_0^3 a_{ik}dk \leq H$; for unenthusiastic consumers, the attention budget constraints are $a_{ik} \geq 0$ and $\int_0^3 a_{ik}dk \leq L$. Note $a_{ik} = 0$ if opinion producer k is at a position not reachable by i or i chooses $a_{ik} = 0$ when k is reachable. Let $\int a_{ik}di = A_k$, the total amount of attention received by the opinion producer k (who rests at point k).

Consumers at the central position in the social space (at the point 1.5) get no utility from opinion producers from [0,1)∪(2,3] and their utility function is

$$\int_1^2 (1 + A_k)^{1-\rho}(a_{ik})^{\rho}dk \ .$$

Note consumers at the point 1.5 have the same utility function, whether they are willing to pay H or L units of attention.

We have a few reasons to use this utility function form. First, we use this utility function form to capture social influence: the higher A_k, the higher is the marginal benefit from paying attention to opinion producer k. Note A_k is the total amount of attention received by opinion producer k.[4] Second, we use $(1 + A_k)^{1-\rho}$ rather than

[4] Actually in the utility function for consumer i, A_k should be the total amount of attention excluding i's attention received by the opinion producer, but since there is a continuum of knowledge consumers, we can safely drop the "excluding i's attention" in defining A_k and assume each consumer takes $\{A_k \mid k \in [0,3]\}$ as given, though $\{A_k \mid k \in [0,3]\}$ is endogenously determined in equilibrium.

$(A_k)^{1-\rho}$ because we assume there is still some utility for a consumer to follow an opinion producer who is followed by nobody else.[5] Third, based on this utility function, with A_k fixed, consumer i has a decreasing marginal utility from paying more attention to k and thus tends to spread attention over many opinion producers. This diversification incentive is reasonable. Finally, this function form makes the model tractable.

Consumers at point 1 get utility from opinion producers from [0,1] or [1,2] but not both. To maximize utility, consumers at point 1 choose whether to pay attention to opinion producers from [0,1] or those from [1,2], and in the meanwhile choose the amount of attention for each opinion producers. We have this assumption because we try to model social influence: consumers at point 1 cannot search or pay attention in every direction and which direction to choose is subject to social influence. The utility function for consumers at point 1 is

$$\max\{\int_0^1 (1 + A_k + s_i)^{1-\rho}(a_{ik})^\rho dk\,, \int_1^2 (1 + A_k)^{1-\rho}(a_{ik})^\rho dk\}\ .$$

Let's explain why there is s_i in the utility function. For each consumer at the periphery an s is drawn from a uniform distribution over [-t, t] with t>0. This parameter magnifies the attention to opinion producers in [0, 1] if s>0, and lessens the effect if s<0. Due to symmetry, it does not matter if the extra (positive or negative) utility is added for the opinion producers near the center [1, 2] or for those near the periphery [0, 1]. Without loss of generality, we set t=1. s plays a role of smoothing equilibria as will be seen later.[6] Also note consumers at point 1 have the same utility function, whether they are willing to pay H or L units of attention.

---

[5] Here 1 can be changed into any positive number. If consumer utility from an opinion producer who is followed by nobody else is zero, there'll be an infinite number of uninteresting equilibria.

[6] Without s, we have the same theoretical conclusion: the number of opinion leaders and their similarity increase in π which, as mentioned above, measures the degree to which enthusiastic knowledge consumers occupy the central network position (point 1.5) (and thus it corresponds to r1 and r2 in our regression analysis). But the increasing function is a step function, with either no increase (for π below a threshold) or increase to the full extent (for π above or equal to a threshold); it's also possible for the threshold to be zero depending on the relative magnitude of H and L. This ad hoc jumpy behavior is avoided by s providing heterogeneity for consumers, which we think is more realistic and consistent with the empirical findings and, more importantly, elucidates our insight better. Note s plays a similar role as does the random part of the utility function in standard discrete choice models.

Using analogous assumptions, the utility function for consumers at point 2 is

$$\max\{\int_2^3 (1 + A_k + s_i)^{1-\rho}(a_{ik})^{\rho} dk\,, \int_1^2 (1 + A_k)^{1-\rho}(a_{ik})^{\rho} dk\}\ .$$

Note consumers at point 1 have the same utility function, whether they are willing to pay H or L units of attention.

2.3 Equilibrium and comparative statics

With the behavioral assumptions and embeddedness assumptions, we solve for the equilibrium, which is summarized in the following.

(1) Solving for the equilibrium, we find A'_k's are the same for all $\mathrm{k} \in [1, 2]$ and let's denote this common value by $\overline{\mathrm{A}}_{\mathrm{c}}$ (c for center); A'_k's are the same for all $\mathrm{k} \in [0, 1) \cup (2, 3]$ and let's denote this common value by $\overline{\mathrm{A}}_{\mathrm{p}}$ (p for periphery). Note A'_k is the total attention received by opinion producer k not the number of consumers following k.

(2) H-consumers (or L-consumers) from point 1.5 pay H (or L) attention to each opinion producers between 1 and 2[7]. H-consumers (or L-consumers) from point 1 with $\mathrm{s} < \mathrm{s}^* \equiv \overline{\mathrm{A}}_{\mathrm{c}} - \overline{\mathrm{A}}_{\mathrm{p}}$ pay H (L) attention to each opinion producer between 1 and 2; H-consumers (or L-consumers) from point 1 with $\mathrm{s} > \mathrm{s}^* \equiv \overline{\mathrm{A}}_{\mathrm{c}} - \overline{\mathrm{A}}_{\mathrm{p}}$ pay H (or L) attention to each opinion producer between 0 and 1. H-consumers (or L-consumers) from point 2 with $\mathrm{s} < \mathrm{s}^* \equiv \overline{\mathrm{A}}_{\mathrm{c}} - \overline{\mathrm{A}}_{\mathrm{p}}$ pay H (or L) attention to each opinion producer between 1 and 2; H-consumers (or L-consumers) from point 2 with $\mathrm{s} > \mathrm{s}^* \equiv \overline{\mathrm{A}}_{\mathrm{c}} - \overline{\mathrm{A}}_{\mathrm{p}}$ pay H (or L) attention to each opinion producer between 2 and 3.

(3) From (2) we know each consumer's decision and we also know that $\frac{\pi}{2}$ H-consumers and $\frac{1-\pi}{2}$ L-consumers are from point 1.5 and that $\frac{1-\pi}{4}$ H-consumers and $\frac{\pi}{4}$ L-consumers are from point 1 and from point 2 respectively. With these two pieces of information combined, the following equations determine $\overline{\mathrm{A}}_{\mathrm{p}}$ and $\overline{\mathrm{A}}_{\mathrm{c}}$:

---

[7] Note the measure of opinion producers share H (or L) units of attention is 1, so each opinion producer gets H (or L) units of attention.

$$\overline{A}_p = \left(\frac{1-\pi}{4} \times H + \frac{\pi}{4} \times L\right)(1 - [\frac{1}{2} \times (\overline{A}_c - \overline{A}_p + 1)])$$

$$\overline{A}_c = \left(\frac{1-\pi}{4} \times H + \frac{\pi}{4} \times L\right)\left[\frac{1}{2} \times (\overline{A}_c - \overline{A}_p + 1)\right] + \left(\frac{\pi}{2} \times H + \frac{1-\pi}{2} \times L\right)$$

H>L>0 and (H+L)/2 =1, as assumed above, ensures $(\overline{A}_c - \overline{A}_p) \in (-1, 1)$.

(4) Combining $(\overline{A}_c - \overline{A}_p) \equiv s^*$ and the two equations in (3), we solve for s^* and get $s^* = 2 - \frac{1}{1 - \frac{1-\pi}{4} \times H - \frac{\pi}{4} \times L}$ which is increasing in π and belongs to (-1, 1). As will be seen soon, s^* determines how many consumers at the periphery pay attention to peripheral (and thus central) opinion producers.

(5) To simplify notation, let's assume the CDF of the uniform distribution over [-1, 1] for s is G(s). Per (2) and the model setup, the number of consumers pay attention to the opinion producers in the central region (between point 1 and point 2) is $\frac{1}{2} + \frac{1}{2}G(s^*)$, while the number of consumers pay attention to the opinion producers in peripheral regions (between point 0 and 1 and between point 2 and 3) are both $\frac{1}{4} - \frac{1}{4}G(s^*)$. With the assumption on opinion producers, the number of opinion leaders is given by $\frac{2}{1+e^{\delta(\frac{1}{2}-\frac{1}{2}G(s^*))}} + \frac{4}{1+e^{\delta(\frac{3}{4}+\frac{1}{4}G(s^*))}}$, which is increasing in $G(s^*)$. Together with $G(s^*)$ increasing in s^* and s^* increasing with π, we have the following comparative statics.

Comparative Statics 1: Increasing π results in more opinion leaders and a higher concentration of opinion leaders between point 1 and point 2.

(6) Based on (5), as π increases, $\frac{1}{2} + \frac{1}{2}G(s^*)$ (the number of consumers pay attention to the opinion producers in the central region) increases, $\frac{1}{4} - \frac{1}{4}G(s^*)$ (the number of consumers pay attention to the opinion producers in peripheral regions) decreases, and the amount of increase is greater than the amount of decrease. Consequently, there are more opinion leaders in the central region, who are

followed by the same set of consumers, and less opinion leaders in the peripheral regions, who are followed by different consumers. So opinion leaders on average have more similarity with each other in terms of who follow them.

Comparative Statics 2: The higher concentration of opinion leaders implies more similarity in terms of who follow them.

Thanks to higher ability of increasing A'_k, enthusiastic consumers generate more social influence to other consumers than unenthusiastic consumers do. Consumers at the central region make choices influencing consumers at peripheral regions by attracting them to pay attention to opinion producers at the central region. Thus, when there are more enthusiastic consumers at the central region, there'll be higher concentration of consumers' attention into the central region. This concentration leads to more and similar opinion leaders as shown in (5) and (6) above.

3 Empirical analysis I: data, text analysis, and variable construction

3.1 Data

We use the data from Hodas and Lerman (2014).[8] Hodas and Lerman (2014) collected tweets over three weeks in the fall of 2010 and then retained tweets containing a URL in the message body. A URL, uniform resource locator, is usually in the form of a string of characters or symbols (e.g., http://gd.is/4nfm) that references a web resource and specifies its location on a computer network. When a twitter user shares a web resource in his tweet, the web address is coded into a short string of characters and symbols and shown in the tweet. Thus, a URL can be regarded as a meaningful key word more informative about the content or topic of a tweet than usual words such as good, the, or, increase, etc. Hodas and Lerman (2014) then retrieved all tweets containing these URLs, ensuring the complete tweeting history of

[8] One can download data and find data details at https://www.isi.edu/~lerman/downloads/twitter/twitter2010.html

all the URLs, resulting in 3 million tweets in total. They also collected the friend and follower information for all tweeting users at the time, resulting in a network with almost 700K nodes and over 36M directed edges (an edge is directed from user A to user B if user A follows user B). Thus we have URLs in the twitter text but no full text contents in the data. Twitter has forbidden using large amounts of text contents by external researchers. From Hodas and Lerman (2014), we thus have as our raw data a directed network structure, i.e., who follows whom with each Twitter user represented by a user id number, and a set of URLs for each user id if the user had URLs in his tweets. Since Twitter has changed their way of coding user accounts, we don't know who a user is from his id.

Next in section 3.2 we do topic modeling based on the data. Later we use the resultant topics to define knowledge markets. We are interested in the characteristics of knowledge markets. In section 3.4 we will define key variables for each knowledge market based on network analysis. The variables will be used in the regression analysis in section 4, with each knowledge market corresponding to a sample point.

3.2 Topic modeling

To use topic modeling, we view each user as a document and associated URLs as words informative about the topics of the document. We then model topics with the correlated topic model (CTM) (Blei and Lafferty, 2006). The CTM is a state-of-the-art topic modeling method.[9] Compared with other commonly used topic models, the CTM models a document generating process that exploits more subtleties in the data to capture topic correlation. For example, a document about sports is more likely to also be about health than international finance (Blei and Lafferty, 2006; Roberts, Stewart, and Tingley, 2014). Below we briefly introduce the CTM before showing our topic modeling outcomes. For more technical details, applications, and comparison of different topic models, please see Blei and Lafferty, (2006) and Roberts, Stewart, and Tingley, (2014).

---

[9] Besides, the CTM can be seen as an improved version of the LDA, a very commonly used topic modeling method; Blei, an author of the CTM, is also one of the researchers independently discovering the LDA.

The CTM models every document as a distribution over topics with every topic modeled as a distribution over words. The CTM thus models a given set of documents as generated by a mixture of distributions as elucidated below, and estimated parameters of these distributions can be used to estimate topic proportions for each document. Specifically, the CTM uses the following generative process for a document with length m:

(1) A K-by-1 vector is drawn. K is the number of topics specified by researchers. The vector is drawn from a logistic normal distribution and represents topic frequencies (proportions) in a document. The use of a logistic normal distribution allows the CTM to capture topic correlation.

(2) The K-by-1 vector from (1) is used for the parameters of a multinomial distribution over K topics. An m-by-1 vector is constructed by randomly drawing topics m times from the multinomial distribution, where m is document length. Word order in a document does not matter. Each element in the m-by-1 vector indicates one topic from the K topics.

(3) A distribution of words is used for each topic. Specifically, an N-by-1 vector is used for each of the K topics and represents the distribution of words in each topic. N is the total number of words in all the documents. The elements in the vector for a topic correspond to the frequency of each word in the topic.

(4) For each element that indicates a topic from the m-by-1 vector in (3), a word is randomly drawn based on the topic's word distribution/frequencies constructed in (2). The m words thus drawn form a document.

Estimates of the parameters of the logistic normal distribution used in (1) are topic proportions for each document (each Twitter user in our case).

The topic modeling outcome

A Twitter user is regarded as a document. URLs in users' tweets are regarded as words informative about the topics of documents. We're interested in the number of opinion leaders in different topics, so in topic modeling we only focus on the documents that are widely read (i.e., Twitter users that are intensively followed). Let's call Twitter users whose numbers of followers are among the top 0.5% as the top 0.5% followees. Specifically, to correspond to the topic model above, we regard the top 0.5% followees as documents and their URLs as words.[10] Note a Twitter user is both a follower and a followee: when we focus on how many users follow a Twitter user, the user is a followee; when we focus on who and how many users are followed by a Twitter user, the user is a follower. The number of the top 0.5% followees is 3498, and the followees among them with the fewest followers have 1118 followers (these two numbers for top 1% followees are 6985 and 648).

The top 0.5% followees provide the observations for the CTM estimation. This means the top 0.5% followees are used as a set of documents and their URLs are used as words to be inputted into the CTM. 45 topics result from the CTM. This number of topics is chosen using the method introduced by Roberts, Stewart, and Tingley (2014). Specifically, we experiment with different numbers of topics in the document generating process described in section 3.2 (with K from 30 to 100 with a step of 5) and randomly leave out a proportion of documents from our sample for out-of-sample evaluation. Among the K's, the numbers of topics, that best or nearly best fit the out-of-sample documents, we choose the one fitting well with the in-sample documents, so the criterion of choosing K takes into account both out-of-sample predictive power and in-sample fit. Using the top 0.5% followees and the top 1% both results in the topic number of 45.

Many documents have very small estimated topic proportions over some topics. We treat a topic proportion as 0 if the topic's proportion in a document is less than 10%.[11] We have two reasons for this censoring: a very small topic proportion is usually not taken seriously in text mining (Robinson and Silge, 2017) since it's largely a model artifact for a better fit. Without getting rid of the small proportions, every document includes all topics, and this means the numbers of opinion leaders (we'll formally

[10] The conclusions are the same if we instead us the top 1% as shown in the appendix.
[11] The conclusions are the same when we use 20% instead of 10%, as shown in the appendix.

define opinion leaders in section 3.4) are almost the same for all topics, which is neither interesting nor realistic.

Based on the topic modeling outcomes and the network structure, we'll define key variables for regression analysis in section 4. But before that, in order to be clear about what corresponds to an observation or a sample point in the regression analysis, let's first introduce the network community detection technique we use to separate networks into different sub-networks: this turns out to give us a panel data structure.

3.3 Network Community Detection

How things work in Hobbits' Shire can be quite different from that among Gandalf's wizard friends (Tolkien, 1955). A very large social network can consist of several nearly isolated sub-networks each of which may be better regarded as a relatively isolated kingdom. A sub-field of network science, community detection, focuses on dividing a network into communities within which nodes are densely connected and between which nodes are sparsely connected (Thurner, Hanel, and Klimek, 2018; Newman, 2018). We use a state-of-the-art network community detection method (Clauset, Newman, and Moore, 2004; Csardi, 2019) with especially good performance for very large networks as is the present case. We divide the Twitter network into communities. We find nearly all the top 0.5% followees (more than 95%) are in four communities. 1494, 439, 798, and 601 of the top 0.5% followees are respectively in the four communities and 174 are in other communities. Approximately 94% of links among all the links directed toward and away from the nodes of these four communities are within communities. We think the four communities can be regarded as relatively isolated kingdoms. Of course, the 94% is not 100%, so we'll use some econometric techniques to mitigate this concern in section 4.

We'll briefly explain the community detection method soon, but let's now explain the purpose of doing community detection. We have two reasons. First, with this community detection outcome and the topic modeling outcome, the variables in our regression analysis are topic-community specific, i.e., an observation or a sample point in our regression analysis corresponds to a topic in a community. This gives us a panel data structure. Second, it's reasonable to assume consumers from different communities behave differently, as is shown by Reddy, Kitsuregawa, Sreekanth, and Rao (2002) who use a similar network community detection method to identify consumers with similar interests and purchasing habits.

Let's now briefly talk about the underlying logic of community detection. For more technical details and applications, see Thurner, Hanel, and Klimek, (2018), Newman (2018), and Clauset, Newman, and Moore (2004). We first transform the Twitter network into an undirected network. So there is an undirected link between Twitter user i and Twitter user j when there is a directed link from i to j or from j to i or when there are two directed links in both directions. L denotes the total number of such undirected links. $\widetilde{A}_{ij} = 1$ if there is an undirected link between i and j. Let's define k_i as the number of network neighbors of node i in the undirected network. (Two nodes are network neighbors of each other if they are connected in the network.) $\delta_{ij} = 1$ if i and j are assigned into the same community; otherwise, $\delta_{ij} = 0$. Note if the network is generated randomly lacking community structure, the probability that $\widetilde{A}_{ij} = 1$ is given by $\frac{k_i k_j}{2L}$. The community detection method we use chooses who are in what communities and the number of communities by maximizing a modularity score, Q, as defined in the following.

$$Q = \frac{1}{2L}\sum_{ij}\left(\widetilde{A}_{ij} - \frac{k_i k_j}{2L}\right)\delta_{ij}$$

Intuitively, this score compares the actual network to random networks that lack community structure. Specifically, the score compares $\widetilde{A}_{ij}$ (the actual presence of a link) and $\frac{k_i k_j}{2L}$ (the probability of the presence of a link if the network is generated randomly lacking community structure) when $\delta_{ij} = 1$. The score is larger if for a community assignment, the number of with-in community links (picked out by $\delta_{ij} = 1$) is larger than the expected number of links generated from a random network that lacks community structure.

3.4 Key variables for regression analysis

Based on the above community detection and topic modeling, let's define key variables used for the regression analysis in section 3.

In order to define the following variables, we now specify what we mean by opinion leaders. If a Twitter user (1) belongs to the top 0.5% followees,[12] (2) has a topic proportion greater than 10% on topic j, and (3) belongs to network community k, he is an opinion leader for topic j in network community k.

It is important to emphasize that the observations used for the CTM and those used for regression analysis are different. In the CTM, each Twitter user in the top 0.5% followees is an observation (i.e., each of these users is regarded as a document and the URLs in his tweets regarded as words). In the regression analysis in section 4, each observation is a knowledge market uniquely pinned down by a community-topic pair. Then **for topic j in network community k** (i.e., for the knowledge market defined by j and k), we define the following variables.

1. N_{jk}: the number of opinion leaders for topic j in community k.

The number of opinion leaders for a topic in a community is the number of top 0.5% followees who are in this community and whose topic proportions are greater than 10% in this topic. (In the appendix, we report outcomes with other choices of these two thresholds, where the conclusions are the same.)

2. S_{jk}^{in}: the average similarity of opinion leaders for topic j in community k in terms of who follow them.

[12] Note the top 0.5% followees are used in CTM to mine topics and topic proportions.

Based on Adamic and Adar (2003), we define the average similarity of opinion leaders in terms of who follow them by the average of the in-degree Jaccard similarity coefficients of all pairs of the opinion leaders for a topic in a community. The in-degree Jaccard similarity coefficient of a pair of opinion leaders for topic j in community k is the number of users in community k following both opinion leaders divided by the number of users in community k following at least one of the opinion leaders.

3. f_{jk}: the total amount of attention (divided by 1000) to all the opinion leaders for topic j in community k paid by their followers.

First, starting from all the opinion leaders for topic j in community k, we find the set of all the Twitter users in this community following them excluding the opinion leaders themselves. Second, for each member of this set, say, follower m, we sum his opinion leaders' topic proportions in this topic (j), and define follower m's attention to this topic (j) in this community (k) as this sum, denoted by f_{mjk}. Finally, we sum the attention of all such followers for this topic in this community and divide it by 1000 to get $f_{jk} = \frac{\sum_m f_{mjk}}{1000}$. The 1000 will make the estimates in section 4 look simple.

4. $r1_{jk}$: for topic j in community k, the weighted average in-degree centrality (divided by 1000) of followers who pay more than median attention to the opinion leaders.

Intuitively, $r1_{jk}$ is a weighted sum of how central the network positions taken by enthusiastic knowledge consumers. It thus corresponds to π in the model, the degree to which enthusiastic followers take central positions in a network. Specifically, we define it in the following way. First, as defined above, the in-degree centrality of a Twitter user is the number of users following him. To measure how central is a network position occupied by a user, we calculate his in-degree centrality in this community k (note each community can be viewed as a separate network.). Second, note we have each follower's attention for a topic in a community (f_{mjk}) when we define f_{jk} in 4. We then define Ω_{jk} as consisting of the followers whose attention is greater than the median follower attention for topic j in community k. Finally, for a follower $m \in \Omega_{jk}$ we denote his in-degree centrality in community k by ID_{mk}. We then define $r1_{jk}$ by averaging the in-degree centralities weighted by each follower's attention per the following equation and then divide the average by 1000 to make the regression estimates in section 4 look simple.

$$r1_{jk} = \frac{\sum_{m \in \Omega_{jk}} (f_{mjk} \times ID_{mk})}{\sum_{m \in \Omega_{jk}} f_{mjk}} \times \frac{1}{1000}$$

Note users who pay more attention receive larger weights in this average. This corresponds to π in the theory (after controlling for f_{jk}) and is our major explanatory variable in the regression analysis. In the rest of this paper, we call followers who pay more than median attention enthusiastic followers or enthusiastic knowledge consumers.

5. $r2_{jk}$: for topic j in community k, the weighted average eigenvector centrality of followers who pay more than median attention to the opinion leaders.

In defining $r1_{jk}$, we use the in-degree centrality of a user that counts how many people follow him. It is intuitive and reasonable, but we can exploit more subtlety in the network: for a user, some of his followers may be followed by many people while his other followers may be followed by a few. The eigenvector centrality extends in-degree centrality and takes into account this heterogeneity among a user's followers to evaluate the user centrality in a network: ceteris paribus, users whose followers are more followed are more central.[13] Below we define $r2_{jk}$ using the eigenvector

[13] Formally, when an user's centrality is regarded as correlated with the sum of his followers' centralities, the eigenvector centrality, x_i, of user i can be defined from $x_i = \kappa^{-1} \sum_j A_{ji} x_j$, where $A_{ji} = 1$ if j follows i (to calculate eigenvector centrality for undirected networks, the same formula is used with $A_{ji} = 1$ if there is a link between j and i). κ is a scalar whose role will be seen soon. Putting the eigenvector centralities of all users into a vector, x, one gets $A^T x = \kappa x$. x is thus a eigenvector of A^T. Per Perron-Frobenius theorem, for an adjacency matrix from a connected network only the leading eigenvector is non-negative. Note each community is a connected network. Eigenvector centralities, usually assumed to be non-negative, are thus given by the leading eigenvector of A^T. However, there is an undesirable property for using eigenvector centrality for directed networks. People without followers have zero eigenvector centrality. Per the definition of eigenvector centrality, anyone who is followed only by such followers has zero eigenvector centrality. Iteratively, users who only receive incoming links from

centrality. We'll use $r2_{jk}$ and $r1_{jk}$ separately as our major explanatory variable in regression analysis in section 4: this strategy can be viewed as a robust check.

With f_{mjk} defined above in 4 and with the eigenvector centrality (defined in footnote 14) for each follower (m) in a community (k) denoted by EC_{mk}. We define $r2_{jk}$ per the following equation.

$$r2_{jk} = \frac{\sum_{m\in\Omega_{jk}}(f_{mjk} \times EC_{mk})}{\sum_{m\in\Omega_{jk}} f_{mjk}}$$

This is the same as $r1_{jk}$ except ID_{mk} substituted with EC_{mk} and without the term $\frac{1}{1000}$.

6. $r1iv_{jk}$: for topic j in community k, the average in-degree centrality (divided by 1000) of followers in the friendship network who pay more than median attention to all the opinion leaders.

This is constructed in the same principle as $r1_{jk}$ with the in-degree centrality calculated based on the friendship network. The friendship network is constructed by dropping any directed link without a reciprocal of it, i.e., a link from i to j is dropped if there is no link from j to i. In the four communities under focus, the numbers of user pairs connected in the friendship networks (undirected networks) are respectively 30%, 44%, 20%, and 55% of the numbers of user pairs connected with at least one directed link in corresponding community networks (directed networks). When we do regression analysis in next section, we'll explain why and how this can serve as an instrument for r1.

7. $r2iv_{jk}$: for topic j in community k, the average eigenvector centrality of followers in the friendship network who pay more than median attention to all the opinion leaders.

zero-centrality users have zero eigenvector centrality. However, in our case the impact from this property can be ignored since such users account for only around 1% of users in each community. For technical details, please see Newman (2018), Thurner, Hanel, and Klimek (2018), and Bonacich (1987).

This is constructed in the same principle as $r2_{jk}$ with the eigenvector centrality calculated based on the friendship network. When we do regression analysis in next section, we'll explain why and how this can serve as an instrument for r2.

With these variables defined and with 45 topics from topic modeling and four communities from network community detection, we do regression analysis in the next section and present our empirical findings. The summary statistics for each variable are given in Table 1.

Table 1: summary statistics for variables used in the regression analysis

	Mean	Standard Deviation
N	50.3	127.9
S^{in}	0.111	0.115
r1	0.212	0.131
r2	0.066	0.061
f	69.4	192.9
r1iv	0.156	0.109
r2iv	0.062	0.059

4 Empirical analysis II: panel data analyses based network communities

From section 3, we have a data structure with each observation corresponding to a topic-community pair. With 45 topics from topic modeling and four communities from network community detection, we construct a panel data set as detailed in section 3. The top 0.5% followees are distributed across different communities and the top 0.5% followees (documents) within a community may not cover all topics, that is, not all topics appear in all communities: 9 topics appear in all of the four communities, 9 topics in only three of the four communities, 13 topics in only two of the four communities, and 10 topics in only one of the four communities.[14] So the data structure turns out to be an unbalanced panel with 99 observations. Compared to standard panel data, topic corresponds to individual and community to time.

4.1 Panel data analysis without instruments

In this section we have regression equations like the following:

$$\mathbf{y_{jk} = \beta \times x_{jk} + \gamma \times f_{jk} + \mu_j + \rho_k + \varepsilon_{jk}.}$$

In all the regression analyses in this paper, each observation is a knowledge market uniquely pinned down by a community-topic pair such as topic j and community k. As normal, β and γ are coefficients, and ε_{jk} is the error term. We run two-way fixed effect regressions controlling for topic fixed effects (μ_j) and community fixed effects (ρ_k). In section 3.4 we have defined the variables to be used in the regressions and we now explain how the variables are to be used.

y_{jk} will be either N_{jk} (the number of opinion leaders for topic j in community k) or S^{in}_{jk} (the similarity of opinion leaders for topic j in community k in terms of who follow them).

x_{jk} is the primary explanatory variable and will be $r1_{jk}$ (a thousandth of the average in-degree centrality of enthusiastic followers for topic j in community k) or $r2_{jk}$ (the average eigenvector centrality of enthusiastic followers for topic j in community k). The similarity and difference between these two variables are discussed above. We separately use these two as robustness checks. Note r1 and r2 correspond to π in the theory in section 2, capturing in different ways how central enthusiastic consumers' network positions are. So our theoretical conclusion in section 2 corresponds to significant positive coefficients of x_{jk} with the dependent variables N_{jk} and S^{in}_{jk}. Recall our theoretical conclusion in section 2: in a knowledge market when

[14] Note the sum of these four numbers is 41 rather than 44. This is because for some topics in some communities, there is only one opinion leader. The values for S^{in}, S^{in}, r1, r2, r1iv, and r2iv are NAs when there is only one opinion leader. So we drop such incomplete cases.

enthusiastic consumers on average occupy more central positions in a network connecting consumers, more opinion producers become opinion leaders and there is more similarity among opinion leaders in terms of who pay attention to them.

We also control for f_{jk} (the total amount of follower attention (divided by 1000) to all the opinion leaders for topic j in community k), because in our theory we fix this amount and only focus on changes in π. Moreover, a high total amount of attention can result in a large number of opinion leaders in a topic due to high demand, while the total amount of attention may also correlate with the network structure in some unknown way that might make it correlate with our key explanatory variables. So we treat it as a potential confounding variable and control for it.

We use the Arellano covariance estimator to estimate standard errors (Croissant and Millo, 2008; Arellano, 1987) for the following reason. The time dimension for a standard panel data analysis corresponds to community membership in our data (while each topic is an individual). The network community detection technique has done its best to divide a network into 4 communities, but there are still links between communities,[15] and error terms can be correlated across communities. So we use the Arellano estimator that takes into account serial correlation of arbitrary form; time has an order where some AR process may be imposed, while community membership cannot be ordered, so an arbitrary form is in need. Besides, for short panel (there are only 4 communities) the Arellano estimator is usually advisable (Croissant and Millo, 2008).

The regression results with three dependent variables and two alternative explanatory variables are shown in Table 1 and Table 2.

[15] As noted above, about 94% of links among all the links directed toward to and away from the nodes in these four communities are within communities.

Table 2: The effect of enthusiastic consumers' centrality on the number of opinion leaders and their similarity: OLS and r1

	Dependent Variables	
	N	S^{in}
r1 (how central enthusiastic consumers are in each knowledge market)	84.6** (p=0.03)	0.238*** (p=0.002)
f (a thousandth of the total amount of attention in each knowledge market)	5.46*** (p<0.001)	0.0003 (p=0.247)

1. Here we use the top 0.5% followees in topic modeling, use 10% as the topic proportion threshold, use median attention to define enthusiastic consumers, and use the in-degree centrality to measure how central enthusiastic consumers are.
2. We control the total amount of attention to a topic in a community (f_{jk}), topic fixed effects, and community fixed effects.
3. The data structure is an unbalanced panel with 99 observations. Compared with standard panel data, topic corresponds to individual and community to time.
4. We use Arellano covariance estimator to estimate standard errors (Croissant and Millo, 2008; Arellano, 1987)
5. Numbers in parentheses are p-values for the null hypothesis of zero effect; significance code: * for 0.1, ** for 0.05, *** for 0.01.

Table 3: The effect of enthusiastic consumers' centrality on the number of opinion leaders and their similarity: OLS with r2

	Dependent Variables	
	N	S^{in}
r2 (how central enthusiastic consumers are in each knowledge market)	175* (p=0.074)	0.889*** (p=0.0001)
f (a thousandth of the total amount of attention in each knowledge market)	5.43*** （p<0.001）	0.000209 (p=0.43)

1. Here we use the top 0.5% followees in topic modeling, use 10% as the topic proportion threshold, use median attention to define enthusiastic consumers, and use the eigenvector centrality to measure how central enthusiastic consumers are.
2. We control the total amount of attention to a topic in a community (f_{jk}), topic fixed effects, and community fixed effects.
3. The data structure is an unbalanced panel with 99 observations. Compared with standard panel data, topic corresponds to individual and community to time.
4. We use Arellano covariance estimator to estimate standard errors (Croissant and Millo, 2008; Arellano, 1987)
5. Numbers in parentheses are p-values for the null hypothesis of zero effect; significance code: * for 0.1, ** for 0.05, *** for 0.01.

The regression outcomes are in table 2 and table 3. The conclusion is summarized below:

Controlling for topic fixed effects, community fixed effects, and the total amount of attention to a topic in a community (f_{jk}), we find there tend to be more opinion leaders (larger N_{jk}) and the opinion leaders tend to be more similar in terms of who follow them (larger S_{jk}^{in}), when enthusiastic followers (who pay more than median attention to all opinion leaders of a topic in a community) occupy more central network positions as measured by r1 and r2.

A remark on the magnitude of the effects

A back-of-envelope calculation based on the information from Table 1-3 gives the following effect magnitudes. Increasing r1 by one standard deviation increases N by approximately 0.1 standard deviations and increases S^{in} by approximately 0.3 standard deviations. Increasing r2 by one standard deviation increases N by approximately 0.1 standard deviations and increases S^{in} by approximately 0.5 standard deviations. The scale of these effects suggests that the estimates capture economically important phenomena. How to evaluate these effect magnitudes needs further investigation because the number of opinion leaders follows a fat tail distribution and we don't know how important an increase of 0.1 standard deviations in the number of opinion leader is for opinion diversity.

4.2 Panel data analysis with instruments: 2SLS

In this section we use r1iv and r2iv as instruments for r1 and r2 to do 2SLS. The second stage regression equations are the same as the regression equations in section 4.1 except that r1 and r2 are substituted with their first stage fitted values. The first stage regression equations are as follows:

$$\mathbf{x_{jk}} = \widetilde{\boldsymbol{\beta}} \times \mathbf{xiv_{jk}} + \tilde{\boldsymbol{\gamma}} \times \mathbf{f_{jk}} \times \tilde{\boldsymbol{\gamma}} + \tilde{\boldsymbol{\mu}}_{\mathbf{j}} + \tilde{\boldsymbol{\rho}}_{\mathbf{k}} + \tilde{\boldsymbol{\varepsilon}}_{\mathbf{jk}}$$

As in section 4.1, x_{jk} will be $r1_{jk}$ (a thousandth of the average in-degree centrality of enthusiastic followers for topic j in community k) or $r2_{jk}$ (the average eigenvector centrality of enthusiastic followers for topic j in community k) with xiv_{jk} being r1iv and r2iv respectively. $\widetilde{B}$ and $\tilde{\gamma}$ are coefficients and $\tilde{\varepsilon}_{jk}$ is the error term in the first stage. $\tilde{\mu}_j$ and $\tilde{\rho}_k$ are topic and community fixed effects in the first stage. Let's now explain why the outcomes from using r1iv and r2iv as instruments are informative.

Though it's hard to imagine a story that more opinion leaders or more similarity among the opinion leaders leads to enthusiastic followers occupying more central network positions, or to imagine a story that a third factor leads to the correlation between our dependent and explanatory variables, r1 and r2 may be subject to an endogeneity concern because the network structure may be endogenous in a way that confounds our estimates, even after we've controlled for topic fixed effects, community fixed effects, and the total amount of attention to a topic in a community. So we use r1iv and r2iv as instruments which, compared with r1 and r2, are defined based on the friendship network rather than the original network (please refer to section 3 for definition details). The logic of using the friendship network as an instrument is explained in the following.

If I follow you but you don't follow me, it's highly possible that I regard you as a knowledge source. If I follow you and you also follow me, it's still possible that I regard you as a knowledge source, but it's also possible that we have a mutual relationship merely because of friendship. Thus the key point is that the friendship network should be exogenous or subject to the endogeneity concern to a lesser degree.[16] To put this in another way, some unknown interaction of topic characteristics and community characteristics might have an impact on both the number of opinion leaders and enthusiastic consumers' network positions determined by the original network, but the position measures based on the friendship network

[16] By a lesser degree, we mean the proportion of exogenous variation in the variation of a variable is smaller.

are unlikely to subject to this endogeneity concern since it is mainly driven by friendship.

Of course, we don't know if the friendship network is actually totally exogenous to the knowledge market activities, but we don't need this absolute exogeneity. Importantly, significant positive effects from 2SLS are not very informative in themselves, but significant positive 2SLS outcomes being insignificantly different from (statistically equivalent to) the outcomes in section 4.1 are informative and tell us that the endogeneity concern can be ignored. Our logic is this: under the assumption that the friendship network is subject to an endogeneity concern to a lesser degree than the original network is, the outcome based on friendship network is significantly different from that based on the whole network if the endogeneity were really a serious concern. The contrapositive of this logic is this: under the assumption that the friendship network is subject to an endogeneity concern to a lesser degree than the original network is, if the outcome based on friendship network is not significantly different from that based on the whole network, the endogeneity is not a concern. Later we do Wald tests testing the statistical equivalence between the coefficients of the primary explanatory variables in OLS and 2SLS and fail to reject the equivalence, so endogeniety is not a problem in our case. More details for this will be provided after we show the 2SLS results in Table 4 and Table 5 below.

Table 4: The effect of enthusiastic consumers' centrality on the number of opinion leaders and their similarity: 2SLS with r1iv instrumenting r1

	Dependent Variables	
	N	S^{in}
r1 (how central enthusiastic consumers are in each knowledge market)	70.2** (p=0.047)	0.243*** (p=0.002)
f (a thousandth of the total amount of attention in each knowledge market)	5.45*** (p<0.001)	0.00031 (p=0.246)
The first stage F-statistics for excluded iv is 1600		

1. Here we use the top 0.5% followees in topic modeling, use 10% as the topic proportion threshold, use median attention to define enthusiastic consumers, and use the in-degree centrality to measure how central enthusiastic consumers are.
2. We control the total amount of attention to a topic in a community (f_{jk}), topic fixed effects, and community fixed effects.
3. The data structure is an unbalanced panel with 99 observations. Compared with standard panel data, topic corresponds to individual and community to time.
4. We use Arellano covariance estimator to estimate standard errors (Croissant and Millo, 2008; Arellano, 1987)
5. Numbers in parentheses are p-values for the null hypothesis of zero effect; significance code: * for 0.1, ** for 0.05, *** for 0.01.

Table 5: The effect of enthusiastic consumers' centrality on the number of opinion leaders and their similarity: 2SLS with r2iv instrumenting r2

	Dependent Variables	
	N	S^{in}
r2 (how central enthusiastic consumers are in each knowledge market)	166* (p=0.094)	0.918*** (p=0.0001)
f (a thousandth of the total amount of attention in each knowledge market)	5.43*** （p<0.001）	0.000207 (p=0.44)
The first stage F-statistics for excluded iv is 1000		

1. Here we use the top 0.5% followees in topic modeling, use 10% as the topic proportion threshold, use median attention to define enthusiastic consumers, and use eigenvector centrality to measure how central enthusiastic consumers are.
2. We control the total amount of attention to a topic in a community (f_{jk}), topic fixed effects, and community fixed effects.
3. The data structure is an unbalanced panel with 99 observations. Compared with standard panel data, topic corresponds to individual and community to time.
4. We use Arellano covariance estimator to estimate standard errors (Croissant and Millo, 2008; Arellano, 1987)
5. Numbers in parentheses are p-values for the null hypothesis of zero effect; significance code: * for 0.1, ** for 0.05, *** for 0.01.

Note as shown in section 3, in the four communities under focus which contain almost all opinion leaders, the numbers of user pairs connected in the friendship networks (undirected networks) are respectively 30%, 44%, 20%, and 55% of the numbers of user pairs connected with at least one directed link in corresponding community networks (directed networks). So the friendship network, though highly "correlated" with the original network, is very different. Finally, the first-stage F statistics for excluded IV is much greater than 10, so the biases toward OLS of the 2SLS estimates can be safely ignored.

Table 6 below shows results from Wald tests testing the statistical equivalence between the coefficients of the primary explanatory variables in OLS and 2SLS. We fail to reject the equivalence, so endogeniety is not a problem in our case. The statistical equivalence is not resulted from imprecise estimations as shown above in Table 3 and Table 4.

Table 6: Equivalence tests with null hypotheses that OLS and 2SLS coefficients of the primary explanatory variables are equivalent.

p-values for Wald tests with H0: OLS = 2SLS	N	S^{in}
r1/r1iv	0.278	0.792
r2/r2iv	0.396	0.629

Note: Here we use the top 0.5% followees in topic modeling, use 10% as the topic proportion threshold, and use median attention to define enthusiastic consumers.

The regression outcomes are in Table 4-6. The conclusion is summarized below:

Controlling for topic fixed effects, community fixed effects, and the total amount of attention to a topic in a community (f_{jk}), and using r1iv and r2iv to instrument for r1 and r2 respectively, we get estimated effects insignificantly different from their counterparts in section 4.2. First-stages are strong. This insignificant difference is not due to imprecise estimation and implies we can ignore endogeneity concern.

4.3 Robustness

As said above, in the main text we report outcomes from using the top 0.5% followees' tweets for topic modeling and using the top 0.5% followees and the 10% topic proportion threshold to defining opinion leaders.[17] The conclusions are the same, when we try every combination of the top 1% and o.5% followees with 10% and 20% topic proportion thresholds. Besides, we have used median to define r1, r2,

[17] When a top followee's topic proportion on a topic is smaller than 10%, we treat the topic proportion as zero; also see the previous section when we define N_{jk}.

r1iv, and r2iv, but the conclusions are the same if we use mean instead of median (again for every combination of the top 1% and o.5% followees with 10% and 20% topic proportion thresholds). So together with r1 and r2 skinning the same cat in two different ways, we've checked robustness across 16 combinations. The only exception is when we regress the number of opinion leaders on r2 with and without instrument using top 1% followee, 20% topic proportion threshold, and median. There the two estimates of interest associate with p-values greater than 0.1. Given that the robustness is checked across many possibilities, we think this single fish won't spoil the whole pond. Details are provided in the appendix.

5 Discussion: methodology reflection and what can be done in the future

So far we have done with the economic and econometric content of this paper. Before conclusion in the next section, we here highlight some methodological points in this paper that we think may offer other researchers something new.

A multi-layer network perspective: IV based on friendship network

In the regression analysis we use measures defined based on a friendship network as instrument variables for those based on the original network. We've talked about our logic in section 4. Now let's try to be general. People live their lives simultaneously in different networks or in a network with many layers, e.g., a network based on friendship (a friendship layer) and a network based on financial relationship (a finance layer). Though different layers can be deeply interacted with each other, each layer should be able to provide exogenous variation for another layer to some degree. Of course, the use of a multi-layer network structure for economics and econometrics is more than providing instrument variables. For technical tools and applications in the field of multilayer networks, please see Bianconi (2018).

Network community and panel data

Network community detection is a well developed sub-field in network science (Thurner, Hanel, and Klimek, 2018; Newman, 2018). But to the best of our knowledge, no one has applied this technique to construct a panel data structure for econometric analysis. We do this not only because what happens in isolated communities should be disentangled, but also because panel data structure improves identification.

An embeddedness perspective

A rudimental disagreement between the paradigm of economics and that of sociology is that economist assumes people do rational calculation and sociologists see people as societal construction whose behavior is a function of their social embeddedness (Grannovetter, 1985; Grannovetter 2017). We think that why and how people do maximization can be a function of social contexts and that social influence can exist due to people's maximization. So, in the sense that "All models are wrong, but some are useful" (Box, 1976), it's pointless to argue which paradigm is more fundamental. To be useful, our model rests at the middle point of the two: we model rational agents socially embedded in a specific way. Moreover, we think it can be fruitful in many research contexts to be specific about agents' social embeddedness, though network effects are usually modeled as well-stirred interactions (Belleflamme and Peitz, 2015).

What can be done in the future?

First, in this paper we focus on the number of opinion leaders. We assume more opinion leaders lead to opinion diversity in knowledge markets and such diversity can enhance adaptive efficiency for a system. To prove "the more opinion leaders the better" requires criteria for qualifying opinions. We haven't seen any such criteria in the academic world. Such evaluation can help answering many interesting questions in the future. Second, we make simple assumptions on the supply side. We think it's important for future research to explore subtleties on the supply side of knowledge markets where suppliers face a very different incentive structure than usual commodity, service, or factor suppliers do. Finally, what we have is a decentralization story, where enthusiastic consumers in central network positions can be loosely regarded as a self-organized centralized propaganda tool that announces which opinion producers are worthy to be followed. We think in the future it can be very interesting to explore what forces drive enthusiastic consumers to and away from central network positions.

6 Conclusion

What we find in this work can be summarized as follows: there'll be more opinion leaders in a knowledge market if the most active knowledge consumers occupy more central positions in a social network connecting consumers. Ceteris paribus, the structure of consumer interactions matters. Compared with existing literature on network effects, the network effects explored in this paper are non-homogenously networked rather than well-stirred. This is a feature of both our theoretical and empirical parts. Moreover, understanding knowledge markets, especially formally understanding what impacts the number of opinion leaders and formally understanding the sources of opinion diversity, is a less explored field. The key variables both in our theory and in our empirical work are dealt with only intuitively in the existing literature. To the best of our knowledge, ours is the first paper to formally define opinion leaders, knowledge consumers, and amount of attention. Last but not least, the paper makes some methodological contributions by constructing a new type of IV based on a multi-layer network perspective, by using network community detection techniques to construct a panel data structure, and by emphasizing the embeddedness perspective.

Appendix: robustness check

Table A1: the main results (top 0.5% followee, 10% threshold, mean)

	N	S^{in}	1st-stage F for excluded IV	Statistically equal to its 2sls
r1	1.28e-01 (0.014**)	3.06e-04 (0.003***)		Yes
r2	2.40e+02 (0.043**)	1.08e+00 (0.00001***)		Yes
2SLS: r1 r1iv as iv	1.09e-01 (0.027**)	3.15e-04 (0.002***)	900	
2SLS: r2 r2iv as iv	2.34e+02 (0.059*)	1.13e+00 (0.000003***)	900	

Table A2: the main results (top 0.5% followee, 20% threshold, median)

	N	S^{in}	1st-stage F for excluded IV	Statistically equal to its 2sls
r1	6.73e-02 (0.031**)	2.08e-04 (0.007***)		Yes
r2	1.34e+02 (0.077*)	7.76e-01 (0.001***)		Yes
2SLS: r1 r1iv as iv	5.70e-02 (0.044**)	2.14e-04 (0.005***)	1700	
2SLS: r2 r2iv as iv	1.26e+02 (0.094*)	7.99e-01 (0.001***)	1000	

Table A3: the main results (top 0.5% followee, 20% threshold, mean)

	N	S^{in}	1st-stage F for excluded IV	Statistically equal to its 2sls
r1	1.17e-01 (0.014**)	3.03e-04 (0.006***)		Yes
r2	2.12e+02 (0.039**)	1.04e+00 (0.0001***)		Yes
2SLS: r1 r1iv as iv	1.04e-01 (0.021**)	3.16e-04 (0.004***)	840	
2SLS: r2 r2iv as iv	2.11e+02 (0.048**)	1.10e+00 (0.00004***)	780	

Table A4: the main results (top 1% followee, 10% threshold, median)

	N	S^{in}	1st-stage F for excluded IV	Statistically equal to its 2sls
r1	6.10e-01 (0.0004***)	3.82e-04 (0.000005***)		Yes
r2	8.48e+02 (0.011**)	6.88e-01 (0.005***)		Yes
2SLS: r1 r1iv as iv	5.20e-01 (0.001***)	3.76e-04 (0.00002***)	1300	
2SLS: r2 r2iv as iv	7.22e+02 (0.022**)	6.28e-01 (0.013**)	530	

Table A5: the main results (top 1% followee, 10% threshold, mean)

	N	S^{in}	1st-stage F for excluded IV	Statistically equal to its 2sls
r1	5.97e-01 (0.001***)	3.63e-04 (0.00001***)		Yes
r2	8.76e+02 (0.011**)	6.96e-01 (0.001***)		Yes
2SLS: r1 r1iv as iv	5.29e-02 (0.001***)	3.67e-04 (0.00001***)	1400	
2SLS: r2 r2iv as iv	7.73e+02 (0.016**)	6.55e-01 (0.004***)	580	

Table A6: the main results (top 1% followee, 20% threshold, median)

	N	S^{in}	1st-stage F for excluded IV	Statistically equal to its 2sls
r1	3.11e-01 (0.004***)	3.00e-04 (0.002***)		Yes
r2	3.94e+02 (0.123)	6.28e-01 (0.047**)		Yes
2SLS: r1 r1iv as iv	2.37e-01 (0.048**)	2.95e-04 (0.006***)	1600	
2SLS: r2 r2iv as iv	2.79e+02 (0.30)	6.33e-01 (0.066*)	520	

Table A7: the main results (top 1% followee, 20% threshold, mean)

	N	S^{in}	1st-stage F for excluded IV	Statistically equal to its 2sls
r1	4.72e-01 (0.00004***)	3.18e-04 (0.002***)		Yes
r2	6.88e+02 (0.016**)	7.18e-01 (0.017**)		Yes
2SLS: r1 r1iv as iv	3.93e-01 (0.0005***)	3.25e-04 (0.002***)	1300	
2SLS: r2 r2iv as iv	5.66e+02 (0.045**)	7.47e-01 (0.023**)	530	

Chapter 2: Assessing External Validity

1. Introduction

In designing any causal study, steps must be taken to address both internal and external threats to its validity (see Campbell, 1957, and Cook and Campbell, 1979). Researchers tend to focus primarily on threats to internal validity, i.e., determining whether it is valid to infer that, within the context of a particular study, the differences in the dependent variables are caused by the differences in the relevant explanatory variables. External validity, on the other hand, concerns the extent to which a causal relationship holds over variations in persons, settings, and time. It is important to underscore the fact at the outset that external validity does not extend to modifications in the treatment, although, in practice, researchers often try to generalize their results by conflating the two levels of generalization into a question of external validity.

Randomized controlled trials solve the problem of selection bias in the identification of causal effects. Thus, theoretically, cause-effect constructs identified by means of randomized controlled trials are internally valid, that is, they permit the identification of causal effects for the population from which the random sample used in the estimation was drawn. The outcomes of such experiments are interesting in their own right, but researchers sometimes explicitly assume external validity (EV), i.e., that the internally valid estimates obtained for one population can be extrapolated to other populations. In fact, it is not uncommon that, after researchers have established a cause-and-effect relationship in a specific population, they proceed to discuss its implications based on the assumption that this relationship is generally valid. In this paper, we formalize the concept of external validity and show that, in general, it is unlikely that any given study will be externally valid in any general sense. This is one reason why Manski (2013) says that the current practice of policy analysis "hides uncertainty".

Once researchers have conducted an internally valid analysis, that analysis yields an established set of findings for the specific case in question. As for the future usefulness of that result, however, what matters is its degree of EV. The most commonly held view in this regard is that the EV problem hinges on assumptions about the relationship between the population for which internally valid estimates have been obtained and another, different population. Apart from researchers who are focusing on EV in a specific context, many researches either ignore the EV problem altogether or approach it subjectively. In this paper, we provide a formal and general reflection on the EV problem and propose a simple and generally applicable method for evaluating the external validity of randomly controlled trials (RCTs).

In this paper we define external validity as the stability of the conditional distribution p(outcome | treatment) across different populations. We then formalize the degree to which we can make judgments about a new population (density) generated as a subpopulation from an overarching population that also
generates the 'original' population studied for which there is an internally valid estimate. Without loss of generality, assume that we have data that allows estimation of the joint distribution p(outcome, treatment). We then have p(outcome, treatment) = p(outcome | treatment) × p(treatment). Then, we say that there is external validity if for other data with a potentially different joint distribution of outcome and treatment, the conditional distribution p(outcome | treatment) stays the same.

Based on our framework, we then propose two alternative measures of external validity. To the best of our knowledge, we are the first to propose formal mathematical definitions of external validity and, on that basis and in the context of an RCT, to propose purely data-driven measures related to theoretical constructs.

Ultimately, the external validity of all causal estimates is established by replication in other datasets (Angrist, 2004).[18] Nevertheless, we would like to determine whether a given specific study can be generalized to other populations.[19] In this paper we propose a method to do precisely that. Our method applies to RCTs, but it should be noted that the issue of external validity is general and not restricted to RCTs, as shown in our formal and general reflection below.

The rest of this paper is structured as follows. In Section 2, we provide a formal and general reflection on the EV problem. Based on the model described in that section, in Section 3 we propose a simple and generally applicable method for assessing the external validity of RCTs. Finally, we present final remarks.

2. External Validity

A single experiment allows us to arrive at a point estimate for the population of cause-effect parameters. Assessing the EV of one causal parameter entails estimating treatment effects as a function of different populations. Thus, evaluating the EV of an internally valid estimate of a cause-effect parameter entails assessing a distribution of cause-effect parameters based on a single draw from it.

In this section, we formalize the concept of EV. We develop our framework in terms of population density functions because nearly all sample analyses are intended to characterize an underlying population. Thus, we assume it is always possible to

---

18 In the areas of labor and development economics, a number of studies use similar multi-country strategies to generalize cause-and-effect constructs. For example, Cruces and Galiani (2007) examine the effects of fertility on labor outcomes in three countries; Dehejia, Pop-Eleches, and Samii (2019) examine the causal effects of sibling sex composition on fertility and labor supply across many countries and years and characterize how its effects vary in terms of available covariates; Banerjee et al. (2015) study microcredit in six countries; Galiani et al. (2017) study the effects of sheltering the poor in three countries; Gertler et al. (2015) study health promotion in four countries; Dupas et al. (2016) examine the effects of opening savings accounts in three different countries; and Galiani et al. (2016) investigate slum upgrading in three countries.

19 For example, Deaton (2010) writes: "*We need to know when we can use local results, from instrumental variables, from RCTs, or from nonexperimental analyses, in contexts other than those in which they were obtained.*"

obtain consistent estimates of the joint density of outcomes and treatment status or the conditional density of outcomes conditional on treatment status. Focusing on population densities might seem unnecessary, since most researchers need to model only the first and second moments of a population (density) to obtain their parameters of interest. The first moment is needed for a point estimate, while the second moment is used for evaluating sampling variability. Nevertheless, both moments are a function of a population density, i.e., a density of outcomes conditional on treatment status. We focus on population densities here because we want to emphasize that the nature and difficulty of assessing EV lies in the differences among populations. In addition, at a conceptual level, this simplifies the analysis, since it is necessary to make comparisons for only one entity (the density) instead of two (the first and second moments).

2.1. From one population to another

Conducting external inference from internally valid estimates entails switching the population under study. Assume that there is an overarching population that consists of all vectors (y, z, w) from the probability density D(y, z, w; θ). Also assume that we obtain a sample from a subpopulation defined by the density D(y, z; $w = w_0, \theta$). When we say something, based on estimates from this sample, about another subpopulation defined by the density D(y, z; $w = w_1, \theta$), for some $w_1 \neq w_0$, we are conducting external inference.

This is the general setup: w defines, in a general way, the difference between populations; internally valid inferences will usually yield different estimates of the cause-effect constructs of interest. θ governs how the differences in w affect those constructs across populations. Alternatively, we could define different populations by assuming that their w's are distributed in different ranges instead of assuming that they take different point values, but this change would not add any insight to the analysis. Actually, the assumptions are conceptually equivalent.

w_0 is a given realization of w and is constant for the subpopulation for which the sample is used to conduct the empirical analysis. For example, if we draw a sample within a country, all different sample points have the same country identity (w_0). With the sample drawn from D(y, z; $w = w_0, \theta$), whether we estimate the joint density of (y, z) or a conditional density of y on z, we always conduct inference under the condition w = w_0. Then, conducting external inference implies assessing whether such estimates are valid for a different population (in our example, from another country), characterized by $w = w_1$.

Researchers generally do not have information about how changing w would change the density D(y, z, w; θ). Usually, they have to make assumptions in this regard in order to conduct external inference. We now explore this question formally. It is informative to express D(y, z; $w = w_0, \theta$) as follows:

$$D(y, z; w = w_0, \theta) = D(y|z; \theta_1(w_0, \theta) \times D(z; \theta_2(w_0, \theta))$$

Assume that the estimand is the conditional density D(y|z; $\theta_1(w_0, \theta)$). In the case of an RCT, the marginal density $D(z; \theta_2(w_0, \theta))$ can be ignored, since z is randomly assigned. Thus, assume we estimate $D\big(y\big|z;\ \theta_1(w_0, \theta)\big)$ and then want to know how well we would assess another population characterized as $w = w_1$ if we rely only on our internally valid estimate. Assuming the conditional density function is differentiable almost everywhere with respect to w and applying the mean value theorem, we obtain:

$$\int \left[D\big(y\big|z;\ \theta_1(w_0, \theta)\big) - D\big(y\big|z;\ \theta_1(w_1, \theta)\big)\right]^2 dy$$

$$= (w_1 - w_0)^2 \int \left(\frac{\partial D\big(y\big|z;\ \theta_1(w_b, \theta)\big)}{\partial \theta_1} \frac{\partial \theta_1}{\partial w_b}\right)^2 dy$$

where w_b takes a value between w_1 and w_0. If w_1 is not known but can be assumed to be close to w_0, we can approximate w_b by w_0 (where the higher order residual in a

Taylor expansion is negligible). Then, for a close w_1, the more sensitive the conditional density with respect to w is, the more we will miss the target while conducting external inference by relying only on an internally valid estimate of the estimand of interest in our sample. Importantly, this is independent of the sample size.

Relying on this setup, we define:

(1) Punctual local external validity when $\mathrm{w}_1 = \mathrm{w}_0$;

(2) Local external validity as $\int (\frac{\partial \mathrm{D}\left(\mathrm{y}\middle|\mathrm{z};\ \theta_1(\mathrm{w}_1, \theta)\right)}{\partial \theta_1} \frac{\partial \theta_1}{\partial \mathrm{w}_1})^2 dy = 0$ for all w_1 within a small interval of w_0;

(3) External validity as $\int (\frac{\partial \mathrm{D}\left(\mathrm{y}\middle|\mathrm{z};\ \theta_1(\mathrm{w}, \theta)\right)}{\partial \theta_1} \frac{\partial \theta_1}{\partial \mathrm{w}})^2 dy = 0$ for all w_1;

(4) Indirect External validity where there exist $f(\mathrm{w}; \theta_1) = \int (\frac{\partial \mathrm{D}\left(\mathrm{y}\middle|\mathrm{z};\ \theta_1(\mathrm{w}, \theta)\right)}{\partial \theta_1} \frac{\partial \theta_1}{\partial \mathrm{w}})^2 dy$, either known or estimable that can be used to adjust $\mathrm{D}\left(\mathrm{y}\middle|\mathrm{z};\ \theta_1(\mathrm{w}_0, \theta)\right)$ to calculate $\mathrm{D}\left(\mathrm{y}\middle|\mathrm{z};\ \theta_1(\mathrm{w}_1, \theta)\right)$.

In the literature, researchers conducting external inference have attempted to either test (1), (2), or (3) or to exploit (4). As our discussion makes clear, the question of external validity rests on the relationship between the population for which we have an internally valid estimate and the population about which we are to make judgments. The function $f(\mathrm{w}; \theta_1) = \int (\frac{\partial \mathrm{D}\left(\mathrm{y}\middle|\mathrm{z};\ \theta_1(\mathrm{w}, \theta)\right)}{\partial \theta_1} \frac{\partial \theta_1}{\partial \mathrm{w}})^2 dy$ formalizes this relationship.

2.2. From one population to any population

We now extend our analysis. The estimand is a conditional model based on data generated from the density $\mathrm{D}(\mathrm{y}, \mathrm{z};\ \mathrm{w} = \mathrm{w}_0, \theta)$. We denote the conditional model by $\mathrm{D}(\mathrm{y} \mid \mathrm{z};\ \mathrm{w} = \mathrm{w}_0, \theta_1)$, where θ_1 is the parameter governing the conditional model and is a function of w_0 and θ. Applying Bayes' law and assuming z is weakly exogenous (see Engle, Hendry, and Richard, 1983) for θ (this is always the case when z is randomly assigned), we have:

$$\mathrm{D}(\mathrm{y} \mid \mathrm{z};\ \mathrm{w}=\mathrm{w}_0, \theta_1) = \frac{\mathrm{D}(\mathrm{y,z;\ w=w_0,\theta_2})}{\mathrm{D}(\mathrm{y,z;\ \theta})} \times \mathrm{D}(\mathrm{y|z;\theta_3}),$$

where $\mathrm{D}(\mathrm{y|z;\theta_3}) = \int \mathrm{D}(\mathrm{y|\ z,w;\theta})\, dw$, $\mathrm{D}(\mathrm{y,z;\ \theta}) = \int \mathrm{D}(\mathrm{y,z,w;\theta})\, dw$, and θ_3 is a function of θ. Note that here we use $\mathrm{D}(\mathrm{y,z;\ w=w_0,\ \theta_2})$, where θ_2 is a function of θ, rather than $\mathrm{D}(\mathrm{y,z;\ w=w_0,\ \theta})$, because we want to emphasize that, when moving from D(y, z, w; θ) to $\mathrm{D}(\mathrm{y,z;\ w=w_0,\ \theta_2})$, the parameter vector θ may change. θ_2 can be seen as a function of θ and w. However, there is nothing wrong with using the general notation $\mathrm{D}(\mathrm{y,z;\ w=w_0,\ \theta})$ instead. External inference in this setting entails assessing $\mathrm{D}(\mathrm{y|z;\theta_3})$ based on the estimation of $\mathrm{D}(\mathrm{y} \mid \mathrm{z};\ \mathrm{w}=\mathrm{w}_0, \theta_1)$ (i.e., from a subpopulation to the overarching population). Note also that θ_2 and θ_3 are not variation-free (Engle, Hendry, and Richard, 1983), so θ_3, estimated by optimizing a loss function based on $\mathrm{D}(\mathrm{y} \mid \mathrm{z};\ \mathrm{w}=\mathrm{w}_0, \theta_1)$, generally does not coincide with the result obtained by estimating on the basis of an optimization of a loss function based on $\mathrm{D}(\mathrm{y|z;\theta_3})$, which researchers cannot optimize in any event, since the available data is generated only from the subpopulation with $\mathrm{w}=\mathrm{w}_0$. Taking these considerations into account, we define:

(1) Overarching External Validity as $\frac{\mathrm{D}(\mathrm{y,z;\ w=w_0,\theta_2})}{\mathrm{D}(\mathrm{y,z;\ \theta})} = 1$.

(2) Indirect Overarching External Validly when the function $\mathrm{g}(\theta) = \frac{\mathrm{D}(\mathrm{y,z;\ w=w_0,\theta_2})}{\mathrm{D}(\mathrm{y,z;\ \theta})}$ is either known or estimable and can be used to adjust a loss function based on $\mathrm{D}(\mathrm{y} \mid \mathrm{z};\ \mathrm{w}=\mathrm{w}_0, \theta_1)$ that makes it possible to estimate $\mathrm{D}(\mathrm{y|z;\theta_3})$. Clearly, this requires a known (or assumed) relationship between the subpopulation under study, $\mathrm{D}(\mathrm{y,z;\ w=w_0,\ \theta_2})$, and the overarching population, $\mathrm{D}(\mathrm{y,z;\ \theta})$, which is given by $\mathrm{D}(\mathrm{y,z;\ \theta}) = \int \mathrm{D}(\mathrm{y,z,w;\theta})\, dw$.

From the above discussion, it is clear that it is more likely that an internally valid estimate will have punctual or local external validity than external validity or overarching external validity. Thus, not surprisingly, the existing literature has focused on specific populations for extrapolation, making specific assumptions about the relationship between the population for which there are internally valid estimates and the target population. Next, we review that literature.

2.3. Literature Review

One reason why internally valid estimates of causal constructs might lack external validity is because of changes in the population over time. We could posit, for example, that w in the above framework varies over time. In such a setup, Rosenzweig and Udry (2018) provide an innovative way of conducting external inference. Using repeated cross-sections, they estimate the causal effect of interest over time, where in each period the vector w is fixed at some specific value. They focus on one dimension of w for which they have a measurement, i.e., rainfall. They then estimate the response of the casual construct of interest to rainfall. Using the empirical distribution of the underlying shock (rainfall), they can infer both how the causal parameter of interest varies with this shock and its average effect. Thus, they also estimate the effect for the overarching population. This method requires that other time-varying unobservable variables in w are not correlated with the observable one, which, in their application, may be the case, since rainfall is determined outside the economic system, although it still might trigger adjustments in some unobservable variables.

Andrews and Oster (2019) propose a method for estimating the average treatment effect (ATE) for a target population based on another population (often a trial population) for which a researcher is assumed to have an internally valid estimate. They assume that the conditional ATE-given covariates and unobservables are the same in the trial and target populations (and that the covariates and unobservables are uncorrelated). First, they adjust the ATE by differences in covariates between the trial and the target populations. Second, they model how unobservables and covariates simultaneously affect individual treatment effects and the likelihood that individuals in the target population were also in the trial population. Relying on this model, they derive a formula to adjust the ATE for differences in unobservables.

Other papers have dealt with the issue of non-compliance in instrumental variables estimation. One line of discussion about external validity relates to the local average treatment effect (LATE) (see Angrist, Imbens, and Rubin, 1996). The standard setup assumes the presence of a binary endogenous treatment variable instrumented with a binary ignorable variable (for example, random assignment to treatment). Assuming monotonicity, the population is divided into three groups: (1) compliers, whose treatment status is affected by the instrument; (2) always-takers, who always receive treatment regardless of the value of the instrumental variable; and (3) never-takers, who never receive treatment regardless of the value of the instrumental variable. The second and third groups are usually combined and labeled as non-compliers. The variation (information) in the instrument only makes a difference for compliers. Since, in this setting, internally valid estimates are derived from the variation in the instrument, the estimates do not provide a basis for internally valid inferences about the whole population, but only about a hypothetical population of compliers. The question with regard to external validity that usually arises in this setting has to do with when and how estimates for compliers can be used to infer the parameter of interest for the whole population. Naturally, the first step in answering this question is to understand the relationship between compliers and non-compliers. Examples in the literature include Angrist (2004) and Angrist and Fernandez-Val (2010). Angrist (2004) examines a few possible relationships between compliers and non-compliers that yield externally valid inferences and then estimates the ATE using information for the LATE under each relationship. Angrist and Fernandez-Val (2010) assume that the instrumental variables, conditional on covariates, are as good as if they were randomly assigned and that observable covariates fully determine covariate-specific treatment effects. The relationship between compliers and non-compliers is then reduced to different compositions of observable covariate values. Since the LATE and ATE are both weighted sums of (observable) covariate-specific treatment effects, the EV problem of differentiating the ATE from the LATE becomes one of modifying the weights used in the LATE to align them with the weights used in the ATE.

Another line of analysis concerning external validity involves regression discontinuity methods. A regression discontinuity estimator is, by definition, a local estimator: it only identifies causal constructs for the subpopulation of subjects whose forcing variable values are near a discontinuity threshold (and are also compliers in the case of a fuzzy design). The external validity question usually asked in such a setting is how estimates for the subjects near the threshold (and that are also compliers) apply to the sample population. Naturally, the first step in assessing external validity in this setting is to understand the difference between the subjects whose treatment effect can be identified and other subjects. Dong and Lewbel (2015) exploit the sensitivity of estimates of the forcing variable to shed light on the relationship between subjects with different forcing variable values. Angrist and Rokkanen (2015) advocate testing whether the forcing variable and the "treatment" outcome are uncorrelated conditional on variables, which, if it were the case, would be informative about the relationship between subjects with different forcing variable values and, in turn, could be of use in addressing external validity questions. Bertanha and Imbens (2018) focus on the fuzzy regression discontinuity design and provide a test to determine whether compliers are systematically different from non-compliers conditional on the forcing variable as well as exogenous covariates. They argue that external validity requires the null hypothesis of no differences in order to be valid.

These studies share a common feature: the pivotal element in approaching the issue of external validity is assumptions about the relationship between the population for which there are internally valid estimates of causal parameters and the population for which the researcher would like to make an external inference. The relationships exploited in the above-cited studies are the one between compliers and non-compliers, the one between subjects with different forcing variable values, and the one between different periods. The effectiveness of any practical evaluation of external validity is determined by these relationships, as we explain in Sections 2.1 and 2.2.

These papers focus on specific populations as a basis for extrapolations and do not explore EV in any general form. This is natural enough, since, for any given population, the relationship between it and the population that was originally studied can be easily assumed or modeled, while it is very hard to undertake an evaluation of EV in general. Thus, in terms of our analysis, the literature has focused mostly on methods relating to the concept of local external validity.

In the next section, we propose a method for assessing EV both for specific populations and generally. Our method is based on the insights derived from Sections 2.1 and 2.2. The focus is on determining the likelihood that an internally valid estimate could be generalized not only to the overarching population (overarching external validity) but also to specific populations, starting from those close (local external validity) to the one studied and then moving away from it to more different populations (external validity).

3. Assessing external validity

We propose a method for evaluating the degree to which a conclusion based on a given population applies to populations represented by samples that have been formed by randomly reweighting the original sample. Our method is therefore both data-driven and generally applicable.

In order to maintain consistency with the theoretical framework outlined in Section 2, we assume that each reweighting corresponds to a value of w. After defining and constructing new populations, we propose a way of measuring the extent to which the conclusion reached about the original sample holds true for the reweighted samples (or the new populations). This is a global measure of EV (based on the concept of overarching external validity). We also propose a local measure of EV (based on the concepts of local external validity and external validity) by grouping new populations based on a specific criterion and then measuring the degree to which the conclusion for the original sample still holds for each group.

3.1 Defining new populations and constructing representative samples

We assume that a researcher starts from a randomized controlled trial (RCT) dealing with a given population and then wants to assess the degree of EV of the causal constructs that have been estimated. We provide a general method for doing this. Our method, inspired by our formal and general reflection above, takes advantage of RCTs as shown below. However, it should be noted that the problem of external validity is general and not restricted to RCTs.

Assume that the sample size of the control group is m and the sample size of the treatment group is l, so the sample size analyzed is n=l+m. Assume also that there is baseline pre-treatment information (Y, X)’, which is usually the case. We pair each observation in the treatment group with its nearest observation in the control group in terms of the Mahalanobis distance[20] using the baseline information (the Mahalanobis distance between vectors of baseline information (Y, X)’). We thus choose the nearest neighbor in the control group to each observation in the treatment group as a counterfactual and pair the two observations. Repetitive use of observations from the control group are allowed, and unused observations from the control group are discarded after all observations from the treatment group are paired. For each pair, we assign an index $i \in \{1,2,\ldots l\}$.

[20] The Mahalanobis distance can be defined as the Euclidean distance with each variable rescaled to have unit variance. Though this distance is the most commonly used measure in the literature, there are many alternative matching criteria (Rosenbaum, 2010) that researchers can use for their specific purposes. In addition, researcher can match one sample point for a group (treatment or control group) with many sample points from the other group.

We define a reweighting vector P for the indices: $P = (p_1, p_2, \dots, p_l)' = \frac{(G_1, G_2, \dots, G_l)'}{\sum_1^l G_i}$, with $p_i \geq 0$ and $\sum_1^l p_i = 1$. $(G_1, G_2, \dots, G_l)'$ is a random vector, with each element drawn independently from the Gama(1,1) distribution. Reweighting the original sample based on the Gama(1,1) generates reweighting vectors uniformly distributed over all possible reweighting vectors (Efron and Hastie, 2016), so our exploration of how the original conclusions apply to new populations treats every possible population evenly. For each reweighting $P = (p_1, p_2, \dots, p_l)' = \frac{(G_1, G_2, \dots, G_l)'}{\sum_1^l G_i}$ of the original sample, we create a reweighted sample in which p_i is the weight for the pair indexed by i=1, 2, … , l. We do multiple reweighting (as many as 1,000 times), and we regard each reweighting outcome as a sample representing a new population. Weighting the actual sample by $\left(\frac{1}{l}, \frac{1}{l}, \dots, \frac{1}{l}\right)'$ is a consistent estimator of the original population. Therefore, each weighting vector corresponds to a consistent estimate of a new population. The new populations arrived at by reweighting the original sample are deemed to be conceptually valid in the light of the following citation:

"[…] the only populations that can be referred to in a test of significance have no objective reality, being exclusively the product of the statistician's imagination […]."(Fisher, 1956)

3.2 A global measure of EV

We now propose a global measure of EV for the average treatment effect (ATE), but our measure can be applied to any estimand obtained by contrasting treatment and control groups in an RCT. For each reweighted sample, we calculate the ATE and its standard error. Our proposed global measure of EV for the original internally valid analysis is based on the percentage of new populations for which the original conclusion still holds true. We say that the original conclusion holds when one of the following holds: if, in the original sample, the estimate was not statistically significant at a certain level, then the same is true for the reweighted sample or, if the estimate for the original sample was statistically significant at a certain level, then, in the reweighted sample it is also significant, at least at that level, and with the same coefficient sign as in the original population.

We start from an RCT and estimate the new populations by means of a matching estimator with replacement. Abadie and Imbens (2006) set out conditions for the consistency of this matching estimator.[21] However, in finite samples, there are two sources of bias due to imperfect matching on observables and non-matching on unobservables. For the second source of bias, researchers can perform a sensitivity analysis as introduced by Rosenbaum (2010) and Imbens and Rubin (2015).

Correcting the first source of bias can be done using the methods discussed in Imbens and Rubin (2015, sect. 18.8). We use one of those methods to adjust the estimates for new populations and therefore, from now on, when we refer to the estimated ATE for new populations, we are referring to bias-adjusted estimates, as follows:

1. With the data from the original control group, we have regressed the outcomes on observable variables and recorded coefficients for baseline variables as *B*.
2. We have adjusted each pair's treatment-minus-control outcome by – *(observable variables of the treated in the pair)*B + (observable variable of the control in the pair)*B*.

Up until now, we had assumed that all the subjects were compliers in the original sample or we had focused on the intention-to-treat effect (ITT). However, there is no problem with including non-compliers in the analysis and focusing on a parameter such as the LATE since the reweighted samples are expected to be balanced for non-compliers, especially when the sample size is large.

[21] The data and treatment assignment from an RCT satisfy the conditions required for consistency given in Abadie and Imbens (2006). The matching estimator here is a weighted sum, but given our way of generating reweighting vectors, when the sample size goes to infinity, the probability that a finite number of pairs receive all the weight goes to zero, so by applying a law of large numbers, such as Chebychev's weak law of large numbers, consistency is proved.

Alternatively, one possibility is to estimate the average effect on compliers. Under standard assumptions, we know who the never-takers in the treatment group are. In their case, we can improve on the matching with the control group because we can add the residuals of the treatment effect analysis (in addition to the baseline variables already used) to the matching variables. Thus, we can also match on unobservables. This matching strategy does not work for compliers, since their residuals are affected by their heterogeneous treatment effects. Always-takers can be matched in the same fashion. After matching, researchers can restrict their analysis to compliers both in the original sample and in the EV exercises proposed in this paper.

3.3 Local measures of EV

With a large number (e.g., 1,000) of reweighted samples, each representing a new population, these populations can be grouped based on a given criterion. We now have a vector of treatment-control matched pairs from the original sample. Each pair yields a treatment-minus-control outcome, adjusted as proposed in the previous section based on Imbens and Rubin (2015). First, we calculate the correlation between the vector of these adjusted treatment-minus-control outcomes and each reweighting vector. The higher this correlation is, the more weight a reweighting vector gives to pairs with high treatment-minus-control outcomes. Second, we calculate 1 minus this correlation for each reweighting vector (i.e., for each new population); this calculation gives the distance of a new population from the population with the largest effect magnitude. Note that the original population has a distance of 1 because the above correlation for the original population is zero, so the extent of the difference between other populations and the original population can be summarized by how their distance measure differs from 1.

Intuitively, populations with a distance measure close to 1 give similar weights to pairs with high or low adjusted treatment-minus-control outcomes, so they are “near” the original population, which gives equal weight to every pair. Populations with a distance measure greater (smaller) than 1 have a weighting vector that is negatively (positively) correlated with the vector of the adjusted treatment-minus-control outcomes, so they give more weight to pairs with small (large) effect magnitudes.

With the above definition of distance, we then also propose using the EV curve to measure the degree to which the conclusion regarding the original population holds for new populations as their distance from the population with the largest effect magnitude (moving away from distance zero) or from the original population increases (moving away from distance 1). The EV curve is defined below, with distance denoted by d.

$$\mathrm{EV(d)} = \frac{\text{Number of new populations for which the original result holds at distance} \in [\mathrm{d}, \mathrm{d} + \epsilon]}{\text{Number of new populations at distance} \in [\mathrm{d}, \mathrm{d} + \epsilon]}$$

In the definition of the EV curve, the original result holds when, as above, one of the following holds: if, in the original sample the estimate was not statistically significant at a certain level, then the same is true for the reweighted sample or, if the estimate was originally significant at a certain level, then it is also significant, at least at that level, and with the same coefficient sign as the new population.

Before we provide examples illustrating the two methods proposed above, we need to add a caveat that applies to both of them. If the sample at hand or the population represented by the sample being studied does not contain characteristics that would generate a new population and are relevant for the statistical inference, then our method is moot. One cannot make bricks without straw. Any statistical method can be useful only up to the point that its information inputs allow. Our method is designed to provide the best possible assessment of EV based on a single sample (a single population), which is usually all that researchers have.

3.4. Two Simulated Examples

We now provide two simulated examples to illustrate our method for assessing EV. First, we start from an internally valid analysis assumed to have a significantly positive ATE. We then calculate the global measure of EV and present the EV curve to be used to assess EV locally. Second, we do the same exercise starting from an internally valid analysis that is assumed to have an ATE that is not significantly different from zero.

3.4.1 Assessing EV for an internally valid significant positive result

In the first simulated example, the sample size is 100, with 50 observations in the treatment group and 50 in the control group. There are two observable variables, x1 and x2, and one unobservable variable, u. For each of these three variables, 100 values are drawn independently from the standard normal distribution. The (potential) heterogeneous treatment effect for each observation, whether it is in the treatment group or control group, is given by $\tau_i = x_{1i} + x_{2i} + u_i + v_i + 10 * 1\{c_i > 0.8\}$, where v_i is drawn from a unit-variance normal distribution with a mean=-1, and c_i is drawn from the 0-1 uniform distribution.

We define the treatment status vector, a vector with a length of 100, as D, of which the first 50 elements are equal to 1 and the other 50 elements are equal to 0. D_i=1 indicates that observation i is in the treatment group; D_i=0 indicates that observation i is in the control group. The outcome variable is thus defined as: $y_i = 1 + \tau_i * D_i + x_{1i} + x_{2i} + u_i$. The internally valid estimate of ATE in our simulation is 1.52 with a standard error equal to 0.677 and a p-value for a test of the null hypothesis (ATE equals 0) equal to 0.027.

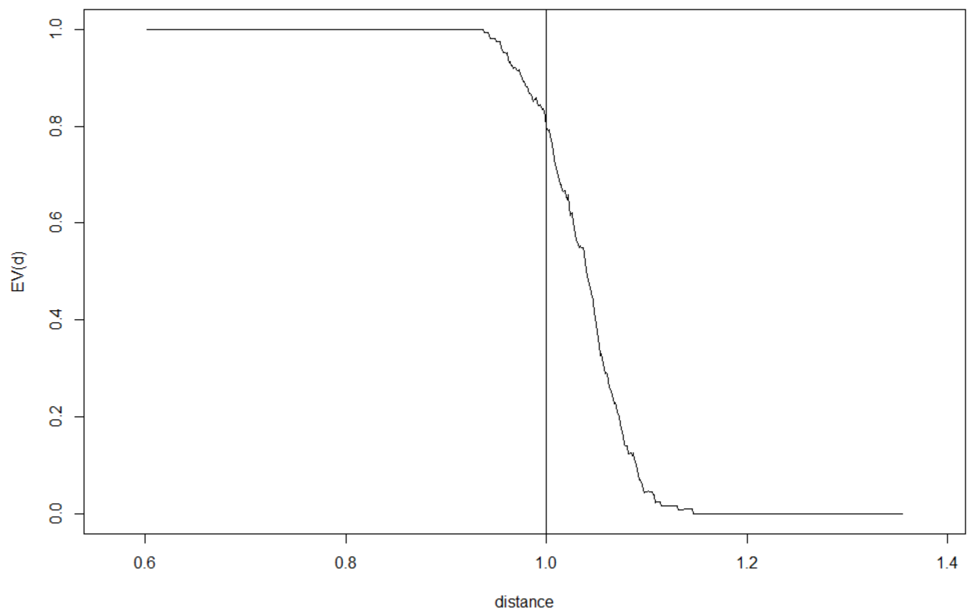
We generate 1,000 reweighting vectors over the pair indices as explained in Section 3.1 and calculate the global measure of EV introduced in Section 3.2. In this case, this measure is the proportion of these 1,000 new populations for which the lower bound of the confidence interval is greater than zero. Since the original analysis is significant at the 95% level, we choose the lower bound associated with two standard errors. The value of the global measure of EV in our simulated example is 0.584. This means that the conclusion reached in the original internally valid analysis holds for 58.4% of the uniformly generated new populations.

Now we compute the EV curve, EV(d), introduced in Section 3.3. To apply the definition of EV(d)[22], we note that when the original result holds for a new population, this means that the confidence interval lower bound for the new population is greater than zero. Since the original analysis is significant at the 95% level, we choose the lower bound for new populations as point estimates - two standard errors. If we look at the curve starting from a distance equal to 1, we see the positions of the new populations relative to the original population, whose distance measure equals 1. (Remember that the distance is a measurement of the distance of a new population from the population with the largest possible effect magnitude.) In Figure 1 below, we see that the original conclusion of positive significance is very likely to hold at a small distance (around 1 and smaller than 1) and it is very unlikely to hold at a large distance. Intuitively, new populations with small distances have more weights on pairs with large effect magnitudes and those with large distances have more weights on pairs with small or even negative effect magnitudes. In this example, EV is assessed locally by seeing how quickly EV(d) drops as the distance moves to the right and away from distance 1 (the distance for the original population),

[22] With respect to the choice of ϵ in the definition of EV(d), we choose a value of 0.05 in both this and the next example in order to make the curve smooth. This works like a moving average that smooths out a graph. Note that if ϵ is too small, the curve will be very rugged locally; if it is too large, the curve will not be locally informative, since it will simply be an overall average.

around which EV(d) is about 80%.

Figure 1



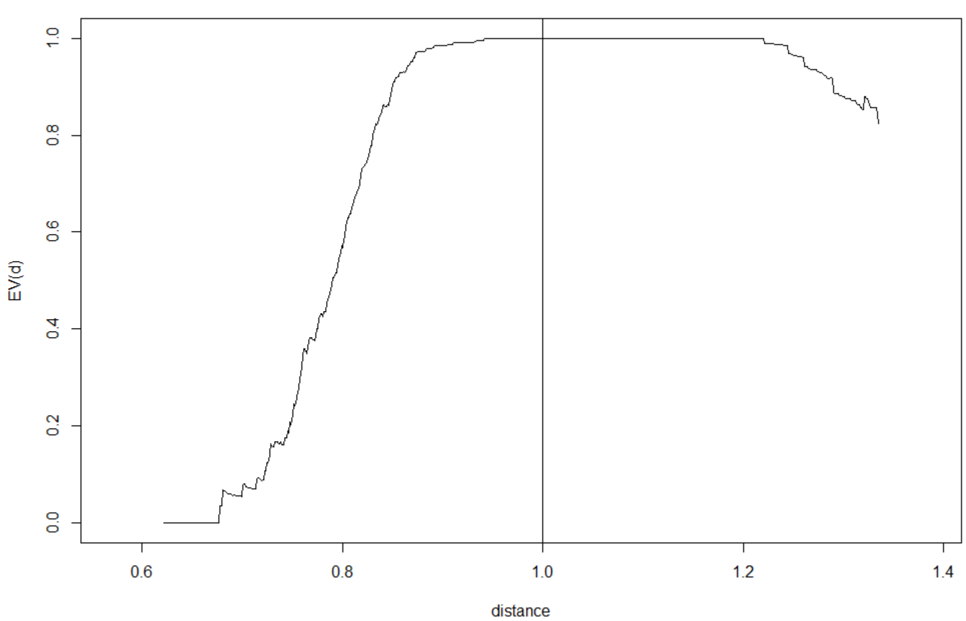
3.4.2 Assessing EV for an internally valid result without a significant difference from zero

In the second simulated example, the setup is the same as in the first example except that the (potential) heterogeneous treatment effect is instead given by $\tau_i = x_{1i} + x_{2i} + u_i + w_i + 10 * 1\{c_i > 0.8\}$, where w_i, instead of v_i as in the previous example, is drawn from a unit-variance normal distribution with a mean=-2. x_{1i}, x_{2i}, u_i, c_i, D_i, and y_i are generated in the same way as in the previous example. The internally valid estimate of ATE in our simulation is then 0.52 with a standard error equal to 0.677 and a p-value for a test of the null hypothesis (ATE equals 0) that equals 0.446.

We generate 1,000 reweighting vectors over the pair indices as explained in Section 3.1 and calculate the global measure of EV introduced in Section 3.2. In this case, this measure is the proportion of the new populations with confidence intervals including zero in these 1,000 new populations. Since the original analysis yields an estimate not significantly different from 0 at the 95% significance level, we choose the lower bound of the confidence interval associated with two standard errors. The value of the global measure of EV in our simulated example is 0.908. This means that the conclusion reached in the original internally valid analysis holds for 90.8% of the uniformly generated new populations from reweighting the original sample.

Now we compute the EV curve, EV(d), introduced in Section 3.3. Applying the definition of EV(d), we note that when the original result holds for a new population, this means that the new population's confidence interval includes zero. Since the original analysis yields an estimate not significantly different from 0 at the 95% significance level, we choose the range of confidence intervals for new populations as point estimates $\pm$ two standard errors. In Figure 2 below, as expected, we see that the original conclusion of no significance (neither significantly positive nor significantly negative) is very unlikely to hold at very large or very small distances. We also see that, as new populations move closer to the original population, whose distance measure is equal to 1, EV(d) increases. In Figure 2, we see that EV(d) is 100% in the small neighborhood of the original population and that EV(d) eventually drops as the distance measurement moves away from 1.

Figure 2



4. Final Remarks

Our method of evaluating EV is based on our theoretical definitions of external validity. It has become clear that, in order to achieve external validity in a practical sense, we need to identify new populations whose relationship with the population represented by the original sample is reasonable and workable. In particular, we assume that each specific w which defined new populations in Section 2 corresponds to a new weighting.

Our method of evaluating external validity is purely data-driven, but theory can play an important role in valid extrapolation (Getchter et al., 2018; Deaton, 2010; Wolpin, 2013). As discussed above, our method goes only as far as the information contained in the sample (original population) allows it to go. When a researcher wants to say something about a new population with inference-relevant characteristics that are absent from the original population, he or she needs to make further assumptions and to model certain mechanisms. One fruitful line of future work could be to use a combination of theoretical and experimental approaches to measure the generalizability of those mechanisms.

Once researchers have conducted an internally valid analysis, that analysis yields an established set of findings for the specific case in question. As for the future usefulness of that result, however, what matters is its degree of EV. To design for EV, what is wanted is a sample that includes as many different subjects as possible, ones that do not necessarily represent the original population. Specifically, if, for the population studied in an internally valid analysis, very small weights or no weight at all are assigned to some kinds of subjects, then the sample at hand may include very few such subjects or even none at all; if this is the case, such subjects will have very little chance of being represented in new populations. This limitation of the original sample limits the assessment of EV. Thus, stratification at sampling may enhance EV analysis. Similarly, the use of non-representative samples may also facilitate EV analysis. This issue requires further investigation, however.

Chapter 3: Development Spillover and Institutional Changes: A Trade Economy Political Perspective

1. Introduction

The importance of good institutions for economic development has been well argued and documented (Acemoglu, Johnson and Robinson, 2012; Acemoglu, Johnson and Robinson, 2005a; North, 1990). Better protected property rights lead to more expected payoffs from economic activities (Besley and Ghatak, 2010); with increased expected payoffs, people tend to invest and trade more, and consequently more gains of trade promote economic development (Yang, 2003). That trade and specialization stimulate growth has been emphasized since Adam Smith, and it has been well argued that more can be produced with the technology of increasing returns to scale, if people specialize and trade instead of producing everything by themselves. Such gains of trade can affect and be affected by institutions: improved institutions can increase expected payoffs for people who trade and induce more people to trade as pointed out above, while potential gains of trade could serve as stimuli for people to push for better institutions (Acemoglu, Johnson and Robinson, 2005b).

Some empirical observations suggest potential relationships between trade and institutions. Kelejian, Murrell and Shepotylo (2014) show spatial spillover in institutional development between neighboring countries that also tend to trade more with each other (Limao and Venables, 2001). Other empirical evidence shows specific interactions between institutions and trade in historical contexts. Acemoglu, Johnson and Robinson (2005b) show the positive impact of Atlantic trade on institutional changes in early modern Europe; Keller and Shiue (2016) show how French rule in 19th century Germany expanded trade. Theories have been provided to explain interactions between institutions and trade, and each emphasizes different mechanisms, e.g., international contractibility (Antras, 2015) and the political economy of trade policies (Galiani and Torrens, 2014). This paper differs from the existing literature with a political economy perspective of trade costs and provides a theoretical framework formalizing mutual promotion between trade and institutional changes, as well as how such interdependency matters for economic development. It should be noted that it's possible for a multitude of mechanisms to coexist or different mechanisms to be conjured up under different conditions.

In particular, I build a theoretical model that comprises an economic equilibrium capturing how institutional quality affects trade and a concatenated political equilibrium capturing how trade affects institutional changes, as well as the implication of the interdependency on economic development. The model features an integration of economic and political equilibria, which is inspired by Galiani, Schofield, and Torrens (2014). What directly follows is an intuitive summary of my model. First, gains from specialization encourage trade, while trade costs arising from poor institutional quality discourage trade. I define trade costs as a combination of transportation costs and elite expropriation on trade. Thus, more people participate in trade under lower trade costs or better institutions. Furthermore, the better the institutional quality of a trade partner, the more prosperous an economy is in terms of both trade and per capita utility/income, because more people in the partner country will trade and demand the economy's goods. Additionally, more people dedicated to trade imply a larger commercial class in an economy, and a larger commercial class has a higher bargaining power against the local extractive elite whose expropriation partially explains trade costs. The extractive elite trades off between potential social conflicts and avoiding conflicts by decreasing expropriation (reform) when challenged by the commercial class, while the commercial class trades off between fighting for a better institution (lower expropriation) with a losing probability and leaving the reality as it is. These tradeoffs and common knowledge of the elite's anticipation of the commercial class's response generate strategic interaction, which is to be modeled in a political game. Finally, the interaction leads to institutional changes if initial institutional quality is high enough but not very high: when initial institutional quality is very low (high), the commercial class has no capability (incentives) to fight. In Section 2 I build a model to formalize these intuitive arguments and to be specific with the economic and political implications and their required conditions.

In Section 3 I provide major comparative statics of the model and in Section 4 I provide historical evidence consistent with my results. The historical cases correspond to early modern Europe and 19th century Germany, where I find stories of trade, institutions and economic development that are consistent with my model. Briefly speaking, Acemoglu, Johnson, and Robinson (2005b) tell an empirical story about institutional changes induced by Atlantic trade in early modern Europe, which inspires and supports this work, namely, Keller and Shiue (2016) tell an empirical story about expanded trade induced by a trade partners' institutional changes (French rule) in 19th century Germany, which supports this work from another perspective. In Section 5 I conclude with a few comments on how this work is relevant to current issues as well as on methodology.

2. A Theoretical Framework of Political Trading Economy

Acemoglu, Johnson and Robinson (2005b) and Acemoglu and Robinson (2012) argue that Atlantic trade stimulated commercial class development for post-medieval European countries with relative better initial institutional quality (mainly in terms of more constraints on rulers' expropriation). These developments lead to institutional improvement that finally put the countries onto the track of modern economic growth. The theoretical framework below is inspired but not limited by this argument. In particular, I explore economic agents' tradeoffs between self-provision (without trade costs but with a fixed learning cost) and trade (without the fixed cost but with trade costs). I construct a model to formalize the idea that trade contributes to an economic unit's prosperity thanks to specialization, and such contribution is greater when institutional quality of the economic unit or its trade partner is better. Moreover, this model also clarifies a mechanism about how increased commercial community size induced by expanded trade can lead to institutional changes through reform or conflicts.

Next, we discuss the theoretical framework that comprises an economic equilibrium connecting trade, economic development, and institutions; and a political equilibrium connecting commercial community size, initial institutional quality and institutional changes.

2.1 The Economic Equilibrium

Inspired by Yang (2003, Chapters 4 and 7), I start with the decision problem of a consumer-producer who chooses whether to produce by herself all her consumption or trade for some consumption with her productions as a price taker. I do not adopt the commonly used neoclassical dichotomy of consumers and firms. By focusing on the consumption, production, and trade decisions all made by one economic agent, the model sheds insightful lights on tradeoffs that endogenize labor division and development of the commercial class. Consequently, the model can provide a perspective on why and how institutions matter. These will be made clear as I go on.

I assume there are two economic units (economy i and i') and two goods (good j and j') for consumption.Each economy has a unit mass of people (consumer-producers). I assume the two economies differ in two aspects. First, people in economy i (i') have advantages in the production and commercialization of good j (j')in the sense thatpeople in economy i (i') have to pay a fixed learning cost to produce and cannot commercialize good j'(j), while they need not pay a fixed cost to produce and can commercialize j (j').[23] I denote the person-specific learning cost with A and assume it is distributed uniformly between 0 and 1 in each economy. Second, if a consumer-producer in economy i chooses to trade, she has to pay a trade cost in the sense that she loses a fraction, $1 - \mathrm{K_i}$, of the goods she trades in.

[23] In a more general setup, people within the same economy could also trade with each other. I assume away this possibility mainly because it complicates mathematics a lot without changing model insights. As long as people from different economies trade with each other under the trade cost and the production (dis)advantage setup, the mechanism modeled and the model insights survive. If we allow for the possibility that people within the same economy could also trade with each other, only added is a new feature that inter-economy trade would disappear (or there would be an autarky) if trade costs are high enough. Since this is not the purpose of this model and complicates mathematics too much, I assume away the possibility. Besides, it could be assumed that when some people in economy i have learned to produce good j′, what they produce are simply home production and consumption satisfying the same desire as the bought exotic good j′ does, but could not be sold in the market. For a toy example, I can cook by myself or eat in a restaurant and have the same desire satisfied, but it’s highly costly for me to commercialize what I cook, though I may sell my carpentry work daily.

K_i (for people in economy i) is crucial for this model. I define $K_i = (1 - t_i)k_i$. t_i denotes physical trade costs, such as transportation costs, and$1 - t_i$ is the proportion left after physical trade costs are paid. I assume in each economy that there is an extractive elite who does not assume any productive role and expropriates a fraction, $1 - k_i$, of traded-in goods after physical trade costs are paid. So, the $(1 - t_i)k_i$ is a proportion of traded-in goods finally consumed by an economy i's consumer-producer who trades. k_i serves as a measure of institutional quality of economy i and is higher with less expropriation.

The decision problem of a consumer-producer

I start from the decision problem of an economy i's consumer-producer with her learning cost to produce good j' being A, which is person-specific and follows the same distribution in each economy. Economyi's consumer-producers do not need to pay learning costs to produce good j, as they are assumed to have advantage in it. Next, I focus on economy i (things are defined and derived analogically for economy i').

Utility to be maximized by a consumer-producer: $\mathbf{u_{iA} = x_j x_{j'}}$

Subscript **iA** means this consumer-producer is from economy i and her learning cost to produce good j' is A. $\mathbf{x_j}$ and $\mathbf{x_{j'}}$ respectively denote the amount of goods j and j' consumed by the consumer-producer.

Subject to Constraints (in addition to non-negativity of all variables):

(1) $\mathbf{x_j = K_i x_j^d + x_j^p - x_j^s;\ x_{j'} = K_i x_{j'}^d + x_{j'}^p;}$

x_j^d denotes the amount of good j demanded by the consumer-producer from trade, and the superscript "d" stands for "demand from trade." As discussed above, only a fraction of traded-in amount, $K_i x_j^d$, are consumed by the consumer-producer because of trade costs. x_j^p denotes the amount of good j produced by the consumer-producer, and the superscript p stands for "production." x_j^s denotes the amount of good 1 supplied by the consumer-producer for trade, and the superscript s stands for "supply for trade," so the amount of good j that the consumer-producer consumes is equal to the amount she buys discounted by trade costs plus her production minus the amount she sells. Things are similarly defined for good j' except that there is no $x_{j'}^{s}$, since people in economy i are assumed un able to commercialize their productions of j'.

(2) $\mathbf{x_j^p = L_j; x_{j'}^p = L_{j'} - A}$

This describes the consumer-producer's production of goods j and j'. $L_j(\mathbf{L_{j'}})$ denotes the amount of labor used by the consumer-producer to produce good j (j'), since the consumer-producer in economy i has advantage in producing good j, $x_j^p = L_j$: one unit of labor for one unit of output. The consumer-producer has to pay a learning cost to produce good j', so $x_j^p = L_j - A$. $\mathbf{A}$ is the fixed learning cost that this consumer-producer in economy i has to pay to if she chooses produce good j', which she is not good at, and is assumed to be uniformly distributed across [0, 1].

(3) $\mathbf{L_j + L_{j'} = 1}$(each consumer-producer is endowed with 1 unit of labor)

(4) $\mathbf{p_j x_j^s = p_{j'} x_{j'}^d + p_j x_j^d}$

This is sale=expenditure. $p_j, p_{j'}$ are prices of goods j and j' that are exogenous to individuals but endogenous in general equilibrium. Since I assume costs to commercialize good j' in economy i are prohibitively high, any good j' produced in economy i are not for sale but for self-consumption.

The above optimization problem involves comparing corner solutions, i.e., comparing maximized utilities achieved respectively under self-provision and trade. An analogy mentioned in Yang (2003) makes this methodology clear: if we are studying college education decision, we frame the problem such that a person first decides whether to take physics or economics as her major and then decides how to allocate time among subfields of the chosen major. Two corner solutions of interest for the above consumer-producer decision in economy i correspond to:

(S)Self-provision: $\mathbf{x_j^s = x_j^d = x_{j'}^d = 0}$

(T)Trade: not producing but buying goodj':$\mathbf{x_j^s > 0, x_{j'}^d > 0, x_j^d = 0, x_{j'}^p = 0}$

There are other possible corners, such as people both selling and buying good j in economy i. But, all the possibilities other than (S) and (T) can be safely ignored, given the model assumption and Wen's Theorem in Wen (1998). For model structures including the one adopted here, Wen (1998) proves an individual does not simultaneously buy and sell the same good, nor does she simultaneously buy and self-provide the same good; she sells at most one good.

One point should be clear now: if an economy i's consumer-producer chooses to trade, she sells good j and trades with people from economy i'selling goodj'. Although labor division does not necessarily mean inter-economic-unit trade, such as international trade, I only focus on inter-economic-unit trade without modeling intra-unit trade that only complicates the model without providing more insights for the current purpose.

A consumer-producer's optimized decisions given prices and a learning cost

Under (S) Self-provision with $x_j^s = x_j^d = x_{j'}^d = 0$ and constraint (4) irrelevant, optimization leads to$x_j = x_{j'} = \frac{1-A}{2}$ for the consumer-producer with person-specific learning cost A. Then, the maximized utility under (S) Self-provision is $\frac{(1-A)^2}{4}$. Under (T) Trade with people in economy i selling good j and buying good j' from economy i' and $x_j^s > 0, x_{j'}^d > 0, x_j^d = 0, x_{j'}^p = 0$, optimization leads to $x_j^s = \frac{1}{2}, x_{j'}^d = \frac{p_j}{2p_{j'}}$ for given prices p_j and $p_{j'}$. Thus, the maximized utility under (T) is $\frac{K_i p_j}{4p_{j'}}$. We then have below results for economy i.

Comparing maximized utilities under (S) and (T), **a consumer-producer in economy i chooses trade if $\frac{K_i p_j}{4p_{j'}} > \frac{(1-A)^2}{4}$, which is equivalent to $1 - \sqrt{\frac{K_i p_j}{p_{j'}}} < A$; otherwise, the consumer-producer chooses self-provision.** It follows that the number/measure of people in economy i choosing to trade is $\sqrt{\frac{K_i p_j}{p_{j'}}}$, since A is assumed to be uniformly distributed on [0,1]. This is the **commercial class size** in economy i for given prices and trade costs.

Above is the solution for economy i, and the solution for economy i′ is derived in the same way with subscripts i and i′ swapped, j and j′ swapped.

Equilibrium price given institutional quality $\mathbf{k_i},\ \mathbf{k_j}$

From above, the measure of people in economy i participating in trade is $\sqrt{\frac{\mathbf{K_i p_j}}{\mathbf{p_{j'}}}}$, and each of them has a supply of good j $x_j^s = \frac{1}{2}$ and a demand of good j′ $x_{j'}^d = \frac{p_j}{2p_{j'}}$. Only people in economy i sell good j and only they buy good j′. So the aggregate supply of good j is $X_j^S = \frac{1}{2}\sqrt{\frac{K_i p_j}{p_{j'}}}$ (individual supply in economy i times the number of individuals in economy i who supply), and the aggregate demand of good j is $X_j^D = \frac{p_{j'}}{2p_j}\sqrt{\frac{K_{i'} p_{j'}}{p_j}}$ (individual demand in economy i′ times the number of individual in economyi′ who demand). Aggregated demand and supply for good j′ are derived analogically. I use capital letter X for aggregate demand (the superscript D stands for Demand) and supply (the superscript S stands for supply), while lowercase x stands for individual demand and supply. With market clearing conditions that equalize aggregate demand and supply for both goods, I solve for **equilibrium (relative) price:** $\frac{\mathbf{p_j}}{\mathbf{p_{j'}}} = (\frac{\mathbf{K_{i'}}}{\mathbf{K_i}})^{\frac{1}{4}}$.

Economic equilibrium outcomes

Plugging the solved relative price to individual decisions and the commercial class size expression, I get below economic results or comparative statics. Recall $\mathbf{K_i} = (\mathbf{1} - \mathbf{t_i})\mathbf{k_i}$, where k_i measures institutional quality.

Economic Result 1: The commercial class size (the measure of people who trade) in economy i is $K_{i'}^{\frac{1}{8}} K_i^{\frac{3}{8}}$, which positively depends not only on economy i’s own institutional quality k_i but also on its trade partner's institutional quality $k_{i'}$. Clearly, the size is a number between 0 and 1.

Economic Result 2: The average utility (welfare per consumer-producer) in economy i is $\frac{1}{12}+\frac{K_{i'}^{\frac{3}{8}}K_i^{\frac{9}{8}}}{6}$. The total welfare of existing merchants (the aggregated utility of the consumer-producers who trade) in economy i is $\frac{K_{i'}^{\frac{3}{8}}K_i^{\frac{9}{8}}}{4}$. These positively depend not only on economy i's own institutional qualityk_i but also on its trade partner's institutional quality $k_{i'}$.

Economic Result 3: The elite rent in economy i from expropriating trade is $\frac{K_{i'}^{\frac{3}{8}}(1-t_i)^{\frac{1}{8}}k_i^{\frac{1}{8}}}{2}(1-k_i)$, since the extractive elite expropriates a fraction, $1-k_i$, of every unit traded-in after physical trade costs are paid. Endogenous determination of k_i will be modeled in a political game below.

Intuitively, with higher k_i(better institutional quality in terms of less expropriation), trade expands, increased specialization reduces total fixed learning costs paid, and total output consumed in each economy increases without increasing labor endowments. That is, TFP increases due to an institutional improvement.

However, the above depicts an incomplete picture. According to North (1981) and Acemoglu, Johnson and Robinson (2005a), institutions are endogenously determined and changed in the political interaction between parties that have different economic benefit-cost calculations. So, in the following, I endogenize k_i with a political gamebased on above economic outcomes that stipulate each party's benefit-cost calculation.

2.2 The Political Equilibrium

In this part, I focus on economy i and construct a political game between the commercial class and the local elite in whichk_i (the institutional quality of economy

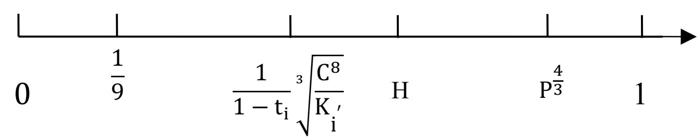
i) is endogenized given exogenously $K_{i'}$, t_i and other parameters discussed below. I provide comparative statics of changing t_i and $K_{i'}$ in Section 3, and relate them to historical evidence in Section 4. Here, the political game is intended to provide insights on how trade interacting with an economy's initial institutional quality (this is a path dependence story) can have implications for institutional changes. The resulting political equilibrium combined with the above economic equilibrium that depicts how institutional quality affects trade complete my logic of mutual promotion of institutional improvements and trade expansion.

Setup conditions and assumptions for the political game

First and most crucially, there is an initial constraint on the extractive elite of economy i. The constraint is denoted as $\overline{k_\imath}$. I assume the elite is free to choose k_i above $\overline{k_\imath}$, but not able to choose k_i below $\overline{k_\imath}$. I also assume if the elite wants to change the institutional constraint on him to $\overline{\overline{k_\imath}}$, he can only choose $\overline{\overline{k_\imath}} \geq \overline{k_\imath}$. $\overline{k_\imath}$ is an exogenous and crucial parameter for the political game, where the interaction of $\overline{k_\imath}$ and trade drives institutional changes. I have a few reasons behind this setup. (1) Economic and political phenomena never happen from nowhere, and, in many cases, a political economy has its context that has been determined historically, socially, or culturally; I parameterize such a context with $\overline{k_\imath}$ and interpret it as a constraint on the elite's expropriation. (2) This interpretation leads to an asymmetry assumption that the elite is free to choose k_i above $\overline{k_\imath}$, but not able to choose k_i below $\overline{k_\imath}$. This assumption is a little extreme, but softening it by assuming that it is less costly for the elite to choose k_i above $\overline{k_\imath}$ than below $\overline{k_\imath}$ does not change intuitions and unnecessarily complicates mathematics. Please note the elite's rent is concave in k_i and achieves its maximum at $\frac{1}{9}$ according to Economic Result 3, so the elite under the constraint chooses k_i=max{the constraint, $\frac{1}{9}$}. (3) This setup echoes the concept of institutional drifts in (Acemoglu and Robinson, 2012), which argues for the necessity of a good initial institutional quality for institutional improvements when there are opportunities, and a good initial institutional quality can be a context that resulted from institutional drifts. However, $\overline{k_i}$ is not necessary to be interpreted as a result from drift.

Second, I assume a threshold C between 0 and 1 such that if the size of the commercial class in economy i is below C, the probability for the commercial class to win a conflict against the elite is zero. Then, the commercial class would not fight for a better institution. If the size is above C, the commercial class has a winning probability **P** after choosing to fight against the elite for a better institution. Further, I assume the institutional quality $k_i$becomes 1 and the elite gets zero rent if the commercial class wins the fight against the elite. If the commercial class loses, its members have zero consumption and nothing changes for the elite.[24]

Third, the outcome depends on parameters and I restrict my discussion for the set of parameters satisfying the below order. H is defined by $H^{\frac{1}{8}} - H^{\frac{9}{8}} = \frac{P^{\frac{1}{6}}\left(1-P^{\frac{4}{3}}\right)}{1-P}$.



It can be proved that $H < P^{\frac{4}{3}}$. I impose no restriction on $\overline{k_i}$, so it could be any real number between 0 and 1. This third assumption leads to all possible cases for institutional stagnation and changes. At the end of Section 2, I briefly discuss other possible orderings that tend to shut down one or more cases. Besides, $\frac{1}{9}$ is the best institutional choice for the elite without any meaningful constraint and I put $\frac{1}{9}$ at a relatively extreme position because it is rare for an elite to be totally free to choose, even when the commercial class fails to form a threat. This third assumption also smoothes mathematics.

In sum, $K_i{'}$, t_i, $\overline{k_i}$, C and P are the exogenous parameters underlying the political game for institutional changes in economy i.

[24] Alternatively, we can assume that when the commercial class loses, the ruler can reset k_i regardless of the constraint $\overline{k_i}$, but this provides no extra insight for institutional changes, except for adding straightforwardly a possibility of institution deterioration.

The political game structure

I construct a 2-stage game to formalize interactions between the commercial class and the extractive elite. I assume away mixed strategies. The payoffs are based on the economic results above.

Stage 1: Starting from an initial institutional constraint $\overline{k_1}$, the extractive elite chooses whether to reform (i.e., to choose a $\overline{\overline{k_1}} > \overline{k_1}$ as a new constraint on his expropriation) **or** to do nothing (equivalently to choose$\overline{\overline{k_1}} = \overline{k_1}$). Choice of constraint$\overline{\overline{k_1}}$ is a credible institutional change not a policy change. Note $\overline{\overline{k_1}}$ is an endogenous choice and $\overline{k_1}$ is an exogenous parameter.

Stage 2: After $\overline{\overline{k_1}}$ is chosen in stage 1, the commercial class size is $K_i{}'^{\frac{1}{8}}(1 - t_i)^{\frac{3}{8}}k_i^{\frac{3}{8}}$, where k_i=max$\{\overline{\overline{k_1}}, \frac{1}{9}\}$ as discussed above,[25] and this commercial class chooses whether to fight for a better institution **or** to leave the reality as it is. Whatever the result is, the commercial class after stage 2 is no longer able to initiate a fight.[26]

Expected payoffs: The commercial class and the elite collect payoffs after stage 2. Given $\overline{\overline{k_1}}$ andthat the commercial class chooses to fight, the elite gets $\frac{1}{2}K_i{}'^{\frac{3}{8}}(1 - t_i)^{\frac{1}{8}}k_i^{\frac{1}{8}}(1 - k_i)(1 - P)$ (the commercial class's losing probability times the elite's rent according Economic Result 3), where k_i=max$\{\overline{\overline{k_1}}, \frac{1}{9}\}$,and the commercial class

[25] With the assumptions I make for the parameters, k_i is always equal to $\overline{\overline{k_1}}$ in the below discussion. Other cases are possible with the assumptions relaxed, involve similar arguments, but lead to trivial outcomes and insights.

[26] There are a few ways to interpret this model setup. There could be a window of opportunity that ends after stage 2 for the commercial class to take collective actions. Alternatively P, the winning probability if the commercial class chooses to fight, can be seen as a subjective probability such that when the commercial class loses, the class turns to disbelieve a fight.

gets $P\frac{K_{i'}^{\frac{3}{8}}(1-t_i)^{\frac{9}{8}}k_i^{\frac{3}{8}}}{4}$ (winning probability times the total utility of members in the commercial class when k_i becomes 1). Given$\bar{\bar{k}}_1$ and that the commercial class chooses not to fight, the elite gets $\frac{1}{2}K_{i'}^{\frac{3}{8}}(1-t_i)^{\frac{1}{8}}k_i^{\frac{1}{8}}(1-k_i)$, and the commercial class gets $\frac{K_{i'}^{\frac{3}{8}}(1-t_i)^{\frac{9}{8}}k_i^{\frac{9}{8}}}{4}$.

Intuitions: The commercial class weighs fighting with a losing possibility against leaving the reality as it is. Although the elite likes $\bar{\bar{k}}_1$ to be as low as $\bar{k}_1$, he has to consider the negative impact of social conflicts on his expected payoff and may want to offer a higher $\bar{\bar{k}}_1$ (reform) to avoid social conflicts.

In the model, the commercial class decides as a whole. This model actually ends up with the same conclusion if each merchant is assumed to make his own decision: the inequality governing the benefit-cost calculation of the whole commercial class is mathematically equivalent to the one for each individual, since in my setup people who trade are homogenous in their maximized utilities. To be more specific, the commercial class chooses to fight if $P\frac{K_{i'}^{\frac{3}{8}}(1-t_i)^{\frac{9}{8}}k_i^{\frac{3}{8}}}{4}$ (the class's expected payoff if it fights) $>\frac{K_{i'}^{\frac{3}{8}}(1-t_i)^{\frac{9}{8}}k_i^{\frac{9}{8}}}{4}$ (the class's payoff if it does not fight); if this inequality is divided by $K_{i'}^{\frac{1}{8}}(1-t_i)^{\frac{3}{8}}k_i^{\frac{3}{8}}$ (the size of the commercial class) on both sides, the inequality becomes $P\frac{K_{i'}^{\frac{1}{4}}(1-t_i)^{\frac{3}{4}}}{4}$ (an individual merchant's expected payoff when the commercial class fights) $>\frac{K_{i'}^{\frac{1}{4}}(1-t_i)^{\frac{3}{4}}k_i^{\frac{3}{4}}}{4}$ (an individual merchant's payoff when the commercial class does not fight). So, when it is better for the commercial class to fight, a member of the class also finds it better to fight if he believes other members find it better to fight. This belief can be supported by the above inequality governing individual decisions that are homogenous among the commercial class members.[27]

[27] It's also possible that every merchant believes no other merchants would like to fight and chooses not to fight;

The elite's and commercial class's equilibrium strategies

In this part, I show equilibrium strategies of the elite and the commercial classes in the above game, leaving solution details in the next part. With these equilibrium strategies (mutually best responses), later, I can present political equilibrium results about how the initial institutional constraint ($\overline{\mathrm{k}}_1$) matters for institutional changes throughinteracting with $\mathrm{K}_{i'}$ (e.g.,i's trade partner's institutional changes) and t_i (e.g., transportation technology improvement or location advantages).

The elite's equilibrium choice ($\overline{\overline{\mathbf{k}}}_1$) in anticipation of the commercial class's response and given the initial institutional constraint ($\overline{\mathbf{k}}_1$):

$\overline{\overline{\mathbf{k}}}_1 = \overline{\mathbf{k}}_1$ **if** $\overline{\mathbf{k}}_1 > \mathbf{P}^{\frac{4}{3}}$ **or** $\frac{1}{9} < \overline{\mathbf{k}}_1 < \boldsymbol{H}$;

$\overline{\overline{\mathbf{k}}}_1 = \mathbf{P}^{\frac{4}{3}}$ **if** $\mathbf{P}^{\frac{4}{3}} > \overline{\mathbf{k}}_1 > \boldsymbol{H}$;

any $\overline{\overline{\mathbf{k}}}_1 \in [\overline{\mathbf{k}}_1, \frac{1}{9}]$ **if** $\overline{\mathbf{k}}_1 > \frac{1}{9}$ **(or without loss of generally,** $\overline{\overline{\mathbf{k}}}_1 = \frac{1}{9}$ **for this case).**

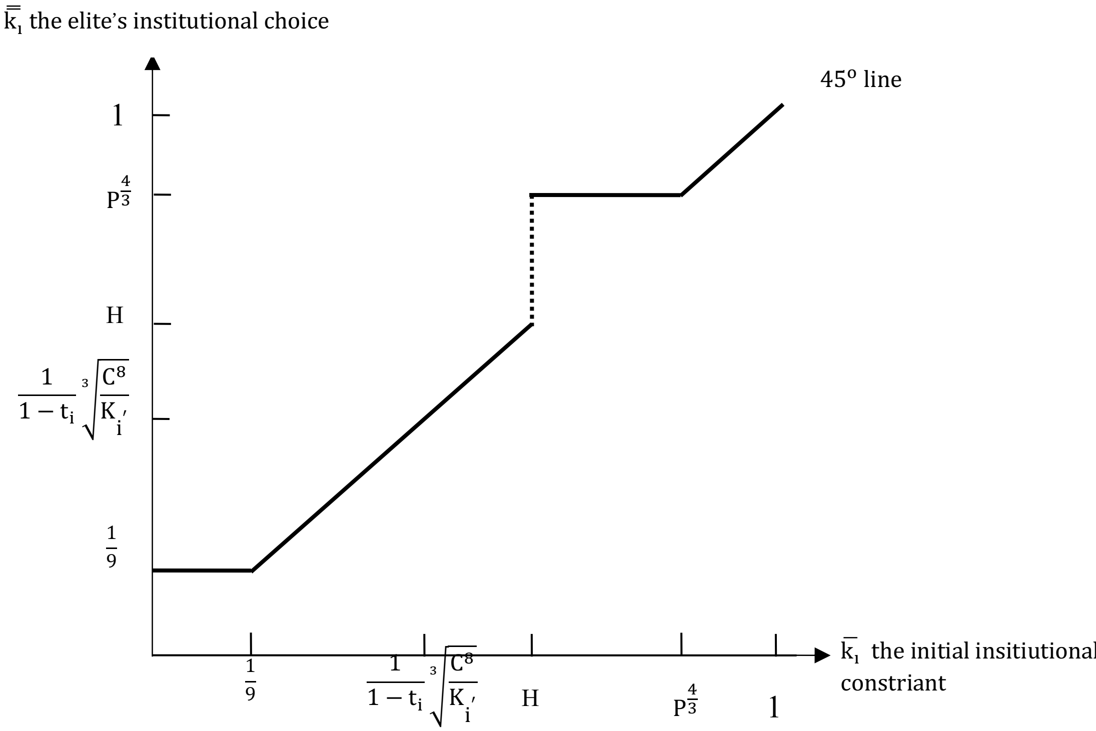
(His defined by $\mathrm{H}^{\frac{1}{8}} - \mathrm{H}^{\frac{9}{8}} = \frac{\mathrm{P}^{\frac{1}{6}}\left(1-\mathrm{P}^{\frac{4}{3}}\right)}{1-\mathrm{P}}$. It is easy to prove $\mathrm{H} < \mathrm{P}^{\frac{4}{3}}$. I assume $\mathrm{H} > \frac{1}{9}$ to avoid trivial discussions.)

Intuitively, when the initial institutional constraint is not very high or very low ($\mathrm{P}^{\frac{4}{3}} > \overline{\mathrm{k}}_1 > H$), the elite would like to reform to avoid conflicts; the elite would otherwise make no changes, either because it is better for the elite to undergo conflicts or because the commercial class has no incentive or ability to struggle for a better institution. This will be seen more formally later in solution details. Figure 1 visualizes how the elite's institutional choice (choice of new institutional constraints) depends on initial institutional constraints when the elite takes into account the commercial class's equilibrium strategy. Note the elite's institutional choice is not

thus a Nash equilibrium without collective actions is resulted in. However, I disregard this possibility because in the setup collectives actions should be more probable and interesting if each merchant knows other merchants' payoffs, which I assume. Furthermore, as discussed in Note 4, the window of opportunity or the subjective winning probability interpretation allows for limited coordination.

identical to the equilibrium outcome for institutional changes, since there is a probability for the institutional outcome to be different from the elite's choice when a social conflict happens.

Figure 1: The elite's institutional choice as a function of initial institutional constraints



The commercial class's equilibrium choice after observing $\bar{\bar{k}}_1 \geq \bar{k}_1$ chosen by the elite:

To fight if $\frac{1}{1-t_i}\sqrt[3]{\frac{C^8}{K_{i'}}} < \bar{\bar{k}}_1 < P^{\frac{4}{3}}$;

Not to fight if $\bar{\bar{k}}_1 > P^{\frac{4}{3}}$ or $\bar{\bar{k}}_1 < \frac{1}{1-t_i}\sqrt[3]{\frac{C^8}{K_{i'}}}$.

(I assume $\frac{1}{1-t_i}\sqrt[3]{\frac{C^8}{K_{i'}}} < P^{\frac{4}{3}}$ to avoid trivial discussions.)

Intuitively, the commercial class has no incentive to fight if the institutional constraint is high enough ($\bar{\bar{k}}_1 > P^{\frac{4}{3}}$), and the commercial class has no ability to fight (the class size is very small) if the institutional constraint is very low. This will be seen more formally later in solution details.

The cost-benefit calculations in the political game are based on trade outcomes; I therefore use "political trading economy" in the title. The solution details are explained next. Readers uninterested in such mathematical details can safely skim them, although the solution details make each party's strategic consideration clearer and more formal.

Solution details

When the commercial class size is smaller than C, the commercial class will not fight for a better institution, since the wining probability is then zero. The elite without threat simply chooses $\bar{\bar{k}}_1 = \bar{k}_1$(no institutional changes) and correspondingly k_i=max$\{\bar{k}_1,\frac{1}{9}\}$.That the commercial class size is smaller than C is equivalent to $K_{i'}^{\frac{1}{8}}(1-t_i)^{\frac{3}{8}}k_i^{\frac{3}{8}} < C$ (iff $k_i < \frac{1}{1-t_i}\sqrt[3]{\frac{C^8}{K_{i'}}}$) based on Economic Result 1. So, when $\bar{k}_1 < \frac{1}{1-t_i}\sqrt[3]{\frac{C^8}{K_{i'}}}$, there are no institutional changes with the commercial class incapable of fighting. This is a straightforward case.

When $\bar{k}_1 > \frac{1}{1-t_i}\sqrt[3]{\frac{C^8}{K_{i'}}} > \frac{1}{9}$, k_i=max$\{\bar{\bar{k}}_1, \frac{1}{9}\}$=$\bar{\bar{k}}_1$, because $\bar{k}_1 > \frac{1}{1-t_i}\sqrt[3]{\frac{C^8}{K_{i'}}} > \frac{1}{9}$ and $\bar{\bar{k}}_1 \geq \bar{k}_1$. I solve the game by backward induction starting from the second stage.

The commercial class does not fight if $\frac{K_{i'}^{\frac{3}{8}}(1-t_i)^{\frac{9}{8}}\bar{\bar{k}}_1^{\frac{9}{8}}}{4} > P\frac{K_{i'}^{\frac{3}{8}}(1-t_i)^{\frac{9}{8}}\bar{\bar{k}}_1^{\frac{3}{8}}}{4}$ (iff $\bar{\bar{k}}_1 > P^{\frac{4}{3}}$) based on the payoff schedule in the game structure and given the first stage choice $\bar{\bar{k}}_1$. This is because $\bar{\bar{k}}_1$ is good enough that the commercial class does not want to take risk to fight. Consequently, the commercial class fights when $P^{\frac{4}{3}} > \bar{\bar{k}}_1 \geq \bar{k}_1 > \frac{\mathbf{1}}{\mathbf{1}-t_i}\sqrt[3]{\frac{C^8}{K_{i'}}}$. In summary, given $\overline{\overline{k_1}} \geq \bar{k}_1$ chosen by the elite at the first stage, the commercial class's best response at the second stage is

$$\text{Fight if } \overline{\overline{k_1}} < P^{\frac{4}{3}}\text{; Not to fight if } \overline{\overline{k_1}} > P^{\frac{4}{3}}.$$

Now, having been clear about the commercial class's best response at the second stage, we can figure out the ruler's optimal strategy at the first stage, given the initial constraint $\overline{k_1}$ and in anticipation of the commercial class's response. First, without threat the elite's best strategy is $\overline{\overline{k_1}} = \overline{k_1}$ when $P^{\frac{4}{3}} < \overline{k_1}$, because the commercial class always chooses not to fight with $P^{\frac{4}{3}} < \overline{\overline{k_1}}$, $\overline{\overline{k_1}} \geq \overline{k_1} > \frac{1}{9}$, and $\frac{1}{9}$ is the maxima of the elite's concave utility under no constraint.

Second, let us solve the elite's optimal strategy when $P^{\frac{4}{3}} > \overline{k_1} > \frac{\mathbf{1}}{\mathbf{1}-t_i}\sqrt[3]{\frac{C^8}{K_{i'}}}$. From the payoff schedule in the game structure, the elite's optimization boils down to a choice between the maximized payoff when the fight is avoided and the maximized payoff when the elite does not avoid conflicts. The former is achieved by choosing $\overline{\overline{k_1}} = P^{\frac{4}{3}}$ and is equal to $\frac{1}{2}K_{i'}^{\frac{3}{8}}(1-t_i)^{\frac{1}{8}}P^{\frac{1}{6}}\left(1-P^{\frac{4}{3}}\right)$, since $\overline{\overline{k_1}}$ should be at least as large as $P^{\frac{4}{3}}$ for the elite to avoid conflicts, according to the commercial class's optimal response at the second stage. The latter is achieved by choosing $\overline{\overline{k_1}} = \overline{k_1}$ and is equal to $\frac{1}{2}K_{i'}^{\frac{3}{8}}(1-t_i)^{\frac{1}{8}}\overline{k_1}^{\frac{1}{8}}(1-\overline{k_1})(1-P)$. So, the elite's maximized payoff when the fight is avoided is smaller than his maximized payoff when the he does not avoid conflicts if $\frac{1}{2}K_{i'}^{\frac{3}{8}}(1-t_i)^{\frac{1}{8}}P^{\frac{1}{6}}\left(1-P^{\frac{4}{3}}\right) < \frac{1}{2}K_{i'}^{\frac{3}{8}}(1-t_i)^{\frac{1}{8}}\overline{k_1}^{\frac{1}{8}}(1-\overline{k_1})(1-P)$.

I define H by $H^{\frac{1}{8}} - H^{\frac{9}{8}} = \frac{P^{\frac{1}{6}}\left(1-P^{\frac{4}{3}}\right)}{1-P}$ (It could be proved that $H < P^{\frac{4}{3}}$, though H could not be solved analytically—I'll discuss its existence later). The inequality at the end of the above paragraph is equivalent to $\overline{k_1}^{\frac{1}{8}} - \overline{k_1}^{\frac{9}{8}} < \frac{P^{\frac{1}{6}}\left(1-P^{\frac{4}{3}}\right)}{1-P}$, of which the left-hand side decreases in $\overline{k_1}$. Then the elite chooses to reform ($\overline{\overline{k_1}} = P^{\frac{4}{3}}$) when $\overline{k_1} >$ H (iff $\overline{k_1}^{\frac{1}{8}} - \overline{k_1}^{\frac{9}{8}} < \frac{P^{\frac{1}{6}}\left(1-P^{\frac{4}{3}}\right)}{1-P}$), and chooses $\overline{\overline{k_1}} = \overline{k_1}$toendureconflicts when $\overline{k_1} <$ H. So, when $P^{\frac{4}{3}} > \overline{k_1} > \frac{\mathbf{1}}{\mathbf{1}-t_i}\sqrt[3]{\frac{C^8}{K_{i'}}}$, the elite will reform to avoid social conflicts ($\overline{\overline{k_1}} = P^{\frac{4}{3}}$) if

$P^{\frac{4}{3}} > \overline{k_1} > H$, and will do nothing and endure conflicts ($\overline{\overline{k_1}} = \overline{k_1}$) if $\frac{1}{1-t_i}\sqrt[3]{\frac{C^8}{K_i{}'}} < \overline{k_1} < H$.

Finally, below is the best strategy of the elite at the first stage in anticipation of the commercial class's response and given $\overline{k_1}$:

$$\overline{\overline{k_1}} = \overline{k_1}, \text{ if } P^{\frac{4}{3}} < \overline{k_1} \text{ or } \overline{k_1} < H;\ \overline{\overline{k_1}} = P^{\frac{4}{3}} > \overline{k_1}, \text{ if } P^{\frac{4}{3}} > \overline{k_1} > H.$$

Political equilibrium results

Based on each party's equilibrium strategy, three political results follow directly:

Political Result 1: When the initial institutional constraint is low ($\overline{k_1} < \frac{\mathbf{1}}{\mathbf{1}-t_i}\sqrt[3]{\frac{C^8}{K_i{}'}}$), there are no institutional changes since the commercial class has no capability to fight for a better institution. This threshold positively depends on t_i and negatively on $K_i{}'$.

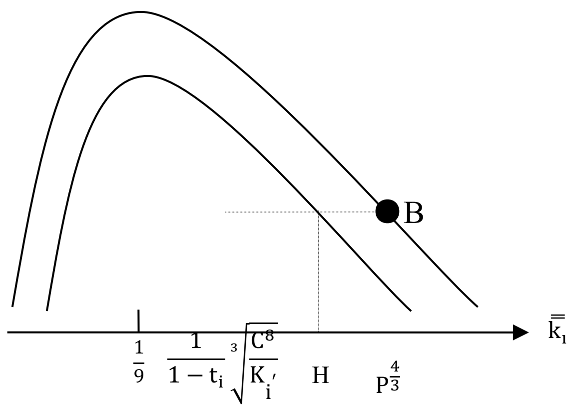
Political Result 2: When the initial institutional constraint is high ($\overline{k_1} > P^{\frac{4}{3}}$), the commercial class has no incentive to fight for a better institution and thus there are no institutional changes.

Political Result 3: When $P^{\frac{4}{3}} > \overline{k_1} > H$, the commercial class has both ability and an incentive to struggle for a better institution if the elite makes no institutional changes, and the elite under threat chooses to reform (to improve institutional constraint to $\overline{\overline{k_1}} = P^{\frac{4}{3}}$) to avoid conflicts.

Political Results 4: When $\frac{\mathbf{1}}{\mathbf{1}-t_i}\sqrt[3]{\frac{C^8}{K_i{}'}} < \overline{k_1} < H$, the commercial class has both ability and an incentive to struggle for a better institution if the elite makes no institutional changes, and the elite chooses to endure the challenge, since loss from reform is relatively large for him in this case.

Proof of the Political Results 1-4: These results directly follow the solved equilibrium strategies of the commercial class and the extractive elite.

Figure 2: The elite's institutional choice problem



The upper concave curve is the elite's utility curve without conflict; the lower concave curve is the one when the commercial class initiates a fight.

Political results for institutional changes have been derived. To understand these results more intuitively, I visualize the elite's optimization problem in Figure 2. What follows is how to interpret the figure.

The upper concave curve is the elite's utility curve without conflict; the lower concave curve is the one when the commercial class initiates a fight. It has been proved that the commercial class chooses to fight when the level of institutional quality is between $\frac{1}{1-t_i}\sqrt[3]{\frac{C^8}{K_{i'}}}$ and $P^{\frac{4}{3}}$. So, the lower curve is relevant for any institutional quality strictly between $\frac{1}{1-t_i}\sqrt[3]{\frac{C^8}{K_{i'}}}$ and $P^{\frac{4}{3}}$, while the upper one is relevant otherwise. When the initial institutional quality $\bar{k}_i$ is outside the range between $\frac{1}{1-t_i}\sqrt[3]{\frac{C^8}{K_{i'}}}$ and $P^{\frac{4}{3}}$, $\bar{\bar{k}}_i = \bar{k}_i$ is best for the elite: this corresponds to Political Results 1 and 2. The elite chooses Point B ($\bar{\bar{k}}_i = P^{\frac{4}{3}} \geq \bar{k}_i$) in the figure if H<$\bar{k}_i$<$P^{\frac{4}{3}}$, since the elite's payoff at B (when the upper curve is relevant) is higher than any payoff associated with $H<\bar{k}_1<P^{\frac{4}{3}}$ (and $\bar{\bar{k}}_1 \geq \bar{k}_1$): this corresponds to Political Result 3.

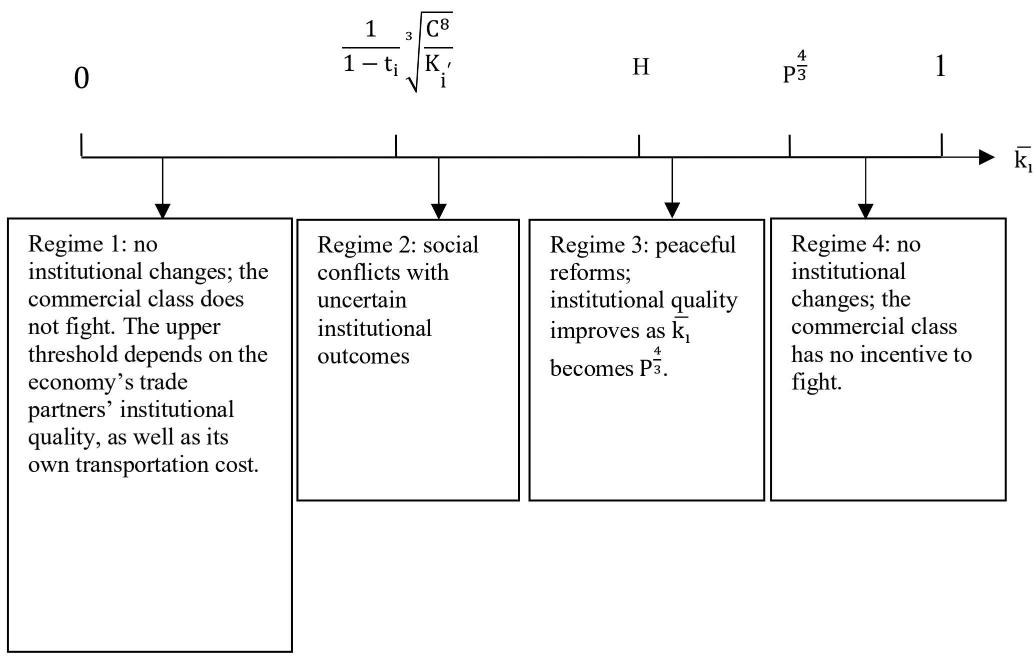
Consistent with Political Result 4, the elite chooses $\bar{\bar{k}}_1 = \bar{k}_1$, if $H>\bar{k}_1 > \frac{1}{1-t_i}\sqrt[3]{\frac{C^8}{K_{i'}}}$.

Summary of the equilibrium outcomes for institutional changes

Based on the above political results, Figure 3 summarizes the main insights in this work about institutional changes as equilibrium outcomes.

Institutional changes usually happen when the initial institutional constraint is neither very high nor very low, for otherwise the commercial class is either too weak or too satisfied to fight for a better institution. Given that the initial institutional constraint is interpreted as something developed historically, this is path dependent. More interestingly, how high (low) is "very high (low)" depends on transportation costs and the economy's trade partners' political trade costs (t_i and $K_{i'}$): this is to be discussed in next section.

Figure 3: Institutional changes as a function of initial institutional constraints ($\bar{k}_1$)



Historically, there are both radical and progressive reforms without social conflicts (more in Section 4). I model different types of reforms not qualitatively but quantitatively. According to the model implications, there is going to be a reform if an economy's initial institutional constraint is within in Regime 3, when the commercial class has incentives to fight and the extractive elite finds it best to avoid social conflicts by reform. How radical a reform is depends on the distance between an economy's initial institutional constraint and $P^{\frac{4}{3}}$, with initial institutional quality above which the commercial class no longer has incentives to fight. Different countries may have different initial positions or different $P^{\frac{4}{3}}$, so we may observe radical and progressive reforms in different countries.

Two technical remarks end Section 2. First, I assume H exists and $\frac{1}{9} < \frac{1}{1-t_i}\sqrt[3]{\frac{C^8}{K_{i'}}} < H$. ($H < P^{\frac{4}{3}}$ holds mathematically if H exists.) If part of the assumption fails, a regime in Figure 2 might be lost but the rest are interpreted similarly. For example, when H doesn't exist (actually because of a very high P), Regime 2 in Figure 2 does not exist and Regime 3 is adjacent to Regime 1: the elite when threatened always reforms since P is very high.

Second, although P is a constant in this setup, the probability for the commercial class to win a fight against the elite for a better institution does depend on the class size, since the wining probability is zero when the community size is smaller than C and the winning probability is P when the community size is larger than C.

3. Comparative Statics for institutional changes

In this section, I discuss comparative statics of the model to understand a mechanism behind economic development and institutional changes. I only focus on comparative statics with respect to the threshold $\frac{1}{1-t_i}\sqrt[3]{\frac{C^8}{K_{i'}}}$ which, according to the above political equilibrium, governs movement of an economy from institutional stagnation to institutional changes by conflicts or reform.

Comparative Statics with changes in t_i and $K_{i'}$ (t shock and K shock):

Without loss of generality, I focus on economy i with threshold $\frac{1}{1-t_i}\sqrt[3]{\frac{C^8}{K_{i'}}}$ as depicted in the political equilibrium.

Comparative Statics 1 (t shock): As t_i decreases (e.g., because of a drastic improvement in marine technology or geography knowledge), trade expands and per capital utility increases in economy i according to the economic results above; the threshold $\frac{1}{1-t_i}\sqrt[3]{\frac{C^8}{K_{i'}}}$ decreases and thus it is more possible for the economy moves from Regime 1 (institutional stagnation) to Regime 2 (conflicts) or 3 (reform), according to the political results above.

In particular, as t_i decreases (e.g., due to a drastic improvement in marine technology or geography knowledge), trade costs per unit of trade decrease, since the per unit trade costs are $1-K_i$ with $K_i=(1-t_i)k_i$. This means the people who trade consume more traded-in goods for any given amount they pay. Then, more people in economy i will switch from self-provision to trade: this is seen from the above solution to the decision problem of a typical consumer-producer. With more people to trade, gains from specialization increase and consumption per person increases for an unchanged amount of labor endowment, seen from above as Economic Result 2 (welfare increases as t_i decreases). The efficiency is improved because lower trade costs induce more specialization. Moreover, the size of commercial community also increases according to Economic Result 1, and this connects the economic equilibrium to the political one. According the political equilibrium, if the economy's initial institutional quality is high enough, the expanded commercial class constitutes a threat to the local extractive elite. Then, through strategic interaction modeled in the political game, the economy might move from Regime 1 (institutional stagnation) to

Regime 2 (conflicts) or 3 (reform); it is equivalent to say these work through decreased threshold $\frac{1}{1-t_i}\sqrt[3]{\frac{C^8}{K_{i'}}}$.

Comparative Statics 2 (K shock): As $K_{i'}$ increases, trade expands and there is economic development (consumption per person increases) in economy i according to the economic results above; the threshold $\frac{1}{1-t_i}\sqrt[3]{\frac{C^8}{K_{i'}}}$ decreases and thus it is more possible for the economy moves from Regime 1 (institutional stagnation) to Regime 2 (conflicts) or 3 (reform), according to the political results above.

In particular, with higher $K_{i'}$ and according to Economic Results 1 and 2, economy i experiences economic development(consumption per person increases) and the commercial class expands, though the degree of changes also depends on economy i's own institution. The intuition is that more people in economy i′will demand goods form economy i with increased $K_{i'}$ and consequently profitability of trade increases for economy i. If economyi's initial institutional quality is high enough, it moves from Regime 1 to 2 or 3, due to decreased threshold $\frac{1}{1-t_i}\sqrt[3]{\frac{C^8}{K_{i'}}}$.

4. Historical Evidence

In this section I apply the above economic results and political comparative statics to two historical cases that are consistent with the theory.

4.1 The rise of Europe from Atlantic trade

This part is mainly based on Acemoglu and Robinson (2012) and Acemoglu, Johnson and Robinson (2005b). Below is a t shock story consistent with the above comparative statics 1.

The rise of Europe started from Atlantic trade, i.e., European trade with the New World, Africa, and Asia through Atlantic sea routes between 1500 and 1850. This is a t shock story. The crucial factor behind Atlantic trade was gradual marine technology improvement that reduced transportation costs for overseas trade. These contributed a decrease in physical trade costs such as transportation costs for western European countries that participated in Atlantic trade. The location of western European countries also rendered them a transportation advantage in Atlantic trade (Acemoglu, Johnson and Robinson, 2005b). Consistent with the above Economic Result 2, there was a direct effect of trade on economic development due to decreased transportation costs. Trade also had an indirect impact by altering the balance of political power and enhancing growth-friendly institutions, especially for trading countries with good initial institutions, and this mechanism is consistent with the above political equilibrium. In particular, the reduction of physical trade costs decreased the threshold $\frac{1}{1-t_i}\sqrt[3]{\frac{C^8}{K_{i'}}}$ through development of commercial class, and if economy i initially had better institutions, it was more possible for the economy to move from Regime 1 to Regime 2 or 3 (i.e., to improve their institutions through either reform by the extractive elite or conflicts between merchants and the elite) as depicted in the political equilibrium. Finally, the improved institution brought about enhanced economic development. These growth outcomes are documented in Acemoglu, Johnson and Robinson (2005b) (also see their Figure 1A, 1B, 2A, 2B).

As emphasized in Acemoglu and Robinson (2012) and Acemoglu, Johnson and Robinson (2005b), the development of the commercial class is the pivot. Consistent with the Economic Result 1 and the political equilibrium, the commercial class (mainly including overseas merchants) expansion induced by Atlantic trade was both a driving force behind and manifestation of economic development for countries participating in Atlantic trade; for countries participating in Atlantic trade with relatively better initial institutions, their commercial classes were more likely to fight for better institutions (low expropriation). This explains the difference in development and institutional improvement between a few western European countries.

The history is also consistent with the Political Results that formalize the implication of initial institutional quality on institutional changes. Great Britain, the

Netherlands, Spain, Portugal and France all have access to Atlantic trade and enjoyed trade induced economic development. However, Great Britain and the Netherlands have merchant groups being active not only in economic activities, but also in political arenas in struggling for and maintaining better institutions, because they are relatively better initial institutions (less royal control on merchants). In Great Britain during the (English) Civil War (1462-1649) and the Glorious Revolution (1688-1689), Atlantic merchants using their profits and through Parliamentarian forces successfully constrained royal power and strengthened merchant rights. In the Netherlands, commercial interests mainly that were involved in Atlantic trade won independence from the Habsburg monarchy and thus established a more secure commercial environment that lead to the Netherlands' economic prosperity in the early modern era. The cases of Great Britain and the Netherlands echo Political Result 3. In contrast, Spanish, Portuguese and French merchants each were subject to much more royal influence, or, to put it in another way, merchants who traded in these countries had to pay a lot to the local extractive elite (higher trade costs); with worse initial institutions, the commercial classes in these countries were not developed enough to drastically influence institutional changes (Acemoglu, Johnson and Robinson, 2005b). The cases of Spain, Portugal and French echo Political Result 1.

In sum, the economic development directly induced by Atlantic trade in early modern Europe is consistent with the Economic results. The political equilibrium results also replicate the institutional changes in countries with better initial institutions and the role played by the commercial class in institutional changes.

4.2 19th century Germany under French rule

This part is mainly based on Keller and Shiue (2016) and Acemoglu, Cantoni, Johnson and Robinson (2011), and is a K shock story consistent with the above comparative statics 2.

The French Revolution was exported mainly by French armies in late 18th century to many European countries. The export was in form of radical, "designed" and externally imposed reforms. Acemoglu, Cantoni, Johnson and Robinson (2011) show that such export had positive economic impacts for economic units affected, specifically for German cities with institutional reforms imposed by French rule around 1800. Keller and Shiue (2016) empirically test a mechanism behind such positive impacts. They use an instrumental-variable approach and conclude that most impact of institutional improvement on city development is through market integration that I interpret as expanded trade. This echoes the model here.

The institutional improvements induced by French rule were mainly imposition of the civil legal code, the abolition of guilds and the remnants of feudalism, according to Acemoglu, Cantoni, Johnson and Robinson (2011). These improvements relate to changes in k in the model here, and according to the above Economic Result 2, improvement in k (thus increase in K) contributes to economic development through expanded trade for both economic units with k improvement and economic units trading with the unit whose k is improved. This model implication or mechanism is empirically supported by Keller and Shiue (2016).

In particular, Keller and Shiue (2016) use (1) the abolition of guilds, (2) the possibility of redeeming feudal lands, and (3) a guarantee of equality before the law to measure institutional improvement. These measures relate to changes in k in my model: (1) and (3) directly reduced trade costs and made it easy for a wide range of people to operate their business; (2) reduced expropriation risk. The econometric analysis of Keller and Shiue (2016) shows not only these institutional improvements expanded trade and consequently stimulated economic development, but also trade was the major channel through which institutional improvements stimulated economic development. This is consistent with my model. More interesting and specific, they also show both a city's and its trade partner's institutional improvements had a significantly positive impact on the city's economic development through trade expansion, which support this work further. This is especially consistent with Economic Result 2, per which the development of an economic unit depends on both its own trade costs and those of its trade partner's.

Finally, Keller and Shiue (2016) also show market integration (expanded trade) did not significantly affect institutional changes in their historical context. This does not support the political implications of my model, but does not negate it, either.

Institutional changes in their historical context were externally driven by outside forces as French rule, which is different from the political game in this work that emphasizes endogenous forces driving institutional changes.

5. Discussion and Conclusion

In the previous 4 sections, I have introduced model intuitions, formalized the mechanism, discussed model comparative statics, and provided historical evidence. In this last section, I'll provide a few discussions related to this work, with the hope that they may stimulate further thinking or at least be interesting.

Although more empirical evidence is needed, I have used historical evidence to support my model, but I think this model also sheds insightful light on understanding current affairs. Currently, we see a lot of trade protections especially from developing countries (Kee, Nicita and Olarreaga, 2009). The presence and degree of such protections determine trade costs and can only be well understood from a political economy perspective. Here, by a political economy perspective, I mean political groups interact with each other under economic benefit-cost calculations. The model in this work contributes to our understanding of these issues by a micro-foundation emphasizing people's tradeoff between trade costs and specialization gains, as well as how economic potentials can feed political interactions that lead to further economic implications. The model emphasizes a possible mechanism that may be applicable to issues involving trade costs, specialization, institutions, political interactions, and economic development.

This model begins from the decision problem of a typical consumer-producer who produces, consumes, and decides whether to trade with her production. This is different from the commonly used neoclassical consumer-producer dichotomy. There is no best choice since purposes define methods. The advantages of starting from a consumer-producer rest at its convenience and insightfulness in modeling endogenous labor division. Such an approach puts trade costs directly against gains from specialization, as shown by this work and Yang (2003). This is a central theme of Adam Smith. To economically help underdeveloped countries or districts, it is sometimes less feasible to widely inject financial resources or high-quality human capitals, nor is the technology frontier always relevant. However, reducing trade costs by social or economic reorganization and letting specialization to boostrap development from given resources may be a more economical and effective choice in most cases. Thus, I think modeling beginning with a typical consumer-producer should and will provide more useful insights in development problems.

This work is about mechanism. Many works (e.g., Acemoglu, Johnson and Robinson, 2012; Acemoglu, Johnson and Robinson, 2005a; North, 1990) show the importance of good institutions for economic development. But, the mechanism problem, i.e., channels through which institutions matter and how institutions change, is relatively not well cooked. As James Robinson said, "possibly the biggest thing we don't understand in social science is how and why a society moves from one institutional equilibrium to another." This work approaches the mechanism problem by formalizing the mutual promotion between institutional improvement and trade expansion, as well as how such mutuality stimulates economic development. It emphasizes two methodological aspects: first, an integration of economic and political equilibria is usually necessary to understand institutional changes (North, 1981); second, echoing to Greif (2006), we sometimes need an equilibrium structure that comprises two layers. The parameters controlling the upper layer is usually an equilibrium outcome of the lower layer governed by another set of parameters or their ranges.

Finally, this work as mainly a theoretical work needs to be confronted with more supporting or denying evidence as well as data analysis testing mechanisms. Assumptions plus logic are always a half story.

References

References for Chapter 1

Adamic, L, and Adar, E., 2003. Friends and neighbors on the Web. Social Networks, 25(3):211-230.

Akerlof, G.A., 2018. Sins of Omission and the Practice of Economics. Journal of Economic Literature.

Arellano, M., 1987. PRACTITIONERS' CORNER: Computing robust standard errors for within - groups estimators. Oxford bulletin of Economics and Statistics, 49(4), pp.431-434.

Barabási, A.L. and Albert, R., 1999. Emergence of scaling in random networks. science, 286(5439), pp.509-512.

Belleflamme, P., and Peitz, M., 2015. Industrial Organization: Markets and Strategies. Cambridge University Press

Blei, D.M. and Lafferty, J.D., 2006, June. Dynamic topic models. In Proceedings of the 23rd international conference on Machine learning (pp. 113-120). ACM.

Bianconi, G., 2018. Multilayer Networks: Structure and Function. Oxford University Press.

Bonacich, P., 1987. Power and Centrality: A Family of Measures. American Journal of Sociology,92, 1170-1182.

Bruner, J., 2013. Tweets Loud and Quiet. https://www.oreilly.com/ideas/tweets-loud-and-quiet

Clauset, A., Newman, M.E. and Moore, C., 2004. Finding community structure in very large networks. Physical review E, 70(6), p.066111.

Coase,R., and Wang,N., 2012. How China Became Capitalist. Palgrave Macmillan.

Croissant, Y. and Millo, G., 2008. Panel data econometrics in R: The plm package. Journal of statistical software, 27(2), pp.1-43.

Csárdi, M.G., 2019. Package 'igraph'.

Eagle, N., Macy, M. and Claxton, R., 2010. Network diversity and economic development. Science, 328(5981), pp.1029-1031.

Easley, D., and Kleinberg, J., 2010.Networks, Crowds, and Markets.Cambridge University Press.

Gentzkow, M., Kelly, B.T. and Taddy, M., 2017. Text as data (No. w23276). National Bureau of Economic Research.

Grajzl, P. and Murrell, P., 2019. Toward understanding 17th century English culture: A structural topic model of Francis Bacon's ideas. Journal of Comparative Economics, 47(1), pp.111-135.

Grannovetter, M., 2017.Society and Economy: Framework and Principles. Belknap Press: An Imprint of Harvard University Press.

Granovetter, M., 1985. Economic action and social structure: The problem of embeddedness. American journal of sociology, 91(3), pp.481-510.

Granovetter, M., 1973. The Strength of Weak Ties. American Journal of Sociology, 78(6), pp.1360-1380.

Hayek, F.A., 1945. The use of knowledge in society. The American economic review, 35(4), pp.519-530.

Hodas, N.O. and Lerman, K., 2014. The simple rules of social contagion. Scientific reports, 4, p.4343.

Jackson, M., 2019. The Human Network.Pantheon.

Katz, E, and Lazarsfeld, P, 1955. Personal Influence. Routledge.

Newman, M., 2018. Networks. Oxford University Press

North, D., 1990. Institutions, Institutional Change and Economic Performance. Cambridge University Press

North, D., Wallis, J.J., and Weingast, Barry, 2009. Violence and Social Orders: A Conceptual Framework for Interpreting Recorded Human History. Cambridge University Press

Page, S., 2018. The Model Thinker. Basic Books.

Reddy, P.K., Kitsuregawa, M., Sreekanth, P. and Rao, S.S., 2002, December. A graph based approach to extract a neighborhood customer community for collaborative filtering. In International Workshop on Databases in Networked Information Systems (pp. 188-200). Springer, Berlin, Heidelberg.

Roberts, M.E., Stewart, B.M. and Tingley, D., 2014. stm: R package for structural topic models. Journal of Statistical Software, 10(2), pp.1-40.

Robinson, David, and Silge, Julia, 2017. Text Mining with R: A Tidy Approach. O'Reilly Media.

Salganik, M.J., Dodds, P.S. and Watts, D.J., 2006. Experimental study of inequality and unpredictability in an artificial cultural market. science, 311(5762), pp.854-856.

Shelling, T.C., 2006. Micromotives and Macrobehavior. W. W. Norton & Company.

Thurner, S., Hanel, R., and Klimek, P, 2018.Introduction to the Theory of Complex Systems. Oxford University Press.

Tolkien, J.R.R., 1955. The Lord of the Rings. Allen & Unwin.

Yang, X., 2003. Economic Development and the Division of Labor. BLACKWELL PUBLISHERS

References for Chapter 2

Abadie, A., and Imbens, G.W., 2006. “Large Sample Properties of Matching Estimators for Average Treatment Effects.” Econometrica, 74(1): 235-267.

Andrews, I., and Oster, E., 2019. “Weighting for External Validity.” NBER Working Paper No. 23826.

Angrist, J. D., 2004. “Treatment Effect Heterogeneity in Theory and Practice.” Economic Journal, 114(494): C52–83.

Angrist, J.D., and Fernandez-Val, I., 2010. "ExtrapoLATE-ing: External Validity and Overidentification in the LATE Framework." NBER Working Paper No. 16566.

Angrist, J.D., Imbens, G.W., and Rubin, D.B., 1996. "Identification of Causal Effects Using Instrumental Variables." Journal of the American Statistical Association, 91(434): 444-455.

Angrist, J.D., and Rokkanen, M., 2015. "Wanna Get Away? Regression Discontinuity Estimation of Exam School Effects Away from the Cutoff." Journal of the American Statistical Association, 110(512): 1331–1344.

Banerjee, A.V., Karlan, D., and Zinman, J., 2015. "Six Randomized Evaluations of Microcredit: Introduction and Further Steps." American Economic Journal: Applied Economics, 7(1): 1-21.

Banerjee, A.V., Chassang, S., and Snowberg, E., 2017. "Decision Theoretic Approaches to Experiment Design and External Validity." In Handbook of Economic Field Experiments, (A.V. Banerjee and E. Duflo, eds.) pp.141-174.

Bertanha, M., and Imbens, G.W., 2018. "External Validity in Fuzzy Regression Discontinuity Designs." NBER Working Paper No. 20773.Cameron, C., and Trivedi, P., 2005. Microeconometrics: Methods and Application. Cambridge University Press. Cook, T., and Campbell, D., 1979. Quasi-Experimentation: Design & Analysis Issues for Field Settings. Houghton Mifflin.

Cruces, G., and Galiani, S., 2007. "Fertility and Female Labor Supply in Latin America: New Causal Evidence." Labour Economics, 14: 565-573.

Deaton, A., 2010. "Instruments, Randomization, and Learning about Development." Journal of Economic Literature, 48(2): 424–455.

Dehejia, R., Pop-Eleches, C. and Samii, C., 2019. "From local to global: External validity in a fertility natural experiment." Journal of Business & Economic Statistics, pp.1-27.

Dong, Y., and Lewbel, A., 2015. "Identifying the Effect of Changing the Policy Threshold in Regression Discontinuity Models." Review of Economics and Statistics, 97(5): 1081-1092.

Dupas P., Karlan, D., Robinson, J., and Ubfal, D., 2106. "Banking the Unbanked? Evidence from Three Countries." NBER Working Paper No. 22463.
Efron, B., and Hastie, T., 2016. Computer Age Statistical Inference. Cambridge University Press.

Engle, R., Hendry, D.F., and Richard, J.F., 1983. "Exogeneity." Econometrica, 51(2): 277-304.

Fisher, R.A., 1956. Statistical Methods and Scientific Inference. Oliver and Boyd.

Galiani, S., Gertler, P., Undurraga, R., Cooper, R., Martinez, S., and Ross, A., 2017. "Shelter from the Storm: Upgrading Housing Infrastructure in Latin American Slums." Journal of Urban Economics, 98: 187-213.

Gechter, M., Samii, C., Dehejia, R. and Pop-Eleches, C., 2018. Evaluating Ex Ante Counterfactual Predictions Using Ex Post Causal Inference. arXiv preprint arXiv:1806.07016.

Gertler, P., Shah, M., Alzua, M.L., Cameron, L., Martinez, S. and Patil, S., 2015. How Does Health Promotion Work? Evidence from the Dirty Business of Eliminating Open Defecation. NBER Working Paper No. 20997.Hahn, J., Kuersteiner, G. and Mazzocco, M., 2015. "Estimation with Aggregate Shocks." arXiv preprint arXiv:1507.04415.

Imbens, G.W., and Rubin, D.B., 2015. Causal Iference in Statistics, Social, and Biomedical Sciences. Cambridge University Press.

Manski, C.F., 2013. Public Policy in an Uncertain World: Analysis and Decisions. Harvard University Press.

Rosenbaum, P.R., 2010. Design of Observational Studies. Springer Press.

Rosenzweig, M., and Udry, C., 2018. "External Validity in a Stochastic World: Evidence from Low-Income Countries." Working Paper, Economic Growth Center, Yale University.

Wolpin, K., 2013. The Limits of Inference without Theory. MIT Press.

Gechter, M., Samii, C., Dehejia, R. and Pop-Eleches, C., 2018. Evaluating Ex Ante Counterfactual Predictions Using Ex Post Causal Inference. arXiv preprint arXiv:1806.07016.

References for Chapter 3

Acemoglu, Daron, Johnson, Simon, Robinson, James, 2005a. Institutions as the fundamental cause of long-run growth. In: Aghion, Philippe, Durlauf, Steven (Eds.), Handbook of Economic Growth, Vol. 1A. Elsevier, North-Holland, pp. 385-472.

Acemoglu, Daron, Johnson, Simon, Robinson, James, 2005b. The rise of Europe: Atlantic trade, institutional change, and economic growth. American Economic Review 95(3), 546-579.

Acemoglu, Daron, Cantoni, Davide, Johnson, Simon, Robinson, James, 2011. The Consequences of Radical Reform: The French Revolution. American Economic Review 101(7), 3286-3307.

Acemoglu, Daron, Robinson, James, 2012. Why Nations Fail. Crown Publishing Group.

Acemoglu, Daron, García-Jimeno, Camilo, Robinson, James, 2015. State capacity and economic development: A network approach. American Economic Review 105(8), 2364-2409.

Antras, Pol, 2015. Global Production: Firms, Contracts, and Trade Structure. Princeton University Press.

Besley, Timothy, Ghatak, Maitreesh, 2010. Property rights and economic development. In: Rodrik, Dani, Rosenzweig, Mark (Eds.), Handbook of Development Economics, Vol. 5. Elsevier, North-Holland, pp. 4525-4595.

Galiani, Sebastian, Schofield, Norman, Torrens, Gustavo, 2014. Factor endowments, democracy and trade policy divergence. Journal of Public Economic Theory 16(1), 119-56.

Galiani, Sebastian, Torrens, Gustavo, 2014. Autocracy, democracy and trade policy. Journal of International Economics 93, 173-193.

Greif, Avner, 2006. Institutions and the Path to The Modern Economy: Lessons from Medieval Trade. Cambridge University Press.

Limao, Nuno, Venables, Anthony, 2001. Infrastructure, geographical disadvantage, transport costs, and trade. World Bank Economic Review 15(3), 451-479.

Kee, Hiau Looi, Nicita, Alessandro, Olarreaga, Marcelo, 2009. Estimating trade restrictiveness indices. The Economic Journal 119, 172-199.

Kelejian, H., Murrell, Peter, Shepotylo, Oleksandr, 2014. Spatial spillovers in the development of institutions. Journal of Development Economics 101, 297-315.

Keller, Wolfgang, Shiue, Carol, 2016. Market integration as a mechanism of growth. CESifo working paper.

North, Douglass, 1990. Institutions, Institutional Change and Economic Performance. Cambridge University Press.

North, Douglass, 1981. Structure and Change in Economic History. W. W. Norton.

Wen, Mei, 1998. An Analytical Framework of Consumer-Producers, Economies of Specialization and Trade costs. In: Arrow, K., Ng, Y-K., Yang, X. (Eds.), Increasing Returns and Economic Analysis. Palgrave Macmillan, pp. 170-185.

Yang, Xiaokai, 2003. Economic Development and the Division of Labor. Blackwell Publishing.