Religious Orders and Growth through Cultural Change in Pre-Industrial England

Andersen, Thomas Barnebeck; Bentzen, Jeanet Sinding; Dalgaard, Carl-Johan Lars; Sharp, Paul Richard

Publication date:
2011

Document Version
Early version, also known as pre-print

Citation for published version (APA):
Andersen, T. B., Bentzen, J., Dalgaard, C-J., & Sharp, P. (2011). Religious Orders and Growth through Cultural Change in Pre-Industrial England. Department of Economics, University of Copenhagen.

Discussion Papers
Department of Economics
University of Copenhagen

No. 11-07

Religious Orders and Growth through Cultural Change in Pre-Industrial England

Thomas Barnebeck Andersen, Jeanet Bentzen,
Carl-Johan Dalgaard, and Paul Sharp

Øster Farimagsgade 5, Building 26, DK-1353 Copenhagen K., Denmark
Tel.: +45 35 32 30 01 – Fax: +45 35 32 30 00
http://www.econ.ku.dk

ISSN: 1601-2461 (E)

Religious Orders and Growth through Cultural Change in Pre-Industrial England*

Thomas Barnebeck Andersen† Jeanet Bentzen Carl-Johan Dalgaard
Paul Sharp‡

February 16, 2011

Abstract

We advance the hypothesis that cultural values such as high work ethic and thrift, "the Protestant ethic" according to Max Weber, may have been diffused long before the Reformation, thereby importantly affecting the pre-industrial growth record. The source of pre-Reformation Protestant ethic, according to the proposed theory, was the Catholic Order of Cistercians. Using county-level data for England we find empirically that the frequency of Cistercian monasteries influenced county-level comparative development until 1801; that is, long after the Dissolution of the Monasteries. The pre-industrial development of England may thus have been propelled by a process of growth through cultural change.

Keywords: Protestant ethic, Malthusian population dynamics, economic development

JEL Classification codes: N13; O11; Z12

*We thank Daron Acemoglu, Philippe Aghion, Quamrul Ashraf, Roland Benabou, Raquel Fernández, Oded Galor, Thomas Markussen, Omer Moav, Stelios Michalopoulos, Ola Olsson, Enrico Perotti, Karl Gunnar Persson, Christoffer Sonne-Smith, Holger Strulik, Fabrizio Zilibotti as well as seminar participants at the 2011 AEA meeting in Denver, DEGIT XV in Frankfurt, the 2010 NBER Summer Institute, Gothenburg University, University of London Royal Holloway, University of Oslo and University of Zürich for comments and suggestions. Errors are ours.

†Andersen is affiliated with the Department of Business and Economics, University of Southern Denmark, Campusvej 55, DK-5230 Odense M. Contact: Thomas Barnebeck Andersen (barnebeck@sam.sdu.dk).

‡Bentzen, Dalgaard, and Sharp are affiliated with the Department of Economics, University of Copenhagen, Øster Farimagsgade 5, building 26, DK-1353 Copenhagen, Denmark. Contact: Jeanet Bentzen (jeanet.bentzen@econ.ku.dk), Carl-Johan Dalgaard (carl.johan.dalgaard@econ.ku.dk), and Paul Sharp (paul.sharp@econ.ku.dk).

1 Introduction

In what is surely one of the most famous works in all of social science, Max Weber (1905) argued that the Protestant Reformation was instrumental in facilitating the rise of capitalism in Western Europe. More specifically, Weber argued that Protestantism, in contrast to Catholicism, commends the virtues of hard work and thrift. These values, which Weber famously refers to as the "Protestant ethic", laid the foundation for the eventual rise of modern capitalism. A noteworthy study by Becker and Woessmann (2009) suggests that Weber was right, albeit for the wrong reasons: Protestants did not prosper as a result of their work ethic, say; rather, they prospered because instruction in reading the Bible generated the human capital so crucial to economic prosperity. Using data for Prussia, where some regions converted to Protestantism while others remained Catholic, Becker and Woessmann document the strength of the human capital channel. In fact, Becker and Woessmann argue that the human capital mechanism can account for most of the difference in comparative development between the Protestant and Catholic regions of Prussia.[1]

Nevertheless, one does not have to reject the original Weber thesis in order to support the human capital story. Landes (1999) is a case in point. While acknowledging the human capital mechanism, he maintains the importance of the Protestant ethic, both on empirical grounds (Protestant merchants and manufacturers played a leading role in banking, industry and trade) and on theoretical grounds (the Reformation created a new kind of man: rational, ordered, diligent and productive). More formally, using the World Values Survey, McCleary and Barro (2006) make probable that Weber was in fact right in emphasizing a link between religion on the one hand and work ethic on the other.[2] Nevertheless, it is clearly a difficult task to separately identify the importance of the human capital mechanism and the traditional "Weber mechanism", as both have arguably been at work simultaneously in the wake of the Reformation.

The present paper offers an attempt to separate the impact of the Weber mechanism and the human capital mechanism. We document below that the cultural virtues emphasized by Weber had a pre-Reformation origin in the religious Order of the Cistercians; a Catholic order which spread across England during the 12th century. We hypothesize that the Cistercians had a long term impact on development by encouraging a greater appreciation of hard work and wealth accumulation in local populations. That is, we argue the Cistercians encouraged growth by instigating the kind of cultural change that Weber attributed to Protestantism. Using cross-county data for England for the period 1377-1801, we find strong empirical support for a growth enhancing impact of Cistercian presence. Since the Cistercians did not encourage human capital accumulation, our findings suggests that the original Weber thesis, stressing the importance of cultural values like hard work and thrift, holds considerable explanatory power with respect to the pre-industrial growth

[1] Cantoni (2009), however, finds no effects of Protestantism on city growth across the German Lands of the Holy Roman Empire prior to industrialization.

[2] Thrift is not significant in their specification.

record of England.

The Cistercian order, a Benedictine offshoot, was established in France in 1098 as a reformist movement with the aim of returning to the literal observance of the "Rule of St. Benedict". They rejected the developments the Benedictines had undergone and tried to reproduce life exactly as it had been in St. Benedict's time; in fact, they often ventured beyond it in austerity. Put differently, the salient feature in the reform was a return to hard manual labor and the restrain from consumption (Kieser 1987). This meant that within the walls of the Cistercian monasteries one would find cultural values similar to those which, promulgated by the Protestant Reformation centuries later, is thought to have assisted the rise of capitalism outside the monastic walls. Several scholars have noted that the simplicity of the Order's lifestyle and their pursuit of wealth were in fact early manifestations of "the Protestant ethic" (e.g., Baumol 1990, p. 906; Collins 1986, p. 54; Kieser 1987, p. 116); Weber (1958, p. 118-19) himself singled out the Cistercians as encompassing values with a clear antecedent to the Protestant ethic.

We hypothesize that the cultural values embedded in the Cistercian order diffused to the local populations. Hence we argue that virtues associated with the Protestant movement started to spread in England long before Martin Luther posted his theses on the door of the All Saints' Church in Wittenberg. Of course, the cultural influence from the Cistercians was not immediate. Initially, the Cistercians may only have "convinced" a (potentially very) small group of people to "adopt" their attitudes towards hard work and thrift. But in a Malthusian setting work ethic and thrift translates into economic success, and ultimately into reproductive success. To the extent that cultural values carry over from parent to offspring, a cumulative process of growth through cultural change can be envisioned. If the pervasiveness of Protestant-type cultural values increases, this will stimulate work effort, investments and technological change; in turn, this works to encourage population growth and, as a consequence of selection, cultural change.

We construct a model that illustrates this cumulative process. To fix ideas, we focus on how Cistercians may have influenced the attitude towards hard work and thereby macroeconomic development. Using the model, we demonstrate that an initially small group of dynasties featuring a relatively strong preference for work effort could plausibly have come to dominate the population within the span of 500 years. Moreover, we show that small differences in the initial rate of "conversion" to a high work ethic could result in considerable cross-county variation in cultural values in the course of centuries. Finally, we derive an estimable equation from the model.

In order to proxy the initial cultural influence from the Cistercians on local populations, we employ information on the historic location of English Cistercian abbeys at the county level. With this data in hand, we proceed to document that the intensity of Cistercian presence left an important imprint on comparative development across English counties until 1801; that is, long after the Dissolution of the Monasteries, which took place between 1536 and 1540. As we focus on the pre-industrial period, and in keeping with our theoretical model, we rely on population density as our measure of productivity; in doing so we follow the lead of, among others, Ashraf and Galor (2009). Specifically, we show that, conditional on relevant exogenous controls, English counties with a higher share of Cistercian monasteries (as a fraction of all religious houses) experienced faster population growth during the period 1377-1801.

We believe the most plausible explanation for this finding is that the Cistercians influenced local cultural values, which subsequently took hold in the population. These new values in turn stimulated growth through attendant changes in work effort, investment behavior and technological progress. While we cannot document a link between Cistercian presence and *pre-industrial* cultural values across England, we are able to present evidence that the historic share of Cistercian abbeys is strongly correlated with *contemporary* work ethic and thrift at the regional level in England, as measured by the World Values Survey.

Naturally, a priori there could be other viable explanations for the observed link between intensity of Cistercian presence and population growth over the period in question. Perhaps this particular religious order simply managed to locate in areas with high growth potential; perhaps they influenced growth via international trade; or maybe the observed association is best motivated by technological change or human capital accumulation. We address these alternative narratives below. But we are led to the conclusion that they are unable to account for the observed relationship between the intensity of Cistercian presence and county-level population growth.

Our analysis contributes to several strands of literature. By demonstrating an impact from religious orders on economic development, we contribute to a literature which examines the religion-prosperity nexus (e.g., Landes 1999; Barro and McCleary 2003; McCleary and Barro 2006; Cavalcanti, Parente and Zhao 2007; Becker and Woessmann 2009; Cantoni 2009). In addition, by documenting a long lasting impact from Cistercian monasteries, our work contributes to a recent literature which suggests that past events (treatments) can permanently affect economic outcomes if they influence norms of behavior and/or culture (e.g., Guiso, Sapienza, and Zingales 2008; Nunn and Wantchekon 2009; Tabellini 2010).

The closest precursor to the argument developed below is the work of Clark (2007), which also takes as point of departure that cultural attributes, such as a high work ethic, breed economic success, and ultimately reproductive success in a Malthusian setting. Clark's theory is based on endogenous factors: The rich became rich because of certain favorable traits (cultural or perhaps even genetic); their children inherited these traits, and because the rich had such staggering reproductive success, their offspring were forced to move downward in the social hierarchy, implying that the "positive" traits eventually spread to the entire population. In a similar vein, Doepke and Zilibotti (2008) develop a theory of endogenous preference formation whereby cultural virtues conducive to growth flourish in parts of the population and facilitate a growth take-off. In contrast to Clark, however, Doepke and Zilibotti argue that the new cultural values emerged in the middle class, and not among the initially rich. We differ from both Clark and Doepke and Zilibotti in emphasizing a *shock* to cultural values: the settlement of the Cistercians. This allows us to test our argument statistically. In contrast to Doepke and Zilibotti, but similarly to Clark, we emphasize the reproductive advantage of high work ethic dynasties in explaining the diffusion of cultural values. In practise, deliberate investments and differential fertility probably both contributed to the diffusion of work ethic and thrift.

Finally, our analysis is related to evolutionary growth theory, as pioneered by Galor and Moav (2002). Galor and Moav demonstrate how dynasties with greater preference for child quality, relative to child quantity, hold a selective advantage in a Malthusian setting and come to dominate the population. Moreover, the theory predicts that the positive selection of quality oriented individuals stimulated long-run economic development. Similarly, the theory advanced below predicts that the epoch of Malthusian stagnation involved selection of individuals with a high work ethic, thus importantly influencing comparative development in England during the pre-industrial era.

The rest of the paper is organized as follows: Section 2 presents the theory, including the formal model, while Section 3 contains the empirical analysis. Section 4 provides a conclusion.

2 Theory

This section develops a theory of how Cistercians may have left a lasting imprint on comparative development in England. The following subsection provides details on Cistercian monks. We discuss their values, how these values manifested themselves in terms of work effort, capital accumulation and technological change, and how their values may have spread to the local population.

We argue that a major reason why the presence of Cistercian monasteries is apparent in comparative development long after the Order's disbandment is that they instigated a process of cultural change. To clarify how this process may have played out, we develop a model of growth through cultural change in Section 2.2. The model elucidates how an initially modest "cultural shock" to a population cumulates over time in a Malthusian setting, ultimately leaving a significant imprint on the growth record. The model also allows us to gauge the speed of the process of cultural change.

2.1 Cistercian Values

The Cistercian order was founded in 1098 in France; the first Cistercian monastery in England was founded in 1128 (Cooke 1893; Donkin 1963). During the 12th Century the Order spread rapidly across England, cf. Figure 1. By the end of the 14th century the expansion of the Order had essentially ceased. Hence, from the perspective of our regression analysis below, which involve the time period from 1377 onwards, we can treat Cistercian settlements as predetermined.

There is little doubt that the Cistercians held beliefs which were later to be associated with the Protestant

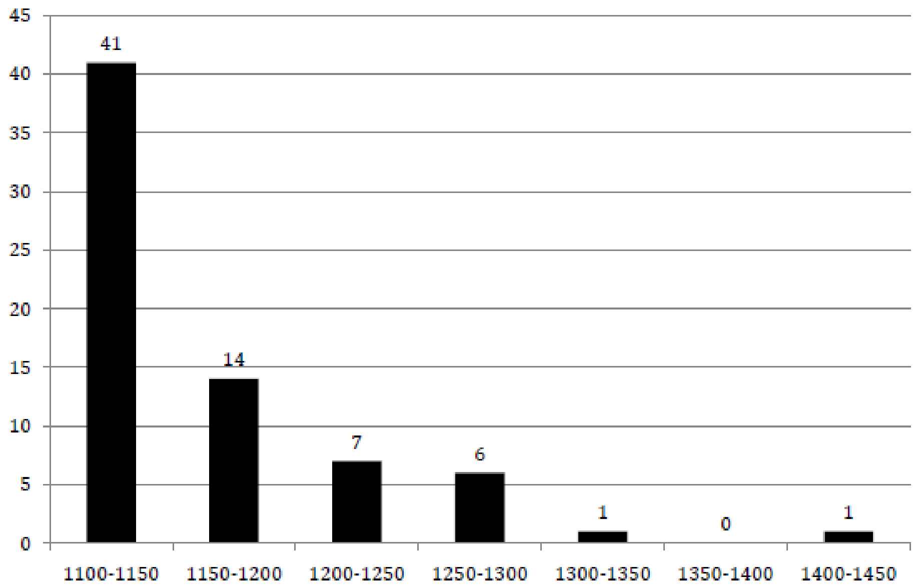

Figure 1: Frequency of founding years of Cistercian monasteries In England.

ethic. By seeking to return to a literal interpretation of the Rule of St. Benedict, the small book written in the sixth century by its namesake, they stressed the trinity of prayer, work and study, as well as the values of practicality, adaptability, simplicity and moderation (Hill 1968, p. 3). The Exordium Cistercii, written in the 1120s, and the statutes promulgated at the general chapter of 1134, stated that the monks were to work hard and live "*from the labour of their own hands, from cultivation and from their flocks*". They were also to live frugally, and were not permitted to have any possessions "*contrary to monastic purity*" such as parish churches, the tithes of other men's labour, dependent peasants, mills, ovens, or other income sources attached to the land. Hence, it is no surprise that Baumol (1990, p. 906) suggests that the monks of the Order of Cistercians may have embodied an earlier "Protestant ethic": "*Puritanical, at least in the earlier years, in their self-proclaimed adherence to simplicity in personal lifestyle while engaged in dedicated pursuit of wealth, they may perhaps represent an early manifestation of elements of 'the Protestant ethic'*". Collins (1986, p. 54) is slightly more direct when he notes that the Cistercians: "*had the Protestant ethic without Protestantism*".[3]

The simplicity of the Cistercians was thus only a liturgical simplicity, replacing long days of ritual with short prayers that could be said in pauses from labor (Bouchard 1991; Hill 1968). Moreover, "useless" labor, such as painting pictures, decorating books, breeding useless animals, etc. was banned (Kieser 1987). Some

[3] Kiefer (1987, p. 116) makes the same observation.

have suggested that they were attempting to reduce the need for manual labor in order to leave more time for prayer (Bloch 1935; Gimpel 1976; Ovitt 1986; Landes 1999). Whatever the case, from the very beginning the Cistercians were involved in the rapidly developing economic practices of the 12th century, and were in some cases initiators of these practices. Moreover, the monks' asceticism, by keeping down consumption, drove up levels of investment (Kiefer 1987; Baumol 1990).

Kaelber (1998) points out that Weber himself saw monastic asceticism as a clear precursor to ascetic Protestantism: the key driving force behind European capitalism according to Weber. More specifically, as argued by Weber (1958, p. 118-19): "*In the rules of St. Benedict, even more so in the case of the monks of Cluny and the Cistercians...[Christian asceticism] has become a systematically developed method of rational life conduct, with the goal to overcome the status naturae, to free man from the power of irrational impulses and his dependence on the world and on nature...It attempted to subject man under the supremacy of purposive will, to bring his action under constant self-control with a careful consideration of their ethical consequences. Thus it trained the monk, objectively, as a worker in the service of the Kingdom of God, and thereby further, subjectively, assured the salvation of his soul. . . [T]he end of this asceticism was to be able to lead an alert, intelligent life: the most urgent task the destruction of spontaneous, impulsive enjoyment, the most important means was to bring order into the conduct of its adherents. All these important points are emphasized in the rules of Catholic monasticism as strongly as in the principles of conduct of the Calvinists.*" Hence the idea that the Cistercians held values close to those promulgated by the Protestant Reformation has a long and distinguished tradition.[4]

The emphasis on hard work and thrift made the Cistercians entrepreneurial and ultimately very successful economically (Baumol 1990). They contributed much as agriculturists and as horse and cattle breeders. Their major contribution was the introduction of the grange system, whereby land was held in compact blocks, in contrast to the usual fragmented and unenclosed village holdings (Donkin 1963). Another contribution seems to have been advanced irrigation techniques, thus predating Rowland Vaughan's famous popularization of these methods by centuries.[5] Moreover, their high level of agricultural technology was matched by their industrial technology. Every monastery had a model factory, often as large as the church, with waterpower to drive the machinery (Gimpel 1976). This power was used for crushing wheat, sieving flour, fulling cloth and tanning (Baumol 1990). The Cistercians are known to have been skilled metallurgists (Gimpel 1976).

The Cistercian monastic system was one based on the principle of kinship, and thus Cistercian work practices and technology seem to have spread easily from house to house (Donkin 1978). These values

[4] As Weber points out, similar values were found among the Cluniacs. The impact of the Cluny order has received scant attention in the literature in comparison with the Cistercians. Yet, as we shall see, they too seem to have left a mark (albeit not as statistically robust as the Cistercians) on pre-industrial growth in England, conceivably for the same reasons the Cistercians influenced growth.

[5] Vaughan's Golden Valley was actually located in an area where the Cistercians had held extensive estates prior to the Dissolution (Cook, Stearne and Williamson 2003).

in turn spread into the local area partly due to the Cistercian practice of incorporating illiterate peasant lay brothers (known as conversi) for agricultural labor (Berman 2000). Lay brothers were bound by vows of chastity and obedience to their abbot, but were otherwise permitted to follow a less demanding form of Cistercian life. Work on Cistercian granges were also carried out by various classes of secular laborers. These included servi (servants), mercenarii (hired laborers), familiares (workers with intermediate status between hired workmen and lay brothers) and donate or oblate (pious laymen exchanging work for support). The exact fraction of lay brothers to these other types of labor is difficult to determine, but the latter seem to have become increasingly important at the turn of the 13th century (Noell 2006). Another important group of settlers in the abbeys were the corrodians, who spent their years of retirement there. Moreover, settled communities, including shopkeepers, formed outside the monasteries (Williams 1970). In this manner, the ways of the Cistercians spread beyond the Order itself.

An ideal check of this would be to study the correlation between Cistercian presence and preindustrial ethical values, like work ethic and thrift. Unfortunately, data constraints prevent us from carrying out such a check. What we can do instead is to study the relationship between the intensity of historical Cistercian presence and the pervasiveness of (proxies for) contemporary Protestant ethic across England.

In order to quantify differences in work ethic across countries, McCleary and Barro (2006) use the fraction of World Values Survey (WVS) respondents who indicated that they think that valuing "hard work" is an important trait for children to learn at home. To measure thrift they calculate the frequency of respondents indicating that "thrift, saving money and things" is an important trait for children to learn at home. These variables are also available for the United Kingdom in the WVS 2005. Unfortunately, it is only possible to disaggregate down to the regional level.[6] The "intensity" of Cistercian presence in a geographical area is proxied as the number of Cistercian monasteries relative to the total number of religious houses in the area.

Figures 2 and 3 depict the correlation between the intensity of Cistercian presence and work ethic and thrift, respectively. As expected, the three variables are positively correlated; Cistercian presence has a correlation of 0.62 with "work ethic" and 0.42 with "thrift". With only eight observations, an OLS regression returns a statistically significant (at the five percent level) correlation between work ethic and Cistercian presence. Statistical significance is not attained (at conventional levels) in the context of thrift; but the positive association between Cistercian presence and fraction of regional respondents emphasizing thrift is visually discernible and positive. Taken together, this exercise provides some support of our hypothesis that the Cistercian "treatment" influenced cultural values across England.

[6] The data only have regional identifyers. We have also contacted the British Values Survey, and the same holds for this survey.

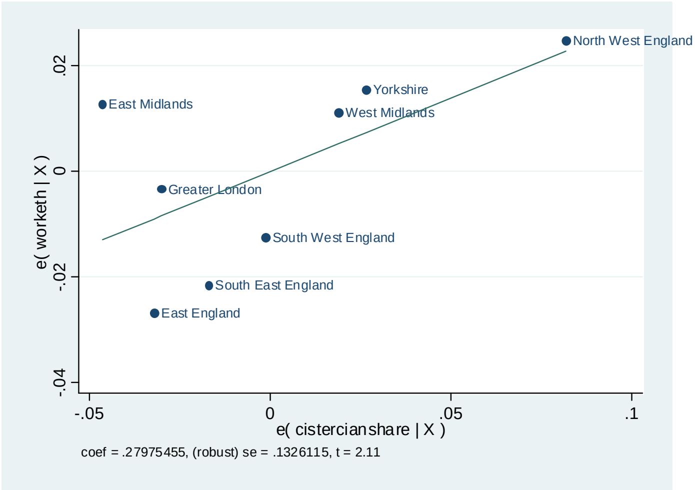

Figure 2: Cross regional correlation between Cistercian intensity and contemporary work ethic (source World Values Survey).

2.2 A Model of Growth through Cultural Change

In order to think more formally about how the ways of the Cistercians spread beyond the Order itself, and the ensuing macroeconomic impact, consider the following overlapping generations model for a closed economy in the process of development. Time is discrete and extends to infinity, $t = 0, 1, 2...\infty$.

People live for at most two periods: childhood and adulthood. During childhood individuals receive consumption from their unique parent; for simplicity of exposition, we assume that individuals only consume during period one.[7] If adulthood is reached, an individual decides on work effort and then reproduces. While each parent produces a fixed number of offspring, nutritional intake (deterministically) determines how many survive to adulthood. Hence, while fertility is exogenous, the number of surviving offspring (and thus population growth) is endogenous on account of the link between nutritional intake and the fraction of children that makes it into adulthood.

Cultural values are crudely represented by utility weights. Specifically, "high work ethic" dynasties are identified as dynasties that attach relatively low disutility to effort, and the impact from the Cistercians are conceptualized as a shock to the utility weights of a subset of the individuals. Moreover, we assume that in the absence of a cultural shock the offspring adopts the preferences of their unique parent. Needless to say these are strong assumptions. Yet the point of the model is not to assess the generality of the hypothesized

[7] This could be viewed as an implicit assumption that parental consumption only involves a fixed minimum consumption requirement, which we then for expositional simplicity normalize to zero. With a positive (exogenous) level of parental subsistence consumption we would have to take into account that there might be situations (i.e., parameter configurations and population levels) under which dynasties die out. We have no particular interest in studying this sort of scenario, which motivates the normalization.

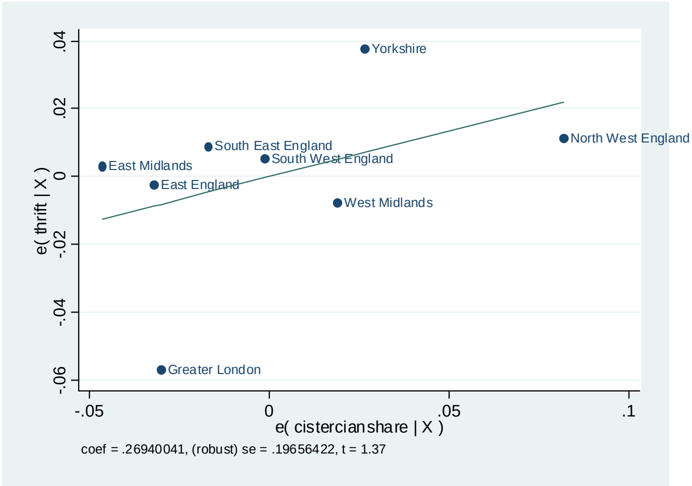

Figure 3: Cross regional correlation between Cistercian intensity and contemporary thrift (source World Values Survey).

trajectory, whereby a shock to the preferences of a small subset of individuals eventually causes a proliferation of these new preferences in the population at large. The point is rather to examine the implications of such a trajectory, which we a priori hypothesize is relevant for pre-industrial England, and then subsequently confront with data, and potentially reject.[8]

By focusing on the impact of changes in the attitude to hard work we suppress changes in cultural attitudes towards saving and investment; i.e., thrift. It is worth observing, however, that while the model focuses on the work ethic of individuals, similar results would likely arise if we instead examined thrift. As long as thrift implies a greater earnings potential, groups with high thriftiness will be selected in the Malthusian setting.[9]

A final observation worth making is that in the canonical Malthusian macro model fertility is endogenously determined, while child mortality is suppressed or implicit (see Ashraf and Galor 2009). Here it is the other way around. We follow this alternative route because it allows us to capture the taste for hard work in a

[8] In a more detailed model one would want to study the conditions under which preferences are in fact maintained over time within a dynasty, and distinguish between men and women in order to study matching. But it should be clear that the reduced form outcome studied below may well be viable in this richer environment if differences prevail across dynasties in perceived benefits of high work ethic values (for which reason such values are promoted by some dynasties but not by others; Doepke and Zillibotti 2008), and if there is sufficient assortative matching in the marriage market such that individuals with identical values choose to set up a family together (Becker, 1973). In such an environment one can think about the "Cistercian shock" as being represented by a change in perceived benefits to a high work ethic rather than by directly modifying preferences of some individuals. In order to maintain tractability, while studying the dynamic general equilibrium of the economy in the presence of cultural heterogeneity, we ignore this kind of micro-level behavior however interesting it may be in its own right.

[9] Becker (1980) explores a dynamic economy where agents differ in terms of the rate of time preference; i.e., in terms of "thrift". In the long run the most patient dynasty ends up "owning" the economy. Below we demonstrate a similar result in that dynasties with greater work ethics will end up dominating the population.

convenient way, while at the same time retaining the basic properties of the Malthusian framework as well as simplicity and tractability.

2.2.1 Individual's optimization problem

Individuals derive disutility from work effort and utility from the number of surviving offspring, with utility given as

$$u_t = \omega \log(1 - e_t) + \log(\sigma_t n). \tag{1}$$

In equation (1) $e_t \in (0,1)$ is work effort, $n > 1$ is the (fixed) number of children coming into existence and σ_t is the fraction that survives to adulthood. The parameter $\omega > 0$ captures individuals' distaste for hard work. It can thus be seen as a simple manifestation of the work ethic of the individual, which we will think of as a cultural value that typically is fixed but may undergo occasional "mutation" in a population. In the absence of shocks to preferences, we assume children inherit the cultural value from their parent. For now we consider a population where everyone shares the same work ethic, ω. Finally, observe that since $\sigma_t n$ is ultimately linked to consumption the utility function is in a reduced form sense defined over consumption and work effort.

A higher level of effort allows for more income, which in turn facilitates a higher level of child consumption, c_t, according to

$$c_t = e_t y_t, \tag{2}$$

where y_t is potential per capita income. Finally, we assume that the survival rate to adulthood depends on child consumption in the following way

$$\sigma_t = \min\left\{\frac{c_t/n}{\eta}, 1\right\}, \tag{3}$$

where the parameter $\eta > 0$ is a reference consumption level.[10] As consumption per child falls below η, child mortality rises. Equation (3) is meant to capture that insufficient nutrition during early childhood weakens the offspring and thereby (deterministically) elevates the mortality rate. In this way we capture the notion of a Malthusian "positive check": In periods of plenty c_t/n will rise thus allowing for more surviving offspring; vice versa when income falls.[11]

[10] Assuming that maximum survival is 100% is a simplification. Nothing would change if we, at the costs of additional notation, were to assume a maximum survival rate of, say, $\bar{\sigma} < 1$ instead; that is, assume that $\sigma_t = \min\{[c_t/n]/\eta, \bar{\sigma}\}$.

[11] Kelly and O'Grada (2008) find that low real wages caused by bad harvests led to increased mortality in England during the 14th and 15th century. This link is however weakened from the 16th century onwards, quite possibly due to the introduction of the Poor Law.

Maximizing (1) subject to (2), (3) and $e_t \in (0,1)$, we find that

$$e_t^* = \begin{cases} \frac{\eta n}{y_t} & \text{if} \quad y_t \geq (1+\omega)\, n\eta \\ \frac{1}{1+\omega} & \text{if} \quad y_t < (1+\omega)\, n\eta \end{cases}, \quad c_t^* = \begin{cases} \eta n & \text{if} \quad y_t \geq (1+\omega)\, n\eta \\ \frac{1}{1+\omega} y_t & \text{if} \quad y_t < (1+\omega)\, n\eta \end{cases}. \tag{4}$$

The solution in (4) shows that for all t, $e_t^* \in \left(0, \frac{1}{1+\omega}\right]$. That is, it can never be optimal to supply an effort level larger than the upper bound $\frac{1}{1+\omega}$. This is so because with $y_t > (1+\omega)\, n\eta$ potential income is so high that providing an effort level at $\frac{1}{1+\omega}$ would make $\frac{c_t/n}{\eta} > 1$. This can never be optimal since individuals (by assumption) have no interest in consumption *per se*; consumption only matters insofar as it increases σ_t. Indeed, when $\frac{c_t/n}{\eta} > 1$ utility can be increased by lowering e_t (and thus c_t) without affecting σ_t. On the other hand, when $y_t < (1+\omega)\, n\eta$ the effort level is constant; i.e., $e_t^* = e^* = \frac{1}{1+\omega}$.

2.2.2 Production

There is a unique consumption good, Y_t, which is produced using labor input, L_t, as well as a fixed supply of land, X. Technology (or aggregate efficiency), A, is also parametrically fixed. Potential income is

$$Y_t = AL_t^{\alpha} X^{1-\alpha} \Rightarrow y_t = A\,(L_t/X)^{\alpha-1}, \tag{5}$$

where $y_t \equiv Y_t/L_t$.

Actual income of individuals depends on effort. Effort is thought to scale income up or down but is not subject to diminishing returns. This assumption ensures that dynasties who exert more effort will hold a permanent earnings advantage; if effort is subject to diminishing returns high-effort dynasties would not persistently be able to sustain larger family sizes than low-effort dynasties. The absence of diminishing returns to effort may be reasonable if "effort" is given a broad interpretation. That is, if effort is thought to capture the intensity at which individuals dedicate themselves mentally as well as physically to income enhancing activities, rather than being narrowly defined as the supply of working hours. Assuming labor absorbs all rents, the income of an individual therefore is

$$e_t y_t = e_t A\,(L_t/X)^{\alpha-1}. \tag{6}$$

2.2.3 The Evolution of the Economy

We now characterize the dynamic evolution of the economy. Initially we consider a setting with cultural homogeneity. Having characterized this we then introduce a cultural change, taking the form of a parametric change in preferences, and investigate how this cultural change plays out in the economy.

Cultural Homogeneity When all individuals share the same work ethic, ω, the size of the population evolves according to

$$L_{t+1} = \sigma_t n L_t, \; L_0 \text{ given.} \tag{7}$$

We therefore have the following lemma:

Lemma 1

(i) The time path for population is given by the law of motion

$$\begin{aligned} L_{t+1} &= \begin{cases} nL_t & \text{if} \quad L_t \leq \overline{L} \\ \frac{1}{1+\omega}\frac{1}{\eta}AL_t^{\alpha}X^{1-\alpha} & \text{if} \quad L_t > \overline{L} \end{cases}, \\ &\equiv G(L_t), \; L_0 \textit{ given,} \end{aligned}$$

and $\overline{L} \equiv \left(\frac{(1+\omega)n\eta}{A}\right)^{\frac{1}{\alpha-1}} X.$

(ii) For constant values of ω, A, η *and* X*, the model admits a unique steady state population size,* L^**, given by*

$$L^* = \left(\frac{A}{1+\omega}\frac{1}{\eta}\right)^{\frac{1}{1-\alpha}} X.$$

(iii) Steady state income per capita is

$$ey^* = \eta.$$

Proof. See Appendix.

Proposition 1

(i) The level of income per capita is independent of technology and preferences. (ii) A higher level of technological sophistication and greater preference for work effort increases long-run population size: $\partial L^*/\partial A > 0$, $\partial L^*/\partial \omega < 0$.

Proof. See Appendix.

The dynamic system works as follows. If initially the population is sufficiently small (i.e., if $L_t < \bar{L}$) the level of income per capita in the economy is high enough to ensure that all children survive.[12] Hence $\sigma = 1$, for which reason the population grows at the exogenous rate n. As $n > 1$, a steady state does not exist in

[12] Again, this follows since we have (to conserve on notation) defined the maximum survival to be 1.

the range $0 < L_t < \bar{L}$. Eventually, however, income drops below the threshold needed for an entire cohort to survive to adulthood, due to diminishing returns to labor input. Gradually, therefore, population growth grinds to a halt and the economy ends up in the steady state L^*. In the steady state the level of income per capita is η; *independent* of technology and preferences.

The intuition for the comparative statics is straightforward. Individuals with a greater work ethic will obtain a higher level of income, which in turn translates into more surviving offspring. Similarly, if A rises, income per capita increases, and this again leads to more surviving offspring. However, in both cases a larger population is associated with diminishing returns, which serves to equilibrate the system. In the long-run, therefore, greater work effort or technological advances will not elevate income per capita but are fully converted into a larger population. Aside from the impact of cultural values on long-run population density, these predictions coincide with those of the canonical Malthusian macro model (Ashraf and Galor 2009).

Cultural Heterogeneity Suppose now that a (small) subgroup of the population experience a parametric preference change; disutility from work effort declines. Subsequently there are two groups in society: group one and two, with $\omega_2 < \omega_1$. These new preferences are preserved within the dynasty, and thus transmitted from parent to offspring. By assuming a *unique* shock to preferences we are able to examine the pure selection channel by which the new preferences become more pervasive in the population. That is, by way of higher reproductive success, high effort dynasties grow as a share of the population. Naturally, a process of cultural change occurs more rapidly if low effort dynasties gradually choose to mimic the successful high effort dynasties. But in the present case we assume that there is a unique shock to preferences.

Hence, in the absence of further shocks we have the following lemma:

Lemma 2

(i) With cultural heterogeneity the law of motion for population size is given by

$$\begin{aligned} L_{t+1} &= \begin{cases} n\left(L_{1,t}+L_{2,t}\right) & \text{if} \quad L_t < \overline{L}_1 \\ nL_{2,t} + \frac{1}{1+\omega_1}\frac{1}{\eta}AL_{1,t}^{\alpha}X^{1-\alpha} & \text{if} \quad \overline{L}_1 \le L_t < \overline{L}_2 \\ \left\{ \frac{\omega_1-\omega_2}{(1+\omega_1)(1+\omega_2)} \left[1 + \left(\frac{1+\omega_2}{1+\omega_1}\right)^t \frac{L_{1,0}}{L_{2,0}} \right]^{-1} + \frac{1}{1+\omega_1} \right\} \frac{1}{\eta} A L_t^{\alpha} X^{1-\alpha} & \text{if} \quad \overline{L}_2 \le L_t \end{cases}, \\ &\equiv G\left(L_t, t\right), \ L_{1,0}, L_{2,0} \text{ given,} \end{aligned}$$

where total population $L_t = \sum_i L_{it}$, $\overline{L}_i \equiv \left(\frac{(1+\omega_i)n\eta}{A}\right)^{\frac{1}{\alpha-1}} X$, $i = 1, 2$.

(ii) For constant values of ω_i, A, η *and* X, *the model admits a unique steady state population size,* L^*,

given by

$$L^* = \left(\frac{A}{1+\omega_2} \frac{1}{\eta} \right)^{\frac{1}{1-\alpha}} X,$$

where $L^* > \overline{L}_2$.

Proof. See Appendix

Proposition 2

If initially the population holds a work ethic consistent with ω_1 *and then subsequently an arbitrarily small subgroup of the population changes cultural values to* $\omega_2 < \omega_1$ *then long-run aggregate population size rises.*

Proof. See Appendix

With two groups there are two thresholds to be distinguished, as ω is a co-determinant of the level of population at which $\sigma = 1$. As group two exerts more effort ($\omega_2 < \omega_1$) it is able to ensure survival of an entire birth cohort at a lower level of potential income, y, than group one. This creates the intermediate regime, $\bar{L}_1 < L_t < \bar{L}_2$, where all offspring of high work ethic parents survive until period two, whereas a fraction of the children of the low work ethic group dies during childhood. Eventually, however, income falls to a sufficiently low level to produce $\sigma < 1$ for both groups. The process ultimately stabilizes and allows for a unique steady state as described in Lemma 2.

The dynamic process works as follows. Initially a small group of citizens change cultural values. Since the new group works harder, their income is greater. This works to increase population density. However, if the high work ethic group initially is small then the immediate impact on aggregate population size will be miniscule. But since the high work ethic group can afford higher levels of child consumption, it holds a reproductive advantage in the Malthusian setting. This advantage implies that the group's population share gradually rises over time, thereby increasingly stimulating aggregate population size. Hence, after the initial shock to the economy, the growth process is driven by the changing composition of the population in terms of cultural values; a process of growth through cultural change is occurring.

Eventually the group with high work ethic will dominate the population, and the economy convergences to a steady state where population size reflects the preferences of the high work ethic group. It follows from Proposition 2 that the steady state level of population is higher in the new steady state compared to the original one, as $\omega_2 < \omega_1$.

The model thus shows how a change in a certain cultural attitude in a small subset of the population may rise in importance due to selective pressures and eventually influence the macroeconomy. The source of the change of preferences is left unexplained by the model. But it seems plausible that the Cistercians have influenced county populations in this manner, as argued in Section 2.1. Accordingly, our hypothesis is that Cistercians planted the seeds of change by affecting the cultural attitudes; or, more appropriately, the work ethic of a (in principle arbitrarily) small part of the county population. By so doing, they instigated a process of growth through cultural change.

2.2.4 Speed of Diffusion

A question of some relevance is how fast the cultural diffusion process played out if it only emanates from differential population growth rates across dynasties with different values. Naturally, the process would conceivably occur at a faster rate than what we find below if values gradually diffuse across dynasties as well, following the initial shock to a select group of dynasties. In practise one may well imagine that both mechanisms were at work.

In order to examine the speed of population-growth driven cultural change, note that the fraction of individuals with high work ethic at time t is

$$\frac{L_{2,t}}{L_t} \equiv \pi_t = \frac{1}{1 + \left(\frac{1+\omega_2}{1+\omega_1}\right)^t \frac{L_{1,0}}{L_{2,0}}}, \tag{8}$$

when $L_t > \bar{L}_2$. Hence we focus on the (more realistic) case where not all children survive to adulthood from either group. The speed at which ω_2 becomes dominant in society depends on how much more effort the high work ethic group exerts, $\frac{1+\omega_2}{1+\omega_1} = \frac{e_1}{e_2}$, as well as how many individuals were "persuaded" to change their values as of time $t = 0$. The ratio of ω's is hard to pin down in any precise manner. But suppose group two exerts 20% higher effort than group one.[13] In this case Figure 4 shows how the new cultural values grow in significance over time for different assumptions about the initial "infection rate"; that is, $\pi_0 = 0.1\%$, 1% and 10% of the population, respectively.

The spread of the new cultural values follows an S-shaped trajectory: the process is slow to begin with but accelerates over time and ultimately levels off. Consider the curve in the middle, associated with an initial "infection rate" of 1%. The first 10 generations only raise the fraction with strong work ethic modestly (to about 6%), the next 10 generations increase the share to 30% of the population, and another five to nearly 50%. If a generation is about 20 years, 25 generations (what it takes to go from 1% to 50%) is about 500 years. The point is that, within the window of observation available to us (about 500 years), it is possible for a small (initial) cultural shock from the Cistercians to accumulate into a major aggregate impact on the

[13] Clark and Van der Werf (1998) estimate that the number of days worked per year (standard deviation in parenthesis) rose in England from 266 (4.8) in 1560-99 to 280 (12.9) in 1771. Suppose this increase is attributable to the rise of the Protestant work ethic, resulting from the Cistercian presence and the Reformation. Then the estimated increase over time in work days provides a crude guesstimate for ω. Factoring in the statistical uncertainty we may note that working days in 1771 may have been between 5% lower and 23% higher than in 1560, with a mean around +10%. Hence, assuming a 20% higher work effort may not be outlandish; especially so since our notion of work effort is somewhat broader than the mere number of days worked.

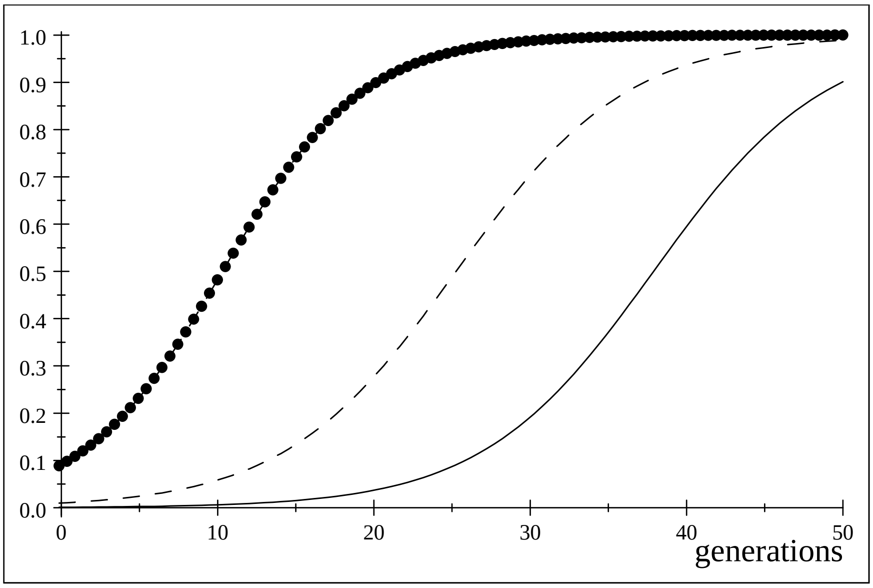

Figure 4: The rise of new cultural values in the population. *Assumptions*: (a) 20% higher work effort among individuals with "high" work effort". (b) Initial "infection" rate: 1/1000 (solid black), 1/100 (dashed), 1/10 (dotted).

composition of the population solely by way of selective pressure.

Another point worth emphasizing is the implied comparative differences in cultural beliefs that seemingly small initial differences translates into. With an initial infection rate, π_0, of one percent, 50 percent of the population holds a high work ethic after 25 generations; but only eight percent have high work ethic after 25 generations if the initial infection rate is 1/10th of a percentage point. This implies that variations in the intensity of Cistercian presence may have generated substantial comparative differences in cultural values across English counties over the period in question, by affecting π_0. It may therefore be possible to detect the legacy of the Cistercians on population dynamics over the period 1377-1801, which we examine below.

3 Empirical Analysis

This section proceeds in a series of steps. We begin by deriving an empirical model based on the theoretical model from the previous section. Subsequently, in Section 3.2, we discuss the relevance and interpretation of the resulting empirical model. Section 3.3 presents the data, while Section 3.4 contains our OLS regression results. Finally, Section 3.5 discusses potential pitfalls in ways of identification and reports our IV results.

3.1 From the Theoretical Model to an Empirical Model

The theoretical analysis established that, upon a small cultural shock to the population, areas with a larger fraction of their citizens having a higher work ethic should see faster population growth. To make this prediction amendable to empirical testing, we use Lemma 2 to write the law of motion for aggregate population as

$$\log(L_{t+1}) - \log(L_t) = -\log(\eta) - (1-\alpha)\log(L_t/X) + \log(\psi_t) + \log(A).$$

where

$$\psi_t = \left[\frac{\omega_1 - \omega_2}{(1+\omega_1)(1+\omega_2)}\pi_t + \frac{1}{1+\omega_1}\right],$$

while $\pi_t \equiv L_{2,t}/L_t$ captures the fraction of the population with high work ethic.[14] If we linearize $\log(\psi_t)$ around $\pi_t = 0$ we obtain $\log(\psi_t) \approx -\log(1+\omega_1) + \frac{\omega_1-\omega_2}{1+\omega_2}\pi_t$, which can be reinserted into the law of motion for population growth so as to obtain (approximately)

$$\log(L_{t+1}) - \log(L_t) = -\log(1+\omega_1) - \log(\eta) + \frac{\omega_1 - \omega_2}{1+\omega_2}\pi_t - (1-\alpha)\log(L_t/X_i) + \log(A).$$

Finally, denoting a county by i and adding an error term, we may write the above as the empirical model:

$$\Delta\log(L_{it+1}) = a_0 + a_1\log(L_{it}/X_i) + a_2\pi_{it} + \mathbf{Z}_i'\mathbf{a} + \epsilon_i,$$

where $\mathbf{Z}_i$ contains time-invariant controls for productivity (A).

Naturally, we do not have data on π_{it}. But, according to the theory, we may proxy it using some measure of the *intensity* of Cistercian presence in the county, as it should influence π_{i0}, and thereby π_{it} (see equation (8)). We define this intensity as the *Cistercian presence relative to other moral influences.* Since the Church was the principal authority in matters of moral in medieval times, we construct π as the ratio of Cistercian monasteries to all religious houses; i.e. $\pi = C/R$. However, the counterfactual we are interested in is that of changing the *composition* of moral influences while at the same time holding constant its *level.* This dictates that we also control for the total number of religious houses, R, separately. Consequently, we take the following specification to the data:

$$\Delta\log(L_{it+1}) = a_0 + a_1\log(L_{it}/X_i) + a_2\frac{C_{it}}{R_{it}} + a_3 R_{it} + \mathbf{Z}_i'\mathbf{a} + \varepsilon_i. \tag{9}$$

Ceteris paribus, areas with more Cistercians saw a larger fraction of the population initially being "persuaded" by the Cistercian work ethic. As seen from Figure 4 this should imply a higher π_{it} at any given

[14] More specifically, we use equation (10) in the proof of the lemma; see Appendix A.3.

point in time. As a result, we expect a_2 to come out with a positive sign. In addition, theory predicts that $a_1 < 0$, capturing convergence effects. The coefficient a_3 is a priori indeterminate.

3.2 Relevance and Interpretation of the Empirical Model

In the context of empirical relevance there are two issues worth considering. First, is the Malthusian perspective relevant for the period in question? Second, equation (9) is derived under a "closed economy" assumption. That is, the model does not allow for migration flows. Is this a reasonable approximation for the period in question?

Ultimately it is an empirical issue whether the Malthusian population theory has any bearing on developments in England from 1377-1801. However, a priori we believe a reasonable case can be made that Malthusian considerations were relevant. Clark (2007) builds a strong case that Malthusian dynamics were relevant until about 1800 in England. In a similar vein, Møller and Sharp (2008), employing time series data for England over the period 1560-1760, confirm the relevance of Malthusian population dynamics. Using cross-country data Ashraf and Galor (2009) also confirm the central predictions of the Malthusian model in pre-industrial times.

We also view the closed economy assumption as a reasonable (albeit crude) approximation to reality for the period in question. Although serfdom began to decline with the Black Death, and was practically obsolete in England by the sixteenth century, even as late as the early eighteenth century internal migration was characterized by limited geographical movement (Clark 1979). A contributing factor to the low degree of mobility was the Old Poor Law, which meant to supply relief to the temporarily unemployed, and which was administrated at the Parish level. In particular, the Settlement Act of 1662 imposed that only "established residents" of a Parish could receive relief. In practise this meant that only individuals who were able to prove an affiliation with the Parish through birth, marriage or apprenticeship were eligible for aid. Needless to say, this policy worked to lower mobility. The Poor Law Amendment Act overhauled the existing system. In particular, it established "Poor Law Unions" around groups of Parishes, which then administrated the poor relief. But this amendment did not take effect until 1834.

This is not to say that individuals did not move at all during the period in question. They did, and increasingly so over time. London, in particular, enjoyed a special status and always experienced substantial immigration. But it was the industrial revolution which saw the major break with the past and large scale migration in particular to the new industrial centres in the Northwest and the Midlands (Nicholas and Shergold 1987). As a result, suppressing internal migration seems like a reasonable approximation for most of the period in question.

Nevertheless, one may speculate what the implications would be if the assumption is not met. Suppose that counties characterized by individuals with high work ethic are (for this reason) more innovative. That is, suppose A is affected by culture (more on this below). If so, one would expect people to migrate to the high work ethic counties, which would stimulate growth in population density in areas where individuals have high work ethic. In terms of population dynamics this outcome is therefore observably equivalent to the no-migration scenario that we examined theoretically. The only way to distinguish the migration scenario from the no-migration scenario would be to study the impact of Cistercian presence on income per capita. With migration innovations would induce rising income per capita, wheras this is not the case when no migration is taking place (see Ashraf and Galor, 2009). Unfortunately, county level data on income per capita is not available for England during this period. Hence we are unable to distinguish whether the Cistercians, through instigation of cultural change, induced higher population growth either by increasing the number of surviving offspring at the county level (as suggested by our model), by stimulating inward migration, or by some combination of the two. However, regardless of the precise source of rising population density, a positive impact from Cistercians on population growth implies, in a Malthusian setting, a productivity enhancing effect from the Order. Hence, the empirical test of the impact from the Cistercians (using equation (9)) is meaningful whatever the "truth" may be about internal migration in England prior to 1801. But there are other reasons why the interpretation of a_2 might differ from what is implied by the theoretical model.

While equation (9) has a structural foundation in our model, we doubt very much that our estimations of a_2 will map into the structural parameters ω_1 and ω_2, for two reasons. First, the Cistercians almost certainly also influenced attitudes towards saving and investment; i.e., thrift. Hence, dynasties that were influenced by the Cistercian mindset also had an earnings advantage through this channel. A more fully articulated theoretical model (but also a much more complex one) would allow both cultural traits to emerge in the wake of Cistercian influence, and grow in pervasiveness over time. We conjecture that such a model would predict that the relationship between π_{it} and population growth reflects preferences for work effort as well as the willingness to postpone consumption. Second, it is probable that the changing cultural values influenced productivity, A. As explained in Section 2.1, the Cistercian order were at the forefront of technological change prior to the Reformation. Insofar as their cultural attitudes spread throughout local populations, it is conceivable that this also encouraged local technology adoption. If so, the level of A may have been influenced by the frequency of individuals valuing hard work and thrift.[15] While our regressions below do involve several controls for A, our estimates for a_2 may nevertheless also be capturing the indirect influence of cultural values on population growth through productivity.

In sum, one should be cognizant of the fact that the cultural values emphasized by the proposed theory represents "ultimate determinants" of prosperity, which served to stimulate key proximate sources of growth: labor effort, capital accumulation and technological change. Realistically, the estimate for a_2 is therefore

[15] Endogenous technological change is not inconsistent with Malthusian population dynamics; see Aiyar, Dalgaard and Moav (2008).

best viewed as the reduced form impact of cultural values on growth mediated by these individual channels.

3.3 Data

3.3.1 Cistercian presence (π_{it})

Researchers at the University College London (UCL) have constructed a database of 776 religious houses in England from the 10th to the 16th century. The database includes the name of the particular religious houses, the order of the monks, nuns etc., year of foundation and dissolution, and the county in which the monastery was located.[16] We gathered these data into one dataset, which we then used to calculate the number of religious houses in each county (*relhouses*) and the number of Cistercian monasteries as a share of total religious houses in each county (*cistercianshare*). In order to gauge robustness, we also construct the share of other major religious orders. We made one correction to the data with respect to the city of York, which was listed by UCL as a county. York was a walled city situated in North Yorkshire. To be able to match the data with the data on population density, we re-coded it as part of the county North Yorkshire. Table 1 lists the frequency distribution of the various religious houses in the UCL database, while Figure 5 maps the spatial distribution of the *cistercianshare*. In the analysis below we focus on the main religious orders: Benedictian monks, Augustinian canons, Cistercian monks, as well as the Premons and Cluniacs.

[Table 1 about here]

3.3.2 Population and population density (L_{it} and L_{it}/X_i)

We obtained data on population density for the year 1377 from Campbell (2008) (*popdens1377*). Campbell also provides the area of the counties; we transformed them from square miles into square kilometers. The distribution of the population in 1377 is based on 1.38 million adult males and females who contributed to the poll tax of 1377.[17] The level of the population is based on an estimate by Campbell (2000) of 4 million.[18] Campbell only reports population numbers for the aggregate of London and Middlesex, not for the two counties separately. In order to match the data, all data on all variables is aggregated in this way. Yet we end up excluding London and Middlesex in all regressions, since it is an outlier. We note for completeness, however, that including London and Middlesex makes no difference to our results.

Wrigley (2007) provides population estimates for 1761, 1771, 1781, 1791, and 1801. These are based on registered marriages, which were more completely recorded than baptisms and burials on which previ-

[16] The data are available online at: <http://www.ucl.ac.uk/history2/englishmonasticarchives/religioushouses/index.php>.
[17] These numbers are available in Dobson (1983).
[18] Campbell (2008) also reports population data for 1300. But since about 10% of Cistercian settlement occur around that year, or after, the risk of reverse causality tainting our estimations would be enhanced if we used 1300 as our initial year. As a result we stick with 1377 as the initial date. However, we will use the 1300 numbers in the context of our IV regressions below.

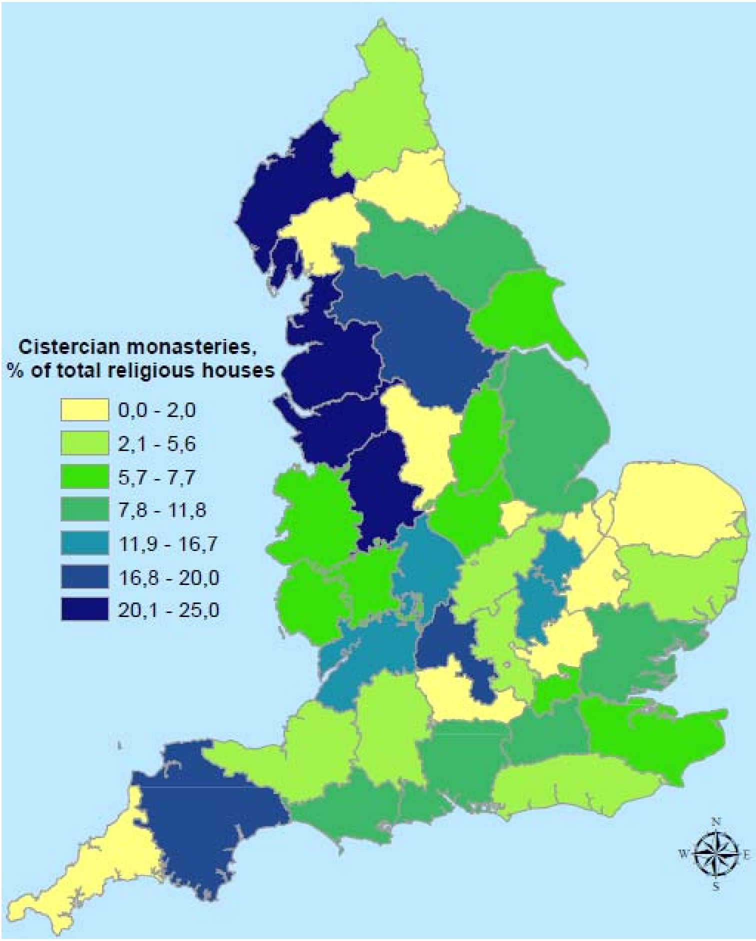

Figure 5: Cistercian monasteries as a share of all religious houses across England, 1098-1540

ous population estimates were based (Rickman 1802). While our preferred variable is population in 1801 (*popdens1801*), the choice of end-year is inconsequential to our results.

3.3.3 Time invariant productivity controls (Z_i)

Agricultural Land Classification Natural England provides a measure of agricultural land classified into five grades plus classifications for non-agricultural and urban land. Grade one is best quality and grade five is poorest quality, grade six is non-agricultural land and grade seven is urban. The measure is calculated by Natural England using information on climate (temperature, rainfall, aspect, exposure, frost risk), site (gradient, micro-relief, flood risk) and soil (depth, structure, texture, chemicals, stoniness). The source of the data is Raster Digital mapping with a scale of 1:250,000.[19] The data was gathered with coordinate precision of 1 meter. We used these data to create a measure of agricultural land quality within each county.

The earliest digital map of English counties is from 1851. These data were kindly provided to us by University of Portsmouth and the Great Britain Historical GIS Project. Combining the shapefile including the agricultural land quality and the shapefile including English county borders, we were able to create measures of the area in a county with agricultural land of quality level 1-5, each as a share of total county area.[20] Our preferred variable is the combination of qualities 1 and 2, which we shall denote *agrquality1_2*.[21]

Waterways As noted in Section 2.1, the Cistercian were strong exponents of water powered production and they employed advanced irrigation techniques, which could be responsible for their influence on English population growth. To control for this kind of influence from Cistercian presence we therefore add controls for waterways.

The German company Geofabrik freely provides shapefiles on various geographic features.[22] Of our interest is their data on waterways in Great Britain, where waterways are divided into canal, dock, drain, moat, river, and stream.[23] As with the data on agricultural land quality, we merge the shapefile describing waterways with the shapefile describing the county borders of England. The outcomes of interest from this procedure is the total length of rivers as a share of the total area in a county (*rivershare*) and the total length of streams as a share of county area (*streamshare*). The variable (*riverstreamshare*) measures the total length of rivers and streams as a share of the county area.

Regional fixed effects In an effort to control more rigorously for structural characteristics with bearing on population growth we add a full set of regional dummy variables. The regional classification is based on Government Office regions: East Midlands, East of England, London, North East, North West, South East, South West, West Midlands, and Yorkshire and the Humber. Observe that in a sample consisting of 40 counties it is a rather strong check of the relevance of Cistercian presence to allow for nine regional identifiers.

Table 2 provides summary statistics and a correlation matrix on the variables discussed above.

[19] Available online at: <http://www.gis.naturalengland.org.uk/pubs/gis/gis_register.asp>. Data description available online at: <http://www.magic.gov.uk/datadoc/metadata.asp?dataset=2&x=16&y=10> and <http://naturalengland.etraderstores.com/NaturalEnglandShop/product.aspx?ProductID=88ff926a-3177-4090-aecb-00e6c9030b29>.

[20] The total county area was here calculated by summing over the land quality variable, since this variable spans the entire area.

[21] None of the results change if we instead include agrquality1 and agrquality2 together or separately. If we include a variable measuring the aggregate agricultural quality over grades 1, 2, and 3, results are unchanged, except column 9 of Table 4 below, where the t-value on *cistercianshare* drops to 1.16.

[22] These shapefiles are based on maps created by the OpenStreetMap project using data from portable GPS devices, aerial photography, other free sources, or simply from local knowledge.

[23] Available online at: <http://download.geofabrik.de/osm/europe/great_britain/>

[Tables 2a and b about here]

3.4 Results

Table 3 reports our baseline results. In all columns of the table we control for initial population density, the total number of religious houses, the share of all religious houses which are Cistercian, and the productivity control agricultural land quality. The regression in column 1 shows that these variables collectively hold significant explanatory power with respect to population growth over the period 1377-1801; the regression explains roughly two thirds of the variation in the dependent variable.

The following eight columns add the additional controls discussed above, one by one. Finally, column 10 includes all the controls simultaneously. Several features of the results are noteworthy. First, the share of Cistercians is statistically significant in all columns. This means that the composition of religious houses matters, with more Cistercians being associated with higher population growth rates. In addition, Cistercian point estimates are fairly stable, always situated in the interval [1.67, 2.07].

Second, while initial population density displays the expected conditional convergence feature, it is surprising that agricultural land quality has a negative impact on population growth. However, the explanation for this is partly found in the high positive correlation between initial population density and agricultural land quality (corr. coef. = 0.43), cf. Table 2b. This means that initial population density picks up some of the effect of land quality on population growth.

Third, land area adds significant explanatory power. Yet the fact that the physical infrastructure of rivers and/or streams did not seem to matter for population growth suggests that neither irrigation nor aqua-based transportation were significant binding constraints to growth.

Fourth, in columns 6-9, where we add the share of the other dominant religious orders, only the Cluniac order adds significant explanatory power. This is in itself an interesting finding since the Cluniacs can be viewed as carriers of the same sort of cultural values embedded in the Order of the Cistercians. The Cluniacs were an earlier attempt (from the tenth century) to return to a more strict observance of the Rule of St. Benedict, although within the Benedictine order of monks (Southern 1970). The significance of the Cluniacs is all the more interesting in light of the fact that Weber highlighted this particular Order alongside the Cistercians as precursors of "Protestant values" (cf. Section 2.1). Accordingly, their influence on pre-industrial development in England represents further support for a pre-Reformation origin of the "Protestant ethic".

Finally, in column 10, where we include all control variables simultaneously, the Cistercian share remains significant and situated in the aforementioned interval. The association between Cistercian presence and population growth is therefore quite robust.

[Table 3 about here]

In Table 4 we add regional fixed effects to all columns of Table 3. To the extent that we have omitted certain time-invariant regional productivity factors, regional fixed effects will alleviate this problem provided the regional classification captures these omitted confounders. One case in point could be proximity to coal production, which Allen (2009) and Pomeranz (2000) argue was critical for British industrialization because it supplied an inexhaustible supply of cheap energy. Since the location of coal mines is a fixed effect, regional dummies will pick up this effect.

The first thing to notice in Table 4 is that the share of Cistercian houses remains significant in all columns save for columns 7 and 10. However, in both columns the regional dummies are jointly insignificant, for which reason these columns add nothing to columns 7 and 10 of Table 3. Consequently, disagreement between Tables 3 and 4 only arise in column 5 with respect to initial population density.

Another way to appreciate the findings reported in Table 4, as compared to those reported in Table 3, is by observing that the point estimate for Cistercians is virtually unaffected by adding regional fixed effects. The occasional change in statistical significance is thus solely due to a reduced precision in estimation, which may well be due to multicollinearity; multicollinearity almost inevitably becomes an issue when we introduce nine regional identifiers, on top of the other controls, in a $N = 40$ sample.

Overall, the results from Table 3 continue to hold up fairly well when regional dummies are added. In particular, only the Cistercians and the Cluniacs seem to exert a significant impact on population growth during the period; it is, however, only the Cistercian impact that survives inclusion of all controls. Finally, it appears that the simple baseline model associated with column 1 of Table 3 is sufficient for purposes of accounting for the association between Cistercian presence and comparative English population growth.

[Table 4 about here]

Figure 6 provides a visual depiction of the relationship between the share of Cistercians and the growth rate in population as estimated by column 1, Table 3. Inspection of the figure shows that Westmorland and Lancashire may exert some leverage on the estimated coefficient. Yet, neither of these two counties is driving the result: exclusion of either one changes nothing. In addition, the share of Cistercians stays significant (slope est. = 1.47 and std. err. = 0.76) when we perform robust regression analysis (more detailed results are available upon request).[24]

What is the economic effect of changing the composition of religious houses in the direction of one more Cistercian abbey, holding the total number of religious houses constant? To answer this question we differentiate (9) with respect to C to get a_2/R. Evaluated at the mean of R, we get that the said change will increase the proportional difference in population by approximately 0.1 log point. This change would have lifted the population of Cambridgeshire in 1801, the median county in terms of population in 1377, from an

[24] This is the **rreg** option in Stata 10.

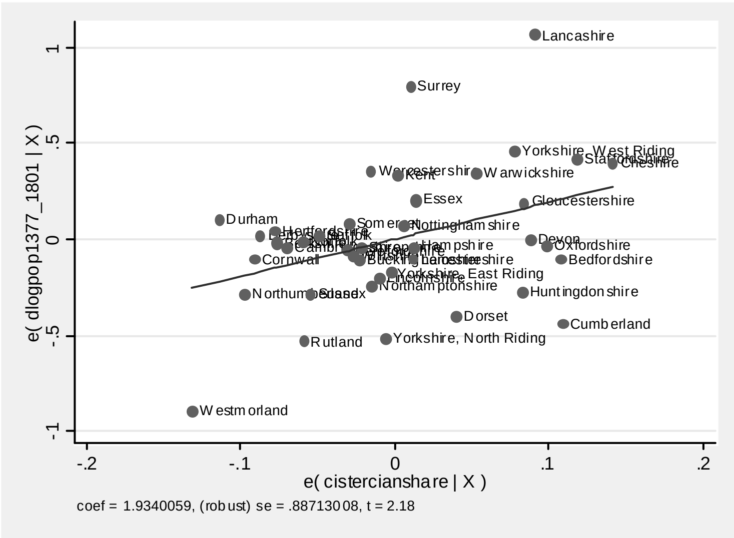

Figure 6: The impact of Cistercians presence on population growth.

actual size 93,440 to a counterfactual size of about 103,000.

3.5 Threats to Identification

3.5.1 Location of Cistercian Monasteries

An objection to the preceding results is that they could be spurious. That is, perhaps the Cistercians simply chose to locate in areas with a pronounced productive potential. Based on the historical evidence, however, this possibility does not seem likely. The Order had a stated preference for situating their monasteries in remote, even devastated, locations (Cooke 1893; Donkin 1963). Indeed, it has long been accepted by scholars that the Cistercians acted as transformers of wastelands into fertile farms, as mirrored in the poet Wordsworth's *Cistercian Monastery.*[25]

This conventional wisdom receives quantitative support in Table 2b, from which it is clear that Cistercian presence was lower in areas featuring high population density in 1377. Also supportive of the traditional view of the spatial distribution of monasteries is the negative correlation between the intensity of Cistercian presence and soil quality. Since we control for both initial population density and soil quality in our regressions, these regularities are unlikely to bias the parameter of interest. To this one may add that since the monasteries were largely in place from the beginning of our period of observation, as observed in Section 2.1, it is impossible that their location is endogenous to population growth during the ensuring centuries on

[25] "Where'er they rise, the sylvan waste retires, And aery harvests crown the fertile lea."

account of reverse causality.

Of course, one may worry that by chance Cistercian monasteries just happened to be located in areas that ultimately proved to be high growth regions. For instance, looking at Figure 5 it is clear that there is a cluster of Cistercian monasteries in the North West of England; areas that long after the Dissolution of Monasteries turned out to be rich on coal (Allen 2009). Yet coal played a relatively modest role *vis-à-vis* the growth process prior to industrialization, which is the time period we examine above. More generally, our regression analysis introduces nine regional dummies, which should ensure that omitted time-invariant confounders do not bias results. In spite of these consideration doubt may linger. As a consequence we provide a final check by invoking instrumental variables estimation.

The Cistercians had a particular preference for locating in secluded and sparsely populated areas, as explained above. At the time of arrival the most secluded areas may well have been the forests owned by the Crown. As Donkin (1963, p. 184) observes: "*..there is a really significant connection with the Royal Forests; one-third of all the English [Cistercian] houses lay at first within or very near their bounds [...]. In these areas there was a good deal of land of low value for endowments; nonroyal landowners were gravely hampered by the forest laws; and, as elsewhere, prospective founders undoubtedly responded to the willingness of the early generations of monks to exploit rough, undeveloped country.*" Thus, there may well have been a double coincidence of wants. Nonroyal landowners, wanting to save their souls from eternal damnation, had an interest in allowing Cistercians to settle at or near Royal Forests, which were of limited value beyond the occasional hunt with the Monarch. At the same time, this location satisfied the ascetic needs of the Cistercian settlers.[26] Hence the presence of a Royal Forest in a county could be a potentially viable instrument for Cistercian settlements. But is the presence of Royal Forest a plausibly excludable instrument for Cistercian presence?

The concept of Royal Forest was introduced in England by the Normans in the 11th century. They were protected areas of land (not necessarily woodland) where the king had privileged hunting rights under the "forest law", which offered strict penalties to anyone using these areas for hunting or farming. The system was at its height in the late 12th and early 13th centuries, but already in 1215 Magna Carta laid down limits to the power of the monarchy in the forests, and the "Great Perambulation" of 1300 vastly reduced the scale of the forests. Generally the system decayed after this time, and Henry VIII placed the forests under the Court of Augmentations in 1547, the body which was set up to administer the land and finances of the Roman Catholic Church after the Dissolution of the Monasteries. Although the designation "Royal Forest" still exists in contemporary England, Royal Forests were considered anachronistic after the Tudor period and the enforcement of any rights had completely died out by the mid-17th century (Grant 1991).

[26] Finally, the monarch may also have had an incentive to encourage the practise. Madden (1963) notes that the king likely granted rights of pasture over wide tracts of the royal lands and forests because the Cistercians were willing to pay for this service using revenue from sale of wool; wool which derived from sheep using the royal lands for grazing.

Towards the end Royal Forests were mainly associated with giving privileged access to timber, but even these rules were poorly enforced. Most of the protection for wooded areas within the Royal Forests was broken with the massive disafforestments of 1327 and the generally less effective means of enforcement (Young 1978, pp. 102-103). Accordingly, since the Royal Forest as an institution was of little importance after the 14th century, it seems plausible that while the settlement pattern of the Cistercians is partially explained by the location of Royal Forests, the location of Royal Forests as such cannot explain cross-county population growth into the nineteenth century.

We obtained data on the location of Royal Forests in the 13th Century from Bazeley (1921). Based on the maps constructed by Bazeley, we constructed a dummy variable, "Rforest", which is equal to one if a royal forest were to be found in the county in the 13th century.[27] Accordingly, we expect to find a positive partial effect of Royal Forest on the intensity of Cistercian settlements.

In order to better explain Cistercian settlements across areas where a Royal Forest was found in the 13th century, we also employ (log) population density in 1300; a point in time where not all monasteries had been founded (see Figure 1). The theory is that Cistercians would prefer to settle near Royal Forests, and even more so if the county in question was sparsely populated. As a result, we expect to find a negative effect from population density in 1300 in the first stage.

We also expect that population density in 1300 is excludable in our regressions. It should be of little relevance to growth in population density between 1377 and 1801 (i.e., above and beyond its influence via Cistercian settlements) since we control for initial population density in 1377 in the regressions below.

In sum, we believe the presence of Royal Forest and county level population density in 1300 both plausibly fulfill the exclusion restriction. Of course, with two instruments and one endogenous variable this prior can be subjected to formal tests, thus offering an opportunity to reject the identifying assumption.

Table 5 reports a summary of our results. In column 1 we estimate a stripped down model were the only independent variables are the intensity of Cistercian presence and initial population density in 1377. The two instruments have the expected sign in the first stage, and the Cistercian influence is estimated with high precision in the second stage. The obtained result is consistent with our OLS findings: more Cistercians seems to foster faster population growth from 1377 to 1801. Moreover, data does not allow us to reject the exclusion restriction.[28]

In the remaining columns we add more controls. In particular, in column 3 we add *all* the controls featured in Table 3 simultaneously; i.e., we estimate the equivalent of column 10 in Table 3. The impact of Cistercians remains significant and positive but the size of the point estimate shrinks. This is no surprise. As explained in Section 2, one may view Cistercian cultural values as an ultimate determinant of productivity,

[27] We also experimented with using forest area in the 13th century, but this instrument turned out to be weak.

[28] In column 1 the instruments are not statistically strong. But the Anderson-Rubin test (not shown), which is robust to weak identification, reveal that the Cistercian share is significant thus suggesting a causal impact.

which influences population growth through several more proximate pathways (e.g., technology, labor effort, and perhaps savings). The more controls we add, the more of these proximate determinants of population density are likely controlled for, which should cause the impact from Cistercian presence to decline. Of course, absent direct measurement of cultural values (or of all the proximate sources of growth), it should not be possible to eliminate the impact from the Order entirely, which is what we find. In columns 2 and 3 our instruments of choice are statisically strong and the exclusion restriction is not rejected by the data.

As a quick comparison between column 3 in Table 5 and column 10 in Table 3 makes clear the estimate rises somewhat when the intensity of Cistercian presence is instrumented. This is consistent with the underlying theory that Cistercian presence is capturing cultural values; if so, then Cistercian presence is a *proxy* variable for the fraction of the population carrying Protestant values, which therefore should be associated with an attenuation bias in the OLS setting.

[Table 5 about here]

Taken together, we believe that these results strengthen the case for a causal link between Cistercian presence and population growth 1377-1801. A greater Cistercian presence worked to promote population growth, consistent with the proposed theory of growth through cultural change. Still a causal impact could also arise for non-cultural reasons, as discussed next.

3.5.2 International Trade

There is an influential strand of literature which asserts that international interaction profoundly influences economic growth. For instance, the work of Frankel and Romer (1999), Alcalá and Ciccone (2004) and Andersen and Dalgaard (2011) suggests that geographic features that facilitate international international interaction (e.g., access to sea, small country size, etc.) hold a significant impact on actual international interaction (either via travel or trade) and ultimately on contemporary prosperity. But why might interaction, in ways of trade for instance, have stimulated growth in the very long run?

One possibility is that the intensity of international trade has influenced the type of policies and institutions that were implemented, and thereby long-run economic outcomes. More concretely, Acemoglu et al. (2005) argue that the Atlantic trade was a key driving force behind the rise of Western Europe after 1500. The argument is that the Atlantic trade enriched and strengthened commercial interests outside the royal circle in countries with non-absolutist initial institutions (such as England); this in turn shifted the balance of power away from the Crown, ultimately instigating significant pro-growth institutional reform.[29]

Naturally, this literature focuses on the development of nation states, whereas we are examining regional

[29] Another potential benefit from trade is that it enables knowledge spillovers between countries. We return to the issue of whether technology can account for our results in the next section.

development. Nevertheless, inspired by the theory of Acemoglu et al., a hypothesis that suggests itself is that counties particularly involved in international trade adopted *local* policies and *local* institutions more beneficial to growth than what was the case elsewhere.

Could such considerations impinge on the apparent link between Cistercians and population growth? It is certainly well known that the Cistercians were active in international trade; they were particularly involved in the trade of wool and cloth. Hence the Cistercians may have contributed to the establishment of *early* trade centers, which would then (partly) account for a persistent effect on *later* population growth. That is, long after the Cistercian presence ended, trade continued at high intensity in the places the Order supported early on. Observe that this theory would still support a causal impact from Cistercians on long-run growth, consistent with our 2SLS estimates above. However, according to this line of reasoning, cultural change is unimportant. In sum, the non-cultural trade story behind our findings would be

$$\text{Cistercians}_t \Rightarrow \text{Trade}_t \Rightarrow \text{Trade}_{t+T} \Rightarrow \text{Population density}_{t+T},$$

where the last arrow could be motivated via policy or institutional changes, following the logic of the Acemoglu et al. (2005) argument. We unfortunately lack data on early international trade flows at the county level. But what we can do instead is to examine whether the influence from Cistercians is reduced once we control for key geographical features that should support international trade; if actual historical trade, much like contemporary international trade, depend on geography then this is a viable proxy variable approach.

Specifically, we add a coastal dummy to our baseline regressions in Table 3. To be sure, if the Cistercians indeed instigated the rise of trade centers, one would expect this effect to be strong precisely along the coast, which is where goods arrive and are shipped off (recall that our regressions already control for inland waterways).

Table 6 presents the results from introducing the coastal dummy. Columns 1-9 reproduce Table 3 while adding on top the coastal dummy. In Column 10 we add all the controls simultaneously (11 in total). The basic message from the table is that the inclusion of the coastal dummy does not affect the partial correlation between the intensity of Cistercian presence and population growth. Hence, although there are some hints at a correlation between Cistercian presence and coastal areas (see Figure 5), and although their strong trading traditions may have left a lasting mark on economic growth, it does not seem that this can account for the reduced form link between the Order and long-run population growth.

[Table 6 about here]

3.5.3 Technology and Human Capital

As discussed in Section 2.1, the Cistercians were mediators of technological change, possibly driven by the desire to free up time for prayer. In addition, the principle that abbots from all Cistercian monasteries had to congregate annually at Cîteaux (the founding monastery) may have implied that new successful innovations (even those not originally due to the Cistercians themselves) were quickly communicated and diffused. While we view technological change as a plausible consequence of the values proliferated by the Cistercians, it cannot be ruled out *a priori* that innovations are in fact the full story. That is, perhaps the legacy left by the Cistercians was technological, and not cultural.

The question is therefore whether the following causal chain, ignoring any cultural effects, may in fact account for our previous findings:

$$\text{Cistercians}_t \Rightarrow \text{Technology}_t \Rightarrow \text{Technology}_{t+T} \Rightarrow \text{Population density}_{t+T}.$$

A theoretical difficulty with this account of our findings above is that the Cistercian impact exhibits such strong persistence. The Monasteries were all dissolved during the fourth decade of the 16th century; nevertheless, we find an effect on population growth reaching as far as the beginning of the 19th century. This requires strong persistence in technology across counties, as reflected in the postulated link between Technology_t and Technology_{t+T}. Recent empirical evidence does in fact suggest that 20th century cross-country technology diffusion may be slow when the intensive margin is considered.[30] However, by any stretch of imagination, it seems hard believe that pre-industrial technological diffusion *within* a country could proceed sufficiently slow so as to account for our results spanning several centuries after the Dissolution of the Monasteries, without some additional source of persistence. Nevertheless, it is worth examining the issue in some detail.

As a first exercise we attempt to control for a key technology that the Cistercians were renowned for diffusing and improving: watermills. If "technology" itself is the main "story", while the cultural channel which we have emphasized all along is largely irrelevant, one would expect the impact of the Cistercian presence to disappear (or at least diminish in a major way) once we control for the intensity of watermills at the county level.

We obtained data on the location of watermills during the 14th century from Campbell and Bartley (2007). Naturally, watermills are seldom preserved across the centuries. Hence, the data originates from records of so-called inquisitions post mortem (IPM) following the passing of lay tenants in chief of the Crown. Specifically, we assign intensity of watermill presence according to the fraction of the IPMs where watermills are mentioned. It should be understood that this is a very crude proxy for the presence of watermills

[30] That is, diffusion appears to be slow when one does not simply consider whether a particular technology is present or not (the extensive margin), but rather focus on the intensity of its use in an area. See Comin et al. (2008).

in local areas; in particular, it excludes watermills belonging to the Church. Nevertheless, if Cistercians were instrumental in increasing the intensity of this important technology prior to the Dissolution of the Monasteries, this variable may still serve as a useful proxy for this technology dimension.

Table 7 shows the effect of introducing watermills. Columns 1-9 reproduce Table 3 with watermill intensity as an added control, whereas column 10 shows the results when all controls, including watermills, are added simultaneously. The recurring theme of the table is that the impact from Cistercian presence is unaffected by the inclusion of the intensity of watermill presence. In fact, comparing the results of Table 7 to those of Table 3 makes clear that the OLS estimate from Cistercian presence is essentially unaltered by the inclusion of watermills. This is fully consistent with the cultural theory proposed above, while hard to reconcile with the notion that technological change in itself accounts for the relationship between Cistercian presence and growth.

[Table 7 about here]

An alternative interpretation of these results is that our measure of technology, watermills, is simply too crude and imprecisely recorded. Hence, in an effort to pursue the matter a bit further, we introduce human capital. The motivation is twofold. First, it seems possible that if the Cistercians managed to instigate technological change during their tenure then this may have increased the return to skill accumulation in the local area. By accounting for the end result of early technological change, i.e. human capital accumulation, we may implicitly be controlling for early technological change more fully than what is admitted by the watermill control. Second, if initial technological change stimulated early human capital accumulation then the latter could be the source of persistence. Perhaps what our results above are suggesting is the following causal chain:

$$\text{Cistercians}_t \Rightarrow \text{Technology}_t \Rightarrow \text{Skill accumulation}_{t+T} \Rightarrow \text{Population density}_{t+T}.$$

That is, perhaps Cistercian presence stimulated early technological change, which led to comparative differences in the speed of skill accumulation in the centuries following the Dissolution of the Monasteries. Insofar as early skill accumulation led to greater earnings it could have worked to elevate population density during the ensuing centuries, prior to the demographic transition. Still, human capital accumulation might also lead to *slower* population growth due to the presence of a quantity-quality trade-off (e.g., Galor and Weil, 2000). Whether the above chain motivates our findings is an empirical matter. Accordingly, by including measures of historical human capital accumulation alongside Cistercian presence we may try to gauge the viability of the cultural change hypothesis.

We obtained data on county-level literacy rates from Hechter (2001). The earliest year from which data on literacy rates is available is 1851. Although this is after the "closing" of our observation window on population growth we hope that 1851 is a sensible proxy for comparative differences across English countries circa 1800 as well. Notice that by introducing the literacy rate at the end of the period we are in effect proxying the rate of human capital accumulation over the preceding centuries, assuming literacy rates were close to zero in 1377. Naturally, one might expect most of this change to have occurred towards the end of the period in question.

Table 8 provides the results from adding the literacy rate. In column 1 we add literacy to our baseline specification which involves (aside from the Cistercian share) the total number of religious houses, initial population density, and land quality. As can readily be seen, the Cistercian share remains significant in spite of the human capital control.

[Table 8 about here]

Human capital enters with a negative sign, suggesting that areas where individuals invested relatively more in their children, at least in terms of basic reading and writing skills, were the areas with the slowest growth in population between 1377 and 1801. This is consistent with an operative quantity-quality trade-off. Naturally, one cannot rule out that the correlation is due to reverse causality. That is, areas with fast population growth may have been areas that later (in 1851) ended up with lower human capital levels. Whatever the right interpretation, the main point is that controlling for human capital does *not* eliminate the significance of the Cistercian share. The basic pattern from the first column is repeated when we add our additional controls one by one, and when we add all controls simultaneously. In every case the Cistercian share remains significant at conventional levels of significance, and the size of the point estimate appears stable. These results are consistent with the theory advanced above, which proposes that the Cistercian influence on Pre-industrial growth in England did not manifest itself through accelerated human capital accumulation but rather via cultural change.

Finally, observe that the results reported in Table 8 also suggest that the impact of Cistercian presence cannot be accounted for by the mechanism featured in Becker and Woessmann (2009): The lasting impact of the Cistercians does not seem to have involved *all* the same proximate sources of growth as Protestantism. Human capital accumulation is unlikely to be part of the story in the present case.

4 Concluding Remarks

The present paper documents that Cistercian monks left a persistent imprint on long-run comparative development across English counties during the pre-industrial era. In counties with greater Cistercian presence population growth was faster during the period 1377-1801, suggesting that the Cistercians stimulated local earnings. The remarkable aspect of this finding is that the Catholic monasteries were dissolved by 1540. Hence the influence from the Order was felt more than 300 years after they had disappeared from England.

These results are robust to a considerable number of controls for productivity, and our IV estimates suggest the correlation can be given a casual interpretation.

We have also offered a potential explanation for these facts, namely that the Cistercians ignited a process of growth through cultural change. That is, a gradual change in local populations in terms of taste for thrift and hard work; much like what Max Weber suggested was the end result of the Protestant Reformation. We believe this theoretical explanation is plausible for two reasons. First, a cultural concordance between the Cistercians and the Protestants, in the dimensions of work ethic and thrift, has already been observed by several scholars including Weber himself. Second, the cultural explanation has the virtue of being able to plausibly account for the long-term persistency of Cistercian influence on growth. Consistent with the cultural mechanism we find, using data from the World Value Survey, that regions in England which historically were influenced relatively more by the Cistercians tend to have populations with greater taste for hard work and thrift today.

Naturally, there are other potentially viable explanations beyond cultural change. For instance, we have examined whether the above facts alternatively can be accounted for by technology, human capital or international trade. While all three channels may be plausible alternatives to the cultural mechanism, we find that none of them are able to account for the Cistercians influence in the data. As a result, we are led to the conclusion that the long term Cistercian impact was most likely caused by a change in cultural values, which stimulated earnings at the local level. Hence, our research suggest that the original Weber thesis, stressing the importance of cultural values like hard work and thrift to economic growth, holds considerable explanatory power with respect to the pre-industrial growth record of England.

A Appendix

A.1 Proof of Lemma 1

(i): Note first that $L_t \lesseqqgtr \overline{L}$ iff $y_t \gtreqqless (1+\omega)\, n\eta$. Consequently, with $L_t \leq \overline{L}$ we have by (4) that $\sigma_t = 1$ in which case (7) gives that $L_{t+1} = nL_t$. With $L_t > \overline{L}$ we have that $\sigma_y = \frac{c_t/n}{\eta}$. Inserting this, the appropriate c^* from (4), and equation (5) into equation (7) then gives $L_{t+1} = \frac{1}{1+\omega}\frac{1}{\eta} A L_t^{\alpha} X^{1-\alpha}$.

(ii): When $L_t \leq \overline{L}$ we have that $L_{t+1} = nL_t$, which cannot cross the 45-degree line in (L_t, L_{t+1})-space since $n > 1$. However, when $L_t > \overline{L}$, $G'(L_t) > 0$, $G''(L) < 0$, and $\lim_{L\to\infty} G'(L) = 0$. This ensures a unique and globally stable steady state. The steady state population, L^*, is found by solving $L^* = \frac{1}{1+\omega}\frac{1}{\eta} A (L^*)^{\alpha} X^{1-\alpha}$ for L^*.

(iii): Steady state income per capita is obtained by inserting steady state population, L^*, and $e^* = \frac{1}{1+\omega}$ into (6). ■

A.2 Proof of Proposition 1

Follows from differentiating the relevant expression from Lemma 1. ■

A.3 Proof of Lemma 2

(i) The law of motion for the first two regimes follows from straightforward application of Lemma 1 to the two groups: ω_1 and ω_2. The law of motion for the final regime, $\bar{L}_2 < L_t$, is obtained as follows. The time path for group size is given by

$$L_{i,t+1} = \frac{1}{1+\omega_i}\frac{1}{\eta} y_t L_{i,t} \qquad i = 1, 2.$$

Using $L_t = L_{1,t} + L_{2,t}$ and the production function, we obtain after some rearrangements

$$L_{t+1} = \left[\frac{\omega_1 - \omega_2}{(1+\omega_1)(1+\omega_2)}\frac{L_{2,t}}{L_t} + \frac{1}{1+\omega_1}\right]\frac{1}{\eta} A L_t^{\alpha} X^{1-\alpha}. \tag{10}$$

Observe now that $L_{2,t}/L_t = (1 + L_{1,t}/L_{2,t})^{-1}$. Since the laws of motions for the individual groups are symmetrical, save for the value of ω, we have that $L_{1,t+1}/L_{2,t+1} = [(1+\omega_2)/(1+\omega_1)]\, L_{1,t}/L_{2,t}$. Solving this difference equation yields $L_{1,t}/L_{2,t} = \left(\frac{1+\omega_2}{1+\omega_1}\right)^t L_{1,0}/L_{2,0}$. Substituting this solution into the law of motion for L, yields the expression stated in the Lemma.

(ii) Let group i grows according to

$$L_{it+1} = G_i L_{it} \tag{11}$$

where G_i, $i = 1, 2$ equals n or $\frac{1}{1+\omega_i}\frac{1}{\eta} y_t$ depending on L_t.

The law of motion for total population is given as $L_{1t+1} + L_{2t+1} = G_1 L_{1t} + G_2 L_{2t}$, and so

$$\frac{L_{t+1}}{L_t} = G_1\left(1 - \frac{L_{2t}}{L_t}\right) + G_2\frac{L_{2t}}{L_t} \Leftrightarrow \frac{L_{t+1}}{L_t} = G_1 + (G_2 - G_1)\frac{L_{2t}}{L_t}. \tag{12}$$

Note also that $\frac{L_{2t}}{L_t} = \left(1 + \frac{L_{1t}}{L_{2t}}\right)^{-1} \equiv \frac{1}{1+z_t}$. We can therefore write (12) as

$$\frac{L_{t+1}}{L_t} = G_1 + \frac{G_2 - G_1}{1 + z_t}. \tag{13}$$

Finally, note that by (11) we have

$$z_{t+1} = \frac{G_1}{G_2} z_t, \tag{14}$$

where we recall that G_1 and G_2 are functions of L_t.

In this setting, a *steady state* is a pair (L^*, z^*) such that $0 < L^* = L_{t+1} = L_t$ and $z_{t+1} = z_t = z^*$, with $0 \leq z^* \leq 1$, and where (L^*, z^*) fulfills equations (13), (14).

We need to consider existence of a steady state in the three regimes stated in the lemma.

Case I: $L < \bar{L}_1$, where (it is recalled) $\bar{L}_1$ is defined as $y(L_1) = (1 + \omega_1) n\eta$. When $0 < L < \bar{L}_1$, $G_1 = n$ and $G_2 = n$. Hence, relative group size (and thus $\frac{L_{2t}}{L_t}$) is constant. But the aggregate population is rising, since (inserting into (13)) $\frac{L_{t+1}}{L_t} = n > 1$. Accordingly, for $L < \bar{L}_1$ there does not exist a steady state with $L_{t+1} = L_t = L^*$ and L^* fulfilling (13).

Case II. $0 < \bar{L}_1 \leq L_t < \bar{L}_2$, with $\bar{L}_2$ is defined as $y\left(\bar{L}_2\right) = (1 + \omega_2) n\eta$. Note that $\bar{L}_2 > \bar{L}_1$ since $\omega_2 < \omega_1$. Relative group size in this interval.

$$z_{t+1} = \frac{G_1}{G_2} = \frac{\frac{1}{1+\omega_1}\frac{1}{\eta}y_t}{n} z_t \tag{15}$$

while aggregate growth is given by (13).

Working towards a contradiction, assume that a steady state exist. This requires (from (13)) that

$$1 = G_1(L^*) + (G_2(L^*) - G_1(L^*))\frac{1}{1 + z^*}.$$

or, since $\bar{L}_1 \leq L_t < \bar{L}_2$,

$$\frac{1 - \frac{1}{1+\omega_1}\frac{1}{\eta}y(L^*)}{n - \frac{1}{1+\omega_1}\frac{1}{\eta}y(L^*)} = \frac{1}{1 + z^*}. \tag{16}$$

That is, existence requires constancy of relative group size z^*. But this demands, from (15), that

$$n = \frac{1}{1 + \omega_1}\frac{1}{\eta}y(L^*),$$

which contradicts $0 \leq z^* \leq 1$, as seen from (16). Hence, a steady state cannot exist in the interval $0 < \bar{L}_1 \leq L_t < \bar{L}_2$.

Case III. $\bar{L}_2 < L_t$. Observe that $G_t' > 0$ for $L_{1,0}$, $L_{2,0} > 0$. This follows from differentiation and noting that $\omega_2 < \omega_1$. Next, note that $\lim_{t\to\infty} \left(1 + \left(\frac{1+\omega_2}{1+\omega_1}\right)^t \frac{L_{1,0}}{L_{2,0}}\right)^{-1} = 1$, implying $\lim_{t\to\infty} G = \frac{1}{1+\omega_2}\frac{1}{\eta}AL_t^{\alpha}X^{1-\alpha}$. Given this result and continuity of $G(L_t, t\ )$, the same considerations as those laid out in the proof of Lemma 1 leads to existence and uniqueness of the steady state.■

A.4 Proof of Proposition 2

Before cultural mutation occurs the law of motion for population size is given by Lemma 1, with $\omega = \omega_1$. After the change, the law of motion is given in Lemma 2. Since $\omega_2 < \omega_1$ Proposition 2 follows from Proposition 1 and Lemma 2.■

References

[1] Aiyar, S., C.-J. Dalgaard and O. Moav, 2008. Technological Progress and Regress in Pre-Industrial Times. *Journal of Economic Growth*, 13, 125-44

[2] Alcalá, F. and A. Ciccone, 2004. Trade and Productivity. *Quarterly Journal of Economics,* 119, 613-646.

[3] Allen, R., 2009. *The British Industrial Revolution in Global Perspective.* Cambridge University Press

[4] Andersen, T.B. and C-J. Dalgaard, 2011. Flows of People, Flows of Ideas, and the Inequality of Nations. *Journal of Economic Growth*, forthcoming

[5] Angrist, J. and J.-S. Pischke, 2009. *Mostly Harmless Econometrics.* Princeton University Press

[6] Ashraf, Q. and O. Galor, 2009. Dynamics and Stagnation in the Malthusian Epoch: Theory and Evidence. Working Paper (Brown University). *American Economic Review* (forthcoming)

[7] Barro, R. and R. McCleary, 2003. Religion and Economic Growth across Countries. *American Sociological Review*, 68, 760-781

[8] Baumol, W., 1990. Entrepreneurship: Productive, Unproductive, and Destructive. *Journal of Political Economy*, 98, 893-921

[9] Bazeley, M.L., 1921. The Extent of the English Forest in the Thirteenth Century. Transactions of the Royal Historical Society, 4 , pp. 140-172.

[10] Becker, G.S., 1973. A Theory of Marriage: Part I. *Journal of Political Economy* 81, p.881-46.

[11] Becker, R.A., 1980. On the long-run steady state in a simple dynamic model of equilibrium with heterogeneous households. *Quarterly Journal of Economics*, 95, 375–82

[12] Becker, S. and L. Woessmann, 2009. Was Weber Wrong? A Human Capital Theory of Protestant Economic History. *Quarterly Journal of Economics*, 124, 531-596

[13] Berman, C.H., 2000, *The Cistercian Evolution.* Penn Press

[14] Bloch, M., 1935. *Avènement et conquêtes du moulin a eau. Annales d'Histoire Èconomique et Sociale,* 7, 538–563

[15] Bouchard, C., 1991. *Holy Entrepreneurs.* Cornell University Press

[16] Cameron, C. and P. Trivedi, 2005. *Microeconometrics.* Cambridge University Press

[17] Campbell, B., 2000. *English seigniorial agriculture, 1250-1450.* Cambridge University Press

[18] Campbell, B., 2008. Benchmarking medieval economic development: England, Wales, Scotland and Ireland, c. 1290. *Economic History Review*, 61, 896-945

[19] Campbell B. and K. Bartley, 2007. *England on the Eve of the Black Death: An Atlas of Lay Lordship, Land and Wealth, 1300-49.* Manchester University Press

[20] Cantoni, D., 2009. The Economic Effects of the Protestant Reformation: Testing the Weber Hypothesis in the German Lands. Manuscript (Harvard University)

[21] Cavalcanti, T.V., Parente, S.L., and R. Zhao, 2007. Religion in macroeconomics: a quantitative analysis of Weber's thesis. *Economic Theory*, 32, 105–123

[22] Clark, P., 1979. Migration in England during the Late Seventeenth and Early Eighteenth Centuries. *Past & Present*, 83, 57-90

[23] Clark, G., 2007. *A Farewell to Alms*. Princeton University Press

[24] Clark, G. and Y. Van Der Werf, 1998. Work in progress? the industrious revolution, *Journal of Economic History*, 58, 830–43

[25] Collins, R., 1986. *Weberian Sociological Theory*. Cambridge University Press

[26] Comin, D. and Hobijn, B. and Rovito, E., 2008. Technology usage lags. *Journal of Economic Growth*, 13, 237–256

[27] Cook, H., Stearne, K., and T. Williamson, 2003. The origins of water meadows in England. *Agricultural History Review*, 51, 155-162

[28] Cooke, A., 1893. The Settlement of the Cistercians in England. *English Historical Review*, 8, 625-676

[29] Dobson, R., 1983. *The Peasants' Revolt*. Palgrave Macmillan

[30] Doepke, M. and F. Zilibotti, 2008. Occupational choice and the spirit of capitalism. *Quarterly Journal of Economics*, 747-93

[31] Donkin, R., 1963. The Cistercian Order in Medieval England: Some Conclusions. *Transactions and Papers (Institute of British Geographers)*, 33, 181–198

[32] Donkin, R., 1978. *The Cistercians: Studies in the geography of medieval England and Wales*. Toronto: Pontifical Institute of Medieval Studies

[33] Galor, O. and D. Weil, 2000. Population, technology, and growth: From Malthusian stagnation to the demographic transition and beyond. *American Economic Review*, 90, 806–828.

[34] Galor, O. and O. Moav, 2002. Natural Selection and the Origin of Economic Growth. *Quarterly Journal of Economics*, 117, 1133–91

[35] Gimpel, J., 1976. *The medieval machine*. Penguin

[36] Grant, R. (1991). *The Royal Forests of England*. Stroud, UK: Alan Sutton Publishing

[37] Guiso, L., Sapienza, P., and L. Zingales, 2008. Long Term Persistence. EUI Working Paper 2008/30

[38] Hill, B.D., 1968. *English Cistercian monasteries and their patrons in the twelfth century*. University of Illinois Press

[39] Kaelber L., 1998. *Schools of Asceticism: Ideology and Organization in Medieval Religious Communities*. Penn State Press

[40] Kelly, M. and C. O'Grada, 2008. Living Standards and Mortality in England since the Middle Ages: The Poor Law versus the Positive Check. Manuscript (University College Dublin)

[41] Kieser, A., 1987. From Asceticism to Administration of Wealth. Medieval Monasteries and the Pitfalls of Rationalization. *Organization Studies*, 8, 103-123

[42] Landes, D., 1999. The Wealth and Poverty of Nations. Abacus

[43] Madden, J.E. (1963). Business Monks, Banker Monks, Bankrupt Monks: The English Cistercians in the Thirteenth Century. *The Catholic Historical Review*, Vol. 49, No. 3, pp. 341-364

[44] McCleary, R. and R. Barro, 2006. Religion and Economy. *Journal of Economic Perspectives*, 20, 49-72

[45] Møller, N. and P. Sharp, 2009. Malthus in Cointegration Space: A new look at living standards and population in pre-industrial England. Discussion Paper No. 08-16, Department of Economics, University of Copenhagen

[46] Nicholas, S. and P.R. Shergold, 1987, Intercounty labour Mobility during the Industrial Revolution: Evidence from Australian Transportation Records. *Oxford Economic Papers*, 39, 624-40

[47] Noell, B., 2006. Expectation and unrest among Cistercian lay brothers in the twelfth and thirteenth centuries. *Journal of Medieval History*, 32, 253-274

[48] Nunn, N. and L. Wantchekon, 2009. The Slave Trade and the Origins of Mistrust in Africa. Manuscript (Harvard University)

[49] Ovitt Jr, G., 1986. The Cultural Context of Western Technology: Early Christian Attitudes toward Manual Labor. *Technology and Culture*, 27, 477–500

[50] Ovitt Jr, G., 1987. *The Restoration of Perfection: Labor and Technology in Medieval Culture*. Rutgers University Press

[51] Pomeranz, K., 2000. *The Great Divergence: China, Europe, and the Making of the Modern World Economy*. Princeton University Press

[52] Southern, R.W., 1970. *Western Society and the Church in the Middle Ages*. Penguin

[53] Tabellini, G., 2010. Culture and institutions: economic development in the regions of Europe. *Journal of the European Economic Association,* 8, 677-716

[54] Young, C.R. (1978). 'Conservation Policies in the Royal Forests of Medieval England'. *Albion: A Quarterly Journal Concerned with British Studies*, 10, 95-103

[55] Weber, M., 1905. *The Protestant Ethic and the Spirit of Capitalism.* Allen & Unwin

[56] Weber, M., 1958. *The Protestant Ethic and the Spirit of Capitalism.* Charles Scribner's Sons

[57] Williams, D.H., 1970. *The Welsh Cistercians: Aspects of Their Economic History.* The Griffin Press

[58] Wrigley, E., 2007. English county populations in the later eighteenth century. *Economic History Review*, 60, 35-69

Table 1: Frequency distribution of monastic orders	Number	Share of total
Benedictine monks	239	30.8
Augustinian canons	208	26.8
Benedictine nuns	77	9.9
Cistercian monks	70	9.0
Premonstratensian canons	37	4.8
Cluniac monks	34	4.4
Cistercian nuns	28	3.6
Augustinian canonesses	24	3.1
Gilbertine canons	15	1.9
Trinitarian brothers	10	1.3
Gilbertine canons & nuns	9	1.2
Carthusian monks	8	1.0
Fontevraud nuns	3	0.4
Grandmontine monks	3	0.4
Premonstratensian canonesses	3	0.4
Bonhommes brothers	2	0.3
Cluniac nuns	2	0.3
Brigettine nuns & brothers	1	0.1
Fontevraud monks	1	0.1
Gilbertine nuns	1	0.1
unknown monks or brothers	1	0.1
Total	776	100.0

Table 2a: Summary statistics	**Obs**	**Mean**	**Std.**	**Min**	**Max**
cistercianshare	40	0.09	0.07	0.00	0.25
relhouses	40	19.03	12.93	2.00	73.00
popdens1377	40	31.55	11.83	8.98	52.98
popdens1801	40	60.45	24.82	20.92	143.77
agrquality1_2	40	0.18	0.16	0.01	0.73
riverstream	40	0.41	0.29	0.06	1.09
rivershare	40	0.12	0.03	0.03	0.18
streamshare	40	0.30	0.27	0.01	0.91
area	40	3256.94	1623.37	392.20	7423.75

Table 2b: Correlation matrix								
cistercianshare	1							
relhouses	0.0134	1						
popdens1377	-0.2791	0.3091	1					
popdens1801	0.3161	-0.1644	-0.0579	1				
agrquality1_2	-0.1234	0.2460	0.4287	-0.2149	1			
riverstream	0.0127	-0.1391	-0.6152	-0.1460	-0.2602	1		
rivershare	0.1589	-0.1271	-0.4816	-0.0063	-0.2092	0.6345	1	
streamshare	-0.0044	-0.1341	-0.6025	-0.1553	-0.2542	0.9962	0.5644	1
area	0.2625	0.6603	-0.2802	-0.0381	-0.0865	0.3191	0.2011	0.3181

Table 3: OLS estimation

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
					Dependent variable : dlogpop1377_1801					
logpopdens1377	-0.614***	-0.694***	-0.606***	-0.700***	-0.438**	-0.715***	-0.615***	-0.635***	-0.674***	-0.729***
	(0.171)	(0.207)	(0.186)	(0.206)	(0.173)	(0.172)	(0.169)	(0.155)	(0.160)	(0.249)
cistercianshare	1.934**	1.802**	1.931**	1.779**	1.751**	1.911**	1.982**	2.074**	1.666*	1.773*
	(0.887)	(0.825)	(0.916)	(0.827)	(0.798)	(0.847)	(0.927)	(0.876)	(0.852)	(0.968)
relhouses	-0.007*	-0.007+	-0.007*	-0.006+	-0.017***	-0.005	-0.007*	-0.008**	-0.006*	-0.014**
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.003)	(0.004)	(0.005)
agrquality1_2	-0.634*	-0.633*	-0.634*	-0.632*	-0.536*	-0.543*	-0.631*	-0.488+	-0.675**	-0.279
	(0.313)	(0.321)	(0.315)	(0.320)	(0.297)	(0.288)	(0.326)	(0.301)	(0.326)	(0.324)
riverstream		-0.180								
		(0.233)								
rivershare			0.267							1.481
			(1.940)							(1.987)
streamshare				-0.208						-0.473
				(0.247)						(0.293)
logarea					0.303**					0.280
					(0.113)					(0.176)
augustinianshare						0.771				0.970
						(0.541)				(0.589)
benedictineshare							0.146			0.629
							(0.383)			(0.479)
cluniacshare								2.465**		1.848
								(1.159)		(1.486)
premonshare									-0.674	0.245
									(0.847)	(1.088)
Constant	2.799***	3.148***	2.742***	3.158***	-0.011	2.878***	2.754***	2.744***	3.060***	0.472
	(0.587)	(0.742)	(0.735)	(0.724)	(1.257)	(0.548)	(0.631)	(0.542)	(0.529)	(2.08)
Observations	40	40	40	40	40	40	40	40	40	40
R-squared	0.641	0.646	0.641	0.647	0.676	0.667	0.642	0.683	0.650	0.756

Notes. Robust standard errors in parentheses. Symbols ***, **, * indicate significance at the 1, 5, and 10%, respectively. In column 10, riverstream is omitted due to perfect multicollinearity.

Table 4: OLS estimation

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
					Dependent variable : dlogpop1377_1801					
logpopdens1377	-0.482**	-0.606**	-0.460*	-0.619**	-0.124	-0.521**	-0.469*	-0.287	-0.519*	-0.359
	(0.235)	(0.290)	(0.267)	(0.287)	(0.266)	(0.233)	(0.232)	(0.248)	(0.282)	(0.339)
cistercianshare	1.977*	1.849*	1.975*	1.823*	1.810**	1.662*	1.947+	2.214**	1.659*	1.741
	(1.078)	(1.039)	(1.112)	(1.044)	(0.825)	(0.859)	(1.176)	(0.958)	(0.876)	(1.303)
relhouses	-0.006*	-0.006+	-0.006+	-0.006+	-0.023***	-0.003	-0.006*	-0.007**	-0.005	-0.015*
	(0.004)	(0.004)	(0.004)	(0.004)	(0.006)	(0.003)	(0.004)	(0.003)	(0.004)	(0.009)
agrquality1_2	-0.682*	-0.640*	-0.671*	-0.626+	-0.694**	-0.519	-0.665*	-0.471	-0.717*	-0.309
	(0.350)	(0.366)	(0.352)	(0.368)	(0.299)	(0.390)	(0.340)	(0.426)	(0.361)	(0.586)
riverstream		-0.158								
		(0.286)								
rivershare			0.362							1.294
			(2.128)							(2.204)
streamshare				-0.188						-0.329
				(0.309)						(0.278)
logarea					0.518***					0.328
					(0.182)					(0.237)
augustinianshare						1.137*				1.056
						(0.564)				(0.835)
benedictineshare							-0.154			0.475
							(0.487)			(0.649)
cluniacshare								3.238*		1.853
								(1.653)		(1.861)
premonshare									-0.524	0.342
									(1.374)	(1.665)
Constant	2.295**	2.787**	2.169*	2.815**	-2.649	1.986*	2.288**	1.372	2.462**	-1.274
	(0.971)	(1.153)	(1.200)	(1.126)	(2.151)	(0.977)	(0.980)	(1.072)	(1.152)	(2.709)
Regional dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
F test, H0: Reg. dummies = 0 (p-val)	0.487	0.658	0.510	0.668	0.070	0.374	0.558	0.354	0.616	0.562
Observations	40	40	40	40	40	40	40	40	40	40
R-squared	0.678	0.681	0.679	0.682	0.750	0.723	0.680	0.737	0.682	0.799

Notes. Robust standard errors in parentheses. Symbols ***, **, *, + indicate significance at the 1, 5, 10, and 15%, respectively. In column 10, riverstream is omitted due to perfect multicollinearity.

Table 5: IV (Limited Information Maximum Likelihood) estimation

	IV		
	(1)	(2)	(3)
	Dependent variable : dlogpop1377_1801		
Second stage			
cistercianshare	4.77**	3.37**	2.57**
	-2.07	(1.54)	(1.19)
	Dependent variable: cistercianshare		
First stage			
Rforest	0.08**	0.10***	0.07**
	(0.04)	(0.02)	(0.03)
logpopdens1300	- 0.14**	- 0.17**	- 0.20***
	(0.07)	(0.06)	(0.07)
First stage F-static	4.81	15.68	7.49
Hansen J-stat (p-value)	0.55	0.46	0.75
Additional Controls	logpopdens1377	All Baseline	All controls
Observations	40	40	40

Notes. Robust standard errors in parentheses. Symbols ***, **, * indicate significance at the 1, 5, and 10%, respectively. All regressions include a constant. "Baseline controls" is controls from Table 3, Column 1: logpopdens1377, religious houses and agricultural land quality. "All controls" involve all the controls featured in Table 3: logpopdens1377, religious houses (total), agricultural land quality, rivershare, streamshare, logarea, augustinianshare, clunicshare, and premonshare. To avoid multicollinarity river and stream shares are controled for seperately, while the combined variable "riverstream" is ignored.

Table 6: OLS estimation ***Trade***

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
					Dependent variable : dlogpop1377_1801					
cistercianshare	1.92**	1.77**	1.92**	1.74**	1.77**	1.90**	1.97**	2.08**	1.66*	2.02*
	(0.88)	(0.80)	(0.91)	(0.80)	(0.82)	(0.85)	(0.92)	(0.89)	(0.87)	(1.05)
coastal dummy	0.03	0.03	0.03	0.05	-0.11	0.01	0.01	-0.01	-0.01	-0.25*
	(0.13)	(0.14)	(0.13)	(0.14)	(0.13)	(0.11)	(0.12)	(0.13)	(0.13)	(0.14)
Controls	All Baseline	All Baseline + riverstream	All Baseline + rivershare	All Baseline + streamshare	All Baseline + logarea	All Baseline + augustinianshare	All Baseline + benedictineshare	All Baseline + cluniacshare	All Baseline + premonshare	All controls
Observations	40	40	40	40	40	40	40	40	40	40
R-squared	0.64	0.65	0.64	0.65	0.68	0.67	0.64	0.68	0.65	0.78

Notes. Robust standard errors in parentheses. Symbols ***, **, * indicate significance at the 1, 5, and 10%, respectively. All regressions include a constant. "Baseline controls" are the controls from Table 3, Column 1: logpopdens1377, religious houses, and agricultural land quality. "All controls" means all the controls featured in Table 3: logpopdens1377, religious houses, agricultural land quality, rivershare, streamshare, logarea, augstinianshare, benedictineshare, cluniacshare, and premonshare.

Table 7: OLS estimation					***Technology***					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	[10]
	Dependent variable : dlogpop1377_1801									
cistercianshare	1.92**	1.78**	1.92**	1.76**	1.75**	1.90**	1.98**	2.07**	1.58*	1.79*
	(0.89)	(0.83)	(0.91)	(0.83)	(0.82)	(0.85)	(0.91)	(0.90)	(0.82)	(1.00)
watermills	0.22	0.24	0.22	0.22	0.26	0.14	0.30	0.05	0.39	-0.18
	(0.64)	(0.83)	(0.67)	(0.65)	(0.66)	(65)	(0.70)	(0.61)	(0.64)	(0.60)
Controls	All Baseline	All Baseline + riverstream	All Baseline + rivershare	All Baseline + streamshare	All Baseline + logarea	All Baseline + augustinianshare	All Baseline + Benedictineshare	All Baseline + cluniacshare	All Baseline + premonshare	All controls
Observations	40	40	40	40	40	40	40	40	40	40
R-squared	0.64	0.65	0.64	0.65	0.68	0.67	0.65	0.68	0.66	0.76

Notes. Robust standard errors in parentheses. Symbols ***, **, * indicate significance at the 1, 5, and 10%, respectively. All regressions include a constant. "Baseline controls" are the controls from Table 3, Column 1: logpopdens1377, religious houses and agricultural land quality. "All controls" means all the controls featured in Table 3: logpopdens1377, religious houses and agricultural land quality, rivershare, streamshare, logarea, augstinianshare, benedictineshare, clunyshare, and premonshare.

Table 8: OLS estimation ***Human capital***

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
						Dependent variable : dlogpop1377_1801				
cistercianshare	1.64**	1.58**	1.64**	1.58*	1.53**	1.61**	1.64*	1.79**	1.57*	1.76**
	(0.79)	(0.79)	(0.79)	(0.79)	(0.70)	(0.76)	(0.84)	(0.83)	(0.79)	(0.82)
literacy1851	- 1.93**	- 1.86**	- 1.93**	- 1.86**	- 1.94**	- 1.96**	- 1.93**	- 1.64*	-1.84*	- 1.69*
	(0.91)	(0.88)	(0.94)	(0.88)	(0.80)	(0.93)	(0.94)	(0.83)	(1.07)	(0.97)
Controls	All Baseline	All Baseline + riverstream	All Baseline + rivershare	All Baseline + streamshare	All Baseline + logarea	All Baseline + augustinianshare	All Baseline + benedictineshare	All Baseline + cluniacshare	All Baseline + premonshare	All controls
Observations	37	37	37	37	37	37	37	37	37	37
R-squared	0.69	0.69	0.69	0.69	0.72	0.71	0.69	0.70	0.69	0.77

Notes. Robust standard errors in parentheses. Symbols ***, **, * indicate significance at the 1, 5, and 10%, respectively. All regressions include a constant. "Baseline controls" are the controls from Table 3, Column 1: logpopdens1377, religious houses, and agricultural land quality. "All controls" means all the controls featured in Table 3: logpopdens1377, religious houses, agricultural land quality, rivershare, streamshare, logarea, augstinianshare, benedictineshare, cluniacshare, and premonshare.