

The origins of cultural divergence: evidence from Vietnam

Hoang-Anh Ho¹ · Peter Martinsson² · Ola Olsson²

Accepted: 27 July 2021 / Published online: 20 August 2021 © The Author(s) 2021

Abstract

Cultural norms diverge substantially across societies, often within the same country. We propose and investigate a *self-domestication/selective migration hypothesis*, proposing that cultural differences along the individualism-collectivism dimension are driven by the outmigration of individualistic people from collectivist core regions of states to peripheral frontier areas, and that such patterns of historical migration are reflected even in the current distribution of cultural norms. Gaining independence in 939 CE after about a thousand years of Chinese colonization, historical Vietnam emerged in the region that is now north Vietnam with a collectivist social organization. From the eleventh to the eighteenth centuries, historical Vietnam gradually expanded its territory southward to the Mekong River Delta through repeated waves of conquest and migration. Using a nationwide household survey, a population census, and a lab-in-the-field experiment, we demonstrate that areas annexed earlier to historical Vietnam are currently more prone to collectivist norms, and that these cultural norms are embodied in individual beliefs. Relying on many historical accounts, together with various robustness checks, we argue that the southward out-migration of individualistic people during the eight centuries of the territorial expansion is an important driver, among many others, of these cultural differences.

Keywords Culture · Selective migration · Vietnam

JEL Classification $N45 \cdot O53 \cdot Z1$

Ola Olsson ola.olsson@economics.gu.se

> Hoang-Anh Ho anh.ho@ueh.edu.vn

Peter Martinsson peter.martinsson@economics.gu.se

¹ University of Economics Ho Chi Minh City, Ho Chi Minh City, Vietnam

² University of Gothenburg, Gothenburg, Sweden

1 Introduction

Economic research has uncovered strong associations between many cultural traits and various indicators of individual behavior, and institutional and economic development (e.g., Guiso et al., 2011; Fernández, 2011; Algan & Pierre, 2014; Doepke & Fabrizio, 2014; Alesina & Giuliano, 2014, 2015). Among the cultural traits, the individualism–collectivism dimension has been found to be a powerful predictor of economic and democratic development in a large sample of countries (Gorodnichenko & Roland, 2011, 2015, 2017).¹ These empirical findings lead us to an important question: Why are some societies more or less collectivistic or individualistic than others?

In the present paper, we hypothesize that cultural differences along the individualism-collectivism dimension across modern societies can be traced back to repeated processes of territorial expansion and migration that happened in historical times. In particular, we advance a *selective migration hypothesis*, consisting of three building blocks. First, in regions where settled agriculture and states arose early, collectivist societies emerged through a process of *self-domestication* as communities made the transition from huntergatherer strategies of food procurement, which were characterized by individualism, to agricultural food production, resulting in a gradual strengthening of "civilizing" collectivism. Second, these collectivist societies triggered the out-migration of individualistic members toward peripheral areas. This pattern then repeated itself as the individualistic migrants inhabited and developed these peripheral areas into less collectivistic societies compared to the ones they left behind, which in turn induced more individualistic members to migrate toward more peripheral areas. Eventually, these migration processes gave rise to cultural differences along the individualism-collectivism dimension across societies. Third, owing to the slow-moving nature of culture, these differences have persisted over time and constitute an important feature of the cultural landscapes exhibited in modern times. As a result, the time elapsed since the collectivist transformation can predict the strength of collectivism across modern societies.

Testing the selective migration hypothesis requires a historical setting where there was a large out-migration of people from a collectivist society in the core region to settle down in new regions with a similar biogeography as in the core region, and that this migration repeated over time once collectivism was gradually strengthened in the new regions. We find such an ideal setting in the process of territorial expansion and migration in historical Vietnam. Gaining independence from the colonization of imperial China during the first millennium, historical Vietnam initially governed the region of what is now north Vietnam with a centralized government and a collectivist social organization. At the same time, the territory in the south of historical Vietnam was sparsely populated by many ethnic tribes that did not have a centralized government. From 968 to 1757, historical Vietnam gradually expanded its territory southward to the Mekong River Delta to establish the country as it is today (see Fig. 1). This process happened through successive waves of state conquest followed by civil migration, resulting in the displacement of most of the population of local ethnic tribes. Applying the logic of the selective migration hypothesis, we argue

¹ Some other notable traits include trust, family ties, generalized morality, and attitudes toward work and the perception of poverty. The individualism–collectivism dimension is also the single most relevant dimension in cultural psychology (Gorodnichenko & Roland, 2012). For a survey on the theory and measurement of the individualism–collectivism dimension in social psychology, see Triandis (1995).

that the time elapsed since annexation to historical Vietnam is an important predictor of the strength of collectivism across regions within contemporary Vietnam.

To test the selective migration hypothesis, an ideal empirical strategy would consist of three integral parts. The first part should demonstrate that some early agricultural states were characterized by collectivism, and that people who migrated to the new territories following state expansions were less collectivistic or more individualistic than those who stayed. The second requires historical data to prove that these selective migrations gave rise to early cultural differences along the individualism–collectivism dimension between the initial regions and the new territories. Finally, the third part involves using present-day data to conduct an empirical analysis on the relationship between the time elapsed since the collectivist transformation and the strength of collectivism.

To match this ideal empirical strategy in the context of Vietnam, we first present qualitative accounts to demonstrate that the initial society of historical Vietnam was characterized by strong collectivist norms. Second, we examine available primary records on the territorial expansion of historical Vietnam to identify the categories of people who migrated to the new annexed territories. In addition, we provide qualitative accounts and descriptive statistics to show that cultural differences along the individualism–collectivism dimension across regions were already present in Vietnam in the Seventeenth century. Third, we provide empirical evidence for a positive relationship between the time elapsed since an area was annexed to historical Vietnam and various indicators of collectivism in the present day. Using different robustness checks, we further show that these empirical findings are consistent with the self-domestication/selective migration hypothesis.

To capture the strength of collectivism, we focus on the societal ability to solve collective action problems, which is the main feature of collectivism studied in related economic models (Gorodnichenko & Roland, 2015, 2017). What constitutes a collective action, of course, varies significantly across societies. In Vietnam, labor contribution to public goods production is a typical collective action (Adams & Hancock, 1970). In particular, every year households in a local area send their members to work without payment to build or repair local public infrastructure such as roads, wells, irrigation, schools, and health clinics. Because collectivist societies are considered to be better at solving collective action problems, it should be able to mobilize a larger amount of voluntary labor contribution to public goods production from their in-group members.

Using data on voluntary labor contribution to public goods production from around 30,000 households in the Vietnam Household Living Standard Survey, we aggregate three related indicators at the district level: (1) the percentage of households contributing labor, (2) the average number of persons per household making labor contributions, and (3) the average number of labor days contributed per household. We find that districts annexed earlier to historical Vietnam currently have higher percentages of households contributing labor, more members per household making labor contributions, and more labor days contributed per household. The estimated effects are economically and statistically significant, and robust to the inclusion of potential confounding factors, various sub-samples, and omitted-variable bias, among other checks.

In addition, we also use data from the Population and Housing Census (covering in total 16.5 million households nationwide) to construct two measures of individualism–collectivism traits that have frequently been used in the literature as proxies for such cultural norms (Vandello & Cohen, 1999; Talhelm et al., 2014): *Extended family structure* and *marriage stability*. In line with our hypothesis, we find that households in districts that were annexed earlier have a higher percentage of households with grandchildren living in them and a lower prevalence of divorced households.



Fig. 1 The Vietnamese Southern advance *Note:* The year at which a district was annexed into historical Vietnam. See Online Appendix A for data sources

49

To examine in-group cooperation in more detail, we conduct a lab-in-the-field public goods experiment with high school students from the same districts, including an earlier-annexed district and a later-annexed district. This is a subject pool who are old enough to be aware of the cooperative norms in their communities, but not yet affected by living outside their communities. The advantage of the experiment is that the institutional setting can be kept constant, which helps ruling out the influences of informal institutions on cooperation behaviors. More importantly, the experimental design allows us to examine if the difference in the contribution to public goods between the two chosen districts is driven by a difference in preferences for cooperation or a difference in beliefs about the cooperative substantially more in the public goods experiments compared to subjects from the later-annexed district, and that the result is mainly driven by the belief about the contribution levels of the other subjects. Thus, the experimental findings corroborate the survey data analysis and further suggest that cultural differences across Vietnamese regions are embodied in individual beliefs.

Our research relates to a growing multidisciplinary literature examining the origins of cultural differences along the individualism–collectivism dimension.² Theories based on ecological context posit that some forms of production in subsistence economies (e.g., farming) require more functional interdependence than others (e.g., hunting), which gave rise to collectivism as an adaptation mechanism (e.g., Vandello & Cohen, 1999; Talhelm et al., 2014). In a recent paper, Buggle (2020) documents that societies where irrigation agriculture was practiced tend to have stronger collectivist norms (and a lower degree of innovative activities) even today. In related research, Bentzen et al. (2017) show that historical irrigation is also associated with autocratic governance. Litina (2016) argues further that lower level of natural land productivity increased the return to public agricultural infrastructure, which generated higher incentives for cooperation to solve the problem of collective action.³

Motivated by the history of settlement in the United States and its highly individualist culture, Kitayama et al. (2006, 2009) proposed the *voluntary settlement hypothesis*, asserting that settlers in frontier areas are likely to have highly autonomous, independent, and goal-oriented mindsets. Bazzi et al. (2020) expand on this theme and study the cultural legacy of the nineteenth century westward expansion in the United States. The authors show that contemporary individualism is stronger in historical frontier areas and that a selective migration of individualists to the periphery explains part of this pattern, along with the particular characteristics of wilderness and isolation in the west. Knudsen (2019) finds a similar pattern of selective migration among Scandinavian migrants to the United States in the nineteenth century, using uncommonness of first names as a proxy for the degree of individualism. She also documents that the out-migration of strong individualists made the home regions more collectivist in the long run. Giavazzi et al. (2019) study the evolution of preferences among European immigrants to the United States and find that the persistence in cultural attitudes is substantial across the spectrum of values. Using a similar line of argument as in the current paper, Olsson and Paik (2016) show that collectivism is

² The modernization hypothesis, arguing that societies become more individualistic as they reach higher levels of economic development, essentially focuses on the convergent tendency toward individualism, rather than pre-existing cultural differences across modern societies (Inglehart & Baker, 2000).

³ See also Fincher et al. (2008) who argue that societies with historically greater prevalence of diseasecausing pathogens are currently more prone to collectivist norms, because the emphasis on the in-group/ out-group distinction could serve as an anti-pathogen defense function.

stronger in regions across western Eurasia that adopted agriculture earlier during the Neolithic Revolution.

The present paper builds on and adds to this literature in various ways. To the best of our knowledge, no studies on the origins of cultural differences along the individualism–collectivism dimension have examined the societal ability to solve the problem of collective action, especially using a combination of survey and experimental data. Furthermore, most studies so far have either employed cross-country comparisons or concentrated on currently developed societies. Because these societies have gone through the modernization process to a greater extent, the reduction of the traditional cultural landscapes makes it harder to study the historical origins of cultural differences. By comparing different regions within a single and biogeographically homogenous developing country that experienced a relatively recent economic modernization, our research is able to overcome these limitations.

Our research also fits into a literature in economics examining the persistence of various cultural traits as an important channel through which historical processes could influence contemporary economic development (Nunn, 2012, 2014; Spolaore & Wacziarg, 2013). Some notable traits are gender equality (Alesina et al., 2013; Hansen et al., 2015), trust and cooperation (Nunn & Wantchekon, 2011; Becker et al., 2016; Bigoni et al., 2016, 2018; Guiso et al., 2016; Litina, 2016), anti-Semitic attitudes (Voigtländer & Voth, 2012), time preference (Galor & Özak, 2016), civic values Lowes et al. 2017 and norms favoring hard work (Fouka & Schläpfer, 2020). The theoretical basis for understanding intergenerational and other types of cultural transmission was pioneered by Bisin and Verdier (2001) and Richerson and Boyd (2005).

In a related recent study, Dell et al. (2018) use a border regression discontinuity design along a border segment in southern Vietnam which they claim was a stable demarcation between historical Vietnam and tributary polities to the Khmer Empire from 1698 to the 1830s. The main hypothesis is that the presence of a centralized historical state should crowd in local collective action, which in turn was beneficial for subsequent economic development. The authors show that living standards are currently higher in the border areas governed for a longer period of time by the centralized states of historical Vietnam. As one potential mechanism for this result, Dell et al. (2018) explore whether historical institutions contributed to a greater ability of collective action, measured by participation in civil society organizations.

Our research differentiates from Dell et al. (2018) in the following ways. First, our emphasis is on a mechanism of selective migration rather than on crowding in of norms by a powerful state. In the sections below, we argue that our mechanism has strong support in the historical literature on Vietnam, as well as reflecting a general pattern throughout Southeast Asia (Scott, 2009). It should be recognized though that the two hypotheses are strongly linked and that the true historical process probably had elements of both crowding in and selective migration. Second, our main outcome variable is cultural norms of collectivism-individualism rather than indicators for economic development. We would argue our research is complementary since it investigates a different cultural dimension for understanding long-run economic development. Third, rather than using a border regression discontinuity design as in Dell et al. (2018), our main empirical strategy is to exploit a countrywide sample of districts across all of Vietnam. Our basic rationale for this strategy is that our coding of the official chronicles of historical Vietnam suggests a more or less continuous process of state expansion that was completed in 1757.⁴

⁴ See Online Appendix A for a detailed account and motivation of our coding decisions regarding the Vietnamese southern advance. A further discussion of our coding of the southern advance in relation to Dell et al. (2018) is available upon request. Our investigation does not support an extended delay in this advance at the border of 1698, which is a key identifying assumption in Dell et al. (2018).

The remainder of the present paper is organized as follows. The next section discusses in detail the conceptual framework behind the selective migration hypothesis. Section 3 provides the historical background of the southward territorial expansion of historical Vietnam and the accompanying migration process, with a focus on the three building blocks of the selective migration hypothesis. Section 4 presents the empirical analysis with survey data. Section 5 describes the sample selection, experimental design, and corresponding results. Section 6 closes the paper with some concluding remarks.

2 Conceptual framework

In this section, we first define the individualism–collectivism dimension in the cultural repertoire of a population. We then outline a theory of *selective migration* and cultural divergence along the individualism–collectivism dimension. This theory is the backbone of the selective migration hypothesis.

2.1 Individualism versus collectivism

Research on the individualism–collectivism dimension of culture was first initiated within social psychology. Many of the key insights were summarized by Triandis (1995) and in subsequent cross-country empirical research by Hofstede (2001). In this voluminous literature, *collectivism* is considered to be characterized by a strong focus on the goals of a collective that forms the in-group boundary. In other words, the goals of the individual are subordinate to the goals of the collective, and the individual willingly makes costly sacrifices for the group. The individual typically has low self-expression and self-esteem, and an interdependent sense of agency. Family, duty, honor, and respect for the elders are central for collectivists. On the macro level, collectivist societies are typically characterized by a highly stratified or autocratic leadership (sometimes referred to as *vertical collectivism*) and hostility towards out-groups.

Individualism is the opposite of collectivism on all the features mentioned above. There is a strong focus on the goals of the individual, and the in-group identity is weak. The goals of the individual are superior to the goals of the collective, and the individual is typically unwilling to make costly contributions to the group at the expense of himself or herself. The individual has a strong sense of personal agency, high self-expression and self-esteem. The extended family does not play a central role, and individual preferences and fulfillment are more important than duty and honor. Individualists tend to live in egalitarian societies, are not very loyal to their fellow in-group members, and are willing to cooperate with outgroup members (Triandis, 1995).⁵

How do these cultural norms translate into economic behaviors? This issue has recently been studied in a series of papers by Gorodnichenko and Roland (2011, 2017). The authors outline a hypothesis and demonstrate empirically that societies characterized by individualism are less bound by rules and authority, reward personal achievement, and hence tend to

⁵ Although the characterization above is the most standard one in the literature, it has been proposed that there might also exist a "horizontal collectivism" whereby individuals in a collectivist society consider all their fellow in-group members as equal. Similarly, in a "vertically individualistic" culture, autonomous individuals accept substantial inequality (Singelis et al., 1995).

be associated with fewer constraints and stronger incentives for *innovation*. Analogously, the strong norms towards in-group cooperation, combined with subordination of the self to the goals of the collective, give collectivists a comparative advantage for *collective action* and *public goods production* that under certain circumstances might be necessary for the in-group to survive. Individualistic societies are thus more loosely held together but are on the other hand more dynamic, whereas collectivist societies have tight social ties and effective cooperation but limited growth potential in the longer run.⁶

2.2 Selective migration and cultural divergence

To develop the self-domestication/selective migration hypothesis, we build on Triandis (1995) and Olsson and Paik (2016), and assign a crucial role to the rise and consolidation of the early agricultural states. Before the first transition to agriculture about 12,000 years ago, all societies relied on hunting-gathering-fishing where the household of core family members was often the main unit of social organization. Households only stayed in larger camps during shorter periods but then splintered in order to avoid crowding and social tensions. In some environments, hunting required greater coordination, which sometimes led to larger and semi-sedentary social groups, but whenever possible, the basic tendency in pre-agricultural societies was autonomous households without stronger bonds or obligations to other in-group members (Johnson and Earle, 2000).⁷

The first agricultural societies emerged in regions such as Southwest Asia and China. In these regions, a highly productive irrigation agriculture gave rise to a dense and sedentary population, living in crowded villages and depending on the cultivation of a few domesticated crops and animals. Compared to hunter-gatherer households in pre-agricultural societies, these early farming villages were characterized by a much greater degree of collectivism, where the goals of the collective were far more important than individual aspirations. The survival of such villages often required sophisticated public goods such as irrigation canals, protective walls, military defense, public granaries, and deep wells. Such projects were initiated, coordinated and supervised by a social elite that managed to control the great majority of the population.

By the 4th millennium BCE, the first states arose from such complex farming communities in Mesopotamia, and soon after that, also in Egypt (Borcan et al., 2021; Scott, 2017). As has recently been studied by Mayshar et al. (2017, 2019), the ability of the elite to economically exploit the great majority of farmers depended to a great extent on the transparency and appropriability of agricultural production. Mayshar et al. (2019) argue that crops such as wheat and rice were more appropriable for taxation purposes than tubers like potatoes. In a related paper, Mayshar et al. (2017) contend that the greater reliability of the Nile floods in comparison to those of Euphrates and Tigris, explains why Egypt had a more centralized and more durable state organization during antiquity than the various state formations that repeatedly emerged and collapsed in Mesopotamia. Comparing the attitudes of people from Chinese areas dominated by highly labor- and coordination-intensive rice production with areas dominated by wheat cultivation, Talhelm et al. (2014) find

⁶ Buggle (2020) provides evidence that regions that practiced irrigation agriculture historically have a more collectivist contemporary culture and also tend to have less innovations and more routine-intensive occupations.

⁷ See, for instance, Johnson and Earle (2000)'s description of the contemporary Kung people of the Kalahari, a society which has been considered a modern remnant of prehistorical social organization.

that collectivist norms are stronger in rice areas. Thus, even among communities where a sedentary population practiced a cultivation of domesticated plants, there were great differences in the strength of individualism–collectivism depending on the specific character of the agricultural production process.

The domestication of plants and animals some 12,000 years ago did not only fundamentally change human subsistence production, it also gradually domesticated humans themselves (Johnson & Earle, 2000; Scott, 2017). The term *domestication* denotes a longterm relationship where one species strongly influences and controls the reproduction and care of another species so that the domesticated species loses its capacity to survive in its original habitat. The most often discussed of such relationships is of course modern man's domestication of plants and animals during the Neolithic.

By *self-domestication*, we mean the process whereby a new kind of subsistence food production and its associated social organization, induced humans to adopt a more collaborative and docile social behavior, which in turn implied a loss of capacity to survive on their own. One might for instance argue that the control of fire and the universally adopted inventions of cooking, several hundred thousand years ago, implied a first self-domestication of hominins.⁸

The Neolithic transition to agriculture, was a second self-domestication event. In the permanent early farming villages, food production was more functionally interdependent across households and people had to live together in much larger units than before and without the option of fissioning and setting up camp elsewhere as in most hunting and gathering communities. The natural inclinations towards family-level social units were overcome through a strong selective pressure favoring individuals and groups who could more successfully adapt to the new lifestyle in the farming villages, with a higher pathogen load, more toil and work hours in the fields, a new diet with more carbohydrates and less protein, and more children per woman.

In addition to these changes, we argue that there must have been a very strong pressure towards the adoption of collectivist norms. It is well known that social stratification expanded with agriculture and the goals of the individual were suppressed for the benefit of the common good, involving larger and larger collective action projects such as irrigation, city walls, the construction of cult centers, and even massive burial complexes for divine rulers (Diamond, 1997). This development was not only evident in areas with irrigation agriculture. The fact that people became sedentary also in rain-fed areas implied the need for much stronger investments in permanent physical structures, property rights claims tied to a clan or lineage, and a stronger reliance on protection from a strongman or a proto-state. This kind of social organization would not have been possible without a great increase in the proliferation of collectivist norms. In the official narrative of this process, chroniclers of the early states would typically describe it as the introduction of *civilization* to an environment populated by previously primitive or barbaric tribes.

We argue that the civilizing self-domestication process discussed above first emerged in regions with favorable conditions for agriculture, but it then repeated itself all around the world when farming replaced hunting-gathering and states arose from the dense farming communities. This self-domestication process probably included several related mechanisms. *Evolution* provides a selective advantage for individuals with genes that helped

⁸ See Wrangham and Carmody (2010), for a famous account of how the control of fire led to evolutionary adaptations in the gut so that humans lost the ability to process most raw food. The use of cooking and fire also led to important innovations in social organization and facilitated the expansion of the human brain.

them cope with the physiological, psychological and cultural challenges of intensive farming. In addition, there were presumably *push factors* such as a conscious "weeding out" of individualists who did not adapt to the new collectivist norms. Social exclusion or ostracism might be one mechanism whereby individualists were pushed from the collectivist core to peripheral areas.

There were surely also *pull factors*, inducing individualists to leave the collectivist core voluntarily in order to have a freer life at the periphery. In his description of the history of state formation in Southeast Asia, Scott (2009) describes the very conscious escape of population groups from the rice-growing core areas of the different states as "an art of *not* being governed". For these people, a withdrawal to peripheral areas was a key feature of their strategy of *state evasion*. Ho (2020) further shows that the Vietnamese state often tempted landless migrants with the prospect of obtaining private property rights in the new lands, rights that were sometimes retracted after a few generations when population density had increased.

A similar "pioneer spirit" was emphasized in Turner (1920)'s work on the westward expansion of settlers in North America, and more recently studied by Bazzi et al. (2020). It is also similar to Kitayama et al. (2006, 2009)'s notion of *voluntary settlement* of peripheral areas by individualistic people. As discussed by Bazzi et al. (2020), the adaptation to the living conditions in the "rugged frontier", where a strong sense of individual agency most likely was necessary for survival, surely also contributed to a greater degree of individualism even among people with collectivistic inclinations. As shown by Knudsen (2019), the out-migration of individualists probably contributed to making the culture in the core even more collectivist than before.

Typically, the peripheries to the original agricultural core region were sooner or later colonized by a collectivist farmer-state through territorial expansion. Evolutionary adaptation, push and pull forces then played out in a similar manner, making the peripheral population more collectivistic as well. The exact nature of these adaptations would depend importantly on the biogeographical characteristics of the settled peripheral areas, which in turn would determine the specific technology of agricultural production. But as described by Olsson and Paik (2016) in their application of the selective migration logic to the expansion of Neolithic agriculture throughout the Western hemisphere, the most individualistic people in the periphery would soon once again take off towards more peripheral areas in repeated frontier colonizations.

Scott (2009) provides a narrative account of how rice-based states gradually expanded across Southeast Asia and provoked marginal population groups to settle the highlands or the more peripheral parts of the lowland plains. The core areas of the early states in contemporary Myanmar, Thailand, Cambodia and Vietnam were generally characterized by highly productive irrigated rice cultivation in the lowland valleys of major rivers such as the Irrawaddy (Myanmar), the Mekong (Cambodia) and Red River (Vietnam). Historical Southeast Asia had a much lower population density than India and China, which meant that the periphery was often a feasible alternative for population groups who, for different reasons, wanted to evade the influence of the central governments. There were many different strategies used and reasons for trying to evade the influence of the expanding states in the region. In Scott (2009, p. 326)'s own words:

Those who for whatever reason wished to evade incorporation as subjects had to place themselves out of range either on the plains at greater remove from the core or in the less accessible hills. ... it is clear that the hills were populated increasingly by pulses of migration by state subjects fleeing valley kingdoms for any one of several

reasons - corvée labor, taxes, conscription, war, struggles over succession, religious dissent - all having directly to do with state making.

Despite this unwillingness to be subjects of the expanding states, the populations in the periphery were often willing to engage in mutually beneficial trade. Their societies were non-hierarchical and fluid and often based on foraging or swiddening agriculture. The officials of the rice states typically considered the peripheral populations as uncivilized and barbaric (Scott, 2009). In the terminology of our framework above, we might describe them as non-domesticated individualists.

Since self-domestication, just like evolution, is a function of time, the penetration of collectivist norms was typically also an increasing function of the time elapsed since the "civilizing" collectivist transformation. In this manner, a gradient arose with the greatest degree of collectivism in the oldest regions and the highest degree of individualism in the youngest territories of the farmer-state. The slow-moving nature of culture implied that, centuries or even millennia after the first settlement of individualistic farmers, signals from these early migration processes are still visible in contemporary cultural record.⁹

Nevertheless, as already argued by Triandis (1995), the Industrial Revolution in Europe, with innovation as a key driving factor, once again turned the tables and gave individualism an economic advantage in north European countries such as Britain and the Netherlands. Thus, we might expect that the collectivist legacy of the transition from hunting-gathering to farming should be weaker in countries where an industrial economy has existed for a longer period of time. In addition, Western colonization of regions outside Europe might change the indigenous cultural landscapes to a large extent. In some developing countries that only experienced industrialization recently and had strong indigenous states, the cultural imprint from the historical expansion of the collectivist farmer-state is more likely to be observable in the present day.

3 Historical background

In the previous section, we outlined a theory of self-domestication, selective migration and cultural divergence along the individualism–collectivism dimension. In this section, we survey historical materials to examine three building blocks of our theory in the context of Vietnam: (1) the initial region of historical Vietnam was home to a collectivist society; (2) individualistic people migrated southward as the country expanded its territory, eventually giving rise to cultural differences along the individualism–collectivism dimension; and (3) these cultural differences have persisted to the present day.

3.1 Core region of historical vietnam was a collectivist society

The history of modern humans in Southeast Asia goes back at least 65,000 years and genetic analyses suggest that the region was an important node in out-of-Africa migrations into China, East Asia, and Oceania (Pischedda et al., 2017). Archaeological evidence indicates that ancient populations, probably migrating from southern China, had

⁹ For instance, Olsson and Paik (2016) show that peripheral Scandinavia and other North European countries still have much stronger individualistic norms than people in the old agricultural regions in the Middle East.

settled down in the Red River Delta with rice agriculture around 2000 BCE during the Neolithic Revolution (Nguyen et al., 2004). They gradually assimilated or replaced indigenous hunter-gatherer groups in the area. The farming populations lived together, without a centralized state, in the region that is now north Vietnam (see Fig. 1). From 111 BCE to 939 CE, the whole region was brought under the colonization of the centralized bureaucracy of imperial China. During this period, "the Vietnamese evolved from a preliterate society within a "south-sea civilization" into a distinctive member of the East Asian cultural world" (Taylor, 1983, p. xvii).

After the victory over historical China in 939 CE, the first unified state of historical Vietnam was founded in 968 CE and inherited a centralized bureaucratic system from the Chinese colonizer (Taylor, 2013, p. 51-77). Thus, in terms of the theory discussed above, we might argue that Vietnamese society was, to a large extent, domesticated into a collectivist social organization by an external agent. The dominant ethnic group in this new state formation was the Kinh, speaking an Austroasiatic language with its roots in Southern China. Subsequent dynasties governing historical Vietnam continued to build stronger structures and orders into the society, which emphasized the values of social groups above the needs and desires of its constituent members (Whitmore, 1984, 1997). The collectivist nature of historical Vietnam was best exemplified by its village-based administrative system and family organization. The village was the lowest administrative level, which was responsible for regulating almost all aspects of the daily living of its members (Nguyen, 2003). Two important responsibilities of the village were to allocate public land under its management to its members (Dao, 1993), and to organize unpaid labor for public goods production such as irrigation facilities, roads, and communal buildings (Adams & Hancock, 1970). With respect to the family, parents had absolute authority over their children in almost all aspects of life (e.g., education, marriage, and housing), while children had to serve and obey their parents with the utmost respect throughout their lives (Haines, 1984).

The area bordering historical Vietnam in the south, which is now central Vietnam (see Fig. 1), was inhabited by various ethnic groups that formed the Champa Kingdom. The dominant ethnic group at the time were the *Cham*, who was an Austronesian-speaking group that is believed to have orgininally settled central Vietnam sometime during the first millennium BCE. The settlement was part of a larger expansion of Austronesians, probably originating from Taiwan, across Indonesia, Melanesia and the Pacific Islands. And next to the Champa Kingdom in the south, which is now south Vietnam (Fig. 1), was a large area of swampy forest belonging to the Khmer Empire.

In contrast to the centralized state of historical Vietnam, both the Champa Kingdom and Khmer Empire were basically networks of small political entities (Hall, 2011, p. 67–102, 159–210). The Champa Kingdom in the south was traditionally a trading-oriented nation integrated in the south Asian spice trade and in the broader Austronesian cultural community. Available historical materials do not allow us to draw any comparison between these societies and historical Vietnam along the individualism–collectivism dimension. However, in the terminology of Scott (2009), it is clear that southern Vietnam, over long periods populated mainly by Cham and Khmer ethnic groups, was a periphery to the more centralized states in the core areas of historical Vietnam and the Khmer Empire. The fact that the Champa Kingdom was less centralized and more open to contact with foreigners, probably made it a relatively attractive refuge for more individualist people during the Vietnamese southern advance (see also below).

3.2 Selective migration and cultural divergence

From 968 to 1757, historical Vietnam expanded its territory southward along the coast to the Mekong River Delta. This so-called Vietnamese *Southern Advance (Nam Tien)* took place gradually through various annexations and was completed in 1757, by which time the border of Vietnam was established as it is today (see Online Appendix A). Historical Vietnam first annexed the land from modern Quang Binh to modern Binh Dinh from 968 to 1471. This land was effectively governed by the Nguyen Lords since the early sixteenth century, when the fight to control the throne erupted between them and the Trinh Lords in the initial core region. From 1611 to 1757, the Nguyen Lords continued to expand the country southwards to the Mekong River Delta to establish the border as it is today. Compared to the initial region, the annexed region under the government of the Nguyen Lords was more open towards foreign trade (Tana, 1998, p. 59–98).

After historical Vietnam conquered an area, immigration into the area always followed. There are two types of evidence on this process: anecdotal evidence from historical chronicles and ethno-genetic evidence from historical and contemporary populations. Records from the two official chronicles of historical Vietnam, Dai Viet Su Ky Toan Thu (from 204 BCE to 1675) and Dai Nam Thuc Luc (from 1558 to 1888), indicate that Vietnamese migrants to the annexed region ranged from landless farmers to rich adventurers, who took advantage of the opportunities in the new land, and from exiled criminals to recruited soldiers, who were sent to the new land by the government (see Online Appendix B). There are no records available to identify who were the dominant settlers, let alone their cultural characteristics.¹⁰ Regarding the local ethnic groups, most of their populations moved to more distant peripheries such as the highlands or to the hinterland of other states (Scott, 2009), while those who decided to stay were acculturated to the Vietnamese culture (Wook, 2004).¹¹ The logic of the selective migration hypothesis discussed in the previous section implies that cultural differences along the individualism-collectivism dimension between the annexed region and the core region would emerge as a result of the selective out-migration of individualists. The historical evidence presented below is in support of this prediction.

Within historical Vietnam, cultural differences along the individualism–collectivism dimension between the annexed region and the core region were already remarkable as early as the seventeenth century. For example, Tana (1998, p. 99–116) provides many historical accounts to demonstrate that the social environment in the annexed region was characterized by greater openness, mobility and autonomy compared to the core region. Available statistics of land allocation in the early nineteenth century also illustrate this cultural divergence. In the core region, land was only allocated to or owned by village members (Nguyen, 2003). In the annexed region, however, the in-village/out-village distinction was

¹⁰ Historians seem to be in agreement that these migrants were not recognized as true members of the collectivistic society in the initial region. Such a migrant "lacked standing in a social group, like the family or the village, was less than a full person and could hope for no better future in traditional village society" (Tana, 1998, p. 111), or even was "marginal and undesirable" (Taylor, 2013, p. 211). One may recall the analogous image of European immigrants to the United States during the age of mass migration (Abramitzky et al., 2012). In a related study, Alesina and Giuliano (2010) also find that people with weak family ties are more likely to migrate.

¹¹ In modern times, their descendants only constituted minor fractions in the total population of Vietnam; for example in 1999, the Cham were 0.17% and the Khmer were 1.38% (General Statistics Office of Vietnam, 2001, p. 167).

loosened and land was commonly allocated to or owned by people from other villages. For example, studies on the land registries (cadastres) in the annexed region in the early nineteenth century show that the proportions of land allocated to or owned by people from other villages were 20–30% in the southernmost provinces (Nguyen, 1994) compared to 8–15% in more northern provinces (Nguyen, 1997, 2010, 1996a, b).

The second kind of evidence on the Southern Advance migrations comes from the ethno-genetic configuration of the contemporary population. Although nothing like a census from 968 CE exists, most sources indicate that the territory of current Vietnam was a mosaic of smaller ethnic groups a thousand years ago. Today, Vietnam's population is dominated by the Kinh group (86% of the population) while there are 53 recognized ethnic minorities that make up the remaining 14%. These 54 groups can in turn be subdivided into five different macro language families, of which the Austroasiatic family (henceforth "AA", including the Kinh and Muong) is by far the largest.

The people who migrated southwards and into the highlands during the Southern Advance were of two different kinds: individualistic people from Kinh and other AAspeakers who wanted to evade the burdens of the expanding Vietnamese state, and people from out-groups who moved in order to preserve their cultural and political independence vis-à-vis the dominant Kinh. In a recent study, Liu et al. (2020) collect genome-wide data from 22 ethnic groups in Vietnam, including the Kinh, Muong, and Cham. On the basis of this data, it is possible to reconstruct each group's effective population size during the past 50 generations (equivalent to about 1450 years). It is shown that whereas most groups have gone through a decline in effective population sizes, the closely related AA-speakers Kinh and Muong experienced a rapid expansion about 20 generations back (Liu et al., 2020, Figure 5). This surge coincides closely with the date of a great military defeat of the Champa Kingdom against the Vietnamese state in 1471. As a result of this defeat, a great share of the Cham population were either killed or enslaved and replaced by incoming Kinh. In the 1700s, the last remnants of the Champa Kingdom was dissolved and the Cham group experienced a large out-migration to what is now Cambodia. Those who remained were to a great extent assimilated into the Vietnamese state.

Among the out-group refugees from the Vietnamese state, one would expect that trying to retain their unique ethnic identity was a key motivating factor. Several of these groups indeed still speak a non-AA language, which sets them apart from the dominant Kinh. Somewhat surprisingly, the data from Liu et al. (2020) show that a number of these contemporary ethnic groups are actually genetically proximate to Kinh. A likely reason is that Kinh highland colonists over the centuries have interbred with the highland peoples in a more or less forced manner. Despite this genetic proximity, these groups have managed to maintain a distinct cultural identity.

Besides the selective migration of individualistic people as proposed by our theory, there are certainly other potential explanations for the cultural differences along the individualism–collectivism dimension between the annexed southern and northern core regions of historical Vietnam as described above. First, the characteristics of the frontier environment in the annexed region (e.g., sparsely populated) might have induced Vietnamese migrants to be more individualistic. Second, Vietnamese migrants might have been influenced by the cultural characteristics of the Champa Kingdom and Khmer Empire, which in turn might have been more individualistic than historical Vietnam. Finally, Vietnamese migrants to the annexed region in the south might have picked up individualistic traits from foreigners because of the open trade policy of the Nguyen Lords, which in turn was a continuation of the trade orientation of the Champa Kingdom within a greater cross-border Austronesian community. The main difference between our theory of selective migration and these explanations is that our theory predicts cultural differences even within the annexed region, i.e., areas annexed earlier are predicted to be more collectivistic. The empirical evidence presented below is in support of this prediction.

3.3 Cultural differences have persisted to the present day

The last block of the selective migration hypothesis argues that the cultural differences across regions of historical Vietnam found around the seventeenth century have persisted and made up a key characteristic of the cultural landscape of modern Vietnam.¹² In other words, the time elapsed since annexation to historical Vietnam is an important predictor of the strength of collectivism within contemporary Vietnam. The north-south cultural differences along the individualism–collectivism dimension in modern Vietnam have been documented in details in many anthropological studies, e.g., Hickey (1964), Rambo (1973), and Van Luong (1992). This north-south cultural divergence is also a typical characteristic that is normally mentioned in descriptions about modern Vietnam.¹³

To sum up, the north-south cultural differences along the individualism–collectivism dimension were already in place as early as the seventeenth century and are currently a central theme of Vietnam. Our theory of selective migration discussed in the previous section predicts that areas annexed earlier to historical Vietnam are currently more prone to collectivist norms, and this relationship holds even within the annexed region. We now turn to investigate these predictions empirically using survey data in Sect. 4 and experimental data in Sect. 5.

4 Survey data analysis

4.1 Empirical model

In this section, we use survey data to investigate the proposed theory of selective migration in the context of Vietnam. The key argument of the theory is that collectivist societies triggered the out-migration of individualistic members toward peripheral areas, and, owing to the slow-moving nature of culture, these differences have persisted over time. Our empirical strategy revolves around regressing a measure capturing the strength of collectivism on the time elapsed since a district was annexed to historical Vietnam, while controlling for potential confounding factors. The core regression model takes the following form:

¹² The French colonization started in 1858 and ended with the Vietnamese victory in the First Indochina War (1946–1954), during which the French colonizers concentrated most of their activities in south Vietnam. Following the French defeat was the American intervention in south Vietnam (i.e., the Second Indochina War, commonly known as the Vietnam War), which ended with the reunification of the country in 1975. Meanwhile, Communism started to develop in north Vietnam in the early twentieth century and gained control of this part of the country since 1954.

¹³ Ending his Vietnamese history, Taylor (2013, p. 624) notes that "northerners are more disciplined to accept and to exercise government authority" and "southerners are more individualistic, egalitarian, entrepreneurial, interested in wealth more than in authority". Although regarding collectivism as the main cultural theme in their practical guide to Vietnam, Ashwill and Diep (2005, p. 71–72) note that "northerners are considered to be more intelligent, conservative, austere, serious, and frugal, ..., [and] are more apt to save for a rainy day", while "southerners are perceived as fun-loving, easy-going, open people who rarely think of saving for a rainy day".

$$Collectivism_{i} = \beta TimeSinceAnnexation_{i} + \gamma X_{i} + \epsilon_{i}.$$
(1)

In this equation, *Collectivism_i* is a measure of the strength of collectivism in district *i*, *TimeSinceAnnexation_i* is the time since annexation to historical Vietnam, X_i is a set of potential confounding factors, and ϵ_i is a random error term. Our hypothesis postulates that β is positive with respect to the strength of the collectivist measure, i.e., the longer the time since annexation, the stronger the collectivist norms.

Ideally, our main independent variable should capture historical migrations from the core area. Unfortunately, we have not been able to find such a direct measure of selective migration. Our main variable *TimeSinceAnnexation*_i is an indirect proxy for historical migrations in the sense that we should expect that regions annexed last should host the greatest amount of population groups seeking to evade the influence of the northern state. To what extent would an estimated coefficient of $\beta > 0$ rule out other potential hypotheses regarding the persistent cultural impact of historical states? In particular, does our main explanatory variable allow us to distinguish between (1) selective migration, (2) a crowding-in of collectivist norms by a strong state, and (3) migrants' adoption of individualist norms that were already strong in the southern periphery?

We argue that a $\beta > 0$ would be consistent with our selective migration hypothesis, but that it would not disprove the two other hypotheses. In fact, as discussed earlier in our theoretical framework, we recognize that there is a significant overlap between the three hypotheses, and that they are to some extent reflections of the same underlying process. For instance, a strong crowding-in of norms by a collectivist state will *push* individualists to migrate to the periphery, and the absence of a strong state and a culture of individualism in the south will *pull* an even greater number of individualist migrants to the periphery.

4.2 Variables

The Individualism–Collectivism Trait. In the present paper, we follow the conventional definition of culture in economic research as "decision making heuristics or 'rules of thumb' that have evolved given our need to make decisions in complex and uncertain environments" (Nunn, 2012, p. S109).¹⁴ Based on this definition, many observable outcomes have been used in the literature to capture different aspects of the individualism–collectivism trait, such as extended family structure, marriage stability, and inventiveness (Vandello & Cohen, 1999; Talhelm et al., 2014) or unusual names (Bazzi et al., 2020; Knudsen, 2019).¹⁵ We argue that an outcome variable must satisfy two conditions to be a good measure of the individualism–collectivism dimension. First, it must capture an aspect of the individual ism–collectivism trait that is theoretically relevant to understand individual behaviors or economic development. Second, it must feature as a traditional practice of the society in question, i.e., it captures a decision making heuristic in daily living.

In the present paper, we use voluntary labor contribution to public goods production to capture the strength of collectivism. We argue that this measure satisfies the two conditions

¹⁴ This definition is closely related to another prominent one proposed by Guiso et al. (2006, p. 23): "customary beliefs and values that ethnic, religious, and social groups transmit fairly unchanged from generation to generation."

¹⁵ Other studies also use value survey such as the World Value Survey to capture the individualism–collectivism trait (Olsson & Paik, 2016; Buggle, 2020). Unfortunately, respondents' locations at the district level are not provided owing to data policy. This policy also applies in the case of the Asian Barometer Survey.

mentioned above. First, the ability to solve collective action problems such as public goods production is the main feature of collectivism in related economic models (Gorodnichenko and Roland, 2015, 2017). Because collectivist societies are considered to be better in this respect, they should be able to mobilize a larger amount of voluntary labor contribution to public goods production from their in-group members. Second, labor contribution to public goods production is a traditional activity in Vietnam (Adams & Hancock, 1970). In particular, every year households in a local area send their members to work without payment to build or repair local public infrastructure such as roads, wells, irrigation, schools, and health clinics. These labor contributions are not paid, and hence are arguably voluntary. Thus, the decision to contribute labor to public goods production should capture a decision making heuristic in daily living.

Our main dataset is the Vietnam Household Living Standard Survey (VHLSS) in 2002, which covers almost 30,000 households in 607 (out of 630) districts (roughly 50 households per district) across all 61 provinces in Vietnam and is the only survey round that contains detailed information about voluntary labor contribution to public goods production. We measure cultural norms at the district level by aggregation of household data.¹⁶ In particular, we construct three related variables based on voluntary labor contribution to public goods production. First, we calculate the percentage of households contributing labor in the district to measure the prevalence of labor contributions. Second, we calculate the average number of persons making labor contributions per household. Finally, we calculate the average number of labor days contributed per household. These last two variables capture the intensity of labor contributions. Table 1 shows that, in 2002, around 31% percent of households contributed labor to public goods production, whereas the average number of persons making labor contributions per household is 0.55 and the average number of labor days contributed per household is 4.05. Figure 2 provides a spatial presentation of these data. A visual comparison with Fig. 1 suggests that districts annexed earlier to historical Vietnam currently have higher percentages of households contributing labor, more members per household making labor contributions, and more labor days contributed per household.

To further validate the empirical results, we also employ two other measures of the individualism–collectivism trait that have been used in previous studies: extended family structure and marriage stability (Vandello & Cohen, 1999; Talhelm et al., 2014). Households in collectivist societies should be more likely to include extended family members rather than simply nuclear family members, and less likely to divorce. We follow previous studies to measure at the district level extended family structure by the percentage of households with grandchildren living in them, and marriage instability by the number of divorced households per 100 married households. We use data from the Population and Housing Census in 1999, which cover around 16.5 million households across all districts in Vietnam. Table 1 shows that, on average, the percentage of households with grandchildren in them is 7.2% (ranging from 0 to 27.17%), and the number of divorced households per 100 married households is 2.46% (ranging from 0.21% to 7.82%). As shown in Table D1 in Online Appendix D, the correlation coefficient between labor contribution and extended family is 0.5 (*p* value = 0.000), labor contribution and marriage instability is -0.3 (*p* value = 0.000), and extended family and marriage instability is -0.2 (*p* value = 0.000).

¹⁶ There is certainly individual heterogeneity within a society, but in aggregation one can observe what social psychologists call the "cultural syndrome" of each society (Triandis, 1995). This is also a standard exercise in economics (Alesina & Giuliano, 2015).

| Table 1 Variable description | | | | | | |
|------------------------------|---|-------|-------|-------|-------|-----|
| Variable | Description | Mean | SD | Min | Мах | z |
| Main variables | | | | | | |
| Having contributed | Percentage of households contributing labor in 2002 | 31.30 | 32.90 | 0 | 100 | 607 |
| Members contributed | Average no. of persons per household making contributions in 2002 | 0.55 | 0.65 | 0 | 4.10 | 607 |
| Days contributed | Average no. of days contributed per household in 2002 | 4.05 | 5.63 | 0 | 41 | 607 |
| Extended family | Percentage of households with grandchildren in them in 1999 | 7.20 | 4.82 | 0 | 27.17 | 613 |
| Marriage stability | Number of divorced households per 100 married households in 1999 | 2.46 | 1.04 | 0.21 | 7.82 | 613 |
| Time since annexation | Number of centuries since annexation into historical Vietnam | 6.77 | 3.57 | 2.33 | 10.22 | 607 |
| Control variables | | | | | | |
| Caloric suitability | Average yield (million kilo calories per km ² per year) | 25.08 | 2.50 | 15.88 | 31.74 | 607 |
| Distance to coast | Shortest distance to the coastal line (100 km) | 0.73 | 0.81 | 0 | 4.30 | 607 |
| Elevation | Average height above sea level (km) | 0.20 | 0.30 | 0.001 | 1.56 | 607 |
| Ruggedness | Average topographic heterogeneity (10,000 km) | 1.14 | 1.43 | 0 | 6.27 | 607 |
| Irrigation suitability | Average suitability for irrigation, not suitable (0) and suitable (1) | 0.09 | 0.25 | 0 | 1 | 607 |
| Am climatic zone | Equatorial monsoonal | 0.19 | 0.39 | 0 | 1 | 607 |
| Aw climatic zone | Equatorial winter dry | 0.47 | 0.50 | 0 | 1 | 607 |
| Cfa climatic zone | Warm temperature, fully humid, hot summer | 0.01 | 0.11 | 0 | 1 | 607 |
| Cwa climatic zone | Warm temperature, winter dry, hot summer | 0.32 | 0.47 | 0 | 1 | 607 |
| Distance to Quang Binh | Walking distance to Quang Binh (100 km) | 8.43 | 3.52 | 0 | 13.49 | 311 |
| | | | | | | |

All variables are measured at district level. See the main text for more detail

The Time since Annexation to Historical Vietnam. As previously mentioned, our main explanatory variable is the time elapsed since an area was annexed to historical Vietnam. Following the historical background discussed earlier, we choose the first unified state of historical Vietnam in 968 as the beginning year, while the terminal year is set to 1990. In our analyses, we measure the time since annexation in centuries (100 years) to make the estimated coefficients easy to read in the reported tables. The descriptive statistics in Table 1 show that the annexations took place between 2.33 to 10.22 centuries before the terminal year of 1990.

To construct the time since annexation to historical Vietnam for each modern district, two dimensions are needed: (1) the district's corresponding area in historical Vietnam and (2) the year that this area was annexed. For the year that a historical area was annexed, we rely on two official chronicles of historical Vietnam, Dai (2012) and Dai (2002), recording events from the beginning to 1675 and in the 1558–1888 period, respectively. These chronicles were written by state officials of historical Vietnam to keep track of events and, to the best of our knowledge, constituted the primary sources for Vietnamese histories. We code an area as having been annexed when there is a record in the chronicles demonstrating that this area was under the control of historical Vietnam. To link historical areas to their modern counterparts, we rely on two seminal works of Vietnamese historians: Dao (2005) and Le Phan et al. (2011). All details on the coding procedure are presented in Online Appendix A.

Control Variables. To tackle the endogeneity of the time since annexation into historical Vietnam, we identify a set of potential confounding factors, i.e., factors that might influence both the time since annexation to historical Vietnam and the strength of collectivism. A necessary condition for a variable to be a confounding factor is that it must have existed before the annexation to historical Vietnam. Variables that came to exist after the annexation such as demographic characteristics in the modern day might be caused by the annexation, and hence are bad controls (Angrist & Pischke, 2009). Nevertheless, as shown in the following subsection, our empirical results are also robust to the inclusion of numerous bad controls. Below, we describe the set of potential confounding factors in detail.

First, agricultural suitability might have both attracted historical Vietnam to conquer a region and promoted the development of collectivism. We control for natural land productivity, which has been argued to influence the incentive to cooperate in the production of public infrastructure in the subsistence agricultural economy (Litina, 2016). Second, geographical conditions might affect the difficulty in conquering a region. Geographical isolation is also conducive to cultural assimilation and the development of an in-group identity, giving rise to collectivism (Triandis, 1995). We control for distance to the coast, elevation, and ruggedness to capture geographical isolation. Third, we also control for irrigation suitability, because irrigation agriculture has been shown to be conducive to the development of collectivism (Buggle, 2020). In addition, we control for climatic zones to capture any potential influence of climatic conditions on the development of collectivism.

Natural land productivity is measured by caloric suitability constructed at 5 arc-minute resolution by Galor and Özak (2016), who make their calculation based on data from the Global Agro-Ecological Zones project of the Food and Agriculture Organization. This index measures the average potential yield (million kilo calories per squared kilometer per year) attainable in each grid cell given the set of crops that are suitable for cultivation in the post-1500 period. To capture the natural component of land productivity, the production conditions are set at a low level of inputs and rain-fed agriculture based on agro-climatic conditions, which are unaffected by human intervention. Distance to the coast is measured by the shortest (bird-fly) distance to the coastal line. Elevation is taken from the Global 30





Arc-Second Elevation Dataset (GTOPO30) provided by the Earth Resources Observation and Science Center. The terrain ruggedness index was originally devised by Riley et al. (1999) to quantify topographic heterogeneity in wildlife habitats providing concealment for prey and lookout posts. This index is calculated by Nunn and Puga (2012) based on the GTOPO30 dataset.

The Food and Agriculture Organization of the United Nations (FAO) defines irrigation suitability as the potential increase in agricultural output that can be obtained by fully exploiting irrigation compared to rain-fed agriculture, and classifies irrigation suitability into five classes: (1) only suitable for rain-fed agriculture, (2) output yield increases by 0–20%, (3) output increases by 20–50%, (4) output increases by 50–100%, and (5) output increases by more than 100% (Fischer et al., 2002). Following Buggle (2020), we classify an area to be suitable for irrigation if agricultural output increases by at least 50%. Climatic zones are defined by the Köppen–Geiger classification, and are constructed using precipitation and temperature data in the period of 1901–1925 (Rubel & Kottek, 2010). Descriptive statistics of all variables can be found in Table 1.

4.3 Baseline results

To begin with, we regress the percentage of households contributing labor in a district on the time since a district was annexed to historical Vietnam, controlling for potential confounding factors as discussed above. In all regressions, we report both robust standard errors and standard errors adjusted for spatial autocorrelation following Conley (1999).¹⁷ Table 2 shows that the estimated coefficients of the time since annexation are positive and significant, whether or not all control variables are included. Thus, districts annexed earlier to historical Vietnam today have a higher percentage of households contributing labor on average. Relative to the mean value of the dependent variable, the marginal effect is economically significant. When all control variables are included, for example, a one century increase in the time since annexation is associated with an additional 1.9% points of households contributing labor, which is more than 6% of the mean value of the variable. The reduction in the magnitude of the estimated coefficient of the time since annexation when control variables are added indicates that these variables do confound the impact of the time since annexation on the prevalence of collectivism to some extent. The time since annexation accounts for almost 10% of the variation in the percentage of households contributing labor.

The estimated coefficient of caloric suitability is negative and significant (Column 2), suggesting that a higher natural land productivity is associated with a lower percentage of households contributing labor, which concurs with Litina (2016). In line with Triandis (1995), the estimated coefficients of distance to the coast, elevation, and ruggedness are all significant and positive (Columns 3 to 5), indicating that areas farther from the coastal line, more highly elevated and rugged have higher percentages of households contributing labor. The estimated coefficient of irrigation suitability is negative and significant (Column 6), while only the dummy for Cwa climatic zone has a significant and positive estimated coefficient. When all control variables are added together, their estimated coefficients decrease substantially in magnitude, which is expected given that these variables are correlated (Table D1 in Online

¹⁷ With the latter, we assume that autocorrelation decreases in the distance between district centroids and equals zero for districts that are more than 0.5° (roughly 50 km) apart. A distance of 50 km connects four averaged districts in Vietnam.

| Table 2 The prevalence of collected | tivism Demontrate of | householde contri | thuting Johor | | | | | |
|---|-------------------------|-------------------|-----------------|-----------------|-----------------|-----------------|---------|-----------------|
| | | | | (F) | (2) | (9) | | (0) |
| | (\mathbf{n}) | (7) | (c) | (+) | | (0) | (1) | (0) |
| Time since annexation | 2.819 | | | | | | | 1.924 |
| | $(0.350)^{***}$ | | | | | | | $(0.498)^{***}$ |
| | $[0.627]^{***}$ | | | | | | | $[0.668]^{***}$ |
| Caloric suitability | | -0.036 | | | | | | 0.007 |
| | | $(0.005)^{***}$ | | | | | | $(0.008)^{***}$ |
| | | $[0.007]^{***}$ | | | | | | ***[600.0] |
| Distance to coast | | | 0.184 | | | | | 0.073 |
| | | | $(0.016)^{***}$ | | | | | $(0.031)^{**}$ |
| | | | $[0.018]^{***}$ | | | | | $[0.034]^{**}$ |
| Elevation | | | | 0.483 | | | | 0.175 |
| | | | | $(0.045)^{***}$ | | | | $(0.098)^{*}$ |
| | | | | $[0.054]^{***}$ | | | | $[0.109]^{*}$ |
| Ruggedness | | | | | 0.108 | | | 0.045 |
| | | | | | $(0.009)^{***}$ | | | $(0.020)^{**}$ |
| | | | | | $[0.010]^{***}$ | | | $[0.023]^{**}$ |
| Irrigation suitability | | | | | | -0.266 | | -0.032 |
| | | | | | | $(0.039)^{***}$ | | (0.040) |
| | | | | | | $[0.049]^{***}$ | | [0.041] |
| Aw climatic zone | | | | | | | -0.044 | 0.036 |
| | | | | | | | (0.032) | (0.032) |
| | | | | | | | [0.039] | [0.035] |
| Cfa climatic zone | | | | | | | - 0.079 | -0.116 |
| | | | | | | | (0.118) | (660.0) |
| | | | | | | | [0.125] | [0.092] |
| Cwa climatic zone | | | | | | | 0.146 | -0.004 |

| _ | |
|------|---|
| - T | 1 |
| đ | 5 |
| - 5 | 1 |
| - 2 | |
| | |
| - 14 | 1 |
| - 2 | |
| - 7 | - |
| ~ | 5 |
| | , |
| | |
| 0 | 1 |
| ٥ | J |
| - 7 | 5 |
| | 2 |
| | 2 |
| | - |

| | Percentage (| of households con | INTO ALLES TAUOL | | | | | |
|-------------------------------------|--------------|-------------------|------------------|-------|-------|-------|-----------------|---------|
| | (1) | (2) | (3) | (4) | (5) | (9) | (2) | (8) |
| | | | | | | | $(0.038)^{***}$ | (0.042) |
| | | | | | | | $[0.061]^{***}$ | [0.046] |
| Mean dep. var. | 31.30 | | | | | | | |
| Oster (2019)'s adjusted coefficient | | | | | | | | 1.325 |
| R^2 | 0.094 | 0.075 | 0.203 | 0.195 | 0.219 | 0.040 | 0.067 | 0.286 |
| Observations | 607 | 607 | 607 | 607 | 607 | 607 | 607 | 607 |

the distance between district centroids and equals zero for districts that are more than 50 km apart. All regressions include a constant. The base climatic zone is Am.

p < 0.1; p < 0.0; p < 0.05, p < 0.01

Appendix D). The estimated coefficients of caloric suitability, irrigation suitability, and climatic zones are now not different from zero (Column 8). Together, these control variables account for nearly 20% of the total variation in the percentage of households contributing labor.

Table 3 reports the results of similar regressions for the two other dependent variables measured at the district level, i.e., the average number of persons per household making labor contributions (panel A) and the average number of labor days contributed per household (panel B). With respect to both variables, the estimated coefficients of the time since annexation are significant and positive, whether or not all control variables are included. Districts annexed earlier to historical Vietnam currently have more members per household making labor contributions and more labor days contributed per household on average. For both dependent variables, the marginal effects are economically significant, i.e., more than 7% of the respective mean values. The time since annexation accounts for approximately 10% of the total variations in both dependent variables, while control variables altogether account for another 20%.

Table 4 present the results for two other measures of the individualism–collectivism trait. With respect to the percentage of households with grandchildren in them (panel A), the estimated coefficients of the time since annexation are significant and positive, whether or not all control variables are included. The estimated coefficients of the time since annexation are significant in general and negative with respect to the number of divorced households per 100 married households (panel B), though the standard errors are relatively larger when spatial autocorrelation is adjusted. Districts annexed earlier to historical Vietnam currently have higher percentages of households with grandchildren in them and lower numbers of divorced households per 100 married households. For both dependent variables, the marginal effects are economically significant, i.e., 11% (panel A) and 4% (panel B) of the respective mean values.

4.4 Robustness analysis

Sub-sample Analysis. As discussed earlier in the historical background, we recognize that there might be various characteristics in the annexed region (i.e., harsh living conditions, openness to trade, and existing individualist norms of the Cham and Khmer) that certainly play a role in explaining the cultural differences along the individualism–collectivism dimension between the annexed and the initial regions. In order to examine if our selective migration hypothesis differentiates itself from these channels, we investigate the relationship between the time since annexation to historical Vietnam and the strength of collectivism within the subsample of the annexed areas. As can be seen in Panel A of Table 5, the estimated coefficients of the time since annexation to historical Vietnam remain qualitatively the same with respect to all dependent variables, whether or not all control variables are added. We argue that this finding is consistent with our hypothesis that selective migration of individualistic people in the past is an important driver behind the contemporary cultural differences across Vietnam.

Next, historical Vietnamese immigrants (the Kinh ethnicity) often inhabited the coastal plain with their traditional rice agriculture. At the same time, the highland areas were mainly inhabited by various ethnic groups. After the Reunification in 1975, the Kinh started to migrate to the highland areas on a large scale through state-sponsored programs under the central planning economy to establish new production zones (Hardy, 2003). These later migrations, therefore, might be different from those that happened in historical

| Table 3 The strength of collectivism | | | | | | | | |
|--|-----------------|-------------------|-------------------|-----------------|-----------------|-----------------|---------|-----------------|
| A. | Average numb | er of persons mak | ing contributions | per household | | | | |
| | (1) | (2) | (3) | (4) | (5) | (9) | (1) | (8) |
| Time since annexation | 0.056 | | | | | | | 0.040 |
| | $(0.007)^{***}$ | | | | | | | $(0.010)^{***}$ |
| | $[0.012]^{***}$ | | | | | | | $[0.012]^{***}$ |
| Caloric suitability | | -0.071 | | | | | | 0.016 |
| | | $(0.011)^{***}$ | | | | | | (0.016) |
| | | $[0.013]^{***}$ | | | | | | [0.016] |
| Distance to coast | | | 0.379 | | | | | 0.139 |
| | | | $(0.038)^{***}$ | | | | | $(0.067)^{**}$ |
| | | | $[0.038]^{***}$ | | | | | $[0.065]^{**}$ |
| Elevation | | | | 1.037 | | | | 0.409 |
| | | | | $(0.102)^{***}$ | | | | $(0.194)^{**}$ |
| | | | | $[0.113]^{***}$ | | | | $[0.206]^{**}$ |
| Ruggedness | | | | | 0.230 | | | 0.101 |
| | | | | | $(0.020)^{***}$ | | | $(0.040)^{**}$ |
| | | | | | $[0.021]^{***}$ | | | $[0.045]^{**}$ |
| Irrigation suitability | | | | | | -0.531 | | -0.048 |
| | | | | | | $(0.067)^{***}$ | | (0.067) |
| | | | | | | $[0.091]^{***}$ | | [0.069] |
| Aw climatic zone | | | | | | | -0.095 | 0.088 |
| | | | | | | | (0.063) | (0.064) |
| | | | | | | | [0.075] | [0.067] |
| Cfa climatic zone | | | | | | | -0.179 | -0.269 |
| | | | | | | | (0.201) | (0.181) |
| | | | | | | | [0.206] | [0.160] |
| Cwa climatic zone | | | | | | | 0.278 | -0.020 |

🙆 Springer

| Table 3 (continued) | | | | | | | | |
|---|-----------------|--------------------|-------------------|------------------|-----------------|-------|--------------------------|--------------------|
| A. | Average nun | nber of persons m | aking contributio | ns per household | | | | |
| | (1) | (2) | (3) | (4) | (5) | (9) | (2) | (8) |
| | | | | | | | (0.077)*** [0.121]*** | (0.082) [0.086] |
| Mean dep. var. | 0.55 | | | | | | | |
| Uster (2019)'s adjusted coefficient R^2 | 0.094 | 0.076 | 0.222 | 0.230 | 0.255 | 0.041 | 0.066 | 0.029 0.322 |
| Observations | 607 | 607 | 607 | 607 | 607 | 607 | 607 | 607 |
| B. | Average nun | nber of days contr | ibuted per househ | plot | | | | |
| | (1) | (2) | (3) | (4) | (5) | (9) | (2) | (8) |
| Time since annexation | 0.533 | - | - | | | | | 0.302 |
| | $(0.058)^{***}$ | | | | | | | $(0.077)^{***}$ |
| | $[0.094]^{***}$ | | | | | | | ***[660.0] |
| Caloric suitability | | -0.828 | | | | | | -0.121 |
| | | $(0.104)^{***}$ | | | | | | (0.150) |
| | | $[0.116]^{***}$ | | | | | | [0.158] |
| Distance to coast | | | 3.362 | | | | | 1.297 |
| | | | $(0.382)^{***}$ | | | | | $(0.601)^{**}$ |
| | | | $[0.384]^{***}$ | | | | | $[0.571]^{**}$ |
| Elevation | | | | 7.695 | | | | 1.673 |
| | | | | $(1.034)^{***}$ | | | | (1.676) |
| | | | | $[1.140]^{***}$ | | | | [1.641] |
| Ruggedness | | | | | 1.840 | | | 0.884 |
| | | | | | $(0.195)^{***}$ | | | $(0.331)^{***}$ |
| | | | | | $[0.210]^{***}$ | | | $[0.345]^{***}$ |

70

| B. | Average nu | mber of days con | tributed per house | blod | | | | |
|-------------------------------------|------------|------------------|--------------------|-------|-------|-----------------|-----------------|-----------------|
| | (1) | (2) | (3) | (4) | (5) | (9) | (1) | (8) |
| Irrigation suitability | | | | | | - 3.845 | | 0.217 |
| | | | | | | $(0.556)^{***}$ | | (0.566) |
| | | | | | | $[0.663]^{***}$ | | [0.473] |
| Aw climatic zone | | | | | | | -0.158 | 0.776 |
| | | | | | | | (0.478) | (0.485) |
| | | | | | | | [0.597] | [0.509] |
| Cfa climatic zone | | | | | | | - 2.207 | - 2.606 |
| | | | | | | | $(0.666)^{***}$ | $(0.677)^{***}$ |
| | | | | | | | $[0.723]^{***}$ | $[0.690]^{***}$ |
| Cwa climatic zone | | | | | | | 3.056 | 0.139 |
| | | | | | | | $(0.641)^{***}$ | (0.656) |
| | | | | | | | $[0.942]^{***}$ | [0.709] |
| Mean dep. var. | 4.05 | | | | | | | |
| Oster (2019)'s adjusted coefficient | | | | | | | | 0.151 |
| R^2 | 0.114 | 0.136 | 0.232 | 0.168 | 0.217 | 0.028 | 0.073 | 0.308 |
| Observations | 607 | 607 | 607 | 607 | 607 | 607 | 607 | 607 |

the distance between district centroids and equals zero for districts that are more than 50 km apart. All regressions include a constant. The base climatic zone is Am. p < 0.1; p < 0.05, p < 0.01, p < 0.01

 $\underline{\textcircled{O}}$ Springer

| Table 4 Other measures of individ | ualism-collectivism | | | | | | | |
|---|---------------------|-------------------|---------------------|-----------------|-----------------|-----------------|---------|-----------------|
| A. | Percentage of | households with g | grandchildren in th | hem | | | | |
| | (1) | (2) | (3) | (4) | (5) | (9) | (1) | (8) |
| Time since annexation | 0.916 | | | | | - | | 0.795 |
| | $(0.039)^{***}$ | | | | | | | $(0.061)^{***}$ |
| | $(0.075)^{***}$ | | | | | | | $(0.097)^{***}$ |
| Caloric suitability | | - 0.635 | | | | | | 0.436 |
| | | $(0.097)^{***}$ | | | | | | $(0.120)^{***}$ |
| | | $(0.157)^{***}$ | | | | | | $(0.162)^{***}$ |
| Distance to coast | | | 3.108 | | | | | 1.419 |
| | | | $(0.250)^{***}$ | | | | | $(0.444)^{***}$ |
| | | | $(0.393)^{***}$ | | | | | $(0.579)^{***}$ |
| Elevation | | | | 6.694 | | | | 0.721 |
| | | | | $(0.711)^{***}$ | | | | (0.984) |
| | | | | $(1.199)^{***}$ | | | | (1.088) |
| Ruggedness | | | | | 1.680 | | | 0.733 |
| | | | | | $(0.124)^{***}$ | | | $(0.209)^{***}$ |
| | | | | | $(0.222)^{***}$ | | | $(0.268)^{***}$ |
| Irrigation suitability | | | | | | - 6.243 | | - 0.265 |
| | | | | | | $(0.363)^{***}$ | | (0.206) |
| | | | | | | $(0.607)^{***}$ | | (0.261) |
| Aw climatic zone | | | | | | | -0.463 | 0.880 |
| | | | | | | | (0.301) | $(0.299)^{***}$ |
| | | | | | | | (0.551) | $(0.414)^{***}$ |
| Cfa climatic zone | | | | | | | 1.506 | 0.600 |
| | | | | | | | (2.399) | (1.520) |
| | | | | | | | (2.608) | (1.597) |
| Cwa climatic zone | | | | | | | 5.404 | 1.356 |

| Table 4 (continued) | | | | | | | | |
|---|-----------------|--------------------|--------------------|-----------------|-----------------|-------|------------------------------------|--------------------------|
| A. | Percentage (| of households with | ı grandchildren ir | them | | | | |
| | (1) | (2) | (3) | (4) | (5) | (9) | (7) | (8) |
| | | | | | | | $(0.413)^{***}$ $(0.747)^{***}$ | (0.435)*** (0.593)*** |
| Mean dep. var. Oster (2019)'s adjusted coefficient | 7.20 | | | | | | | 0.459 |
| R^2 , R^2 | 0.454 | 0.107 | 0.266 | 0.172 | 0.252 | 0.100 | 0.309 | 0.612 |
| Observations | 613 | 613 | 613 | 613 | 613 | 613 | 613 | 613 |
| B. | Number of (| livorced household | ds per 100 marrie | d households | | | | |
| | (1) | (2) | (3) | (4) | (5) | (9) | (7) | (8) |
| Time since annexation | - 0.033 | | | | | | | - 0.097 |
| | $(0.012)^{***}$ | | | | | | | $(0.016)^{***}$ |
| | [0.033] | | | | | | | $[0.033]^{***}$ |
| Caloric suitability | | -0.050 | | | | | | -0.129 |
| | | $(0.018)^{***}$ | | | | | | $(0.022)^{***}$ |
| | | $[0.028]^{*}$ | | | | | | $[0.038]^{***}$ |
| Distance to coast | | | -0.199 | | | | | 0.099 |
| | | | $(0.058)^{***}$ | | | | | (0.085) |
| | | | $[0.110]^{*}$ | | | | | [0.128] |
| Elevation | | | | -1.547 | | | | - 1.279 |
| | | | | $(0.123)^{***}$ | | | | $(0.302)^{***}$ |
| | | | | $[0.241]^{***}$ | | | | $[0.361]^{***}$ |
| Ruggedness | | | | | -0.311 | | | -0.159 |
| | | | | | $(0.027)^{***}$ | | | $(0.050)^{***}$ |
| | | | | | $[0.055]^{***}$ | | | $[0.062]^{***}$ |

| B. | Number of | divorced househo | lds per 100 marrie | ed households | | | | |
|-------------------------------------|-----------|------------------|--------------------|---------------|-------|---------|-----------------|-----------------|
| | (1) | (2) | (3) | (4) | (5) | (9) | (1) | (8) |
| Irrigation suitability | | | | | | 0.436 | | - 0.349 |
| | | | | | | (0.168) | | (0.187) |
| | | | | | | [0.338] | | [0.237] |
| Aw climatic zone | | | | | | | 0.878 | 0.413 |
| | | | | | | | $(0.098)^{***}$ | $(0.093)^{***}$ |
| | | | | | | | $[0.197]^{***}$ | $[0.154]^{***}$ |
| Cfa climatic zone | | | | | | | -0.017 | 0.341 |
| | | | | | | | (0.147) | $(0.188)^{*}$ |
| | | | | | | | [0.146] | $(0.177)^{*}$ |
| Cwa climatic zone | | | | | | | 0.518 | 0.687 |
| | | | | | | | $(0.094)^{***}$ | $(0.107)^{***}$ |
| | | | | | | | $[0.158]^{***}$ | $[0.183]^{***}$ |
| Mean dep. var. | 2.46 | | | | | | | |
| Oster (2019)'s adjusted coefficient | | | | | | | | -0.122 |
| R ² | 0.013 | 0.014 | 0.024 | 0.198 | 0.186 | 0.011 | 0.105 | 0.352 |
| Observations | 613 | 613 | 613 | 613 | 613 | 613 | 613 | 613 |

p < 0.1; p < 0.0; p < 0.05; p < 0.01

| | Labor contril | oution | | | | | Other measur | es | | |
|---|--|---|---|---|--|--|--|--|-------------------------------------|---|
| | Percentage | | No. of person | SI | No. of days | | Extended fan | ily | Divorce rate | |
| | (1) | (2) | (3) | (4) | (5) | (9) | (7) | (8) | (6) | (10) |
| A. Annexed region | | | | | | | | | | |
| Time since annexation | 2.833 | 3.775 | 0.065 | 0.078 | 0.389 | 0.435 | 0.608 | 0.521 | -0.228 | - 0.168 |
| | $(1.086)^{***}$ | $(1.285)^{***}$ | $(0.023)^{***}$ | $(0.026)^{***}$ | $(0.150)^{***}$ | $(0.167)^{***}$ | $(0.058)^{***}$ | $(0.067)^{***}$ | $(0.035)^{***}$ | $(0.039)^{***}$ |
| | $[1.218]^{***}$ | $[1.326]^{***}$ | $[0.025]^{***}$ | $[0.025]^{***}$ | $[0.176]^{***}$ | $[0.174]^{***}$ | $[0.078]^{***}$ | $[0.076]^{***}$ | $[0.064]^{***}$ | $[0.056]^{***}$ |
| Mean dep. var. | 0.22 | | 0.37 | | 2.24 | | 4.04 | | 2.49 | |
| Control variables | No | Yes | No | Yes | No | Yes | No | Yes | No | Yes |
| R^2 | 0.030 | 0.214 | 0.043 | 0.246 | 0.032 | 0.139 | 0.304 | 0.564 | 0.103 | 0.384 |
| Observations | 311 | 311 | 311 | 311 | 311 | 311 | 310 | 310 | 310 | 310 |
| | | | | | | | | | | |
| B. Lowland region | | | | | | | | | | |
| Time since annexation | 2.857 | 3.437 | 0.069 | 0.076 | 0.443 | 0.460 | 0.602 | 0.531 | -0.195 | -0.107 |
| | $(1.145)^{**}$ | $(1.453)^{**}$ | $(0.025)^{***}$ | $(0.029)^{***}$ | $(0.158)^{***}$ | $(0.170)^{***}$ | $(0.060)^{***}$ | $(0.068)^{***}$ | $(0.034)^{***}$ | $(0.044)^{**}$ |
| | $[1.239]^{**}$ | $[1.504]^{**}$ | $[0.026]^{***}$ | $[0.028]^{***}$ | $[0.186]^{***}$ | $[0.189]^{***}$ | $[0.076]^{***}$ | $[0.078]^{***}$ | $[0.053]^{***}$ | $[0.055]^{**}$ |
| Mean dep. var. | 0.28 | | 0.46 | | 3.37 | | 3.90 | | 2.43 | |
| Control variables | No | Yes | No | Yes | No | Yes | No | Yes | No | Yes |
| R^2 | 0.039 | 0.198 | 0.064 | 0.237 | 0.059 | 0.122 | 0.361 | 0.635 | 0.132 | 0.378 |
| Observations | 249 | 249 | 249 | 249 | 249 | 249 | 248 | 248 | 248 | 248 |
| Robust standard errors a the distance between dis ruggedness, irrigation su | re in parenthese trict centroids a uitability, Köppe | es, standard erro nd equals zero f n—Geiger clima | ors in squared p for districts that thic zones, and a | arentheses are c t are more than a constant. Pane | calculated follo 50 km apart. C el A only inclue | wing Conley (1 ontrol variables les districts in t | 999) with the a s include caloric he annexed reg | ssumption that c suitability, dis ion. Panel B fur | autocorrelation tance to the cos | decreases in tst, elevation, istricts in Ha |
| Noi and Ho Chi Minh C | ity and districts | whose elevation | ns are above 50 | 0 m. | 3 | |) | | | |

p < 0.1; p < 0.0; p < 0.05; p < 0.01

times. To examine this issue, we further exclude from the estimation all districts in the highland areas, i.e., where the average elevations are above 500 meters (the results are robust to other values such as 400 and 600 meters).¹⁸ Furthermore, we also exclude two provinces, Ha Noi (in the north) and Ho Chi Minh City (in the south), which are the two biggest venues for immigrants in modern times. Panel B of Table 5 shows that the estimated coefficients of the time since annexation remain qualitatively the same with respect to all dependent variables, whether or not all control variables are added.

Omitted-Variable Bias. Although we have considered an extensive list of confounding factors, some factors cannot be included owing to data availability. To examine the potential omitted-variable bias caused by some of these factors, we use an instrumental variable (IV) estimation. We observe that, from the Red River Delta in the north, historical Vietnam could not conquer the Mekong River Delta in the south without annexing all areas located in between, making the north-south geographical order a strong predictor of the time since annexation within the subsample of the annexed areas.¹⁹ Thus, by using the north-south geographical order as an instrumental variable for the time since annexation, we can examine the potential bias caused by unobserved factors that are likely to influence the time since annexation and the individualism–collectivism trait, but do not correlate with the north-south geographical order. In Online Appendix C, we discuss many such factors in detail, including indigenous cultural environments, local economic activities, market access, and the influence of northern culture or institutions.

We propose to use the distance from an annexed area to a northern reference point as a measure of the north-south geographical order. Quang Binh (see Fig. 1), the first area that was annexed to historical Vietnam, is arbitrarily chosen as the northern reference point (the result is robust to other choices as well), and the walking distance along the coast (instead of the geodesic, "bird-fly" distance) is calculated to capture the military route in historical times. Distance is measured in 100 km using the district centroids, where district borders are taken from the Global Administrative Unit Layers. The walking distance from Quang Binh to the farthest district in the south is roughly 1350 km. Panel A of Table 6 shows that the estimated coefficients of the time since annexation remain economically and statistically significant with respect to all dependent variables, whether or not all control variables are added. The first-stage results ensure that the walking distance to Quang Binh is a strong predictor of the time since annexation, i.e., it has significant and negative estimated coefficients and large F statistics (full results are available upon request).

In the spirit of Scott (2009), more distant and rugged areas would make it easier for migrants to escape the influence of the central state. Thus, we can use the interaction between the walking distance from Quang Binh and terrain ruggedness as an instrumental variable to further examine the credibility of the IV estimation. Panel B of Table 6 shows that the estimated coefficients of the time since annexation remain qualitatively the same with respect to all dependent variables, whether or not all control variables are added. The first-stage results also suggest that the interaction term between the walking distance from Quang Binh and terrain ruggedness is a strong predictor of the time since annexation, i.e.,

¹⁸ This definition of highland is taken from Wikipedia: https://en.wikipedia.org/wiki/Highland.

¹⁹ Theoretically, one could conquer an area in the Mekong River Delta by traveling either along the coastline in the east or over the mountainous band in the west separating Vietnam from Laos and Cambodia. Both strategies were infeasible given the logistical and transportation technologies in historical Vietnam. Indeed, we do not find any attempt to do so in the historical accounts.

| | Labor contril | oution | | | | | Other measur | se | | |
|---|---|--|--|--|--|--|---|---|--|---|
| | Percentage | | No. of person | s | No. of days | | Extended far | uily | Divorce rate | |
| | (1) | (2) | (3) | (4) | (5) | (9) | (7) | (8) | (6) | (10) |
| A. IV: Distance | | | | | | | | | | |
| Time since annexation | 6.419 | 5.376 | 0.139 | 0.110 | 0.674 | 0.485 | 0.978 | 0.704 | -0.457 | - 0.275 |
| | $(1.251)^{***}$ | $(1.483)^{***}$ | $(0.026)^{***}$ | $(0.027)^{***}$ | $(0.171)^{***}$ | $(0.198)^{**}$ | $(0.078)^{***}$ | (0.090)*** | $(0.055)^{***}$ | $(0.055)^{***}$ |
| | $[1.613]^{***}$ | $[1.742]^{***}$ | $[0.032]^{***}$ | $[0.029]^{***}$ | $[0.209]^{***}$ | $[0.214]^{**}$ | $[0.111]^{***}$ | $[0.103]^{**}$ | $[0.108]^{***}$ | $[0.072]^{***}$ |
| Mean dep. var. | 0.22 | | 0.37 | | 2.24 | | 4.04 | | 2.49 | |
| Kleibergen-Paap F statistic | 287 | 221 | 287 | 221 | 287 | 221 | 308 | 227 | 308 | 227 |
| Control variables | No | Yes | No | Yes | No | Yes | No | Yes | No | Yes |
| Observations | 311 | 311 | 311 | 311 | 311 | 311 | 310 | 310 | 310 | 310 |
| B. IV: Distance x rugged- ness | | | | | | | | | | |
| Time since annexation | 5.357 | 5.227 | 0.117 | 0.108 | 0.574 | 0.531 | 0.897 | 0.687 | -0.381 | -0.292 |
| | $(1.194)^{***}$ | $(1.467)^{***}$ | $(0.024)^{***}$ | $(0.026)^{***}$ | $(0.160)^{***}$ | $(0.190)^{***}$ | $(0.076)^{***}$ | $(0.086)^{***}$ | $(0.050)^{***}$ | $(0.060)^{***}$ |
| | $[2.942]^{***}$ | $[1.714]^{***}$ | $[0.059]^{***}$ | $[0.029]^{***}$ | $[0.315]^{**}$ | $[0.210]^{**}$ | $[0.236]^{***}$ | $[0.102]^{***}$ | $[0.226]^{***}$ | $[0.074]^{**}$ |
| Mean dep. var. | 0.22 | | 0.37 | | 2.24 | | 4.04 | | 2.49 | |
| Kleibergen-Paap F statistic | 217 | 127 | 217 | 127 | 217 | 127 | 221 | 108 | 221 | 108 |
| Control variables | No | Yes | No | Yes | No | Yes | No | Yes | No | Yes |
| Observations | 311 | 311 | 311 | 311 | 311 | 311 | 310 | 310 | 310 | 310 |
| Robust standard errors are in the distance between district ruggedness, irrigation suitab | parentheses, st centroids and e ility, Köppen–C | andard errors i quals zero for c Jeiger climatic | in squared pare- listricts that are zones, and a constitution form t | ntheses are cald e more than 50 onstant. All reg | culated followi km apart. Con gressions only | ng Conley (19 trol variables i include distric | 99) with the as nclude caloric a ts in the annexe | sumption that a suitability, distr ed region. In Pa | utocorrelation ance to the coa anel A, walking | decreases in st, elevation, g distance to |
| Quang binn is the instrument | tal variadie. III | Panel b, the life | eraction term t | Jetween walkli | ig distance to t | Juang binn an | a terram rugget | Iness is added a | as an insurumer | ital variadic. |

Table 6 Robustness to instrumental variables

p < 0.1; p < 0.05; p < 0.05; p < 0.01

it has significant and negative estimated coefficients and large F statistics (full results are available upon request).

Another way to examine the potential bias from omitted variables is to use the change in the estimated coefficients of the time since annexation into historical Vietnam when the observed confounding factors are included to infer about the potential change caused by the unobserved confounding factors. Under the assumption that the selection on unobserved confounding factors is proportional to the selection on observed confounding factors, Oster (2019) shows that a consistent estimate adjusted for omitted-variable bias can be obtained for each value of this assumed proportional relationship. To examine how the estimated coefficients of the time since annexation responses to omitted variables, we assume that the selection on unobserved confounding factors equals one half of the selection on observed confounding factors.²⁰ Columns 8 of Tables 2, 3, and 4 show that the adjusted estimated coefficients remain more than one half in magnitude compared to their unadjusted counterparts with respect to all dependent variables. Thus, the potential bias from omitted variables, if there is any, should be relatively small.

Bad Controls. There are numerous factors that came to exist after the annexation into historical Vietnam that might influence the strength of collectivism. For example, big cities in the annexed region such as Hoi An and Sai Gon were built after the annexation, the concentration of French colonizers and American intervention in south Vietnam (the latest region to be annexed into historical Vietnam), and various demographic characteristics in the modern day. These factors are bad controls, and should not be included in the regression (Angrist & Pischke, 2009). Nevertheless, Table D2 in Online Appendix D reports that the estimated coefficients of the time since annexation to historical Vietnam remain qualitatively intact when the shortest distance to big cities, a dummy variable for south Vietnam, and various district demographic characteristics such as percentage of households with male head, average head age, average schooling years of head, percentage of urban households, percentage of Kinh households, average household size, and per capita expenditure are added to the regression model.

4.5 Discussion

The survey and census data analysis so far has demonstrated that, on average, districts annexed earlier to historical Vietnam are today more prone to collectivist norms, and this result is robust to a battery of checks. In particular, a novel finding is that voluntary labor contribution to public goods production is both more prevalent and more intensive in districts annexed earlier to historical Vietnam. We have also shown that family structure is more traditional and marriage is more stable in districts annexed earlier. Although the VHLSS provides naturally occurring data that are available for almost all districts across Vietnam, the biggest drawback is that it cannot help us examine further why more collectivist societies in districts annexed earlier to historical Vietnam could mobilize a larger amount of voluntary labor contribution to public goods production from their in-group members. Was that because these societies developed informal institutions that punished non-contributing members (e.g., ostracism)? Or because they have a large fraction of

²⁰ We defend this assumption on the ground that we have been able to identify and control for an extensive set of important observed confounding factors. Thus, if there are any potential unobserved confounding factors, their influence on the time since annexation cannot be too large relative to the influence of the observed confounding factors.

members with strong cooperative preferences? Or because their members share a strong belief that other people would also contribute labor to public goods production?

To complement the survey data analysis, we conduct a lab-in-the-field public goods experiment. Although it is impossible to run the experiment in all districts across Vietnam, the advantage of the experiment is that it allows us to examine deeper why more collectivist societies in districts annexed earlier to historical Vietnam could mobilize a larger amount of voluntary labor contribution to public goods production from their in-group members. First, the experiment holds the institutional setting constant, eliminating the possibility that there are informal institutions that punish non-contributing members. Second, as discussed in detail below, we adopt an experimental design that allows us to measure preferences for cooperation and beliefs about the contributing behaviors of other members. In turn, we can investigate whether preferences or beliefs that drive individual contributions to public goods production.

5 Experimental data analysis

5.1 Sample selection

To capture in-group cooperation, we recruit subjects who come from the same local areas. A crucial aspect is the selection of experimental sites in such a way to minimize differences between the selected sites. First, we focus on the annexed region to rule out differences in the historical frontier environment. Second, we select provinces, and rural districts in them, located along the coast, which was the typical route of migration and settlement of historical Vietnamese. Finally, we choose provinces, and rural districts in them, that were historically inhabited mainly by the Kinh ethnicity (historical Vietnamese) and whose populations have been living there for many generations, i.e., neither any significant immigration nor emigration from these places. Thus, this procedure leaves us with coastal, rural, and Kinh-dominated districts in the annexed region. From this subsample, we randomly select one of the districts with the longest time since annexation to historical Vietnam and one of the districts with the shortest time. This process narrows our selection to randomly choose one rural district in Thua Thien Hue province and one rural district in Ben Tre province; the former is located more to the north and thus has a longer time since annexation (see Fig. 1).

We use high school students as our subjects in the experiment since they are old enough to embody the cultural environments of the places where they grew up, but not yet affected by living outside their communities, which could make it harder to capture the local cultural norms.²¹ Our proposed selective migration hypothesis predicts that subjects in Thua Thien Hue (henceforth the "northern site") share stronger norms of in-group cooperation, and hence on average contribute at a higher level compared to subjects from Ben Tre (henceforth the "southern site"). Each rural district in Vietnam has three to five high schools. To keep similarities between the selected districts, we randomly selected one school located in the center of the district among the schools that had at least six classes for the oldest age cohort, which means that students come from a larger catchment area

²¹ This strategy of focusing on high school students has also been used earlier in the literature on public goods experiments when investigating cultural differences, e.g., Kocher et al. (2012).

where they have attended different secondary schools. The latter requirement was imposed to avoid measuring cooperation norms within a specific class, which might have developed its own norms, when we are aiming at measuring norms in the community in which they lived.

5.2 The public goods experiment

We build our experimental design on the one-shot linear public goods experiment developed by Fischbacher et al. (2001).²² We begin by describing the general features of a public goods experiment before discussing the specific features of the design in Fischbacher et al. (2001).

The basic idea of a public goods experiment is to create a social dilemma situation, where there is a conflict between the social and private optima. In our setting, the subjects are randomly assigned to groups of three, where each member comes from a different class at the high school, and this was clearly stated in the instructions of the experiment. This feature of the design was chosen to avoid having subjects allocated to groups consisting of classmates with whom subjects might have developed a specific norm of behavior, reducing the possibility of measuring norms of cooperation in the places where they reside. All subjects receive an endowment of 20 tokens and must decide simultaneously how much of their endowments to invest in a public good, and the residual is kept for themselves, which is labeled as a private good. The marginal per capita return (MPCR) from the public good is 0.5, which means that each token contributed to the public good by a group member results in 0.5 token to all group members, including the member who contributes the token. If a subject is rational and selfish, then a MPCR below 1 leads to a dominant strategy to free ride (i.e., to contribute zero to the public good), because the return from the public good is lower than the return from the private good. Nevertheless, it is socially optimal to contribute the whole endowment if $MPCR \times n > 1$, where n is the number of group members. Thus, our choice of the MPCR of 0.5 generates the conflict between private and social optima that characterizes a public good. The payoff for subject *i* consists of two components: (1) the amount of the endowment that is not invested in the public good (i.e., what is kept as a private good), and (2) the return from the public good. The payoff function for subject *i* is given by:

$$\pi_i = (20 - c_i) + 0.5 \sum_{j=1}^3 c_j.$$

Each token earned in the experiment is exchanged for money at the exchange rate of one token equals 10000 Vietnamese Dong.

The specific feature of the public goods experiment developed by Fischbacher et al. (2001) is that it is based on the strategy method. In their design, each subject makes two types of contribution decisions to the public good: (1) unconditional contribution and (2) conditional contribution. In the unconditional contribution decision, which is the standard public goods experiment described above, each subject states how much he or she would

²² For other experiments using this specific design, see Kocher et al. (2008), Herrmann and Thöni (2009), Fischbacher and Gächter (2010), Fischbacher et al. (2012) and Martinsson et al. (2015) among others. For a general discussion on public goods experiments, see Zelmer (2003) and Chaudhuri (2011).

like to contribute to the public good from his or her endowment of 20 tokens. The additional feature of the design of Fischbacher et al. (2001) is the introduction of the contribution table in which subjects make contribution decisions conditional on the other group members' average contributions. In a contribution table, which includes all possible average contributions of the two other players in the group, rounded to integers and ranging from 0 to 20 points, a subject indicates how much he or she would contribute to the public good if these were the average contributions to the public good by the other two group members. The contributions reported in the table are referred to as conditional contributions. The final feature of the design of Fischbacher et al. (2001) is to ensure that all decisions, i.e., both unconditional and conditional contributions, are incentive compatible by using the following approach. For two randomly selected group members, it is the unconditional contribution to the public good that is pay-off relevant. For the third member, the average unconditional contribution of the other two group members is calculated, and the contribution of the third member is then determined from her conditional contribution given the average contribution of the other two group members. Thus, when a subject makes his or her decisions, he or she does not know which of all the decisions will be payoff relevant, and hence has no incentive to choose anything other than the preferred option. After the experiment, we also elicited beliefs by asking a subject what he or she thought that the other two group members had contributed unconditionally on average. We pay subjects for the accuracy of their guesses to create incentives for truthful revelation.²³

The strength of the strategy method is that subjects can be categorized into different contributor types based on their 21 conditional contribution decisions to the public good, i.e., how much they decided to contribute to the public good conditional on the average contribution of the other two group members for all integers in the range 0 to 20. These contributor types capture the preferences for cooperation. We use the same classification as proposed in the original paper by Fischbacher et al. (2001). A subject is classified as a "conditional cooperator" if his or her conditional contribution increases weakly monotonically with the average contribution of the other group members or if the relationship between his or her conditional contribution and the average contribution of the others is positive and significant at the 1% significance level, using a Spearman rank correlation coefficient. A "free rider" is a subject who contributes zero to the public good for all levels of the average contribution by others. A "hump-shaped" contributor is characterized by a subject who shows weakly monotonically increasing contributions or a positive Spearman rank correlation coefficient at the 1% significance level, which is the same classification strategy as applied to a conditional contributor, but it only holds up to an inflection point. For average contribution levels by others above this level, the subject's own conditional contributions decrease weakly monotonically or show a significant and negative Spearman rank correlation coefficient at the 1% significance level. Those who cannot be categorized based on any of the above criteria are referred to as "others".

At the beginning of the experiment, subjects received written instructions for the experiment and the instructions were also read aloud.²⁴ Before the experiment began, various

²³ For a discussion on using incentivized guesses in a public goods experiment to increase guess accuracy, see, e.g., Gächter and Renner (2010).

²⁴ Details on the experimental instructions can be found in Online Appendix E. The public goods experiment is the first experiment out of two, and the second experiment is pay-off independent from the first one. To eliminate potential spillover between experiments, the subjects were informed about the second experiment only after completing the first experiment.

examples were given to facilitate understanding of the experiment and the subjects also completed some exercises. When the experiments were finished, subjects answered a short survey about basic socioeconomic information. Finally, subjects were called one at a time for payment done in private. Subjects were recruited by teachers, and the participation rates of students are similar across schools: 70% in the northern site (140 out of 200) and 73% (235 out of 320) in the southern site. In accordance with our expectation, around 97% of the subjects were born in the sampled districts, while the others were born in other districts in the sampled provinces.²⁵ Table D3 of Online Appendix D summarizes other socioeconomic characteristics (gender, household size, and a wealth index), which will be controlled for in the following regression analysis.

5.3 Results

Table 7 shows that subjects from the northern site and southern site on average unconditionally contributed 7.50 tokens and 6.58 tokens respectively out of the endowment of 20.²⁶ Thus, subjects from the northern site on average contribute higher than subjects from the southern site by 0.92 tokens, and this difference is statistically significant (*p* value = 0.024, Mann-Whitney U test). Previous studies have indicated that a large fraction of subjects are conditional cooperators, i.e., their contributions are positively correlated with contribution levels by others. We also elicited guesses about the average contributions by the other two group members, in which subjects from the northern site and southern site on average guessed 8.25 tokens and 7.60 tokens respectively (*p* value = 0.053, Mann-Whitney U test). At the aggregate level, the results indicate conditional cooperative behavior. It is common that guesses about the average contribution by others are higher than own contribution levels because, on average, people are imperfect conditional cooperators, and there is also a fraction of free-riders.

The innovative part of the design developed by Fischbacher et al. (2001) is that it allows us to classify subjects into different contributor types. The lower panel of Table 7 shows the distribution of types. By far, conditional cooperators are the most frequent type both in the northern site (52.17%) and in the southern site (54.04%), while the fractions of free riders are low (3.62% and 5.53% in the northern and the southern site, respectively).²⁷ We indeed cannot reject the null hypothesis that the compositions of contributor types in the northern site and the southern site are drawn from the same distribution (p value = 0.930, Pearson's χ^2 test), indicating that cooperative preferences are not significantly different between the two sites. Furthermore, Table 7 shows that, except for free riders, other types in the northern site on average have higher levels of unconditional contribution and guesses about the average contribution of other group members compared to their counterparts in the southern site. The largest contributor type is the conditional cooperators, and this type is considered the key group for contributions to public goods. Their average unconditional contributions are 7.29 and 6.39 respectively in the northern and the southern site, with guesses of 8.19 and 7.88. These findings suggest that the north-south difference in contribution behaviors is driven by beliefs about the contributing behaviors of other people rather than cooperative preferences.

²⁵ The following results are robust to the omission of subjects who were not born in the sampled district. Details are available upon request.

²⁶ The contribution levels are similar to what has been found in the literature of public goods experiments (Zelmer, 2003; Chaudhur, 2011).

²⁷ This distribution of types is also similar with previous findings in the literature (Chaudhuri , 2011).

We use regression models to examine the unconditional contribution behaviors further and the results are reported in Table 8. In all models, we include a dummy variable if a subject comes from the northern site. In line with the descriptive statistics, the estimated coefficient of the northern site dummy is positive and significant when entering the regression alone (Column 1). In the next regression model, we add belief about the average contribution of the other two group members and find that its estimated coefficient is positive and significant (Column 2). In the same regression, the estimated coefficient of the northern site dummy is reduced substantially in magnitude, indicating that higher levels of contribution in the northern site are mainly driven by higher beliefs about the average contribution of the other two group members. These estimated coefficients remain positive and significant when the socioeconomic characteristics (gender, household size, and a wealth index) are also added to the regression (Columns 3 to 6).

To summarize, the experimental findings corroborate the tendency found in survey data that districts annexed earlier to historical Vietnam currently have stronger norms for cooperation, and further suggest that cultural differences across Vietnamese regions are embodied in individual beliefs.

6 Conclusion

The individualism–collectivism dimension has been found to be a powerful predictor of economic and democratic development in a large sample of countries (Gorodnichenko & Roland, 2011, 2015, 2017). Thus, why some societies have become more collectivistic or individualistic than others is a crucial question in understanding long-run comparative development. In the present paper, we propose and investigate the selective migration hypothesis, stating that cultural differences along the individualism–collectivism dimension are driven by the out-migration of individualistic people from collectivist societies to settle down in frontier areas, and that such patterns of historical migration are reflected even in the current distribution of cultural norms. We use the territorial expansion of historical Vietnam from the eleventh to the eighteenth centuries as an ideal setting to empirically examine this hypothesis. During this period, historical Vietnam gradually expanded its territory southward along the coast from the Red River Delta to the Mekong River Delta through various waves of conquest and migration to form the country as it is today.

Our empirical analysis focuses on the ability to solve collective action problems, which is the main feature of collectivism in related economic models, by using data on voluntary contributions to public goods, which is the most typical collective action in daily living in Vietnam. Using a household survey, we find that areas annexed earlier to historical Vietnam currently have higher levels of voluntary labor contribution to public goods production. Using a population census, we also find that households in districts that were annexed earlier have a higher percentage of households with grandchildren living in them and a lower prevalence of divorced households, which are two other standard measures of individualism–collectivism traits. Conducting a public goods experiment with high school students, we find that subjects from areas annexed earlier to historical Vietnam contribute substantially more to the public good compared to subjects from areas annexed later, and that the result is mainly driven by the belief about the contributions of other subjects. Relying on various Vietnamese historical accounts, together with various robustness checks, we show that the southward out-migration of individualistic people during eight centuries

| | $\frac{\text{Northern site}}{(n=138)}$ | | | $\frac{\text{Southern site}}{(n = 235)}$ | | |
|-------------------------|--|--------------------------|--------------------|--|--------------------------|--------------------|
| | | | | | | |
| | Distribution (%) | Av. Un. Con. (Tokens) | Belief (Tokens) | Distribution (%) | Av. Un. Con. (Tokens) | Belief (Tokens) |
| | | | | | | |
| All subjects | 100 | 7.50 | 8.25 | 100 | 6.58 | 7.60 |
| | | (3.92) | (3.46) | | (4.07) | (4.11) |
| Conditional cooperators | 57.12 | 7.29 | 8.19 | 54.04 | 6.39 | 7.88 |
| | | (3.20) | (3.72) | | (3.35) | (4.07) |
| Free riders | 3.62 | 0.60 | 6.20 | 5.53 | 2.69 | 6.54 |
| | | (0.55) | (4.55) | | (3.88) | (3.55) |
| Hump-shaped cooperators | 5.80 | 10.50 | 8.38 | 4.26 | 6.90 | 6.70 |
| | | (5.45) | (3.42) | | (5.92) | (5.19) |
| Others | 38.41 | 7.98 | 8.49 | 36.17 | 7.42 | 7.44 |
| | | (4.03) | (2.98) | | (4.51) | (4.15) |
| | | | | | | |

Table 7 Distribution of types, unconditional contribution, and belief

Standard errors are in parentheses. Av. Un. Con. = average unconditional contribution. Belief = belief about the average unconditional contribution of other two group members

| | Unconditional contribution | | | | | | | | |
|----------------|----------------------------|----------|----------|----------|----------|----------|--|--|--|
| | (1) | (2) | (3) | (4) | (5) | (6) | | | |
| North | 0.921** | 0.728* | 0.808** | 0.686* | 0.735* | 0.789* | | | |
| | (0.426) | (0.403) | (0.405) | (0.405) | (0.407) | (0.410) | | | |
| Belief | | 0.297*** | 0.299*** | 0.290*** | 0.297*** | 0.294*** | | | |
| | | (0.060) | (0.060) | (0.059) | (0.060) | (0.059) | | | |
| Male | | | 0.376 | | | 0.378 | | | |
| | | | (0.437) | | | (0.440) | | | |
| Household size | | | | 0.118 | | 0.131 | | | |
| | | | | (0.130) | | (0.137) | | | |
| Wealth | | | | | -0.148 | - 0.594 | | | |
| | | | | | (1.613) | (1.715) | | | |
| Constant | 6.579*** | 4.324*** | 4.138*** | 3.877*** | 4.360*** | 3.785*** | | | |
| | (0.266) | (0.494) | (0.517) | (0.745) | (0.666) | (0.836) | | | |
| R^2 | 0.012 | 0.093 | 0.095 | 0.095 | 0.093 | 0.098 | | | |
| Observations | 373 | 373 | 373 | 373 | 373 | 373 | | | |

 Table 8
 Unconditional contribution: regression analysis

Robust standard errors in parentheses. The sample includes 138 subjects from the northern site and 235 subjects from the southern site.

p < 0.1; p < 0.05; p < 0.01; p < 0.01

of territorial expansion of historical Vietnam is an important driver behind these cultural differences.

Despite our efforts, we recognize, however, that in the current study, it is empirically challenging to completely isolate the effects of selective migration from a crowding-in of collectivist norms by a strong state or the pre-existence of individualist norms in the periphery. We leave it for future research on other areas, with access to more detailed data on historical migration patterns and attitudes, to potentially shed further light on the relative contributions of each of these interrelated mechanisms.

We believe that the present paper makes a contribution by offering an extended conceptual framework and an empirical strategy combining survey and experimental evidence for understanding long-run cultural divergence. First and foremost, the migration patterns in the distant past played a crucial role in explaining cultural differences across modern societies. As time goes on, similar processes may continue to enhance cultural differences across societies. These cultural differences may, in turn, have important implications for future levels of comparative development.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s10887-021-09194-x.

Acknowledgements We are grateful for valuable comments from the editor Oded Galor and three anonymous reviewers, as well as Fredrik Carlsson, Carl-Johan Dalgaard, Quoc-Anh Do, Martin Dufwenberg, Masayuki Kudamatsu, Laura Mayoral, Pablo Selaya, Ardeshir Sepehri, Nico Voigtländer, Fabrizio Zilibotti, and seminar participants at the University of Gothenburg, Nordic Conference in Development Economics 2017, ASREC European Conference 2019. We would also like to thank Khanh-Nam Pham, Dang-Thuy Truong, and EfD-Vietnam for kindly supporting our field research and data collection, as well as Luat Do, Trung Nguyen, Cuong Tran, Anh Le, and Chi Le for excellent research assistance. All remaining errors are our own. A research grant from the Swedish Research Council (ref 348 2013-6348) is gratefully acknowledged.

Funding Open access funding provided by University of Gothenburg.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

- Abramitzky, R., Boustan, L. P., & Eriksson, K. (2012). Europe's tired, poor, huddled masses: Self-selection and economic outcomes in the Age of Mass Migration. *American Economic Review*, 102(5), 1832–1856.
- Adams, J., & Hancock, N. (1970). Land and economy in traditional Vietnam. Journal of Southeast Asian Studies, 1(2), 90–98.

Alesina, A., & Giuliano, P. (2010). The power of the family. Journal of Economic Growth, 15(2), 93–125.

Alesina, A., & Giuliano, P. (2014). Family ties. In P. Aghion & S. N. Durlauf (Eds.), Handbook of economic growth (Vol. 2A, pp. 177–215). Elsevier.

- Alesina, A., & Giuliano, P. (2015). Culture and institutions. Journal of Economic Literature, 53(4), 898–944.
- Alesina, A., Giuliano, P., & Nunn, N. (2013). On the origins of gender roles: Women and the plough. *Quarterly Journal of Economics*, 128(2), 469–530.
- Algan, Y., & Pierre, C. (2014). Trust, growth, and well-being: New evidence and policy implications. In P. Aghion & S. N. Durlauf (Eds.), *Handbook of economic growth* (Vol. 2A, pp. 49–120). Elsevier.
- Angrist, J. D., & Pischke, J.-S. (2009). Mostly harmless econometrics: An empiricist's companion. Princeton: Princeton University Press.
- Ashwill, M. A., & Diep, T. N. (2005). Vietnam today: A guide to a nation at a crossroads. Maine: Intercultural Press.
- Bazzi, S., Fiszbein, M., & Gebresilasse, M. (2020). Frontier culture: The roots and persistence of rugged individualism in the United States. *Econometrica*, 88(6), 2329–2368.
- Becker, S. O., Boeckh, K., Hainz, C., & Woessmann, L. (2016). The empire is dead, long live the empire! Long-run persistence of trust and corruption in the bureaucracy. *Economic Journal*, 126(590), 40–74.
- Bentzen, J. S., Kaarsen, N., & Wingender, A. M. (2017). Irrigation and autocracy. Journal of the European Economic Association, 15(1), 1–53.
- Bigoni, M., Bortolotti, S., Casari, M., & Gambetta, D. (2018). At the root of the North–South cooperation gap in Italy: Preferences or beliefs? *Economic Journal*, 129(619), 1139–1152.
- Bigoni, M., Bortolotti, S., Casari, M., Gambetta, D., & Pancotto, F. (2016). Amoral familism, social capital, or trust? The behavioural foundations of the Italian North-South divide. *Economic Journal*, 126(594), 1318–1341.
- Bisin, A., & Verdier, T. (2001). The economics of cultural transmission and the dynamics of preferences. *Journal of Economic Theory*, 97(2), 298–319.
- Borcan, O., Ola, O., & Louis, P. (2021). Transition to agriculture and first state presence: A global analysis. In: Explorations in Economic History (forthcoming).
- Buggle, J. C. (2020). Growing collectivism: Irrigation, group conformity and technological divergence. Journal of Economic Growth, 25, 147–193.
- Chaudhuri, A. (2011). Sustaining cooperation in laboratory public goods experiments: A selective survey of the literature. *Experimental Economics*, 14(1), 47–83.
- Conley, T. G. (1999). GMM estimation with cross sectional dependence. *Journal of Econometrics*, 92(1), 1–45.
- Dai Nam Thuc Luc (2002). (Ngoc Tinh Nguyen, Trans.) (Vol. 1). Ha Noi: Giao Duc.
- Dai Viet Su Ky Toan Thu (2012). (Huy Giu Cao, Trans). Ha Noi: Thoi Dai.
- Dao, Duy Anh (2005). Dat nuoc Viet Nam qua cac doi. Ha Noi: Van Hoa Thong Tin.
- Dao, M.-Q. (1993). History of land tenure in pre-1954 Vietnam. Journal of Contemporary Asia, 23(1), 84–92.
- Dell, M., Lane, N., & Querubin, P. (2018). The historical state, local collective action, and economic development in Vietnam. *Econometrica*, 86(6), 2083–2121.
- Diamond, J. (1997). Guns, germs, and steel: The fates of human societies. W. W. Norton.
- Doepke, M., & Zilibotti, F. (2014). Culture, entrepreneurship, and growth. In P. Aghion & S. N. Durlauf (Eds.), *Handbook of economic growth* (Vol. 2A, pp. 1–48). Elsevier.
- Fernández, R. (2011). Does culture matter? In J. Benhabib, A. Bisin, & M. Jackson (Eds.), Handbook of social economics (Vol. 1A, pp. 481–510). Elsevier.
- Fincher, C. L., Thornhill, R., Murray, D. R., & Schaller, M. (2008). Pathogen prevalence predicts human cross-cultural variability in individualism/collectivism. *Proceedings of the Royal Society B: Biologi*cal Sciences, 275(1640), 1279–1285.
- Fischbacher, U., & Gächter, S. (2010). Social preferences, beliefs, and the dynamics of free riding in public goods experiments. *American Economic Review*, 100(1), 541–56.
- Fischbacher, U., Gächter, S., & Fehr, E. (2001). Are people conditionally cooperative? Evidence from a public goods experiment. *Economics Letters*, 71(3), 397–404.
- Fischbacher, U., Gächter, S., & Quercia, S. (2012). The behavioral validity of the strategy method in public good experiments. *Journal of Economic Psychology*, 33(4), 897–913.
- Fischer, G., van Velthuizen, H., Shah, M., & Nachtergaele, F. (2002). Global agro-ecological assessment for agriculture in the 21st century: Methodology and results. International Institute for Applied Systems Analysis.
- Fouka, V., & Schläpfer, A. (2020). Agricultural returns to labor and the origins of work ethics. *Economic Journal*, 130(628), 1081–1113.
- Galor, O., & Özak, Ö. (2016). The agricultural origins of time preference. American Economic Review, 106(10), 3064–3103.

- General Statistics Office of Vietnam. (2001). Vietnam population and housing census 1999. General Statistics Office of Vietnam.
- Giavazzi, F., Petkov, I., & Schiantarelli, F. (2019). Culture: Persistence and evolution. Journal of Economic Growth, 24(2), 117–154.
- Gorodnichenko, Y., & Roland, G. (2011). Which dimensions of culture matter for long-run growth? American Economic Review: Papers and Proceedings, 101(3), 492–498.
- Gorodnichenko, Y., & Roland, G. (2012). Understanding the individualism-collectivism cleavage and its effects: Lessons from cultural psychology. In M. Aoki, T. Kuran, & G. Roland (Eds.), *Institutions* and comparative economic development (pp. 213–236). London: Palgrave.
- Gorodnichenko, Y., & Roland, G. (2015). Culture, institutions and democratization. Public Choice, 187, 165–195.
- Gorodnichenko, Y., & Roland, G. (2017). Culture, institutions and the wealth of nations. *Review of Economics and Statistics*, 99(3), 402–416.
- Guiso, L., Sapienza, P., & Zingales, L. (2006). Does culture affect economic outcomes? Journal of Economic Perspectives, 20(2), 23–48.
- Guiso, L., Sapienza, P., & Zingales, L. (2011). Civic capital as the missing link. In J. Benhabib, A. Bisin, & M. Jackson (Eds.), *Handbook of social economics* (Vol. 1A, pp. 417–480). Elsevier.
- Guiso, L., Sapienza, P., & Zingales, L. (2016). Long-term persistence. Journal of the European Economic Association, 14(6), 1401–1436.
- Gächter, S., & Renner, E. (2010). The effects of (incentivized) belief elicitation in public goods experiments. *Experimental Economics*, 13(3), 364–377.
- Haines, D. W. (1984). Reflections of kinship and society under Vietnam's Le Dynasty. Journal of Southeast Asian Studies, 15(2), 307–314.
- Hall, K. R. (2011). A history of early Southeast Asia. Rowman and Littlefield Publishers.
- Hansen, C. W., Jensen, P. S., & Skovsgaard, C. V. (2015). Modern gender roles and agricultural history: The Neolithic inheritance. *Journal of Economic Growth*, 20(4), 365–404.
- Hardy, A. (2003). State visions, migrant decisions: Population movements since the end of the Vietnam War. In V. Hy (Ed.), *Postwar Vietnam: Dynamics of a transfroming society* (pp. 107–137). Institute of Southeast Asian Studies.
- Herrmann, B., & Thöni, C. (2009). Measuring conditional cooperation: A replication study in Russia. *Experimental Economics*, 12(1), 87–92.
- Hickey, G. C. (1964). Village in Vietnam. New Haven: Yale University Press.
- Ho, H.-A. (2020). Tying peasants to their land: The rise and fall of private property rights in historical Vietnam. In: EABH Working Paper No. 20-01.
- Hofstede, G. (2001). Culture's consequences: Comparing values, behaviors, institutions, and organizations across nations (2nd ed.). London: Sage.
- Inglehart, R., & Baker, W. E. (2000). Modernization, cultural change, and the persistence of traditional values. American Sociological Review, 65(1), 19–51.
- Johnson, A. W., & Earle, T. (2000). *The evolution of human societies: From foraging group to agrarian state*. Stanford: Stanford University Press.
- Kitayama, S., Hyekyung Park, A. T., Sevincer, M. K., & Uskul, A. K. (2009). A cultural task analysis of implicit independence: Comparing North America, Western Europe, and East Asia. *Journal of Personality and Social Psychology*, 97(2), 236–255.
- Kitayama, S., Ishii, K., Imada, T., Takemura, K., & Ramaswamy, J. (2006). Voluntary settlement and the spirit of independence: Evidence from Japan's 'Northern frontier'. *Journal of Personality and Social Psychology*, 91(3), 369–384.
- Knudsen, A. S. B. (2019). Those who stayed: Selection and cultural change during the Age of Mass Migration. In: Working paper, Harvard University.
- Kocher, M., Martinsson, P., & Visser, M. (2012). Social background, cooperative behavior, and norm enforcement. *Journal of Economic Behavior and Organization*, 81(2), 341–354.
- Kocher, M. G., Cherry, T., Kroll, S., Netzer, R. J., & Sutter, M. (2008). Conditional cooperation on three continents. *Economics Letters*, 101(3), 175–178.
- Le Phan, H., Phan, X. B., Truong, T. K. C., Tran, D. C., Minh Giang, V., Doan, M. H., et al. (2011). *Qua trinh hinh thanh va phat trien vung dat Nam Bo*. Hoi Khoa Hoc Lich Su Viet Nam.
- Litina, A. (2016). Natural land productivity, cooperation and comparative development. Journal of Economic Growth, 21(4), 351–408.
- Liu, D., Duong, N. T., Ton, N. D., Van Phong, N., Pakendorf, B., Van Hai, N., & Stoneking, M. (2020). Extensive ethnolinguistic diversity in Vietnam reflects multiple sources of genetic diversity. *Molecular Biology and Evolution*, 37(9), 2503–2519.

- Lowes, S., Nunn, N., Robinson, J. A., & Weigel, J. (2017). The evolution of culture and institutions: Evidence from the Kuba Kingdom. *Econometrica*, 85(4), 1065–1091.
- Martinsson, P., Villegas-Palacio, C., & Wollbrant, C. (2015). Cooperation and social classes: Evidence from Colombia. Social Choice and Welfare, 45(4), 829–848.
- Mayshar, J., Moav, O., & Neeman, Z. (2017). Geography, transparency, and institutions. American Political Science Review, 111(3), 622–636.
- Mayshar, J., Omer, M., Zvika, N., & Luigi, P. (2019). The origin of the state: Land productivity or appropriability? In: Working Paper. University of Warwick.
- Nguyen, D. D. (1994). Tong ket nghien cuu dia ba Nam Ky Luc Tinh. Ho Chi Minh City.
- Nguyen, D. D. (1996a). Nghien cuu dia ba Trieu Nguyen: Binh Dinh (Vol. 1). Ho Chi Minh City.
- Nguyen, D. D. (1996b). Nghien cuu dia ba Trieu Nguyen: Binh Thuan. Ho Chi Minh City.
- Nguyen, D.-D. (1997). Nghien cuu dia ba Trieu Nguyen: Thua Thien. Ho Chi Minh City.
- Nguyen, D.-D. (2010). Nghien cuu dia ba Trieu Nguyen: Quang Nam. Dai Hoc Quoc Gia TP. Ho Chi Minh.
- Nguyen, K.-S., Pham, M.-H., & Tong, T.-T. (2004). Northern Vietnam: From the Neolithic to the Han period. In I. Glover & P. Bellwood (Eds.), *Southeast Asia: From prehistory to history* (pp. 177–201). Routledge Curzon.
- Nguyen, T.-A. (2003). Village versus state: The evolution of state-local relations in Vietnam until 1945. Southeast Asian Studies, 41(1), 101–123.
- Nunn, N. (2012). Culture and the historical process. Economic History of Developing Regions, 27(sup1), S108–S126.
- Nunn, N. (2014). Historical development. In P. Aghion & S. N. Durlauf (Eds.), Handbook of economic growth (Vol. 2A, pp. 347–402). Elsevier.
- Nunn, N., & Puga, D. (2012). Ruggedness: The blessing of bad geography in Africa. Review of Economics and Statistics, 94(1), 20–36.
- Nunn, N., & Wantchekon, L. (2011). The slave trade and the origins of mistrust in Africa. American Economic Review, 101(7), 3221–52.
- Olsson, O., & Paik, C. (2016). Long-run cultural divergence: Evidence from the Neolithic revolution. Journal of Development Economics, 122, 197–213.
- Oster, E. (2019). Unobservable selection and coefficient stability: Theory and evidence. *Journal of Business* and Economic Statistics, 37(2), 187–204.
- Pischedda, S., Barral-Arca, R., Gómez-Carballa, A., Pardo-Seco, J., Catelli, M. L., Alvarez Iglesias, V., et al. (2017). Phylogeographic and genome-wide investigations of Vietnam ethnic groups reveal signatures of complex historical demographic movements. *Scientific Reports*, 7(12630), 1–15.
- Rambo, A. T. (1973). A comparison of peasant social systems of Northern and Southern Vietnam. Center for Vietnamese Studies, Southern Illinois University.
- Richerson, P. J., & Boyd, R. (2005). Not by genes alone: How culture transformed human evolution. University of Chicago Press.
- Riley, S. J., DeGloria, S. D., & Elliot, R. (1999). A terrain ruggedness index that quantifies topographic heterogeneity. *Intermountain Journal of Sciences*, 5(1–4), 23–27.
- Rubel, F., & Kottek, M. (2010). Observed and projected climate shifts 1901–2100 depicted by world maps of the Köppen–Geiger climate classification. *Meteorologische Zeitschrift*, 19(2), 135–141.
- Scott, J. C. (2009). The art of not being governed: An anarchist history of upland Southeast Asia. Yale University Press.
- Scott, J. C. (2017). Against the grain: A deep history of the earliest states. Yale University Press.
- Singelis, T. M., Triandis, H. C., Bhawuk, D. P. S., & Gelfand, M. J. (1995). Horizontal and vertical dimensions of individualism and collectivism: A theoretical and measurement refinement. *Cross-Cultural Research*, 29(3), 240–275.
- Spolaore, E., & Wacziarg, R. (2013). How deep are the roots of economic development? Journal of Economic Literature, 51(2), 325–369.
- Talhelm, T., Zhang, X., Oishi, S., Shimin, C., Duan, D., Lan, X., & Kitayama, S. (2014). Large-scale psychological differences within China explained by rice versus wheat agriculture. *Science*, 344(6184), 603–608.
- Tana, L. (1998). Nguyen Cochinchina: Southern Vietnam in the seventeenth and eighteenth centuries. Cornell University Press.
- Taylor, K. W. (1983). The birth of Vietnam. University of California Press.
- Taylor, K. W. (2013). A history of the Vietnamese. Cambridge University Press.
- Triandis, H. C. (1995). Individualism and collectivism. Westview Press.
- Turner, F. J. (1920). The frontier in American history. Holt.
- Van Luong, H. (1992). Revolution in the village: Tradition and transformation in North Vietnam, 1925– 1988. University of Hawaii Press.

- Vandello, J. A., & Cohen, D. (1999). Patterns of individualism and collectivism across the United States. Journal of Personality and Social Psychology, 77(2), 279–292.
- Voigtländer, N., & Voth, H.-J. (2012). Persecution perpetuated: The medieval origins of anti-Semitic violence in Nazi Germany. *Quarterly Journal of Economics*, 127(3), 1339–1392.
- Whitmore, J. K. (1984). Social organization and Confucian thought in Vietnam. Journal of Southeast Asian Studies, 15(2), 296–306.
- Whitmore, J. K. (1997). Literati culture and integration in Dai Viet, c. 1430-c. 1840. In: Modern Asian studies (Vol. 31, No. 3, pp. 665–687).
- Wook, C. B. (2004). Southern Vietnam under the reign of Minh Mang (1820–1841). Cornell Southeast Asia Program Publications.
- Wrangham, R., & Carmody, R. (2010). Human adaptation to the control of fire. Evolutionary Anthropology: Issues, News, and Reviews, 19(5), 187–199.
- Zelmer, J. (2003). Linear public goods experiments: A meta-analysis. *Experimental Economics*, 6(3), 299–310.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.