Gutenberg School of Management and Economics

& Research Unit “Interdisciplinary Public Policy”

Discussion Paper Series

Can’t See the Forest for the IVs

Re-examining the Cistercian “Pre-reformation Roots of the Protestant Ethic”

Nico Sonntag

October 6, 2023

Discussion paper number 2316

Johannes Gutenberg University Mainz
Gutenberg School of Management and Economics
Jakob-Welder-Weg 9
55128 Mainz
Germany
https://wiwi.uni-mainz.de/

Contact details

Nico Sonntag
Chair of Sociology and quantitative methods
Johannes-Gutenberg-Universität
Jakob-Welder-Weg 12
55128 Mainz
Germany
nsonntag@uni-mainz.de

Can't See the Forest for the IVs*

Re-examining the Cistercian "Pre-reformation Roots of the Protestant Ethic"

Nico Sonntag†

October 6, 2023

Abstract

I re-examine the claim made by Andersen, Bentzen, Dalgaard, and Sharp that the work ethic of the Cistercian order instigated the kind of cultural change attributed to Protestantism by Max Weber. Following a critical discussion of their historical and theoretical arguments, as well as an assessment of the original study design, I reconsider and expand upon their analyses of the positive associations between past Cistercian presence and early modern economic development as well as contemporary values. Theories about the historical origins of economic development can often only be tested indirectly. Moreover, the theories are often insufficient to deduce the precise specification of statistical models or to choose among competing ways to measure a theoretical construct with available historical data. For this reason, I conduct a systematic robustness check that takes into account a wide range of plausible model specifications. While the correlation between Cistercians and population growth remains robust, all models attempting to identify a causal effect either rely on specific and hard-to-justify choices concerning the operationalization of central constructs or fail to provide strong confirmatory evidence. Furthermore, additional analyses investigating the mediation effect of contemporary value orientations on economic indicators contradict the proposed mechanism. The text concludes by offering recommendations on how to systematically study the cultural and economic impact of Christian orders.

Keywords: Cultural values; Protestant ethic; Economic development

JEL Codes: N13; O11; Z12

*I thank Clara Englert for help with preparing the manuscript. Graham Jones and Dave Balzer provided valuable comments and references during the research process. Participants of the "Causality in the Social Sciences" workshop in Leipzig, the congress "Digital Socieites" of the Akademie für Soziologie in Konstanz, the colloquium of the Institute for Sociology in Wuppertal, and Mark Ebers' research colloquium at the University of Cologne helped to develop the project through critical discussions and helpful suggestions. Data and syntax files are available online: `https://osf.io/cdtn9/`.

†Nico Sonntag, Institute of Sociology, Johannes Gutenberg University Mainz, Germany, Jakob-Welder-Weg 12, D-55128 Mainz. Email: sonntag@uni-mainz.de

1. Introduction

Max Weber's Protestantism thesis (Weber, 2005) is very likely the most famous sociological contribution to the literature on economic development. Economists, and economic historians in particular, increasingly pay attention to the possible influence of cultural and religious factors on the Industrial Revolution as well as economic development in general. Accordingly, there have been a number of attempts to test Weber's suggestion that Protestant asceticism played a role in the emergence of early capitalism (Becker et al., 2016: 3–4, 17–18).

In a recent publication, Andersen, Bentzen, Dalgaard and Sharp (hereafter ABDS) set forth an alternative view on the cultural origin of capitalist work ethics. They hypothesize "that the cultural virtues emphasized by Weber had a pre-Reformation origin in the Order of Cistercians" and that the latter "encouraged growth by instigating the kind of cultural change that Weber attributed to Protestantism" (Andersen et al., 2017a: 1757). The empirical test of their hypothesis combines several pieces of evidence. 1) ABDS show a statistical association between the locations of medieval Cistercian monasteries and population growth of English regions during the early modern period (1377–1801). 2) Furthermore, an instrumental variable approach provides evidence that this is a causal relationship. 3) Additionally, they can show contemporary value orientations regarding hard work and thrift to be correlated with the intensity of historical Cistercian presence throughout European regions. 4) Lastly, historical Cistercian presence in Europe is associated with contemporary variation in regional employment as well.

ABDS's study is remarkable for their hypothesis regarding the wide-spread diffusion of Cistercian values is not just altogether new; the findings, moreover, suggest a rather strong effect that persists today. If we take the study at face value, it is more than a mere footnote in the history of economic development but gives rise to an important research program on the question of how Catholic orders shaped the value orientations of European populaces. Thus, it is not surprising that the paper is well-cited and has already been picked up in books targeted at a popular audience (Henrich, 2020).

Unfortunately, there are reasons to doubt the story as told by ABDS. Just consider Figure 1 which is based on the Figure 3 by Andersen et al. (2017a: 1772).[1] It is descriptive but nonetheless summarizes the main finding with regard to England well. ABDS take population growth as an indicator for unobserved productivity growth based on the assumption that more productive regions were able to sustain a greater population. A change in the "work ethic," therefore, affects productivity which, in turn, affects population growth. As can be seen, however, differential growth is only observable for the 1600–1801

[1] I was not able to replicate the exact numbers using the data and do-file provided in the journal's supporting information (see lines 5–8 in "uk_county.do" in supplement Data S1). Since all other results are easily replicable using the material provided by the authors, I suspect the figure is based on an earlier version of the dataset.

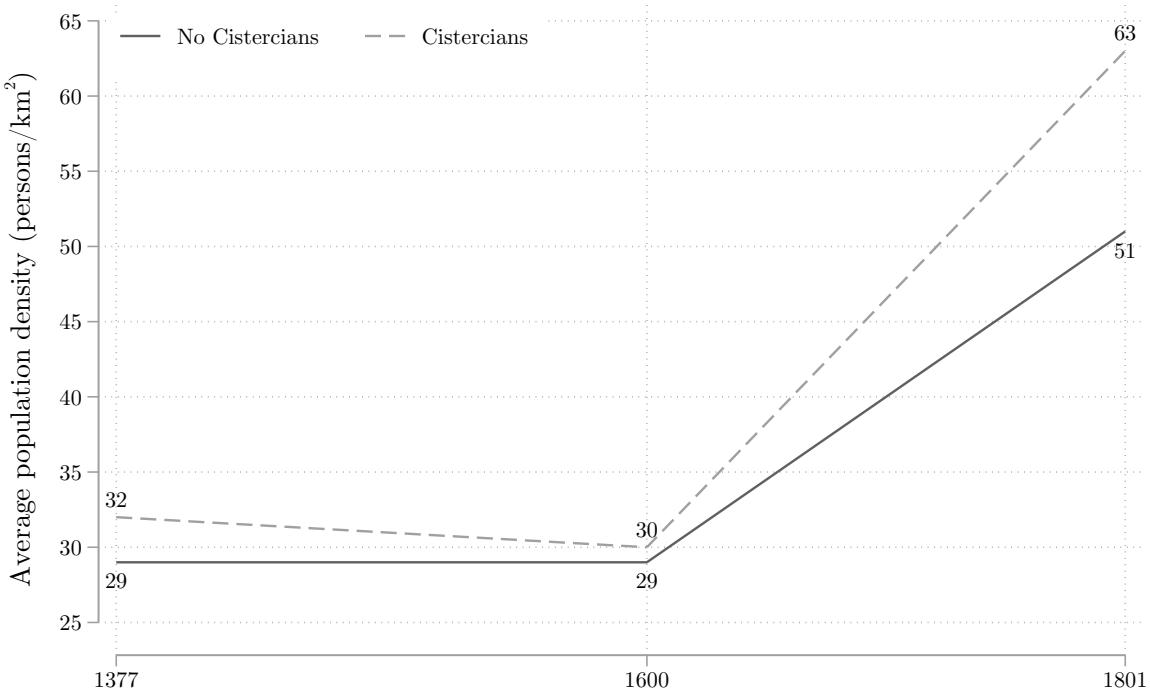


Figure 1: Change in average population density over time. Average population density (Persons/km^2) in areas with at least one Cistercian monastery compared to areas without Cistercian monasteries. Figure and description correspond to Figure 3 by Andersen et al. (2017a: 1772) with minor numerical discrepancies.

period, although the last Cistercian house was disestablished in 1540 during the Dissolution of the Monasteries under Henry VIII. Final exposure to the treatment predates any measurable effect by at least several decades, whereas initial exposure the Cistercians dates back to the 12th century when the majority (ca. 75 %) of all the Cistercian houses in England were founded (Andersen et al., 2017a: 1760, Figure 1). ABDS rationalize this time lag with a model of the diffusion process (Andersen et al., 2017a: 1766-1767). Under somewhat reasonable assumptions, it can take up to 25 generations of 20 years (500 years in total) for the share of the population, that adopted the Cistercian values, to increase from 1 % to 50 %. Furthermore, "[t]he 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." Hence, it could take centuries before the more productive sub-population is large enough to be reflected in average growth differences among regions.

This reasoning is sound but it amounts to an unfalsifiable thought experiment. The interpretation of the divergence as a consequence of historical Cistercian presence is only credible if there are both strong theoretical reasons to suspect Cistercian cultural influence as well as a robust causal effect based on plausible assumptions. Contrary to Andersen et al. (2017a: 1760, Fn 6) response to an anonymous referee's strong judgment that the

"claim that the Cistercians were precursors to the Protestant Ethic is simply incorrect," this disagreement cannot be resolved empirically if "empirical" is, rather narrowly, equated with econometric analyses of available quantifiable data. Historical scholarship is empirical evidence, too, and many of the assumptions underlying ABDS's statistical models are justified with regard to their interpretation of qualitative historical facts. In the following, I will re-assess the theoretical and empirical claims by ABDS. Based on my reading of the relevant scholarship, I regard the hypothesis about the Cistercian cultural influence as highly speculative. The actual content of the Cistercian doctrine did not lend itself to easy adoption by the surrounding population and there is no historical evidence of such value diffusion. Moreover, there are reasons to reconsider their empirical results as well. For instance, the most severe and direct test of their hypothesis – the IV regression analysis – can be criticized for the questionable operationalization of the instrument. The IV analysis rests on the assumption that royal forests aided the foundation of Cistercian settlements. ABDS use a dichotomous instrument. A county, the unit of analysis, is coded as containing royal forests if the share of its area classified as royal forest is greater than zero. This strict cut-off leads to a number of counties with less than one percent forest area as being coded one, leaving only 5 of 40 counties with a zero-coded IV. The main results are not robust to slight variations in the operationalization of the royal forest indicator.

The general problem is a combination of a speculative theory and indirect tests based on scarce historical data, resulting in many "researchers degrees of freedom" (Simmons et al., 2011). This problem of intransparent but consequential analytical decisions has long been recognized. Already Leamer (1983) advised to report the entire range of estimates that result from alternative specifications. In economics, some proponents of the so-called "credibility revolution" hoped that a shift to "design-based studies" would make observational studies more robust because sensitivity analyses could be "targeted at specific threats to validity" (Angrist and Pischke, 2010: 18). "Since the nature of the experiment is clear in these designs, the tack we should take when assessing validity is also clear" (Angrist and Pischke, 2010: 19). However, recent evidence suggests that p-hacking and questionable research practices are worryingly prevalent in studies using the most common research designs for causal identification (Brodeur et al., 2020a). IV designs in particular seem to suffer from substantial p-hacking or selective reporting (Brodeur et al., 2020a: 3642–3643). In line with recommendations from economics (Brodeur et al., 2020b) as well as the growing literature on model uncertainty and specification searching inspired by the social science replication crisis, I conduct a systematic robustness check using a *specification curve analysis* (Simonsohn et al., 2020). In my re-analysis I, first, identify the universe of plausible models based on all analytical decisions that cannot be deduced theoretically. I then transparently present estimates of all models in a way that allows to identify the critical choices leading to confirmatory results. The re-analysis reveals that strong confirmatory results with regard to early modern England can only be obtained via

a very specific set of decisions. These decisions are not derived from theoretical predictions and some – such as the operationalization of the IV – are outright problematic.

Notwithstanding the overall negative assessment, this paper is not a clear-cut falsification. It neither intends nor succeeds in "disproving" the influence of a Cistercian work ethics. It rather urges researchers in the domain of economic history to be more careful in presenting far-reaching conclusions based on bold theoretical arguments and scarce historical data.

2. Theoretical and historical considerations

The following section is a re-assessment of the theoretical argument brought forward by Andersen et al. with the aim to emphasize its speculative nature. Due to the often unsatisfying quality and availability of data in historical settings, research has to be guided by a strong theory in order to make causal claims. The section concludes with a brief discussion of an alternative causal channel of Cistercian influence – the diffusion of technical and managerial skills – that should be tested vis-á-vis the values-based explanation of Andersen and colleagues.

2.1. Weber's Protestant Ethic

Weber (2005) hypothesized that Protestantism, and specifically Calvinism, was conductive to the development of capitalism. According to the Calvinist doctrine of double predestination, God chose some people for salvation and others for damnation. The individual cannot change his or her fate. However, Calvinists started to believe the inner-worldly behavior to be an indicator of otherworldly salvation. Self-confidence, a strongly felt calling, inner-worldly asceticism, and economic success were said to be such credible signals. The tireless ambition of believers to attain assurance of their own salvation fostered a rational and methodological conduct of life as well as the accumulation of wealth. The unintended collective result was the "breakdown of traditionalism" in the economic sphere (Goldthorpe, 2021: 98) and, ultimately, the economic system of modern capitalism.

Although *The Protestant Ethic* is one of the most well-known texts in social science, there is only very limited empirical support for its main hypothesis. Weber's argument has been controversial in both social science and historical scholarship throughout the 20th Century (Becker et al., 2016: 3-4). Any systematic test is complicated by the fact that Weber himself was not entirely clear on the details of the main causal mechanism at work (e. g. Coleman, 1986: 1323). Most recent empirical tests formulate hypotheses about regional differences in economic development between regions with Catholic and (certain) Protestant majorities at some point in time. Studies looking for a correlation of religious denomination and indicators of economic growth in historical data either do not find any statistical relationship (Cantoni, 2015) or assert that the association can be explained

by Protestants' higher investment in human capital (Becker and Woessmann, 2009). However, with the exception of Spater and Tranvik (2019), quantitative tests compared predominantly Lutheran regions with Catholic regions, thus disregarding Weber's emphasis on ascetic Protestant denominations like Calvinism and Puritanism.[2] Spater and Tranvik (2019) estimate a regression discontinuity at the border of two Swiss cantons and find some indirect evidence in favor of Weber's hypothesis. Swiss Calvinists in the canton of Vaud urbanized more rapidly than Catholics in the canton Fribourg, i. e. the cities in the former canton grew faster during the late 19th century and villagers were more likely to move to growing cities. The authors interpret this as eagerness to find productive work.

As Andersen et al. note, Weber also became interested in medieval monasticism as an early predecessor of modern rationality. Importantly, neither in the case of Protestantism nor in the case of monasticism was Weber simply focused on value orientation like "thrift" and "hard work" but on a possible *nexus of asceticism and rationality*[3] (see also Kaelber, 1998: 19–21). Andersen et al. (2017a: 1761) cite the key passage:

> In the rules of St. Benedict, still more so with the monks of Cluny, again with the Cistercians [...], [Christian asceticism] developed a systematic method of rational conduct with the purpose of overcoming the *status naturæ*, to free man from the power of irrational impulses and his dependence on the world and on nature. It attempted to subject man to the supremacy of a purposeful will, to bring his actions 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. (Weber, 2005: 72)

The section in the *Protestant Ethic* on similarities between medieval and Puritan asceticism is at the same time unambiguous about their differences, which Weber deemed "evident." The difference "consisted in the disappearance of the *consilia evangelica* and the accompanying transformation of asceticism to activity within the world" because in traditional Catholic asceticism "the most important thing was the fact that the man who, par excellence, lived a rational life in the religious sense was, and remained, alone the monk" (Weber, 2005: 73). The way of life preached in ascetic teachings "was felt to be something higher than the everyday morality which sufficed as a minimum, and that this latter was not measured by such standards as Puritanism demanded" (Weber, 2005: 74) and hence "had, on the whole, left the naturally spontaneous character of daily life in the world untouched" (Weber, 2005: 101).

[2]Moreover, by the 19th Century, for which Becker and Woessmann (2009) and Spater and Tranvik (2019) have data, capitalism was already fully developed and had very likely long diffused beyond regions with a Protestant majority (Braun et al., 2012: 36).

[3]Rationality *sensu* Weber differs from the more technical definition of rationality in economic decision theory.

While Weber's theory itself is highly contested, it certainly cannot be used to derive the hypothesis that the Cistercian work ethic was conductive to economic development. Moreover, Weber explicitly connected the Protestant Ethic to wider societal trends and the emergence of a specific Capitalist Spirit that, according to him, ultimately spread beyond Protestant regions and milieus. ABDS's argument is more straightforward: Cistercian values induced a change in the individual preferences for leisure and thereby raised economic growth in regions close to Cistercian monasteries. Because the authors do not connect this diffusion process to the abandonment of traditionalist attitudes and the emergence of a Capitalist Spirit, there is only a tenuous connection to Weber's actual theory.

2.2. Monastic "work ethic"

Weber as well as other authors who noted similarities between Cistercian values and "work ethics [...] propagated by Calvinist and other ascetic sects some hundred years later" (Kieser, 1987: 116) never theorized a connection to later economic development. The suggestion is entirely due to ABDS. While the idea is intriguing, how plausible is it in light of historical knowledge about the religious order and its relationship to society at large? There are three interrelated questions. First, what was the exact nature of the Cistercian doctrine? Second, did the Cistercians live up to their own ideal? Third, was the population surrounding monastic estates receptive to Cistercian work ideals?

In the beginnings of Cenobitic monasticism hard work was a means of subsistence. The introduction of division of labor within a community was a precondition for asceticism which initially "had been an elite form of living, suitable only for a talented few," to become "an organized religious movement" (Kieser, 1987: 105). Living from the fruits of their own work allowed ascetics to remain independent of society. The *Regula Benedicti* makes it clear that the monks' primary task is the work of God (*opus dei*) while manual labor is an economic necessity and duty to be performed happily. Securing salvation and increasing the *thesaurus ecclesiae* through spiritual activities were the raison d'être of monasticism. Monastic efforts of economic rationalization or the openness to new technologies are sometimes attributed to a desire for minimizing working hours. According to this view, monks sought to reduce their work load in order to maximize time for prayer (Kieser, 1987: 114).

During the 11th Century, the Benedictine work requirements became purely symbolic. Monastic communities acted like feudal lords of the time, relying on the work of serfs as well as tithes and other income generated from parish churches. At the same time, the share of ordained choir monks increased, causing further emphasis on liturgical duties and spiritual contemplation. The Cistercian reform movement was a direct reaction to the perceived decline of the old monasticism. The Cistercians emphasized manual labor and self-sufficiency. They renounced spiritual income from churches. Their preferred mode of

economic organization was an "isolated and consolidated demesne farm" called a *grange* (Courtney, 1980: 44). In contrast to other orders who employed stewards or bailiffs to manage their estates, many Cistercian granges were directly administered by lay brothers (*conversi*) (Noell, 2006: 265). Though autarky may have been in part motivated by a desire to minimize contact with the outside world, manual labor undoubtedly became again an important component of ascetic practice (Kurze, 1980: 186).

However, there remain important differences between Weber's account of Protestant ethics and the Cistercian world view. The Cistercian attitude towards hard work was far from being unambiguous. Even Bernard of Clairvaux, one of the order's preeminent figures, found it difficult to reconcile monastic work requirements with the religiously higher value of the undisturbed contemplation of God (Kurze, 1980: 186). For Cistercians, work itself is never a "vocational calling" but rather one facet of a thoroughly otherworldly asceticism. As Sundberg (2019: 408) observes about contemporary Cistercians, the "monastic view on work [is] as something to perform indifferently, rather than to master, engage in, and exercise authority within." Monks are often ordered to perform a certain task by their abbot and rotation of duties is common to preserve the desired indifference (Sundberg, 2019: 408).

The accumulation of wealth, likewise, was a matter of concern. At one point, the order's General Chapter legislated (to no avail) that "all already established abbeys were strictly prohibited from the acquisition [...] of any land" as a reaction to public imputations of avarice (Lekai, 1977: 301). Eventually, the dilemma faced by both Puritans and monks – not being allowed to enjoy the wealth accumulated by hard labor and, thus, being forced to re-invest in ever more productive assets – perhaps contributed to the decay of the Cistercians during the Late Middle Ages. Economic success undermined its "ascetic credibility" (Kieser, 1987: 119; Southern, 1970: 260-261). In any case, the attitude towards wealth differed from Weber's characterization of the Puritans according to which "the faithful Christian *must* follow the call by taking advantage of the opportunity" (Weber, 2005: 108; emphasis not in original).

It is important to note that Cistercian spiritual leaders did not assign significance to the work ethic of lay people. For instance, Bernard of Clairvaux, who was also the most influential Cistercian writer in the order's history, in his sermon *Against the Most Wicked Vice of Ingratitude* wrote about the labor of lay people:

> These secular people labor in the fashion of this world that passes away, laboring for their own immediate sustenance and that of their household. At any rate, their labor may not lead to damnation, but it scarcely lends to salvation, so even if they have maintained a moral foundation, in the end they will suffer loss of the transient goods upon which they built. (Bernard of Clairvaux, 2016: 152)

Moreover, Cistercians – in contrast to, for example, mendicant orders – de-emphasized preaching. As one Cistercian writer put it: "The duty of a monk is not to preach but to pray" (Burton and Kerr, 2011: 200). Surviving sermons likewise rarely "direct comments to life in the outside world" (Kienzle, 2012: 245). This is in accordance with the general statement that medieval asceticism was primarily concerned with the spiritual perfection of the monk. Therefore, it is unlikely that both the Cistercians actively promoted "work ethics" outside their monasteries and such elements of Cistercian thought, isolated from the context of monastic life, would have made sense to the lay population at large. This is especially relevant because the authors occasionally write about *moral influences* (Andersen et al., 2017a: 1768). The emphasis on the *moral* dimension would rule out the argument that the surrounding population, observing the economic success of Cistercians, adopted their attitude towards hard work for purely material reasons.[4]

ABDS's (2017a: 1763) proposed mechanism of diffusion is essentially a model of population growth. But cultural change has to start with a small fraction of the population being directly influenced by Cistercian work ethics. Because there is "no evidence as to the exact nature of the contact with the surrounding lay populations" Andersen et al. hypothesize that initially "the ways of the Cistercians spread beyond the Order itself; by power of demonstration, by word of mouth, or both" (Andersen et al., 2017a: 1762). The power of demonstration is a more plausible transmission channel if the Cistercians in fact lived up to their own ideals. However, in later centuries "choir monks and nuns only exceptionally [left] their enclosure to assist in the harvest" (Berman, 2012: 120) and the order gradually became more and more entangled with the feudal economy. Donkin (1963: 181–182) writes about the initial period of rapid expansion in the British Isles until 1152:

> Over most of this short but extraordinarily active period the choir monks were directly involved in field labour; but thereafter, and until about the end of the 13th century, this was left almost entirely to lay brothers (*conversi*) and paid labourers. Later still, the monks began to lease by far the greater part of their estates, becoming mere *rentiers*

and this helps to explain

[4]Historian Thomas (1964: 59) writes about attitudes toward work in the early modern time:

> With the awareness of the economic importance of labour comes a new insistence upon the duty of every man to work. In a sense, this theme had been an element in Christianity from the start. It reached its peak in Protestantism, particularly Puritanism, but its origins lie in the Middle Ages, in the preaching of the Friars and Lollards. It did not, I believe, derive from monasticism, for although the early monastic founders had various motives for including labour as part of their régime, their most characteristic attitude was to regard work as a mortification of the flesh, a remedy for idleness, but not a productive good in itself. [...] It is in the religious teaching of the post-Reformation period, among Catholics perhaps as well as Protestants, that the positive merit of hard work is most cleary asserted.

> [...] the surprisingly rapid change in the reputation and character of the Order—from extreme asceticism and an ability to survive in the most unpromising places, to avarice and great wealth in land.

It is doubtful whether the Cistercians in the Late Middle Ages were involved enough with the actual material production of the monastic estates in order to inspire the adoptions of their supposed "work ethic." Even though the Cistercians employed many lay people as workers and came into contact with the surrounding population in manifold other ways (Lekai, 1977: 378–399), what the latter observed with regard to the economic activity was likely the superior organization and management of land (Lekai, 1977: 295).

2.3. Peasant life

The model of Andersen et al. predicts that once the preference for leisure declines, labor supply expands and fertility increases (Andersen et al., 2017a: 1764). Hence their model assumes 1) labor supply to be a major constrain on medieval population growth; 2) individual preferences for leisure to constrain the labor supply.

In the medieval agrarian economy, the demand for labor was determined by the cycle of the seasons (Thomas, 1964: 52). Pure leisure was confined to the holy days of the Church and these "[f]estivals occur at the slack periods of the agrarian year" (Thomas, 1964: 54). Kaelber (1998: 72), in his extensive study of medieval asceticism, remarks about the lay people working on monastic lands in general "there is little to suggest that their way of life differed from that around them. The people around them [...] were largely the peasantry but peasant life did not lend itself easily to the adoption of methodical and systematic conduct." It seems possible that there was little room for improved work ethics due to the "volatility of conditions in nature and the exigencies of agriculture" (Kaelber, 1998: 72). Hence, the central theoretical assumption of Andersen et al., that output was to considerable extent constrained by preferences for leisure, is conjecture.

In their illustrative models, ABDS assume the high work ethic group to work 20 % more and cite the estimate by Clark and Werf (1998) 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. Factoring in the statistical uncertainty [...] a 20 % higher work effort may not be unrealistic" (Andersen et al., 2017a: 1766, Fn 17). It should be noted that the assumption of a 5 % higher work effort, as implied by the point estimates, changes the prediction of the diffusion model considerably. Assuming initial shares of 0.1 % or 0.01 % with high work ethic results in predicted shares of 3.3 % and 0.3 % of the population, respectively, with a high work ethic after 25 generations.[5] However, the exact numerical value is not as

[5]The optimization problem at the individual level presented by ABDS is dispensable. The relevant macro dynamics (Andersen et al., 2017a: 1766-1767) correspond to a basic model of viability selection (McElreath and Boyd, 2008: 19). Group H exhibits high work effort and group L low work effort. Then

important as the fact that ABDS ignore Clark and Werf's overall conclusion. Weighing in on the debate whether the Industrial Revolution was preceded by an *industrious revolution* "which consisted of increased family labor per year" (Clark and Werf, 1998: 830), they examine historical data on the working hours of threshers and sawyers from the 13th to the 19th century. Although the evidence is "tentative and to some degree contradictory [...] on balance there is little sign of an industrious revolution" (Clark and Werf, 1998: 841). With regard to the early modern period, current research similarly suggests that "what has often been regarded as 'leisure' is better understood as unemployment and underemployment" (Griffin, 2020: 184). Hence, the increase in working hours at the onset of industrialization is likely the result of increasing labor demand.

Whereas the evidence concerning working hours is tentative and contradictory, there are hardly any sources on preferences and the general work ethics of the lower strata of medieval societies, especially where it did not intersect with the concerns of the upper classes. As Thomas (1964: 54) concludes: "What the free peasant working his own plot, as opposed to working for others, thought about it all [i. e. leisure and work, N. S.] I have no idea." Although there are textual sources of complaints about work performance, they were written from the perspective of manorial lords. Most peasants had to perform labor dues to the lord of the manor. "It seems clear at least that labor services were deeply unpopular with those who had to discharge them, and the unsatisfactory nature of their performance was one of the factors leading to commutation" (Thomas, 1964: 54).

In conclusion, it seems likely that individual attitudes towards work were probably not the determining factor of productivity and population growth – especially during the time of the closest observance of the Cistercian rule, i. e. during the 12th and 13th century when the "power of demonstration" was strongest and before the end of traditional manorialism.

2.4. Alternative explanations of Cistercian influence

The point of the preceding discussion was to show that the explanation brought forward by Andersen et al. is not well-grounded in the knowledge about the historical periods under study. Furthermore, many aspects of their theory – the exact mechanism of value transmission, the meaning of terms like "moral authority" etc. – remain elusive and in

---

the share p_t of H at time (generation) t is given by

$$p_t = \frac{1}{1 + \left(\frac{n_{L,0}}{n_{H,0}}\right)\left(\frac{E(L)}{E(H)}\right)^t}$$

where $n_{H,0}$ is the initial share of H, $n_{L,0}$ is the initial share of L, and $E(L)/E(H)$ represents the ratio of average work effort. Thus, the population share of group H assuming $n_{H,0} = 0.01$ and $E(L)/E(H) = 1/1.05$ will be a mere 3.3 % after 25 generations:

$$\frac{1}{1 + \left(\frac{1-0.01}{0.01}\right)\left(\frac{1}{1.05}\right)^{25}} = 0.033.$$

need of specification. Thus, even if there is in fact a robust association between Cistercian settlements and regional population growth patterns, it is necessary to complement this finding with more historical evidence to demonstrate the plausibility of the causal mechanism proposed by the authors. Moreover, there are alternative explanations to be ruled out. The superior economic performance of Cistercians has been explained by their managerial skill, rational planning, and openness to new technology such as water mills (e. g. Berman, 2012: 117–118). Burton and Kerr (2011: 187–188) stress the monks' role as "pioneers, entrepreneurs or disseminators of methods and techniques" bringing about economic change and consider the reorganization of the land as "[p]erhaps the Cistercians' most enduring legacy."

It might be that the spread of technological and managerial knowledge, independent of any ethical and theological doctrines, could explain a possible causal effect of Cistercian settlements on the growth of surrounding areas. Andersen et al. (2017a: 1759) themselves acknowledge this possibility. I therefore propose to look for correlations between the presence of manors or granges owned by the Cistercians and later population growth as well. Although far from decisive, a strong correlation between population growth and grange locations – the centers of economic activity – rather than monasteries – the centers of spiritual activity – would give plausibility to the alternative causal channel as stated above.[6]

3. Empirical considerations

3.1. Operationalization

Andersen's et al. (2017a) preferred operationalization of the main predictor is the Cistercian influence 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, M_c, to all religious houses" (Andersen et al., 2017a: 1768).

On a fundamental level, the assumption that work ethics were primarily shaped by the Church and not by, for instance, the mode of production (Fouka and Schläpfer, 2020) seems doubtful. The authors write that Cistercian work ethics spread to the surrounding population "by power of demonstration, by word of mouth, or both" (Andersen et al., 2017a:

[6]Guerriero (2020) proposes a competing theory about the Cistercian influence on the local economies surrounding their communities. However, her theory is based on a misreading of the historical evidence regarding the motives and mission of the Cistercian order. For instance, in a key passage that she cites in support of her interpretation (Guerriero, 2020: 375), Burton and Kerr (2011: 11) compare two early textual sources – the *Exordium Cistercii* and the *Exordium Parvum* – with regard to the stated reasons for the Cistercian reform initiated by its founder Robert of Molesme. The passage does not summarize the "more recent and substantial historic literature" but the content of one particular source. Moreover, the following discussion of accounts written by Benedictine monks (Burton and Kerr, 2011: 14–15) vindicates the view that the early Cistercians were stressing the importance of strict rule observance.

1762) which seems in line with the fact that Cistercians were not primarily concerned with preaching. But then the restriction to official religious authorities is not an obvious and self-explanatory choice. In principle, the peasant population could have imitated the "ethics" of any hard working group, religious or not.

However, the construct *Cistercian share* is problematic even if we accept the premise that moral influences were primarily ecclesiastical. ABDS use the English Monastic Archives[7] (henceforth EMA) as a source for information on religious houses. However, the EMA only covers monastic orders and regular canons. It does not cover mendicant orders, even though there were at least 54 Franciscan and 48 Dominican houses in medieval England (for comparison: there were 71 Cistercian houses).[8] ABDS also do not explain why they only include data on "major orders"[9] rather than all religious houses in the EMA. Most importantly, apart from religious orders, the Church was represented by the secular clergy such as secular canons and parish priests. ABDS do not consider their influence at all. Thus, in my view, it is not clear what the correct denominator would be. Restricting the denominator to an arbitrary selection of orders could unnecessarily exacerbate measurement error. Lastly, the indicator implicitly assumes the presence of other orders to be a contravening force although there is no reason to expect other religious orders to suppress the spread of an ethic of hard work among the peasant population.

For the reasons stated, I contend that the main predictor should be based on the best available measure of absolute Cistercian presence (not relative to other orders). ABDS offer three operationalizations based on absolute presence: firstly, a dummy variable of Cistercian presence, taking the value 1 if there was at least one Cistercian house in a county, and 0 otherwise; secondly, the absolute number of Cistercian houses in a county; lastly, they use *Cistercian density*, i. e. the number of houses divided by the county area. All three measures seem preferable to the heavily assumption-laden Cistercian share. Moreover, ABDS omit several other equally plausible operationalizations. For example, it could be argued that the influence on the values of the local populace was greater in counties with long-lasting Cistercian presence. One way to construct a corresponding measure would be to compute the "life-span" of all monasteries (years from foundation to dissolution) and then sum those durations at the county-level (see Section 4). Since the theory is silent on the exact initial mechanism of cultural transmission, it is difficult or, perhaps, impossible to decide on the most appropriate measure of Cistercian influence. This has to be taken into account and is a strong motivation to perform a "specification curve analysis" in order to assess the robustness of the results systematically (Section 4.3).

[7] `https://www.ucl.ac.uk/history/research/english-monastic-archives`

[8] The number of Franciscan houses was computed from the dataset provided by Boranbay and Guerriero (2019). The source for the number of English Dominican houses is the *Digital Atlas of Roman and Medieval Civilization* (DARMC; Gibson et al., 2022). The respective DARMC map is based on Jedin et al. (1987: 59) and Bengtson and Milojčić (1995: 30).

[9] Augustinian canons, Benedictines, Cistercians, Cluniac monks, and Premonstratensians.

3.2. Royal forests as instrumental variable

Andersen et al. (2017a: 1777) use the presence of a royal forest in a county as an instrument for Cistercian settlements. They cite Donkin's (1963: 184) observation that

> 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; 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.[10]

They then construct a speculative argument that "[a]t the time of arrival the most secluded areas may well have been the forests owned by the Crown" and that "[t]hus, there may well have been a double coincidence of wants": Cistercians were searching for locations satisfying their "ascetic needs" while royal as well as non-royal landowners were hoping for material and spiritual benefits.

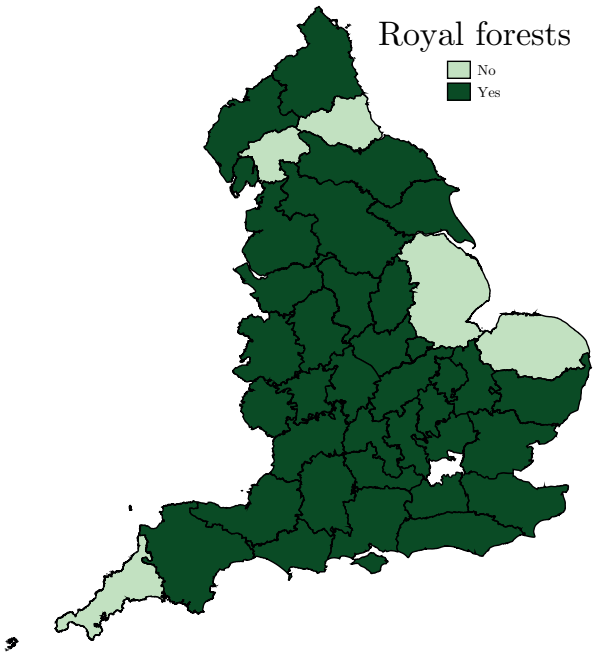
The authors use a map published by Bazeley (1921) to obtain data on the location of royal forests in the 13th century and code the presence of a royal forest in a historic county as a dummy variable (Andersen et al., 2017a: 1777). They use this dummy as an exogenous instrument in the first stage to predict the Cistercian share and presence. The authors also create a measure of the county area that was covered by royal forest in the 13th century as a share of the total county area (*forest share*) which they include as a control variable (as explained below).

According to the authors' coding, only 5 of 40 counties had no royal forests (cf. Figure 2a and Table 1). The dummy coding is very coarse since the authors use a strict cut-off: e. g. 5 historic counties with less than 1% of their area classified as royal forests are coded as 1 (Kent, Suffolk, Sussex, West Riding of Yorkshire, Devon).[11] Such a coding is not warranted by the precision of the hand-drawn map which is explicitly labeled a "Sketch Map." Moreover, without questioning the overall approach of computing forest areas based on Bazeley's map, the dummy coding discards important information on the extent of the forests. Closer inspections of, for example, Sussex reveals that the only Cistercian

[10] In light of more recent research, Donkin's quote is possibly an exaggeration. Jones (2010: 39) corrects the view that royal forests were about little else than "royal pleasure" by illustrating, with many examples, the diversity of the forest economy and its importance for local dwellers. According to Christopher Dyer, the enforcement of royal rights were only a minor impediment to the expansion of settlements. While royal interest in hunting "had some inhibiting effect on assarting," royal officials mainly "collected fines in the forest from those who poached the deer or assarted the woods; their activities annoyed the inhabitants, but did not prevent the clearance of new land" (Dyer, 2001: 20).

[11] Considering that Cistercians were granted rights in at least 19 non-royal forest as well (Langton, 2015: 393, Donkin, 1978: 75–79, 242), it could be argued that their existence likewise fostered Cistercian settlements and hence all forests and chases as identified by the more exhaustive survey by Langton and Jones (2010: Figure 1 reproduced in Appendix B) should be used as the basis for measurement construction. In this case, all historic counties would have to be coded as including forests.

a. Royal forests



b. Royal forest share

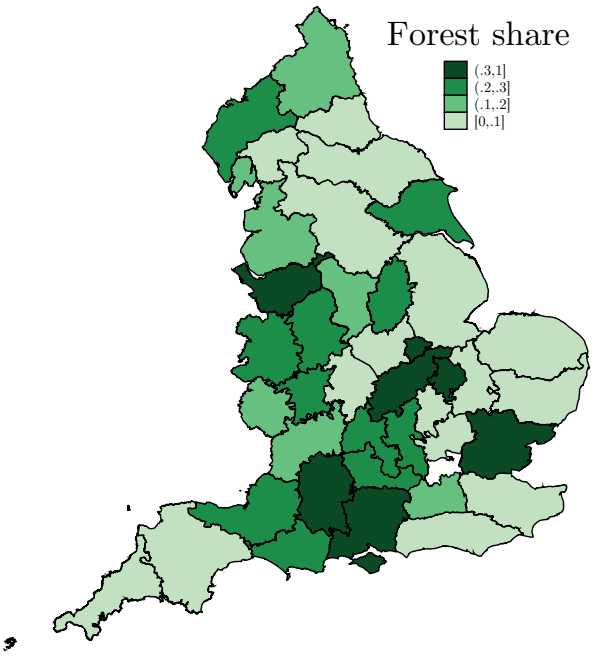


Figure 2: Counties with royal forests according to Andersen et al. (2017a) based on data obtained from Bazeley (1921). The variable Rforest is dummy coded (left figure). The variable royal forest share (right figure) is a measure of the county area designated as royal forests in the 13th century.

house (Robertsbridge Abbey) was in the Eastern part of the county while the royal forest (0.03 % of the land area) was located at the western border (see Figure 17 in Appendix B). There are multiple counties with marginal forest shares and where Cistercians, albeit present, were located not even close to those forests. The first stage is likely to overstate the relationship between royal forests and Cistercian influence. A continuous metric would factor in that small forest shares are unlikely to affect the composition of monastic orders or the absolute number of Cistercian houses.

The rationale of the IV regressions is to alleviate the problem of endogeneity. The geographical distribution of Cistercian houses might correlate with later population growth for reasons altogether unrelated to the diffusion of work ethics. While the authors establish a historical association between Cistercian settlements and royal forests, they circumvent to explicitly state their reasons to assume that the geographical distribution of royal forests on the county level is (conditionally) exogenous to later population growth. The authors themselves are concerned with the possibility that the "use of Rforest as an instrument for the intensity of Cistercian presence [...] might capture resource growth" through deforestation in later centuries (Andersen et al., 2017a: 1777). To alleviate this cause for concern, they add *forest share* as well as a control variable for the suitability of land for pastorialism (Figure 3). The authors argue that "adding forest share to the control set should make the excludability of Rforest in the second stage plausible." However,

Table 1: Cross tabulation of Rforest and dummy treatment (Cistercian presence).

		Royal forest		
		No	Yes	Total
Cistercian Presence	No	3	5	8
	Yes	2	31	33
	Total	5	36	40

since there does not seem to be a strong reason to believe that areas were randomly designated as royal forests, the five counties with a value of zero on Rforest could share many unobserved features. Crucially, a bias analysis reveals that ABDS's main findings are not robust to small hypothetical violations of the exclusion restriction (see Appendix F).

Because the whole analysis relies on a small sample of 40 cases[12] and hence is susceptible to problems due to misclassifications in even a few cases (Table 1), it seems desirable to have a plausibility (or criterion validity) check of the instrument. Fortunately, Donkin (1960, 1978: 121–123) identifies all Cistercian houses that, based on Bazeley's work as well as various other historical sources, are known to have been located in royal forests (Donkin, 1960: 42–43 and Figure 1 therein).[13] I use the information provided in his compilation to construct a variable, a count of forest settlements on the county level, called *inforest*. According to my count, there were 17 Cistercian forest settlements located in 15 of the historic counties that constitute the analysis sample.[14] There is a medium to strong correlation between inforest and the forest share as well as between the share of Cistercian houses located in forests (based on inforest) and the forest share (Table 2). The correlation with Rforest is much weaker. None of the Cistercian forest settlements identified by Donkin (1960) were located in counties with less than 10 % forest area even though a total of 28 Cistercian monasteries were located in those counties. Therefore, I have some reason to believe that the use of Rforest in the first stage overstates the influence of forests on Cistercian presence.[15]

[12]The sample size itself seems quite problematic given the poor small-sample properties of the IV-estimator when the model is exactly identified (for an instructive simulation see the blog post by Millimet, 2019).

[13]It should be noted, however, that Bazeley's (1921) map is somewhat outdated. For a more up-to-date survey of English medieval forests see the edited volume by Langton and Jones (2010), especially John Langton's paper "Medieval forests and chases: Another realm?" (Langton, 2010). He compares Bazeley's work with other maps and discusses the problems with the definition and demarcation of "royal forests." Likewise, Donkins work, according to himself, is preliminary "and more detailed work would be required to show precisely how many houses were founded within the bounds" (Donkin, 1960: 42).

[14]Donkin (1960: 42) gives a total of 20 forest houses. A list is provided in Appendix A alongside an explanation of the inclusion or exclusion of certain houses.

[15]According to ABDS, their preferred interpretation for why the IV estimates exceed the OLS estimates is a reduction of attenuation bias because "indicators of Cistercian presence are imperfect indicators of the fraction of the population with 'Protestant ethics' " (Andersen et al., 2017a: 1780). However, this interpretation seems doubtful given the measurement error in Rforest. The overly strict cut-off criterion for the dummy coding might introduce new sources of spurious findings.

Table 2: Correlation matrix.

	Rforest	Forest share	inforest
Rforest	1		
Forest share	0.344	1	
Cist. houses located in forests (inforest)	0.274	0.701	1
Proportion in forests	0.261	0.653	0.821

Instead of using forest share as a control variable it seems even more, or at least equally, appropriate to substitute it for the Rforest dummy as instrument in the first stage. In light of the proposed causal channel, it seems reasonable to expect a larger forest area to increase the opportunities for Cistercian settlements and hence predict a larger Cistercian share of all houses. This might reduce the plausibility of the exclusion restriction but the minimization of systematic measurement error due to dichotomization is an equally important concern.

Even under the strong assumption that the research design approximates a natural experiment, the power of the research design to detect plausible effect sizes is low (see Appendix E). For example, an effect of 0.5 for the binary treatment variable (Cistercian presence) has to be considered large for it means that the presence of Cistercians before the dissolution of the monasteries increased subsequent population growth by more than 60 % compared to counties without Cistercian presence. However, the power of the IV design to detect such an enormous effect – close to the effect size actually reported by ABDS – is only about 50 % for $\alpha = 0.1$. The power to detect a smaller, but still considerable, effect of 0.2 is reduced to 18 %. Moreover, omission of the covariates in the model likewise greatly reduces the power of the 2SLS estimator. In noisy, small-sample settings, effect sizes have to be large in order to pass the "significance filter," creating the problem of Type M error, i. e. the accumulation of systematically exaggerated results in the research process (Gelman and Carlin, 2014; Felton and Stewart, 2022: 26).

One can question whether the 2SLS approach is in fact necessary. Following my arguments in section 3.1, an absolute measure of Cistercian presence is preferable to the relative measure of Cistercian share. Accepting Andersen's et al. arguments about the association between Cistercian houses and royal forests, inforest can be used as a direct measure of exogenous Cistercian presence at the county level. Using inforest directly in an OLS model compares counties that happened to have "excess" Cistercian houses due to royal forests and counties without such exogenously caused Cistercian presence. Alternatively, if it is judged necessary to use Cistercian share as endogenous regressor, the Rforest IV in the first stage can be replaced by either the forest share of a county or the number of Cistercian forest settlements. Both seem preferable to the coarse, dummy coded variable.

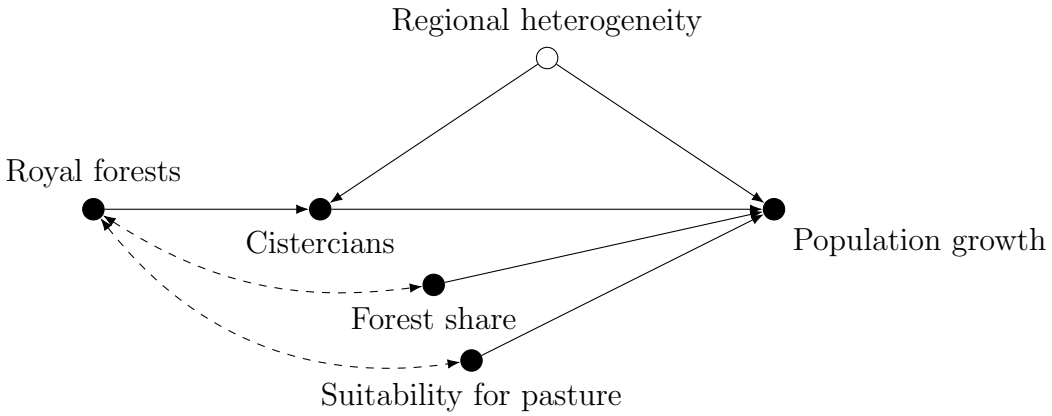


Figure 3: Directed acyclic graph summarizing the causal assumption of Andersen's et al. (2017a) instrumental variable model. Notation following Morgan and Winship (2014: 77–130): Filled circles indicate observed covariates, hollow circles unobservables; solid, single-headed arrows represent directed causal relationships; dashed, double-headed arrows indicate causal dependencies among variables that are not fully spelled out in the graph.

4. Data and analysis strategy

4.1. English county data

The dataset used in the re-analysis is based on the data provided by ABDS (Andersen et al., 2017a: Supplementary information Data S1). ABDS combine data from various sources on English historic counties including information on land quality, the extent of Roman roads, or suitability for pasture. The counties' population numbers are estimates taken from Campbell (2008). For more detailed information on the other variables, I direct the reader to ABDS's appendix (Andersen et al., 2017b). The upper part of Table 3 reproduces the summary statistics of selected variables from the original article (Andersen et al., 2017a: 1771, Table 1). I added the following variables to the English county dataset (for details, see Appendix D).[16]

Inforest. As already described in Section 3.2, I add the number of Cistercian monasteries in every county that were located in royal forests according to Donkin (1960).

Rforest2. I modify ABDS's Rforest variable and assign counties with marginal forest shares of less than 5 % a value of 0.

Manors and granges. Furthermore, to test whether the association between the intensity of Cistercian economic activity and population growth is stronger and more robust than the relationship of the latter with Cistercian religious presence, I add the counts of Cistercian manors for every historic county from the English Monastic Archives. However, not all manors possessed by Cistercian houses were directly managed by monks or lay brothers. The superior economic efficiency and rational planning are primarily ascribed to directly

[16]Data and syntax files are available online: `https://osf.io/cdtn9/`.

Table 3: Summary statistics: English historic counties.

	Mean	SD	Min	Max	N
Original variables					
Cistercian share	0.09	0.07	0.00	0.25	40
Religious houses	19.03	12.93	2.00	73.00	40
Population density 1377	31.55	11.83	8.98	52.98	40
Population density 1600	29.99	6.46	13.97	43.33	40
Population density 1801	60.45	24.82	20.92	143.77	40
Augustinian share	0.28	0.13	0.00	0.62	40
Benedictine share	0.31	0.16	0.00	0.67	40
Cluniac share	0.04	0.05	0.00	0.15	40
Premonstratensian share	0.05	0.09	0.00	0.50	40
Land quality	0.18	0.16	0.01	0.73	40
New variables					
Cistercian houses in forests	0.42	0.59	0.00	2.00	40
Rforest2	0.60	0.50	0.00	1.00	40
Manors (total)	180.82	95.41	11.00	395.00	40
Cistercian manors	37.13	38.43	1.00	168.00	40
Cistercian granges	10.30	10.49	0.00	48.00	40
CCCP/100	5.92	5.24	0.00	24.21	40
Cistercian houses per pop.	0.02	0.02	0.00	0.06	40

managed granges (see Section 2.2). Hence, I create a second variable counting the number of manors with names containing the word "grange." This is a very rough measure likely to undercount granges due to the existence of different naming patterns. I generate equivalent variables for the Augustinians based on the English Monastic Archives in order to carry out comparable analyses. Even though some estates belonging to the latter were also called "granges," they usually lacked the specific qualities of Cistercian granges in the proper sense (Kershaw, 2023).

Cumulative County-level Cistercian Presence (CCCP). As outlined in Section 3.1, CCCP tries to capture the importance of a long and continual Cistercian presence in a given county. For every monastery, its years of duration are calculated from the dates of foundation and dissolution. The years of duration of all monasteries within a county are then summed to produce a county-level variable. The sum is divided by 100 in order to facilitate the interpretation of regression coefficients.

Houses per 100k inhabitants. Another possible operationalization of Cistercian influence pertains to Cistercian presence relative to medieval population size. More monasteries per 100 000 inhabitants in the year 1377 indicate a stronger Cistercian influence.

4.2. European data

The European-wide datasets have been supplemented with additional information on other religious orders. I added influence measures of three orders (Cluniac reform, Carthusians, Premonstratensians)[17] on the level of the NUTS 2 regions. I use the same influence measures that ABDS generated for the Cistercian order: 1) the number of houses in a NUTS 2 region; 2) the natural logarithm of the number of houses[18]; 3) the number of houses divided by the area of the NUTS 2 region (density); 4) a dummy coded variable for presence, taking the value 1 if there was at least one house.

Carthusians. The Carthusians are a monastic order that were founded, like the Cistercians, at the turn of the 12th century. They were not a reform movement within Benedictine monasticism and instead represent a monastic tradition *sui generis* incorporating elements of eremitic life (Hostie, 1983: 53). In contrast to the Cistercians, they have never been associated with the ideal of hard work due to their strict focus on contemplation. Thus they constitute a plausible placebo test: A positive statistical association of past Carthusian influence and contemporary work ethic would indicate that the research design is vulnerable to picking up spurious relationships. The data contains information on 232 Carthusians monasteries that were founded in the sample regions before the 16th century.[19]

Cluniac Reform. The Cluniac Reform (or Reforms) began in Cluny Abbey during the 10th century. It was a movement within Benedictine monasticism to restore traditional monastic life, similar to the Cistercians two centuries later (Hostie, 1983: 48). ABDS write that "similar [to the Cistercians, N. S.] values were found among the Cluniacs" (Andersen et al., 2017a: 1761)[20] but their analyses show no relationship between Cluniac influence and population growth in England. Therefore, it seems worthwhile to include them in the European-wide analyses as well.

Premonstratensians. The Premonstratensians are an order of regular canons, founded in the early 12th century. The governance structure of the order was modeled in part on Cistercian principles (Hostie, 1983: 82–83). However, the Premonstratensians are not commonly linked to a strict work ethic. Like the Cluniacs, ABDS include them in their analyses of English historical data. For completeness, I include them in the European-wide analysis as well. There were 529 Premonstratensian houses founded before 1500 in the sample region.

[17] It would have been interesting to include Augustinian Canons as well because there seems to be a robust relationship with early modern population growth in England (see section 5.1). However, I was unable to find a suitable gazetteer recording Augustinian foundations in Europe.

[18] They add one to the count of Cistercian houses in each region to avoid zero values from dropping out.

[19] I only include monasteries that were founded before the 16th century because the focus is on the *pre-Reformation roots* of work ethics. This effect is most plausible identified by the locations of medieval monasteries. Moreover, the locations of foundations after the reformation are strongly and negatively correlated with the spread of Protestantism.

[20] They cite Max Weber but he refers only to "rational life conduct" in general.

Table 4: Summary statistics: European NUTS 2 regions.

	Mean	SD	Min	Max	N
Original variables					
Thrift (NUTS2 share/mean)	0.43	0.15	0.00	1.00	241
Hard work (NUTS2 share/mean)	0.47	0.25	0.00	1.00	241
Employment 2007 (log)	13.38	0.71	10.93	15.46	241
GDP 2007 (log)	10.44	0.81	8.30	13.14	235
Area (km^2)	15,323.06	15,266.23	173.00	92,961.00	242
Protestant share	0.29	0.36	0.00	1.00	241
Population 2007 (log)	14.21	0.71	11.73	16.27	242
Cistercian Density	0.00	0.00	0.00	0.00	242
Cistercian Houses	2.69	3.87	0.00	26.00	242
Cistercian Houses (log)	0.97	0.78	0.00	3.30	242
Cistercian Presence	0.74	0.44	0.00	1.00	242
New variables					
Carthusian Density	0.00	0.00	0.00	0.01	242
Carthusian Houses	0.87	2.21	0.00	28.00	242
Carthusian Houses (log)	0.41	0.56	0.00	3.37	242
Carthusian Presence	0.43	0.50	0.00	1.00	242
Cluniac Density	0.00	0.00	0.00	0.00	242
Cluniac Houses	4.49	14.77	0.00	152.00	242
Cluniac Houses (log)	0.72	1.10	0.00	5.03	242
Cluniac Presence	0.41	0.49	0.00	1.00	242
Premonstratensian Density	0.00	0.00	0.00	0.01	242
Premonstratensian Houses	1.93	3.12	0.00	27.00	242
Premonstratensian Houses (log)	0.75	0.75	0.00	3.33	242
Premonstratensian Presence	0.60	0.49	0.00	1.00	242

4.3. Analysis strategy

The following analysis replicates and improves upon the research designs used by ABDS. I present four types of analyses: 1) IV estimates of the causal impact of Cistercian presence on population growth in medieval and early modern England 2) LPM/probit estimates of the relationship between Cistercian presence and population growth in medieval and early modern England 3) OLS estimates of the relationship between Cistercian presence and contemporary values regarding the importance of hard work and thrift in Europe 4) OLS estimates testing whether regional differences in value orientation mediate the association between Cistercian presence and contemporary employment in Europe.

However, the speculative nature of the hypothesis under scrutiny, the lack of clear theoretical guidance with respect to operationalizations and model choices as well as the resulting need for *ad hoc* decisions suggest an approach that takes into account model uncertainty and "researcher degrees of freedom" (Simmons et al., 2011). Hence, I conduct a systematic robustness analysis, a so-called descriptive *specification curve analysis*

(Simonsohn et al., 2020). The basic idea is to compute and analyze *all* plausible model specifications with the aim of identifying *critical decisions* that lead to falsification or confirmation. If they correspond to the weakest part of the theory, i. e. confirmatory results can only be obtained with very specific (*ad hoc*) modeling choices, then the reader can weight the evidence accordingly. Similar approaches have been given different names such as "multimodel analysis" (Young and Holsteen, 2017) or "multiverse analysis" (Steegen et al., 2016). What they have in common is a desire to increase transparency given the often vague theories in the social sciences.

It is already common for empirical papers to include numerous robustness checks. Sprawling appendices notwithstanding, robustness checks are usually still an intentional selection from the universe of plausible models. Moreover, sometimes the reasons behind details of specification choices remain opaque. It is, for example, common to vary more than one model property at the same time in robustness checks, even though results may only be robust to specific combinations of deviations from the preferred specifications.[21]

It is important to stress that this is not a rejection of theory guided research. To the contrary, theory still constrains the range of plausible models. The first step is always to identify "the set of theoretically justified, statistically valid and non-redundant specifications" (Simonsohn et al., 2020: 1). However, it is just a matter of fact that theories are often insufficient to deduce the precise specification of statistical models or to choose among competing ways to measure a theoretical construct. The researcher is left with equally plausible subjective decisions. The weaker the theory, the stronger the need for increased transparency. At the same time, nothing prevents researchers from highlighting their preferred specifications.

5. Empirical analysis

The results for the English dataset are presented in Section 5.1. It starts with the multiverse analysis of the instrumental variable and then proceeds with a similar analysis of the OLS models. Section 5.2 consists of a multiverse analysis for the 2008 European Values Study (EVS) data as well as a mediation analysis that tests the mediating role of regional differences in attitudes toward hard work and thrift on economic outcomes.

5.1. England: specification curve analysis

The dependent variable in all regression models for England is the natural logarithm of population growth. In ABDS's notation: $\Delta \log(L_{t+1}) = \log(L_{t+1}) - \log(L_t)$ for population

[21] For instance, when presenting alternative specifications for the effect on contemporary values of European citizens, ABDS use the area (km^2) of the NUTS 2 region as a control for some operationalizations of Cistercian influence and – without providing a theoretical justification – the logarithm of the area for others (Table 6, p. 1786).

Table 5: Original and alternative reasonable specifications of the instrumental variable model.

Decision	Original specifications	Alternative Specifications
(1) Instrument	Rforest	Rforest2, forestshare, inforest
(2) Endogenous regressor	Cist. share, Cist. presence (0/1)	Cistercians (total) Cist. density, CCCP, Houses per population, Cist. manors, Cist. Granges
(3) Dependent variable	1290–1801, 1377–1801, 1600–1801	

level L of county i and observation time t. This is equivalent to the change in the natural logarithm of population density.[22]

Instrumental variable regression

The strongest evidence in favor of a causal relationship between Cistercian influence and regional development are the 2SLS regressions presented by ABDS (Table 4 in Andersen et al., 2017a: 1779). In the first stage an endogenous indicator of Cistercian influence is predicted by the instrument Rforest. If π is the Cistercian influence in historic county i, the basic specification can be written as

$$\pi_i = \beta_1 + \beta_2 \text{Rforest}_i + \beta_3 \log(L_{it}/X_i) + \beta_4 M_i + \mathbf{Z}_i'\beta + \nu_i \tag{1}$$

where M denotes the total number of religious houses, $\log(L_{it}/X_i)$ population density at time t and $\mathbf{Z}_i$ is a vector of time-invariant controls for productivity (see also Figure 3). The second stage estimates the effect of the exogenous part of the indicator $\hat{\pi}_i$ on population growth

$$\Delta \log(L_{t+1}) = \gamma_1 + \gamma_2 \hat{\pi}_i + \gamma_3 \log(L_{it}/X_i) + \gamma_4 M_i + \mathbf{Z}_i'\gamma + \varepsilon_i. \tag{2}$$

ABDS present results for six different IV specifications. Besides the Cistercian share, they also use a dummy of Cistercian presence as endogenous regressor. Furthermore, they vary the time frame of the dependent variable (see Table 5). For this systematic robustness check, I consider all four indicators of Cistercian influence introduced by ABDS as endogenous regressors as well as my four additional indicators (see Section 4.1). I also use three alternative instruments: Rforest2, inforest and forestshare (see Section 3.2). This gives a total of $4 \times 8 \times 3 = 96$ model specifications.

The results of the second stage regressions are shown in Figure 4. The first row are the

[22] $\Delta \log(L_{t+1}) = \log(L_{it+1}/X_i) - \log(L_{it}/X_i)$ for population level L and area X of county i and observation time t.

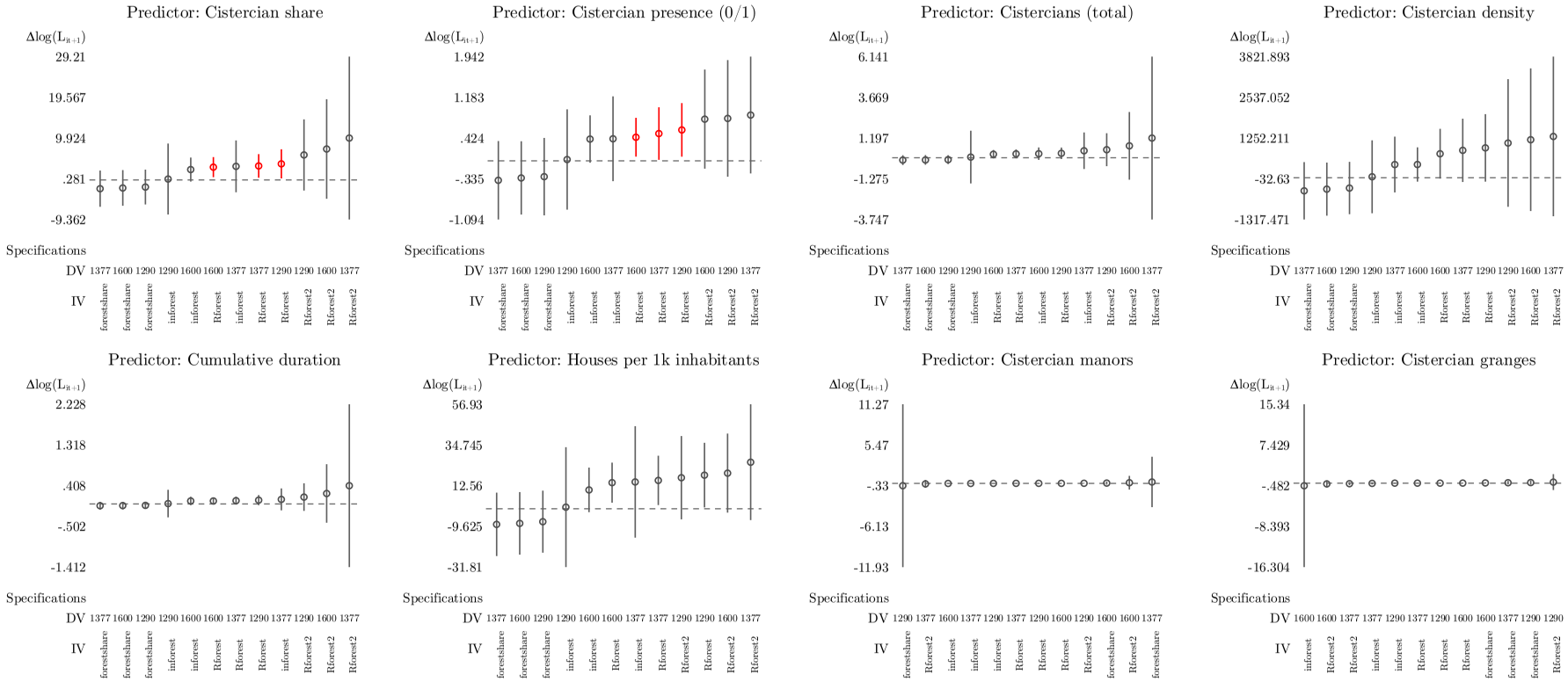


Figure 4: Systematic robustness check of S2LS estimates ($N = 40$). The figure shows coefficients (second stage) from 96 model specifications. Each sub-figure shows results for a different indicator of Cistercian influence (the endogenous regressor). The point estimates are ranked from weakest to strongest effect with bars indicating the 90% confidence interval based on heteroskedasticity-robust standard errors. The dashed line indicates an effect size of zero. Rows below indicate the specification used to obtain the respective coefficient estimate above. DV gives the starting year of the period with the end point being 1801 in all specifications. The specifications in the original ABDS paper are in red. Exogenous controls included in all models: population density at the start of the period, total number of monasteries, land quality, and the suitability for pasture. Models control for forest share when it is not the excluded instrument.

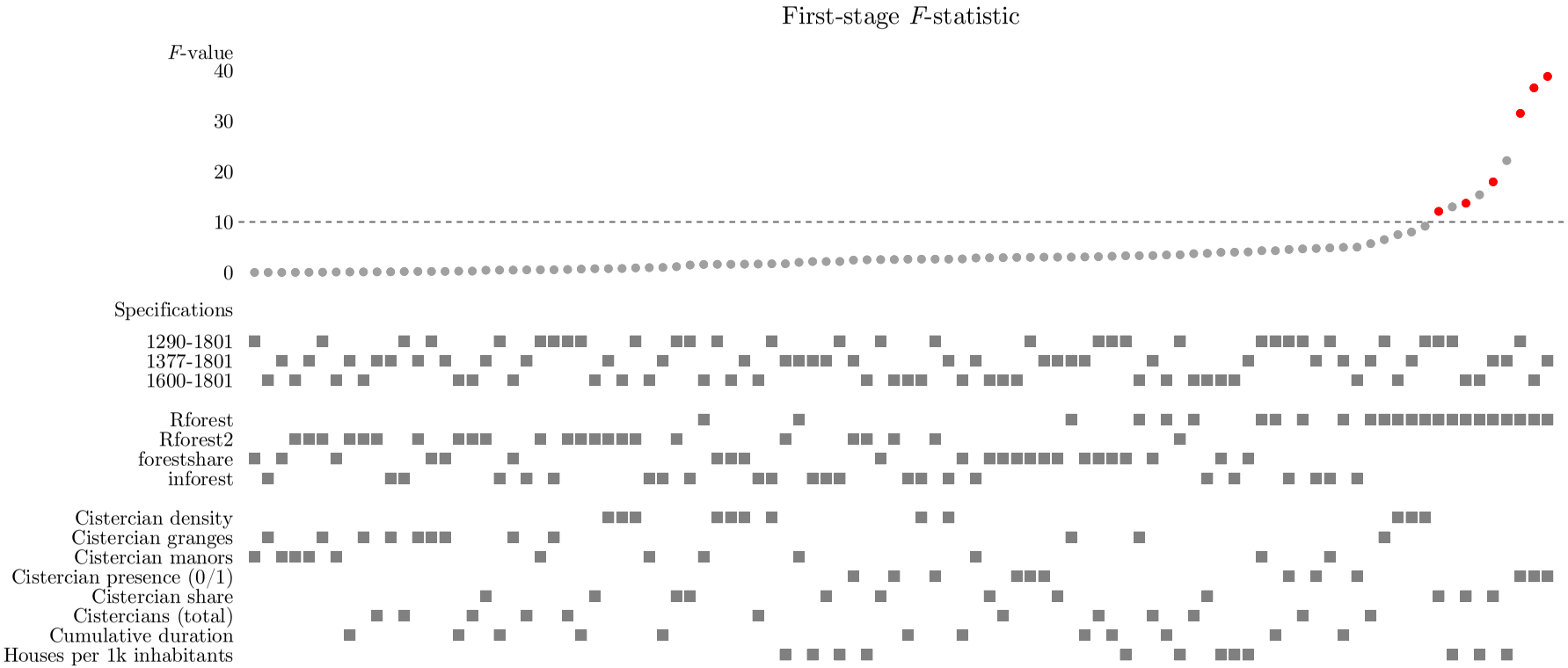


Figure 5: First stage F-statistic of 96 2SLS regressions. The F-values are ranked from smallest to largest. The dashed horizontal line at $F = 10$ corresponds to the rule of thumb suggested by Staiger and Stock (1997). F-values of ABDS models are highlighted in red. The squares below indicate the specification characteristics.

four indicators proposed by ABDS and the second row are my additional four indicators. The 90% confidence intervals correspond to the significance level chosen by ABDS. The six models highlighted in red (Figure 4) are the original specifications.

I start with the four indicators in the first row. Only models using, first, Rforest as an instrument and, second, either Cistercian share or a dummy of Cistercian presence produce significant results. Those are exactly the six specifications presented by ABDS. However, some specifications approach significance, including, in each case, one model using inforest as instrument and Cistercian share, Cistercian presence or Cistercian density as endogenous indicator.

Overall, the most important decision seems to be the choice of the instrument. A clear pattern emerges regarding the IV choice: Using forest share as an IV consistently produces negative, but statistically insignificant, estimates of Cistercian influence on population growth. Using inforest – the actual number of Cistercian houses in forests – leads to effects somewhat smaller or about the size of the original finding but with higher uncertainty. Choosing Rforest and Rforest2 consistently results in estimates with the expected (positive) sign. Rforest2 produces the largest effect sizes – larger than the original Rforest IV – but the estimates are, at the same time, very uncertain as indicated by the wide confidence intervals.

Regarding the second set of indicators in Figure 4, it is clear that specification using CCCP (cumulative duration), the number of manors or the number of granges produce small and very uncertain effects regardless of IV choice. Specifications with the number of Cistercian houses per population as endogenous regressor and Rforest2 as IV have positive effects. Again, only models using Rforest as an instrument have positive effects significant at the 10 %-level.

In summary, the most crucial decision in order to arrive at positive and significant effects is the choice of the Rforest instrument whose problems were discussed in Section 3.2. Figure 5 shows the descriptive specification curve of the first-stage F-statistic for testing the hypothesis that the instrument's coefficient is zero. Only Rforest correlates strongly with some of the endogenous regressors. All other IVs must be considered weak. Most importantly, Rforest2 – albeit leading to some strong positive effects in the second stage – is indeed a weak instrument. Therefore, the strong correlations between Rforest and Cistercian share, the Cistercian presence dummy, and Cistercian houses per population hinge upon the counties with marginal forest shares and no documented forest settlements.

OLS results: inforest

There is yet another possibility to approach a credible estimation of the causal effect. As I have already hinted at in Section 3.2, depending on the assumptions, it may be unnecessary to use 2SLS. If the number of Cistercian forest settlements is exogenous, it

Table 6: Reasonable specification choices of the *inforest* OLS model.

Decision	Alternative Specifications
(1) Controls	Include Cistercian houses outside forests (yes/no)
(2) Operationalization	Number of houses or dummy variables
(3) Dependent variable	1290–1801, 1377–1801, 1600–1801

seems reasonable to simply estimate OLS models with inforest as regressor. This can be rationalized as inforest already representing the best approximation of the desired first-stage outcome, i. e. the exogenous variation in the number of Cistercian houses. Therefore, I estimate the OLS model

$$\Delta \log(L_{t+1}) = \beta_1 + \beta_2 \text{inforest}_i + \beta_3 \log(L_{it}/X_i) + \beta_4 M_i + \mathbf{Z}_i'\beta + \varepsilon_i \tag{3}$$

including the same controls as in the previous specifications. This model involves (at least) three challengeable decisions (Table 6). It could be argued that the models should include the number of Cistercian houses outside forests as well, for their presence may have affected the propensity of Cistercian foundations in forests (see Figure 18 in Appendix C for a DAG presentation of the causal structure). Moreover, since there are only two counties with two settlements and thirteen with one, it could be reasonable to use a dichotomized variable instead. Most trivially, the model can be estimated for all three observation periods.

Figure 6 presents the relevant coefficients of the 12 ($2 \times 2 \times 3$) resulting models. Only one of these (M6) even approached significance at the 10 % level. Comparing the coefficients of models M4–M6 as well as M10–M12 suggests that Cistercian presence outside forests is more reliably associated with later population growth than Cistercian presence inside forests. The former association cannot be interpreted as strong evidence in favor of a causal effect because the possibility of unobserved confounding was the motivation for using IV estimation in the first place.

General OLS results

The bottom line of the previous section is that it proved difficult to convincingly estimate a causal effect of Cistercian presence on population growth. The small sample size and the weakness of the instruments severely limit the usefulness of the IV research design. In light of this, it is worthwhile to briefly revisit the general OLS estimates of ABDS:

First, if the relationship between Cistercian presence and population growth is robust to arbitrary specification choices, the hypothesis is at least not falsified by the best available descriptive historical data. Second, the theory brought forward by ABDS will gain additional plausibility if the relationship is unique to the Cistercian order, i. e. all other orders do not show a robust correlation with population growth. Lastly, ABDS

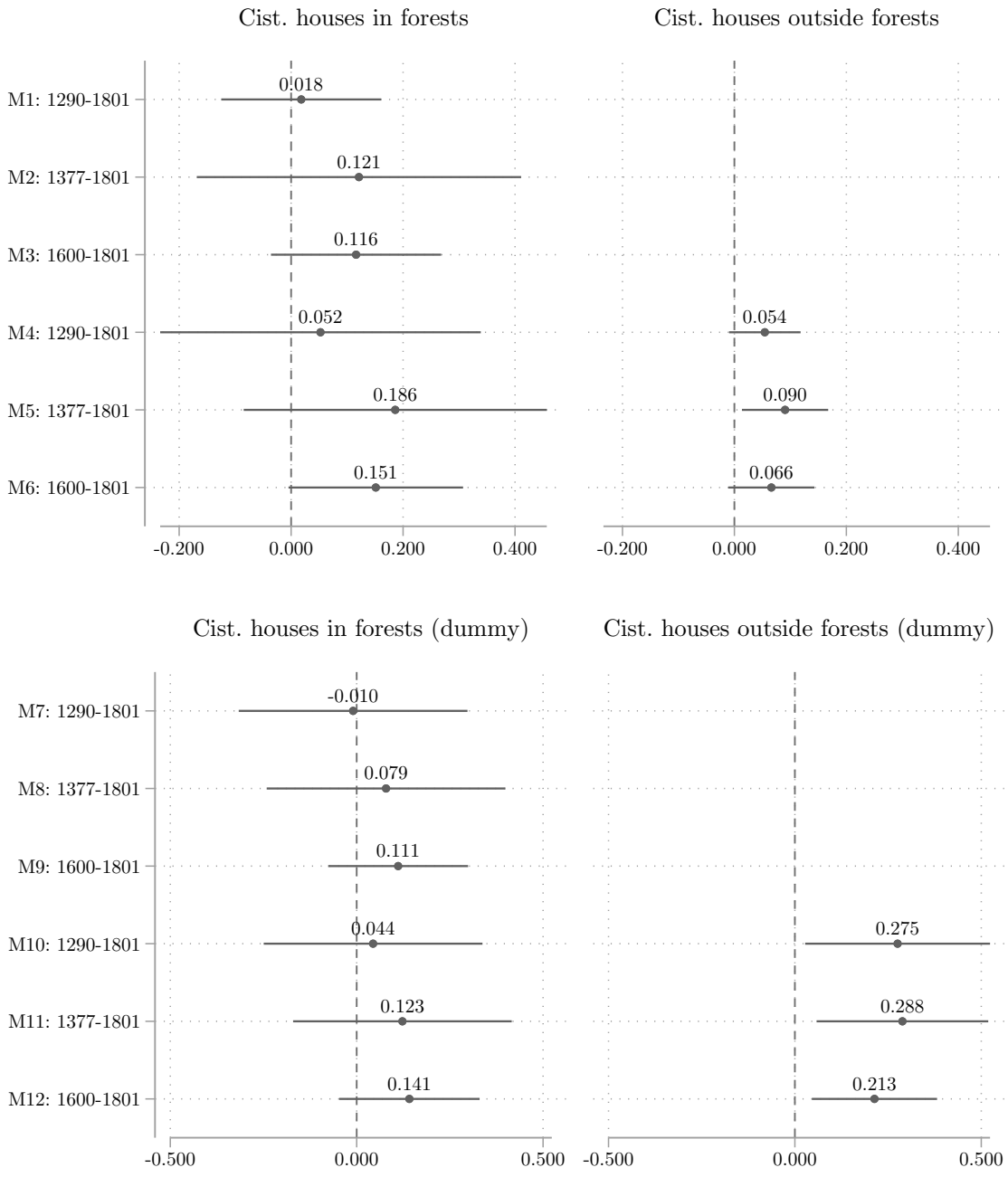


Figure 6: Results of OLS regressions ($N = 40$). The figure shows coefficients of 12 specifications. The upper subfigure shows results for models that use the number of monasteries. The lower subfigure shows results for models that use dummy variables. Bars indicate the 90% confidence interval based on heteroskedasticity-robust standard errors. Exogenous controls included in all models: population density at the start of the period, the total number of monasteries, land quality, the suitability for pasture, and forest share.

Table 7: Reasonable specifications of the OLS model.

Decision	Alternative Specifications
(1) Controls	Land quality, Rivers (length/area), County area (log), Coal, Coastal (= 1), Roman Road density (length/area,) Suitability for pasture (% of area), Literacy rate 1851
(2) Cistercian influence indicator	Cist. share, Cist. presence (0/1), Cistercians (total), CCCP, Houses per population, Cist. density, Cist. manors, Cist. Granges
(3) Dependent variable	1290–1801, 1377–1801, 1600–1801
(4) Region fixed effects	Yes/no (8 regions)

provide evidence for a correlation of contemporary value orientations and former Cistercian settlements in Europe. Again, if this correlation is robust to specification choices as well as unique to the Cistercian order, this, in combination with the historical evidence from England, should at least be considered an interesting puzzle.

The OLS specifications are based on Table 2 and Table 3 (Andersen et al., 2017a: 1773, 1776). ABDS used four indicators of Cistercian presence, ten control variables in various combinations, and region fixed effects.[23]

All models of the specification curve analysis include two basic controls: First, the population density at the start of the growth period. Second, either the total number of religious houses or the total number of manors (if the indicator is Cistercian manors or granges). Table 7 gives an overview of the remaining decisions. In principle there are $2^8 \times 8 \times 3 \times 2 = 12{,}288$ combinations. However, the small sample size limits the usefulness of certain model choices. More specifically, I exclude all combinations that result in fewer than 27 residual degrees of freedom.[24] That means, I do not estimate models with fixed effects that include more than two additional covariates besides the two basic control variables. Moreover, because data on the literacy rate in 1851 is not available for 3 counties, the analysis does not include models with the full set of controls. The effective number of estimated models is 6,648. The descriptive specification curves in Figure 7 and 8 give a very rough overview of the results. I start by focussing on Figure 7 which displays results on ABDS's four indicators.

The figures allow some very general conclusions. First, the sign of the effect is consistently positive. Second, the average size of the point estimates is comparable to the reported effect sizes in Tables 2 and 3 (Andersen et al., 2017a: 1773, 1776). The average effect size of the coefficients of the dummy variable for Cistercian presence is 0.229, close to the 0.262 reported by ABDS in their baseline specification. The average effect size of Cistercian

[23]Admittedly, these choices do not constitute the complete "multiverse" of plausible specifications. ABDS report additional variants in the online appendix of their article, e. g. concerning the functional form of initial population density.

[24]This threshold is somewhat arbitrary. ABDS only report models with at least 28 degrees of freedom.

share is 1.440, almost similar to the effect size reported by ABDS for the model with full controls in Table 3 (1.562). The number of Cistercian monasteries in a county ("Cistercians (total)") – my preferred indicator – has an average effect size of 0.066, slightly lower than the estimates reported by ABDS (0.084 and 0.088). Third, there is variation in the effect size as well as the certainty of the estimate. If statistical significance is used as the decisive criterion for judging the relevance of effects, model choice matters. Fourth, it is possible to identify some patterns determining effect sizes and standard errors. The choice of the dependent variable seems to have the most consistent influence. For example, almost all of the larger and significant estimates for Cistercian share and Cistercians (total) can be found in models using the 1377–1801 growth period as dependent variable. However, these patterns are not consistent across indicators. For example, with respect to the dummy indicator, large effects are found mainly in models using the 1290–1801 period.

The results depicted in Figure 8 are broadly consistent with the general conclusion noted in the last paragraph. The results for Cistercian manors and granges are of special interest because they are intended to test a different causal mechanism (see Section 2.4). By and large, the relationship between these indicators of Cistercian economic (rather than spiritual) activity and later population growth does not seem to be more robust and less variable than the indicators proposed by ABDS. The point estimates of both indicators even turn negative for certain specifications. At the same time, the 133 positive coefficients with the smallest p-values are all from models estimating the effect of Cistercian granges. What these models have in common is the 1600–1801 growth period as dependent variable and that county area is not in the set of controls.[25] Grange coefficients tend to be small or negative for regressions 1) modeling the 1290–1801 growth period 2) controlling for everything except suitability for pasture. Although not more robust than the original indicators, the high rate of significant results for the coefficient of Cistercian granges still suggests the rational management of agricultural resources as a possible alternative channel of Cistercian influence.

Before drawing a final conclusion about the English medieval case, I want to consider one last analysis. The question is: Is this positive, albeit highly variable, relationship unique to the Cistercian order? ABDS present some robustness checks in their appendix (Andersen et al., 2017b). They replace the Cistercian share with the Benedictine share of all monastic houses in Table C3. And in Table C4 they "include the Cistercians alongside the Augustinians, the Cluniacs, the Premonstratensians as well as the Benedictines; only the Cistercians appear to be correlated with population growth over the period" (Andersen et al., 2017a: 1775). However, there is a "multiverse" of possible models for each of these orders that can be systemically studied.

Therefore, I replicate my specification curve analysis except that I replaced the indicators

[25]There are still significant, but slightly smaller estimates for the grange coefficient in models controlling for area.

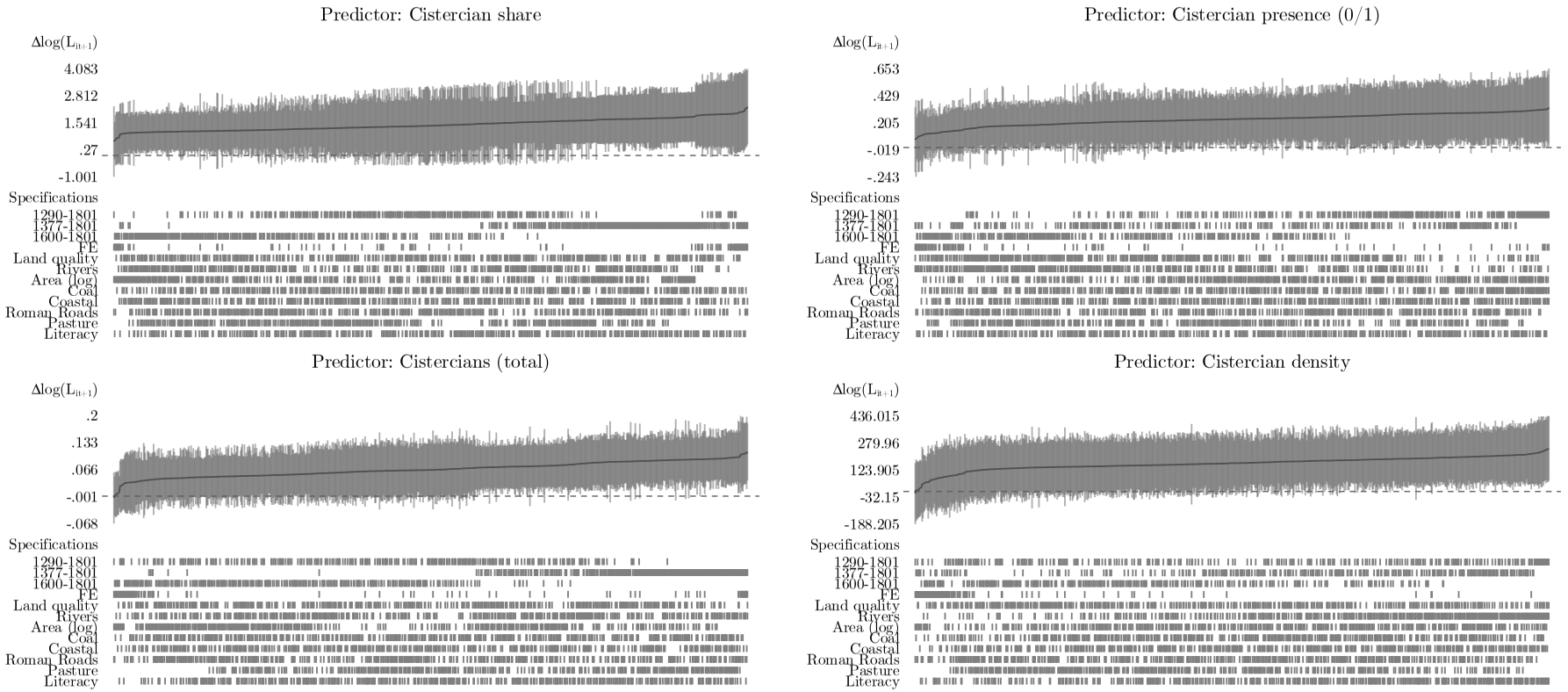


Figure 7: Systematic robustness check of OLS estimates ($N = 40$). The figure shows coefficients from 3324 model specifications. Each sub-figure shows results for a different indicator of Cistercian influence. The point estimates are ranked from weakest to strongest effect with bars indicating the 90% confidence interval based on heteroskedasticity-robust standard errors. The bars below indicate the specification choices for each model (see Table 7). The first three rows indicate the dependent variable. The fourth row indicates whether regional FE are included or not. Rows five to twelve indicate whether a respective covariate was included as a control.

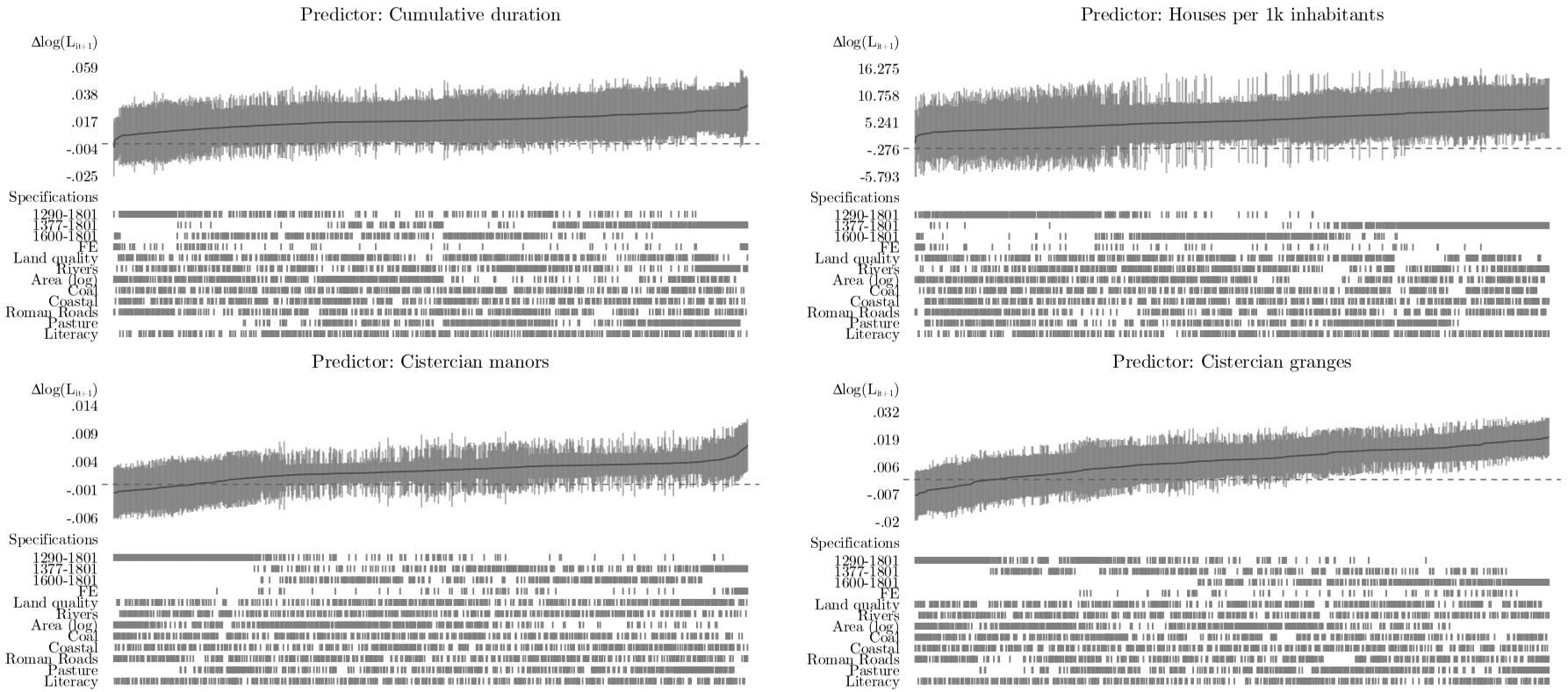


Figure 8: Systematic robustness check of OLS estimates ($N = 40$). The figure shows coefficients from 3324 model specifications. Each sub-figure shows results for a different indicator of Cistercian influence. The point estimates are ranked from weakest to strongest effect with bars indicating the 90% confidence interval based on heteroskedasticity-robust standard errors. The bars below indicate the specification choices for each model (see Table 7). The first three rows indicate the dependent variable. The fourth row indicates whether regional FE are included or not. Rows five to twelve indicate whether a respective covariate was included as a control.

for Cistercian influence with the corresponding indicators for the Augustinian canons. The results can be seen in Figures 9 and 10. Without spending too many words on the details, the overall picture is not too different from the previous finding, i. e. there is a consistently positive effect across all indicators. Some effects are on average smaller, others are larger. The mean coefficient of the Augustinian dummy is 0.569, double the size of the Cistercian dummy, and statistically significant on the 10 %-level in 86 % of the specifications. As Table 8 illustrates, it would be easy to compile a set of confirmatory models including manifold – though selective – robustness checks. Moreover, it is easy to rationalize the choices. For instance, it makes perfect sense to use the 1290–1801 growth period as a baseline because almost all English Augustinian houses were founded before 1290 (203 of the 209 houses in the dataset).

I am not insinuating that ABDS intentionally "p-hacked" their results or were intentionally reporting only confirmatory results. I am merely stressing the difficulty of assessing a general hypothesis with scarce historical data that are very open to different interpretations.

Now it is true that ABDS derived their hypothesis about the Cistercian influence from a (speculative) theoretical argument. Accordingly, one might be willing to assign a higher prior probability to the association of Cistercian presence and population growth to be causal. In contrast, there is no *a priori* reason to believe that the Augustinians were especially conductive to local economic development. Nonetheless, the findings relating to the Augustinian "placebo treatment" could mean that testing the ABDS hypothesis in this way is prone to false positive results.

5.2. Evidence from contemporary Europe

Next, I will partially re-analyze the European-wide data. ABDS proceed in two steps: First, they estimate the relationship between indicators of past Cistercian presence and contemporary values using the 2008–10 wave of the European Values Study (EVS). Further analyses test whether the effect of past Cistercian presence on contemporary employment levels is mediated by regional differences in value orientations.

The EVS models regress *individual* values on *regional* indicators of past Cistercian presence. More specifically, the indicators are computed for the NUTS 2 (sub-)regions where respondents resided at age 14. The dependent variables are answers to two yes/no-questions: whether individuals value hard work and whether they value thrift. It turns out that, controlling for individual characteristics of the respondents, only the relationship with hard work is robust. I focus on the latter in order to keep this section short.

ABDS estimate linear probability models (LPM) with the binary response as dependent variable. In total, there are four indicators of past Cistercian presence at the regional level. The three controls included at the regional level are NUTS 2 units' area (in some models

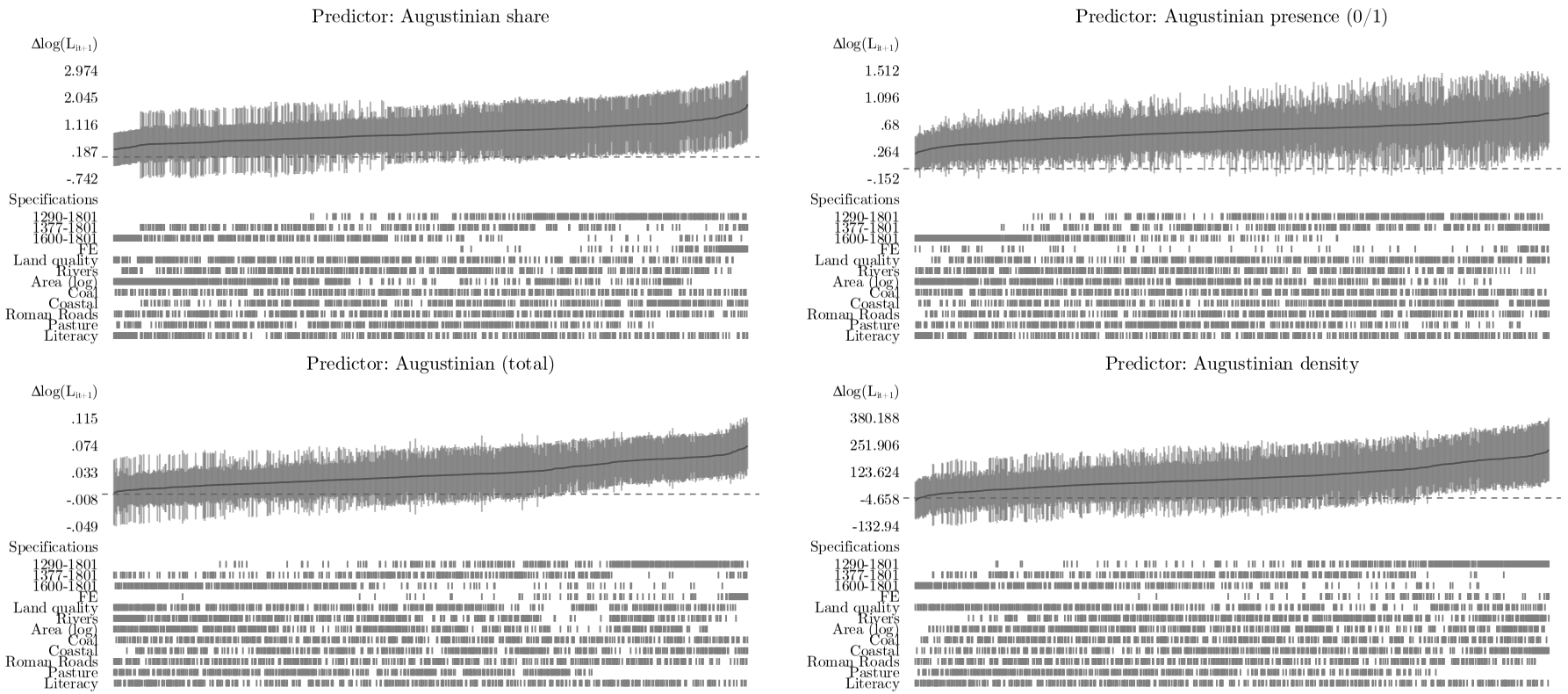


Figure 9: Systematic robustness check of OLS estimates ($N = 40$). The figure shows coefficients from 3324 model specifications. Each sub-figure shows results for a different indicator of Augustinian influence. The point estimates are ranked from weakest to strongest effect with bars indicating the 90% confidence interval based on heteroskedasticity-robust standard errors. The bars below indicate the specification choices for each model (see Table 7). The first three rows indicate the dependent variable. The fourth row indicates whether regional FE are included or not. Rows five to twelve indicate whether a respective covariate was included as a control.

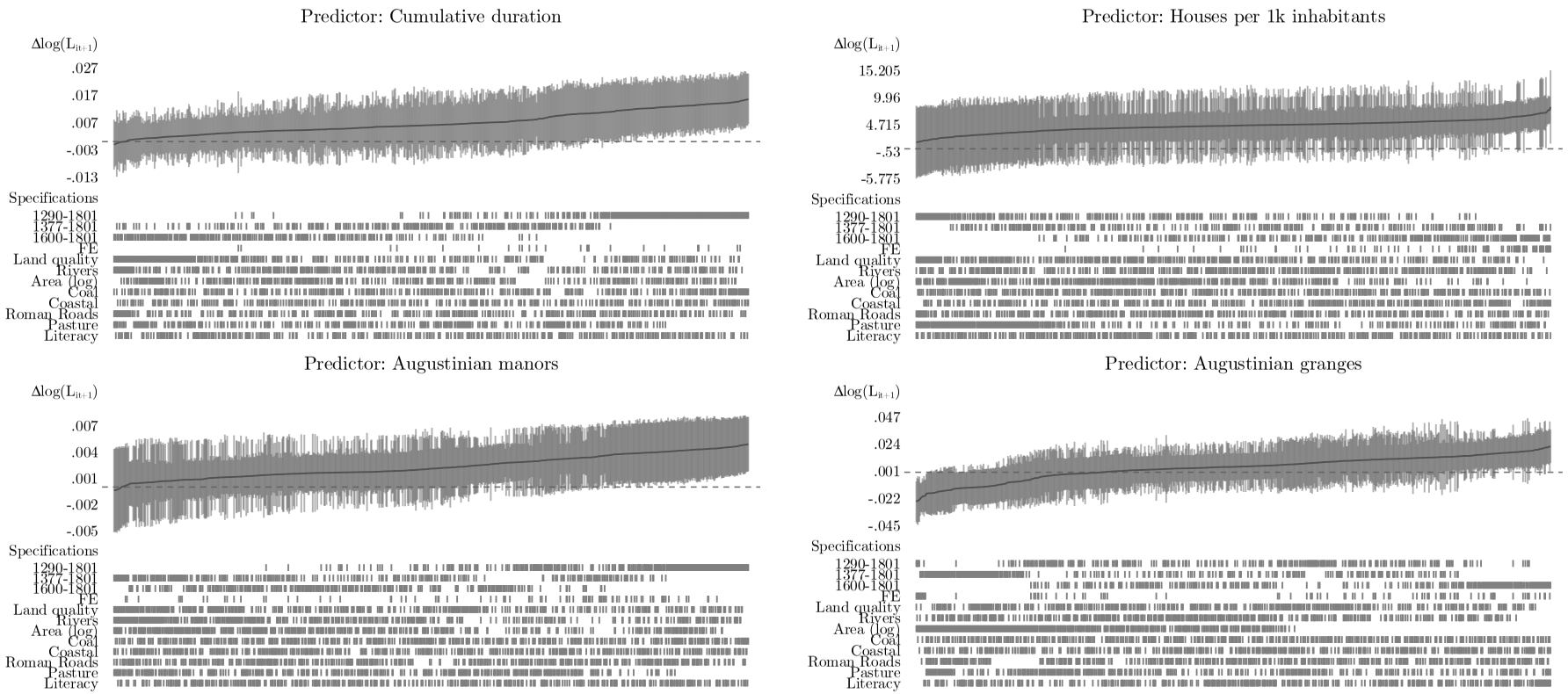


Figure 10: Systematic robustness check of OLS estimates ($N = 40$). The figure shows coefficients from 3324 model specifications. Each sub-figure shows results for a different indicator of Augustinian influence. The point estimates are ranked from weakest to strongest effect with bars indicating the 90% confidence interval based on heteroskedasticity-robust standard errors. The bars below indicate the specification choices for each model (see Table 7). The first three rows indicate the dependent variable. The fourth row indicates whether regional FE are included or not. Rows five to twelve indicate whether a respective covariate was included as a control.

Table 8: Selected OLS models of Augustinian influence. Column names indicate the starting year of the time period under consideration (1290–1801 or 1377–1801). The table lists unstandardized regression coefficients with heteroskedasticity-robust standard errors in parentheses.

	(1) 1290	(2) 1290	(3) 1290	(4) 1290	(5) 1377	(6) 1290	(7) 1290
Augustinian houses	0.057**				0.043+	0.060**	0.078**
	(0.020)				(0.022)	(0.018)	(0.024)
Augustinian share		1.088+					
		(0.579)					
August. presence (0/1)			0.773*				
			(0.342)				
Augustinian density				160.878*			
				(67.122)			
Pop. density 1290 (log)	−1.037**	−1.056**	−1.002**	−1.157**		−1.140**	−1.010**
	(0.175)	(0.171)	(0.168)	(0.176)		(0.190)	(0.196)
Total number of religious houses	−0.016**	−0.002	−0.005	−0.005	−0.015*	−0.013*	−0.028**
	(0.006)	(0.005)	(0.004)	(0.004)	(0.006)	(0.006)	(0.010)
Land quality	−0.326	−0.282	−0.530	−0.334	−0.606*	−0.381	−0.392
	(0.316)	(0.317)	(0.356)	(0.324)	(0.243)	(0.301)	(0.296)
Pop. density 1377 (log)					−0.779**		
					(0.208)		
Coastal (=1)						−0.114	−0.075
						(0.144)	(0.145)
Roadshare (length / county area)						3.753+	3.624+
						(1.990)	(1.964)
Rivers (length/area)						−0.579	−0.633
						(2.201)	(2.061)
Cistercians							0.107*
							(0.049)
Constant	4.223**	3.996**	3.493**	4.431**	3.457**	4.424**	3.981**
	(0.591)	(0.596)	(0.590)	(0.570)	(0.710)	(0.600)	(0.617)
N (counties)	40	40	40	40	40	40	40
R^2	0.709	0.717	0.738	0.712	0.603	0.729	0.758

+ $p \leq 0.1$, * $p \leq 0.05$, ** $p \leq 0.01$. Robust standard errors in parentheses.

Data source: Andersen et al. (2017) and English Monastic Archives.

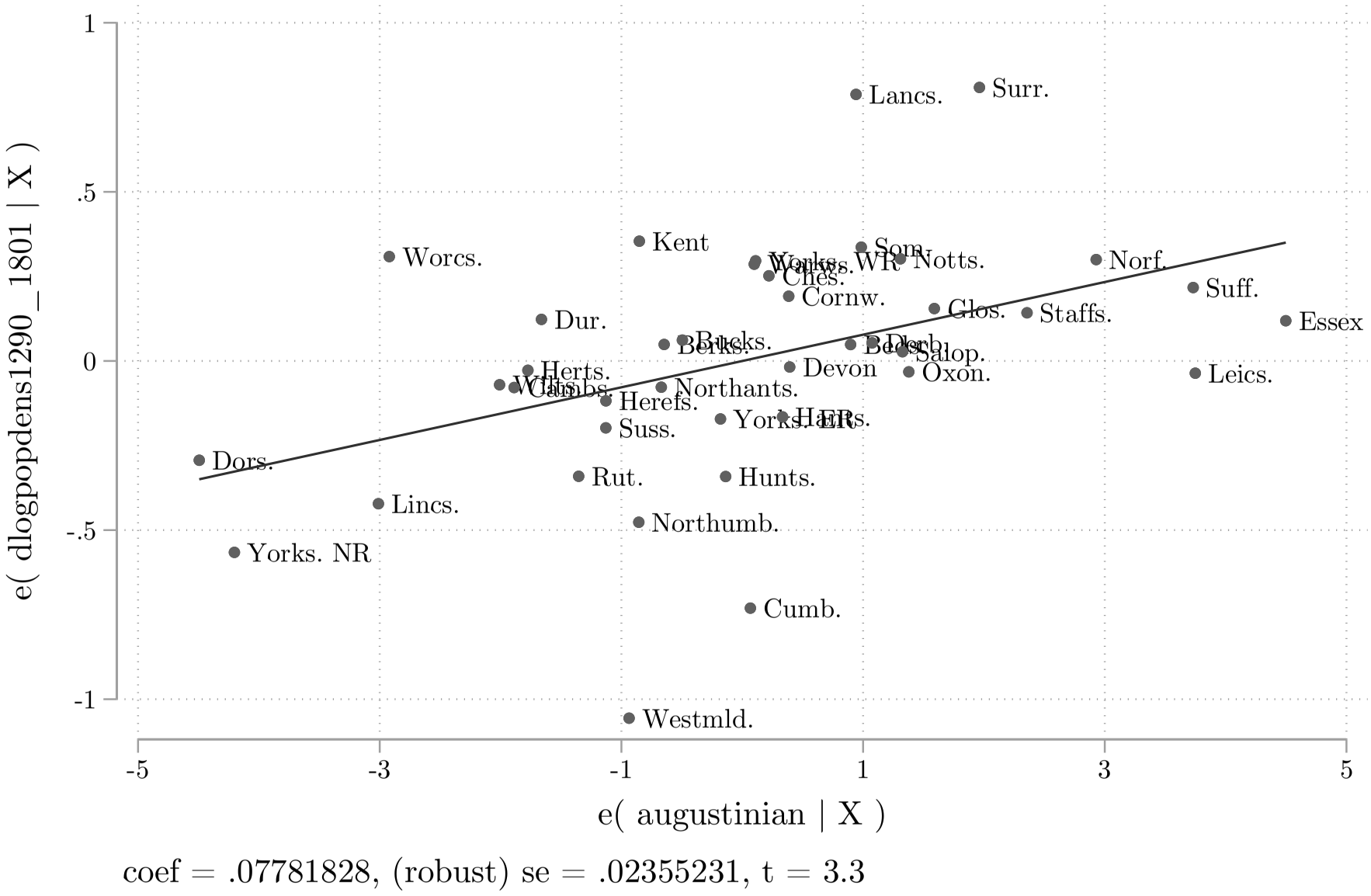


Figure 11: The partial correlation between Augustinians and population growth, 1290–1801. The plot illustrates the regression results reported in Table 8, column (7).

Table 9: Original and alternative reasonable specifications of "Values in Europe" model.

Decision	Original specifications	Alternative Specifications
(1) Influence Indicator	Density, presence (0/1), # houses, # houses (log)	(same but all combinations)
(2) Order	Cistercians	Carthusians, Cluniac, Premonstratensians
(3) Baseline controls	Yes/no	(same but all combinations)
(4) Religion dummys	Yes/no	(same but all combinations)
(5) Area	Area, area (log)	(same but all combinations)
(6) Sample	Full, Catholics and Orthodox	Catholics only
(7) Model	LPM	Probit

the natural logarithm of their area) and the absolute latitude as well as longitude of the NUTS 2 units' centroids. They include individual level controls (*baseline controls*) for age, age squared and dummy variables for the attributes male, married, and educational attainment. Additionally, they include dummies for religious adherence in some regression models. Lastly, they include country fixed-effects.[26]

I modify, add and exclude certain choices for the specification curve analysis, namely: I add data on three other orders as explained in Section 4.2. Moreover, since the dependent variable is binary, the specification curve analysis estimates every specification as an LPM as well as a probit model. I report average marginal effects for the latter kind of models. There is also a small inaccuracy in the original model description. ABDS write they "sometimes restrict the sample to Catholics only" (Andersen et al., 2017a: 1783, notes to Table 5), even though their syntax file reveals that they use a sample compromised of Catholics as well as Orthodox respondents.[27] Therefore, I re-estimate the models with the combined sample as well as the exclusively Catholic sample. However, I do not estimate models with a sample restricted to England and neither do I exclude respondents for Midland "when this is an outlier" (p. 1785).[28] Table 9 summarizes the specification decisions. In total, there are $(4 \times 4 \times 2 \times 2 \times 2 \times 3 \times 2) - 128 = 768 - 128 = 640$ reasonable specifications per dependent variable according to the choices listed in Table 9.

[26]ABDS use standard errors clustered at the country level. Since the treatment is assigned at the NUTS 2 level, it could be argued that standard errors should be clustered at the regional level as well (Abadie et al., 2017). I refrain from doing so in order to stay close to their original approach.

[27]They never discuss this particular choice. What is more, the exclusion of Protestants is based on the idea that their agreement with the values of hard work and thrift might be the results of the diffusion of the "Protestant Ethic" *sensu* Weber (Andersen et al., 2017a: 1782). However, Weber stressed the general diffusion of the Capitalist Spirit beyond the initial adherents of Calvinism. At the same time, ABDS chose England for their historical case study. England is a predominantly Protestant country and ABDS nonetheless claim that regional differences in economic development can be explained by past Cistercian presence. It seems arbitrary to argue that the lasting influence of Cistercian values is not detectable in the contemporary European Protestant population even if the Protestant respondents were socialized in a region whose general culture still reflects past Cistercian influence.

[28]Since the findings for the English case are not robust, this sub-sample analysis is of lesser interest.

128 model variants are redundant because it is unnecessary to include dummies for religious adherence if the sample is exclusively Catholic.

The results are shown in Figure 12. The coefficient of Cistercian influence is positive across all four operationalizations. With the exception of the binary presence indicator, Cistercian presence consistently exhibits the strongest relationship with the value of hard work compared to the other three orders. In the cases of density and number of houses, most coefficient estimates for Cistercian presence are many times larger than for other orders as well as statistically significant at the 10% level. Moreover, effect sizes of those two indicators are stable across models, e. g. the average coefficient estimate for Cistercian houses is 0.0038 within a range of 0.0029 (min.) to 0.0048 (max.). Hence, one additional Cistercian house in the NUTS 2 region of childhood residence increases the probability of valuing hard work by approximately 0.3 to 0.5 percentage points.[29] The coefficient estimates for Carthusians tend to be negative and are strongest for the binary indicator as well as the natural logarithm of houses.

I deem this robust pattern across different specifications as the strongest evidence in favor of the "Cistercian ethics" hypothesis, especially because the Cistercians are the only order that consistently shows such a clear, positive relationship. However, the reader has to keep in mind that the evidence for the other dimension of work ethics, thrift, is much weaker (see Figure 13). The coefficient estimates for the Cistercian influence indicators are not statistically significant, not always positive, and not consistently larger than the estimates for other orders.

Economic Outcomes

The last step in ABDS's line of argument is the connection to contemporary economic outcomes: employment and GDP in European regions. The unit of analysis is not longer the individual but the NUTS 2 region. They hypothesize that Cistercian historical influence should increase the absolute employment of regions but not their GDP (Andersen et al., 2017a: 1790).[30] Here I will not focus on a multiverse re-analysis. Instead I focus on the missing link between contemporary values and employment.

In ABDS's analyses, employment, but not productivity, at the sub-regional (NUTS 2) level is associated with past Cistercian presence, confirming their expectations (Andersen et al., 2017a: 1787–1790) and the authors conclude that the results "paint a coherent picture." However, if higher employment in today's world is still the result of higher productivity in the past, and higher productivity was the result of certain value orientations, it seems logical that the influence of past Cistercian presence on contemporary employment

[29] The average marginal effects produced by the probit models are instantaneous rates of change and only approximate the unit change.

[30] "Incipient labour productivity differences, and thus wage differences, are what theoretically drives mobility, leading to a reallocation of employment. In this process, labour productivity is reduced in the high productivity regions due to diminishing returns" (Andersen et al., 2017a: 1790).

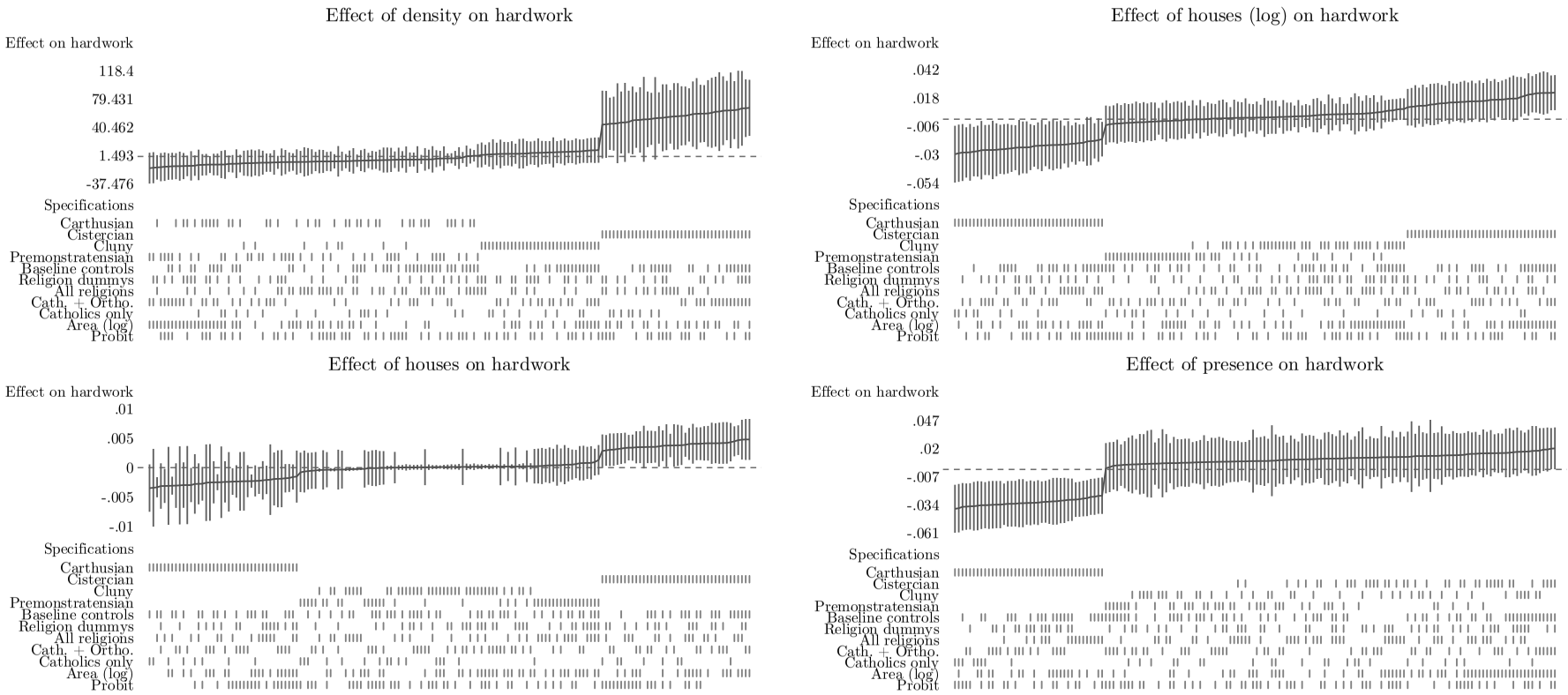


Figure 12: Systematic robustness check of OLS and probit estimates. The figure shows coefficients from 640 model specifications. Each sub-figure shows results for a different indicator of Cistercian influence. The point estimates are ranked from weakest to strongest effect with bars indicating the 90% confidence interval based on clustered standard errors. Estimates for probit models are based on average marginal effects. The sample size varies from 18,083 (subsample with only Catholics, full model with FE, regional, and baseline controls) to 32,358 (full sample, basic model with FE and control variable for NUTS 2 area). The bars below indicate the specification choices for each model (see Table 9). The first four rows indicate the order whose historic influence is being tested. The fifth row indicates whether individual level baseline controls are included. The sixth row refers to the inclusion of dummies for respondents' religious adherence. The seventh to ninth rows indicate the religious composition of the (sub-)samples. The tenth row indicates if the NUTS 2 area is included in natural (km^2) or logarithmic units. The eleventh row indicates whether probit models rather than LPM are estimated.

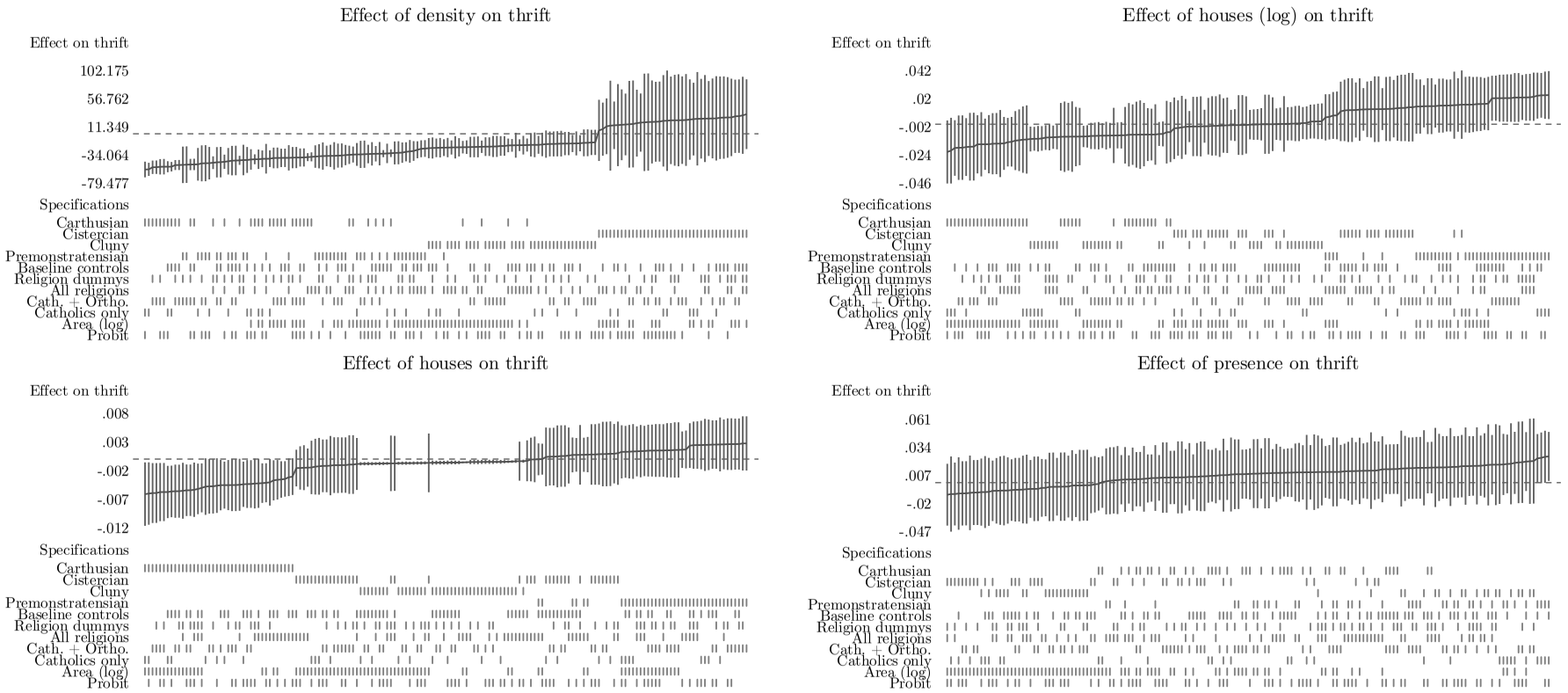


Figure 13: Systematic robustness check of OLS and probit estimates. The figure shows coefficients from 640 model specifications. Each sub-figure shows results for a different indicator of Cistercian influence. The point estimates are ranked from weakest to strongest effect with bars indicating the 90% confidence interval based on clustered standard errors. Estimates for probit models are based on average marginal effects. The sample size varies from 18,083 (subsample with only Catholics, full model with FE, regional, and baseline controls) to 32,358 (full sample, basic model with FE and control variable for NUTS 2 area). The bars below indicate the specification choices for each model (see Table 9). The first four rows indicate the order whose historic influence is being tested. The fifth row indicates whether individual level baseline controls are included. The sixth row refers to the inclusion of dummies for respondents' religious adherence. The seventh to ninth rows indicate the religious composition of the (sub-)samples. The tenth row indicates if the NUTS 2 area is included in natural (km^2) or logarithmic units. The eleventh row indicates whether probit models rather than LPM are estimated.

should at least be partially mediated by regional differences in the contemporary attitude towards hard work and thrift. Such a mediation effect would be much stronger evidence in favor of the proposed explanation. Their appendix contains Table C12, labeled "Horse Race," summarizing models wherein various measures of Cistercian influence are included alongside the shares of EVS respondents within the NUTS 2 regions who affirmed the value of either hard work or thrift (Andersen et al., 2017b: 15). In those models, the indicators of Cistercian influence are strongly and significantly associated with employment but not regressors measuring mean regional work ethic.

Table 10 presents results of a mediation analyses. The selection of covariates is based on the models reported by Andersen et al. (2017b: 13, 15) in Table C9 (mediator model) and Table C12 (outcome model). Cistercian density is strongly associated with the mean value orientation at the sub-regional level (NUTS 2). However, only the association between Cistercian density and the value of hard work is statistically significant at conventional levels, even though the point estimates are roughly similar. In this regard, the regional-level analyses resemble the individual-level results. More importantly, the relationship between Cistercian density and employment in 2007 is unaffected by the inclusion of mean value orientations. Moreover, even when omitting the indicator of past Cistercian presence, value orientations are only marginally and insignificantly associated with employment. Thus, Cistercian density predicts both employment as well as value orientations but the latter do not seem to be mediating variables. Figure 14 illustrates the partial relationship of hard work and employment based on model 6 from Table 10.

Using the dummy coded Cistercian presence as treatment and the mean value of hard work as mediator, Table 11 summarizes the results of a causal mediation analysis that is based on explicit assumptions about the causal structure as well as a flexible estimation strategy which includes possible treatment-mediator interaction effects (see Appendix H for details).[31] The decomposition into a natural direct (NDE) and indirect effect (NIE) is in line with results from Table 10. The NIE is essentially zero while past Cistercian presence increases the total number of employed persons aged 15–64 by 2.6 %. However, it should be obvious that regions with past Cistercian presence might differ from regions without former presence for reasons that could be correlated with the size of a region's contemporary labor force.

6. Conclusion

The general result of my re-assessment is that we should be far less certain about the "pre-Reformation roots" of a strict work ethic than ABDS want us to believe. The evidence is mixed and correlational in nature. The specification curve analysis of the IV results for

[31] Attempts to conduct a mediation analysis using Cistercian density as a continuous treatment failed because the models did not converge.

Table 10: Regression models testing regional differences in mean value orientation as a possible causal channel for differences in employment.

	(1) Hardwork (mean)	(2) Thrift (mean)	(3) log(Empl. 07)	(4) log(Empl. 07)	(5) log(Empl. 07)	(6) log(Empl. 07)
Cistercian density	77.565*	79.304	57.535*	63.478**		
	(31.807)	(65.778)	(21.933)	(21.748)		
Hardwork (mean)				0.023	0.035	0.008
				(0.051)	(0.055)	(0.057)
Thrift (mean)				−0.128	−0.119	
				(0.084)	(0.084)	
Area, km^2	−0.000	0.000	0.000	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Latitude	−0.027**	−0.014	0.016	0.017	0.016	0.016
	(0.008)	(0.017)	(0.016)	(0.015)	(0.014)	(0.014)
Longitude	0.014	0.006	−0.005	−0.005	−0.005	−0.005
	(0.011)	(0.013)	(0.008)	(0.007)	(0.007)	(0.008)
Age (mean)			−0.000	0.001	0.001	0.000
			(0.001)	(0.001)	(0.001)	(0.001)
(Log) Population 2007			0.989**	0.989**	0.992**	0.992**
			(0.013)	(0.013)	(0.014)	(0.014)
Constant	1.599**	0.960	−1.435+	−1.463*	−1.471*	−1.440*
	(0.469)	(0.797)	(0.726)	(0.699)	(0.669)	(0.669)
NUTS1 FE	✓	✓	✓	✓	✓	✓
N (NUTS2 regions)	241	241	241	241	241	241
R^2	0.901	0.586	0.996	0.996	0.996	0.996
No scale effects p-value			0.396	0.385	0.577	0.571

+ p ≤ 0.1, * p ≤ 0.05, ** p ≤ 0.01. The Standard errors in parentheses are clustered at the country level.

Data source: Andersen et al. (2017).

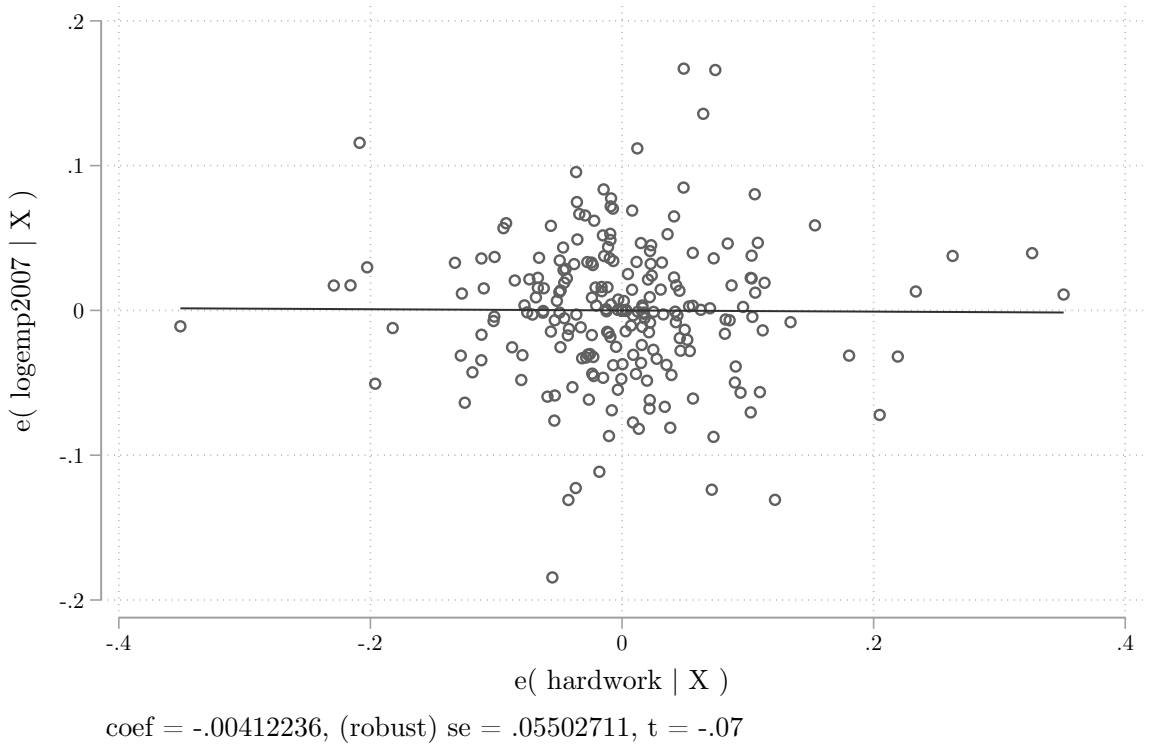


Figure 14: The partial correlation between hardwork and the logarithm of the total number of employed persons aged 15–64 in NUTS 2 regions. The plot illustrates the regression results reported in Table 10, column (6).

England revealed the original results to depend on a disputable operationalization of the instrument. In summary, all attempts to estimate a causal effect using a credible research design have failed to produce strong confirmatory evidence.

At the same time, there is a stable and strong association between Cistercian presence and population growth across most OLS models. But considering that I am able to find equally strong and stable associations for at least one other order – the Augustinian canons –, it seems advisable to be cautious in interpreting the results as evidence for the diffusion of Cistercian values. This is especially true bearing in mind the lack of direct historical evidence. The estimated associations could be the result of confounding by other unobserved factors. The most plausible indicator pertaining to Cistercian economic activity (granges) shows a strong positive relationship with population growth in many plausible specifications. The finding is consistent with the hypothesis that the association between Cistercian presence and population growth is driven by the diffusion of technological and managerial practices rather than work ethics. Since granges were often administered by lay brothers and choir monks, this association could still be the result of attitudes toward work spreading to the surrounding population – it is not decisive evidence – but the Cistercians' more tangible actions to increase the effectiveness of their agrarian estates are well documented (Lekai, 1977: 293–297).

Table 11: Results of a causal mediation analysis (Appendix H). The total effect of the binary treatment (past Cistercian presence) is decomposed into a natural indirect effect and a natural direct effect.

	Outcome: (Log) employment 2007
Natural indirect effect (NIE) via hardwork	−0.000
	(0.001)
Natural direct effect (NDE) of Cistercian presence	0.026**
	(0.009)
Total effect (TE) of Cistercian presence	0.026**
	(0.009)
N	241

+ p ≤ 0.1, * p ≤ 0.05, ** p ≤ 0.01. Standard errors (in parentheses) are clustered at the country level. Unstandardized coefficients based on linear models. Data source: Andersen et al. (2017).

The correlation between past Cistercian presence and the contemporary value of hard work merits further research, though it seems imperative to establish the causal channel more directly, e. g. to connect the contemporary pattern to historical growth processes. Weber motivated his Protestantism thesis, *inter alia*, by pointing out that predominantly Calvinist and Puritan regions in Europe like England, the Netherlands and Geneva were the "most advanced economically" (Weber, 2005: 5).[32] Likewise, one of the reasons ABDS choose to study England's historical development is that it "later turned out to be the epicentre of the Industrial Revolution" (Andersen et al., 2017a: 1757). However, the core area of the Cistercians before and after the Reformation was France with at least 241 foundations until 1500 according to my own data. The question naturally arises why France did not industrialize earlier.

Moreover, a mediation analysis failed to find evidence for a causal path from contemporary values to employment. The share of respondents in a region that value hard work and thrift is not connected to regional employment. At the same time, the coefficient of Cistercian presence, irrespective of operationalization, is undiminished by the inclusion of value orientations. If Cistercian influence on contemporary employment is not mediated by contemporary value orientations, it becomes more difficult to make sense of the totality of findings.

The study also has to be placed in the context of other research. It is my impression that the literature on the interrelation of religious values and economic growth should proceed more systematically. For example, research by Akçomak et al. (2016: 824) found no connection between Cistercian presence and economic growth in the early modern Netherlands (see Appendix G for a brief re-examination with regard to the Cistercians).

[32]The main discussion in the first chapter titled "Religious Affiliation and Social Stratification" focuses on the statistical association between Protestantism and educational as well as occupational choices.

Instead they argue that the religious movement of the *Brethren of the Common Life* (BCL) stimulated human capital accumulation during the early Renaissance. Dutch cities with BCL presence had higher rates of literacy and higher levels of book production. They also find evidence of an impact on city growth. A cynic might wonder whether we can find at least one order or religious movement in every European country whose geographic distribution correlates with regional patterns of economic growth for some time period. Kelly (2019) raises the possibility that many results from the literature on regional economic persistence are the outcome of "fitting spatial noise." However, given the central role of Christian institutions in European history, it still seems possible that their cultural legacy left a measurable imprint on regional development. In my view, this research question has to be addressed in a systematic manner and, importantly, without overstating the theoretical guidance available to the analyst of historical data. The general lessons apply more widely to research in economic history and historical social research.

First, weak theories should be accompanied by systematic robustness checks. Results might be vulnerable to slightly different *ad hoc* decisions on a forking path (see e. g. Steegen et al., 2016). Moreover, it seems advisable to collect data on as many orders in as many regions as possible. A comparative approach reduces the dangers of small samples and selective case studies, in which a correlation can easily be the result of random variation. Furthermore, historical social research has to combine different pieces of evidence. Andersen et al. (2017a) as well as Akçomak et al. (2016) should be applauded for their efforts to back the proposed mechanisms with additional evidence like survey results on values (Andersen et al.) or network data as well as data on literacy and book production (Akçomak et al.). However, this may not be enough and especially the role of qualitative historical evidence should be reconsidered: Social scientists (including economists) have to take historical scholarship seriously. The theoretical and historical discussion revealed that certain aspects of the Cistercian doctrine have been selectively construed and presented in a misleading way. Thus, social scientists should read historical research extensively and be aware of the ambiguity of historic sources, the pitfalls of over-generalization, and changes in the consensus among historians. This is not the same as objecting to the need for abstraction and simplification in quantitative historical research.

Another lesson, in my view, is the legitimate role of exploratory research. Describing an interesting phenomenon or hinting at possible explanations with purely correlational evidence is important work as long as the authors are transparent about the limitations of their theory and evidence. A study that limits itself to establishing an interesting regularity as an *explanandum* for future research is preferable to one with overly confident claims about causation. Withstanding the pressure to produce causal estimates can under certain conditions be conductive to the quality of research. The toolkit of causal analysis is nonetheless indispensable because it allows to identify the limitations of exploratory work. When in fact feasible, causal research designs are invaluable tools. But the combination

of correlational evidence based on high quality data, a strong theory, and qualitative historical case studies might be superior to the merely ritualistic applications of estimation strategies.

References

Abadie, A., S. Athey, G. W. Imbens, and J. Wooldridge (2017). When should you adjust standard errors for clustering? *National Bureau of Economic Research*.

Akçomak, I. S., D. Webbink, and B. ter Weel (2016). Why did the Netherlands develop so early? The legacy of the brethren of the common life. *The Economic Journal 126*(593), 821–860.

Andersen, T. B., J. Bentzen, C.-J. Dalgaard, and P. Sharp (2017a). Pre-reformation roots of the Protestant ethic. *The Economic Journal 127*(604), 1756–1793.

Andersen, T. B., J. Bentzen, C.-J. Dalgaard, and P. Sharp (2017b). Technical appendix to pre-reformation roots of the Protestant ethic. *The Economic Journal 127*(604), 1756–1793.

Angrist, J. D. and J.-S. Pischke (2010). The credibility revolution in empirical economics: How better research design is taking the con out of econometrics. *Journal of Economic Perspectives 24*(2), 3–30.

Bazeley, M. L. (1921). The extent of the English forest in the thirteenth century. *Transactions of the Royal Historical Society 4*, 140–172.

Becker, S. O., S. Pfaff, and J. Rubin (2016). Causes and consequences of the Protestant Reformation. *Explorations in Economic History 62*, 1 – 25.

Becker, S. O. and L. Woessmann (2009). Was Weber wrong? A human capital theory of Protestant economic history. *The Quarterly Journal of Economics 124*(2), 531–596.

Bengtson, H. and V. Milojčić (Eds.) (1995). *Großer historischer Weltatlas. Zweiter Teil: Mittelalter*. München: Bayerischer Schulbuchverlag.

Berman, C. H. (2012). Agriculture and economies. In M. B. Bruun (Ed.), *The Cambridge Companion to the Cistercian Order*, pp. 112–124. Cambridge: Cambridge University Press.

Bernard of Clairvaux (2016). Against the most wicked vice of ingratitude. In M. Casey (Ed.), *Monastic Sermons*, pp. 151–158. Collegeville, MN: Cistercian Publications.

Boranbay, S. and C. Guerriero (2019). A novel dataset on a culture of cooperation and inclusive political institutions in 90 european historical regions observed between 1000 and 1600. *Data in Brief 27*, 1–4.

Braun, N., T. Wolbring, and M. Keuschnigg (2012). *Wirtschaftssoziologie I: Grundzüge* (2 ed.). München: Oldenbourg.

Brodeur, A., N. Cook, and A. Heyes (2020a). Methods matter: p-hacking and publication bias in causal analysis in economics. *American Economic Review 110*(11), 3634–60.

Brodeur, A., N. Cook, and A. Heyes (2020b). A proposed specification check for p-hacking. *AEA Papers and Proceedings 110*, 66–69.

Burton, J. and J. Kerr (2011). *The Cistercians in the Middle Ages*. Woodbridge: Boydell Press.

Campbell, B. M. S. (2008). Benchmarking medieval economic development: England, Wales, Scotland, and Ireland, c.12901. *The Economic History Review 61*(4), 896–945.

Cantoni, D. (2015). The economic effects of the Protestant Reformation: Testing the weber hypothesis in the german lands. *Journal of the European Economic Association 13*(4), 561–598.

Clark, G. and Y. V. D. Werf (1998). Work in progress? The industrious revolution. *The Journal of Economic History 58*(3), 830–843.

Coleman, J. S. (1986). Social theory, social research, and a theory of action. *American Journal of Sociology 91*(6), 1309–1335.

Courtney, P. (1980). The monastic granges of leicestershire. *Transactions of the Leicestershire Archaeological and Historical Society 56*, 33–45.

Donkin, R. A. (1960). The Cistercian settlement and the English royal forests. *Cîteaux: Commentarii Cistercienses 11*, 39–50, 117–132.

Donkin, R. A. (1963). The cistercian order in medieval England: Some conclusions. *Transactions and Papers (Institute of British Geographers)* (33), 181–198.

Donkin, R. A. (Ed.) (1978). *The Cistercians: Studies in the Geography of Medieval England and Wales*. Toronto: Pontifical Institute of Medieval Studies.

Dyer, C. C. (1994 [2001]). 'The retreat from marginal land': The growth and decline of medieval rural settlements. In *Everyday Life in Medieval England*, pp. 13–26. London: Hambledon and London.

Felton, C. and B. M. Stewart (2022). *Handle with Care: A Sociologist's Guide to Causal Inference with Instrumental Variables*. SocArXiv. osf.io/preprints/socarxiv/3ua7q.

Fouka, V. and A. Schläpfer (2020). Agricultural returns to labour and the origins of work ethics. *The Economic Journal 130*(628), 1081–1113.

Gelman, A. and J. Carlin (2014). Beyond power calculations: Assessing type s (sign) and type m (magnitude) errors. *Perspectives on Psychological Science 9*(6), 641–651.

Gibson, K., K. Karterouli, and E. Turnator (2022). Dominican houses 1216–1500. In *Digital Atlas of Roman and Medieval Civilizations*. Online: `https://darmc.harvard.edu`.

Goldthorpe, J. H. (2021). *Pioneers of Sociological Science: Statistical Foundations and the Theory of Action*. Cambridge: Cambridge University Press.

Griffin, E. (2020). Work and leisure. In D. Simonton and A. Montenach (Eds.), *A Cultural History of Work in the Age of Enlightenment*, pp. 169–184. London: Bloomsbury Academic.

Guerriero, C. (2020). Endogenous institutions and economic outcomes. *Economica 87*(346), 364–405.

Henrich, J. (2020). *The WEIRDest People in the World: How the West Became Psychologically Peculiar and Particularly Prosperous*. New York: Farrar, Straus and Giroux.

Hostie, R. (1983). *The Life and Death of Religious Orders. A Psycho-sociological Approach*. Washington, D.C.: Center for Applied Research in the Apostolate.

Jedin, H., K. S. Latourette, and J. Martin (Eds.) (1987). *Atlas zur Kirchengeschichte*. Freiburg im Breisgau: Herder.

Jones, G. (2010). A common of hunting? Forests, lordship and community before and after the Conquest. In *Forests and Chases of Medieval England and Wales c 1000–c 1500: towards a survey and analysis.*, pp. 36–67. Oxford: St John's College Research Centre.

Kaelber, L. (1998). *Schools of Asceticism. Ideology and Organization in Medieval Religious Communities*. University Park: Penn State University Press.

Kelly, M. (2019). The standard errors of persistence. CEPR Discussion Paper No. DP13783.

Kershaw, I. (2023). What was an Augustinian grange? the evidence from Bolton Priory's estates. *Yorkshire Archaeological Journal 95*(1), 124–141.

Kienzle, B. M. (2012). Preaching. In M. B. Bruun (Ed.), *The Cambridge Companion to the Cistercian Order*, pp. 245–257. Cambridge University Press.

Kieser, A. (1987). From asceticism to administration of wealth. Medieval monasteries and the pitfalls of rationalization. *Organization Studies 8*(2), 103–123.

Kurze, D. (1980). Die Bedeutung der Arbeit im zisterziensischen Denken. In *Die Zisterzienser. Ordensleben zwischen Ideal und Wirklichkeit*, pp. 179–202. Bonn: Rheinland-Verlag.

Langton, J. (2010). Medieval forests and chases: Another realm? In J. Langton and G. Jones (Eds.), *Forests and Chases of Medieval England and Wales c 1000–c 1500: towards a survey and analysis.*, pp. 15–34. Oxford: St John's College Research Centre.

Langton, J. (2015). Royal and non-royal forests and chases in England and Wales. *Historical Research 88*(241), 381–401.

Langton, J. and G. Jones (Eds.) (2010). *Forests and Chases of Medieval England and Wales c 1000–c 1500: towards a survey and analysis.* Oxford: St John's College Research Centre.

Leamer, E. E. (1983). Let's take the con out of econometrics. *The American Economic Review 73*(1), 31–43.

Lekai, L. J. (1977). *The Cistercians: Ideals and Reality*. Kent, Ohio: The Kent State University Press.

McElreath, R. and R. Boyd (2008). *Mathematical Models of Social Evolution: A Guide for the Perplexed*. Chicago: University of Chicago Press.

Millimet, D. (2019). IV in Exactly Identified Models. *How the (Econometric) Sausage is Made*. Posted on June 06, 2019. Last vistited on October 30, 2020. `https://dlm-econometrics.blogspot.com/2019/06/iv-in-exactly-identified-models.html`.

Morgan, S. L. and C. Winship (2014). *Counterfactuals and Causal Inference: Methods and Principles for Social Research* (2 ed.). Cambridge: Cambridge University Press.

Noell, B. (2006). Expectation and unrest among Cistercian lay brothers in the twelfth and thirteenth centuries. *Journal of Medieval History 32*(3), 253 – 274.

Simmons, J. P., L. D. Nelson, and U. Simonsohn (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science 22*(11), 1359–1366.

Simonsohn, U., J. P. Simmons, and L. D. Nelson (2020). Specification curve analysis. *Nature Human Behaviour 4*(11), 1208–1214.

Southern, R. W. (1990 [1970]). *Western Society and the Church in the Middle Ages*. London: Penguin.

Spater, J. and I. Tranvik (2019). The Protestant ethic reexamined: Calvinism and industrialization. *Comparative Political Studies 52*(13-14), 1963–1994.

Staiger, D. and J. H. Stock (1997). Instrumental variables regression with weak instruments. *Econometrica 65*(3), 557–586.

Steegen, S., F. Tuerlinckx, A. Gelman, and W. Vanpaemel (2016). Increasing transparency through a multiverse analysis. *Perspectives on Psychological Science 11*(5), 702–712.

Sundberg, M. (2019). Work practices, normative control and ascetic responsibilization in cistercian monasteries. *Journal of Management, Spirituality & Religion 16*(5), 397–412.

Thomas, K. (1964). Work and leisure in pre-industrial society. *Past & Present 29*(1), 50–62.

Weber, M. (1930 [2005]). *The Protestant Ethic and the Spirit of Capitalism*. New York: Routledge.

Young, C. and K. Holsteen (2017). Model uncertainty and robustness: A computational framework for multimodel analysis. *Sociological Methods & Research 46*(1), 3–40.

Appendix

A. Cistercian forest settlements

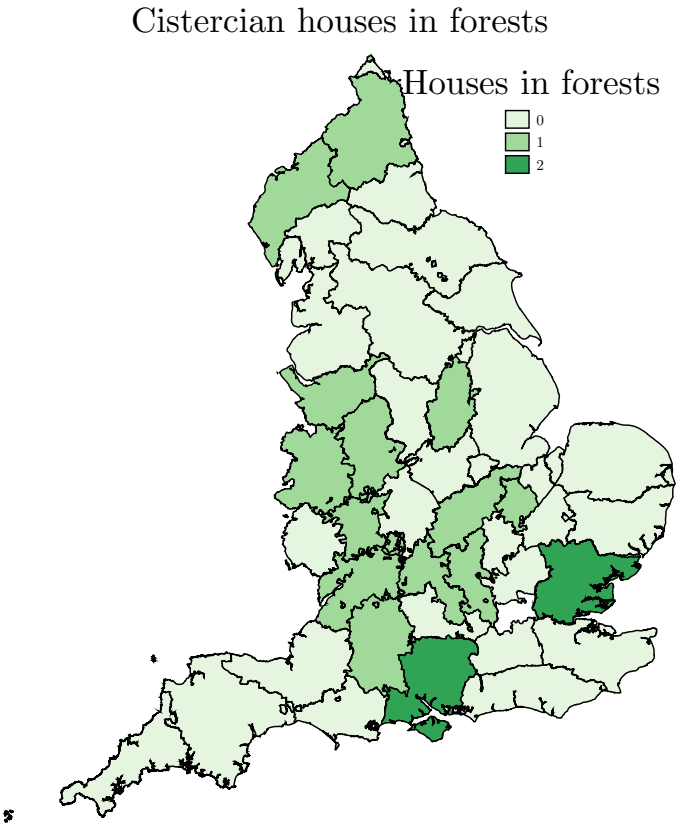


Figure 15: Spatial distribution of Cistercian houses located in forests according to Donkin (1960).

The inclusion of certain houses is based on ABDS sample of monasteries. For example, Radmore Abbey existed for a very brief period (1139–1154) before relocating to another county. However, because ABDS decided to include Radmore in their count, I also code it as a forest settlement. Moreover, Tulketh and Stanlaw were not included by ABDS. The settlement at Tulketh existed for less than five years (1124–1127) but Stanlaw Abbey was populated by Cistercian monks for more than 100 years (1172–1296) before the community relocated to Whalley Abbey and Stanlaw became a grange of the latter. See Donkin (1978: 31–36, 179–180) for a survey of Cistercian site changes in England. Lastly, Donkin (1978: 123) mentions that "Revaulx had marshland rights in the forest of Pickering from the time of Henry II" but does not seem to include it in his list of houses "in or near" royal forests.

Table 12: List of Cistercian monasteries that were located in royal forests according to Donkin (1960, 1978: 121–123).

	Monastery name	Historic county	Comment
1.	Beaulieu	Hampshire	
2.	Biddlesden	Buckinghamshire	
3.	Bordesley	Worcestershire	
4.	Buildwas	Shropshire	
5.	Coggeshall	Essex	
6.	Flaxley or Dean	Gloucestershire	
7.	Holm Cultram	Cumberland	
8.	Netley or Letley or Edwardstow	Hampshire	
9.	Newminster	Northumberland	
10.	Pipewell	Northamptonshire	
11.	Radmore	Staffordshire	Monks relocated to Stoneleigh (Warwickshire)
12.	Rufford	Nottinghamshire	
13.	Sawtry	Huntingdonshire	
14.	Stanley	Wiltshire	
15.	Stratford Langthorne or West Ham	Essex	
16.	Thame	Oxfordshire	
17.	Vale Royal	Cheshire	
Not in analysis sample			
18.	St Mary Graces or Eastminster	London and Middlesex	ABDS exclude county "London and Middlesex" from analyses
19.	Stanlaw or Stanlow	Cheshire	Not in ABDS data, monks relocated to Whalley Abbey (Lancashire)
20.	Tulketh	Lancashire	Priory not in ABDS data, monks relocated to Furness Abbey

B. Medieval forests and chases

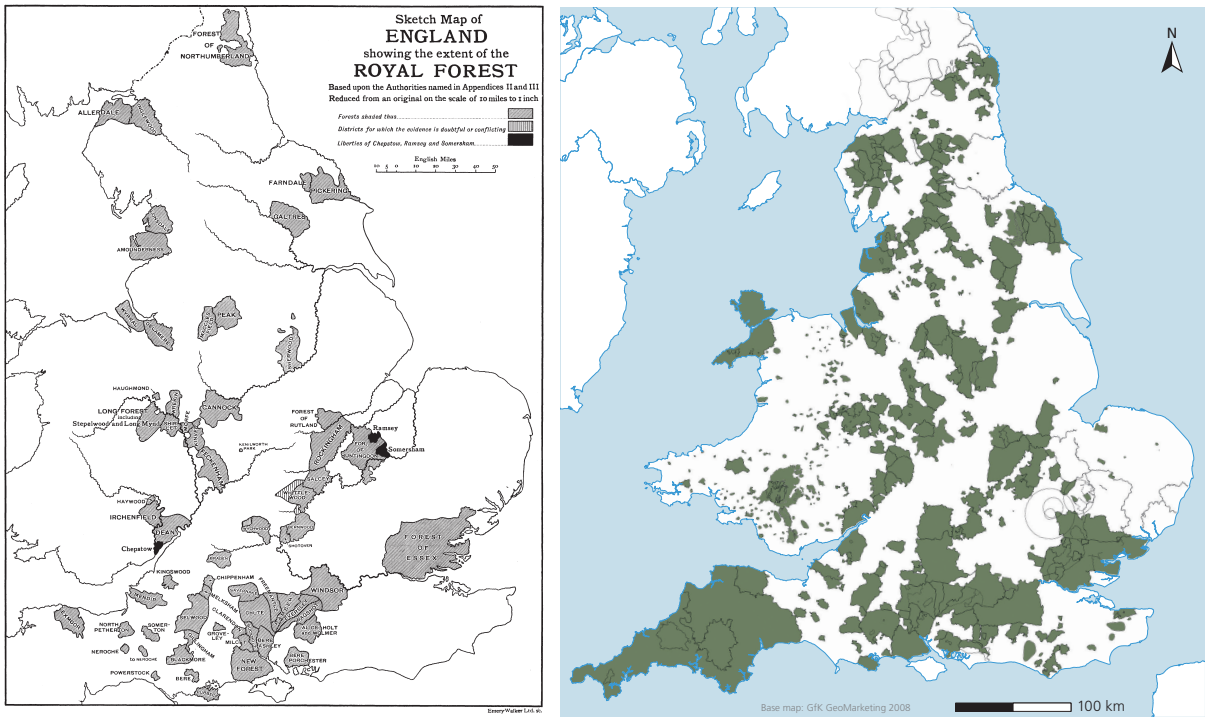


Figure 16: Extent of royal forests according to (on the left) Bazeley (1921) and the extent of medieval forests and chases according to (on the right) Langton and Jones (2010). I thank Graham Jones for making a digital version of his map available to me. Bazeley's work is reproduced with the permission of Cambridge University Press.

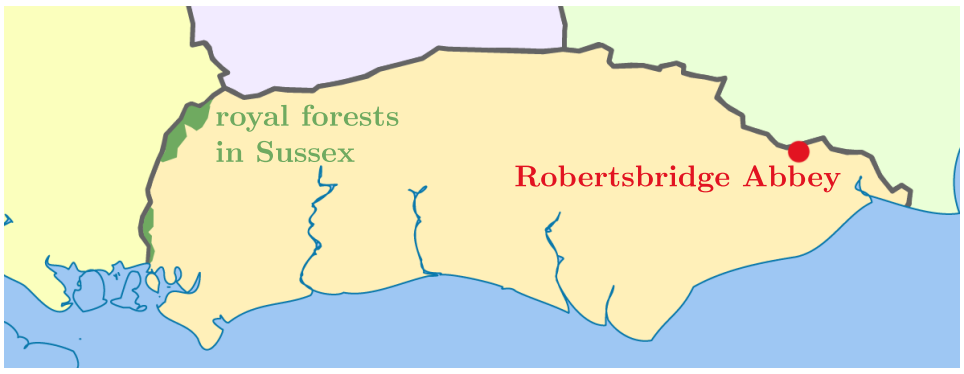


Figure 17: The locations of royal forests and Robertsbridge Abbey in the historic county of Sussex. Own work based on a map provided by Wikimedia user Dr Greg.(`https://commons.wikimedia.org/wiki/File:English_counties_1851_with_ridings.svg`) provided under a Creative Commons Attribution-Share Alike 3.0 Unported license. Forest location drawn according to Bazeley's map (see Figure 16 on the left). The small forest area on the western border is possibly an artifact introduced by the imprecision of a hand-drawn "sketch map" whose county borders do not align perfectly with modern digital maps when overlayed in GIS or graphics software.

C. Alternative DAG

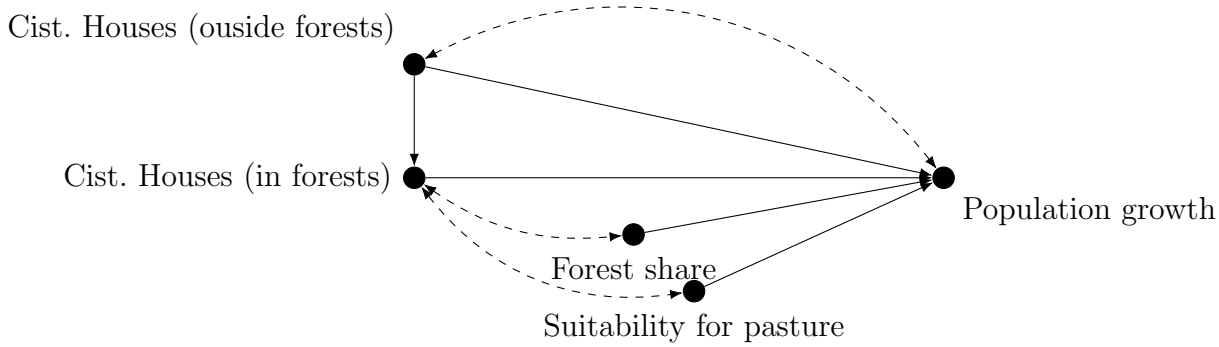


Figure 18: Directed acyclic graph summarizing the causal assumption of the OLS model in Section 5.1. The presence of Cistercian forest settlements is affected by the number of Cistercian houses outside forests. The coefficient of the latter cannot be causally interpreted due to unobserved confounding (the reason for using IV estimation in the first place).

D. Data

Manors and Granges

Information about manors was taken from the English Monastic Archives (`https://www.ucl.ac.uk/history/research/english-monastic-archives`). The EMA contain information on monastic properties. The database distinguishes four types of properties: churches, chapels, manors, and urban property. In total, there are 1,872 Cistercian properties, of which 1,500 are manors. The EMA do not contain complete and systemic information on duration of tenure (e. g. start and end date). Therefore, the figure reflects the total number of manors that have been owned by the Cistercians (or other orders, respectively) at some point in time. The EMA has information about the county in which the manor was located. I use this information to create a count variable at the county level.

I identify granges based on their name, selecting properties containing "Grange" or "grange." 27.87% (418) of the Cistercian manors are coded as granges. As expected, the share of manors whose name contains "grange" is significantly higher among the Cistercians than among all other orders. Only three properties not classified as manors contain the word "grange." These are churches that were located on corresponding manors. I create a count variable of manors named "grange" at the county level.

European Carthusians

The list of all charterhouses was taken from the database on the Analecta Cartusiana webpage (`http://analecta.chartreux.org/catalogue.php`). The Analecta Cartusiana is a series of books on the history and spirituality of the Carthusian monks. Entries on monasteries do not give coordinates. Hence, I geocoded all locations of monasteries founded before the 16th century myself based on the information provided (e. g. by looking up the location in Google Maps). When historical sites were not easily identifiable, I used additional material such as historic maps. In a few cases in which I was unable to research the exact location, I recorded the coordinates of the respective municipality as given by Google Maps. Using a shapefile of the 2006 NUTS regions, I counted the number of charterhouses within the boundaries of each NUTS 2 region (2006 version) based on the geocoded entries.

European Cluniacs

In contrast to the Cistercians, the Congregation of Cluny was a lose federation of monasteries without a clear formal structure. Hence it is difficult to assess the membership of houses. The locations of all "Cluniac sites" were taken from the Clunypedia mapping project (`https://clunypedia.com/map`). I downloaded all database entries (state at 2020/05/06) and kept sites classified as priory, abbey, cell, ermitage, or college. The cleaned dataset has entries on 1,255 monastic communities in the sample region that were founded or reformed by Cluny up until 1500. Using a shapefile of the 2006 NUTS regions, I counted the number of monastic communities within the boundaries of each NUTS 2 region (2006 version) based on the coordinates provided by the Clunypedia project.

European Premonstratensian

The locations of the Premonstratensian houses were taken from the *Digital Atlas of Roman and Medieval Civilization* (DARMC, `https://darmc.harvard.edu/`). The respective DARMC map is based on Jedin et al. (1987: 54) as well as Bengtson and Milojčić (1995: 28). Using a shapefile of the 2006 NUTS regions, I counted the number of monasteries within the boundaries of each NUTS 2 region (2006 version) based on the coordinates provided by the DARMC.

E. Power analysis

The power analysis is conducted using the approach documented by Kang et al. (2021). I focus on the basic 2SLS results but Kang et al. (2021: 11) also provide formulae and

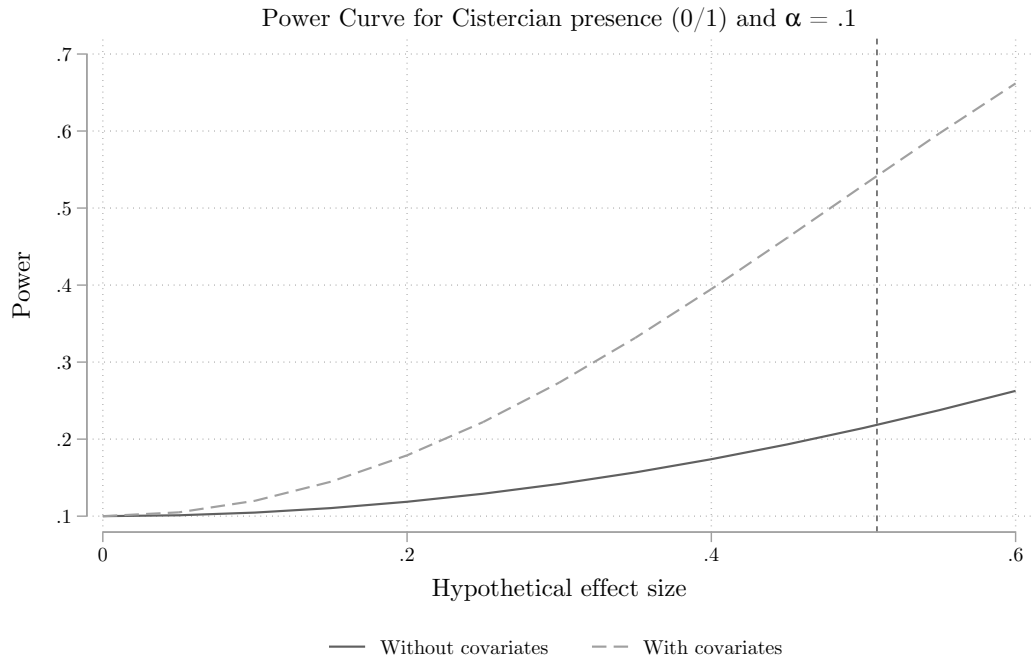


Figure 19: Power analysis for various hypothetical effect sizes of Cistercian presence on the logarithm of population growth 1377-1801. The vertical line indicates the observed effect.

implementations for the Anderson-Rubin test. All computations were done using Stata syntax translated from Kang et al.'s R-command `IVpower`.

In the first stage, an indicator of Cistercian influence π in historic county i is regressed on the binary instrument Rforest, i. e. the presence of royal forests in county i, as well as other covariates $\mathbf{X}$

$$\pi_i = \beta_1 + \beta_2 \text{Rforest}_i + \mathbf{X}_i'\beta + \nu_i.$$

The target quantity is the power of a two-sided t-test in the second stage to reject the null hypothesis that Cistercian presence had no influence on population growth $H_0 : \gamma_2 = 0$

$$\Delta \log(L_{t+1}) = \gamma_1 + \gamma_2 \hat{\pi}_i + \mathbf{X}_i'\gamma + \varepsilon_i.$$

Without covariates, power is calculated using the formula (Kang et al., 2021: 10–11)

$$\text{Power} = \Phi\left(-z_{1-\alpha/2} - \frac{\delta\rho\sqrt{n \cdot \text{Var}(\pi)}}{\sigma}\right) - \Phi\left(z_{1-\alpha/2} - \frac{\delta\rho\sqrt{n \cdot \text{Var}(\pi)}}{\sigma}\right)$$

where $\Phi(\cdot)$ denotes the cumulative standard normal distribution function, $z_{1-\alpha/2}$ the critical value for the desired significance level α, δ the assumed size of the causal effect, σ^2 is the error variance of the second stage[33], and n is the sample size. Furthermore, $\text{Var}(\pi)$

[33] $\widehat{\sigma}^2 = \text{RSS}/(n - L - k)$ where L is the number of instruments (here $L = 1$) and k equals the number

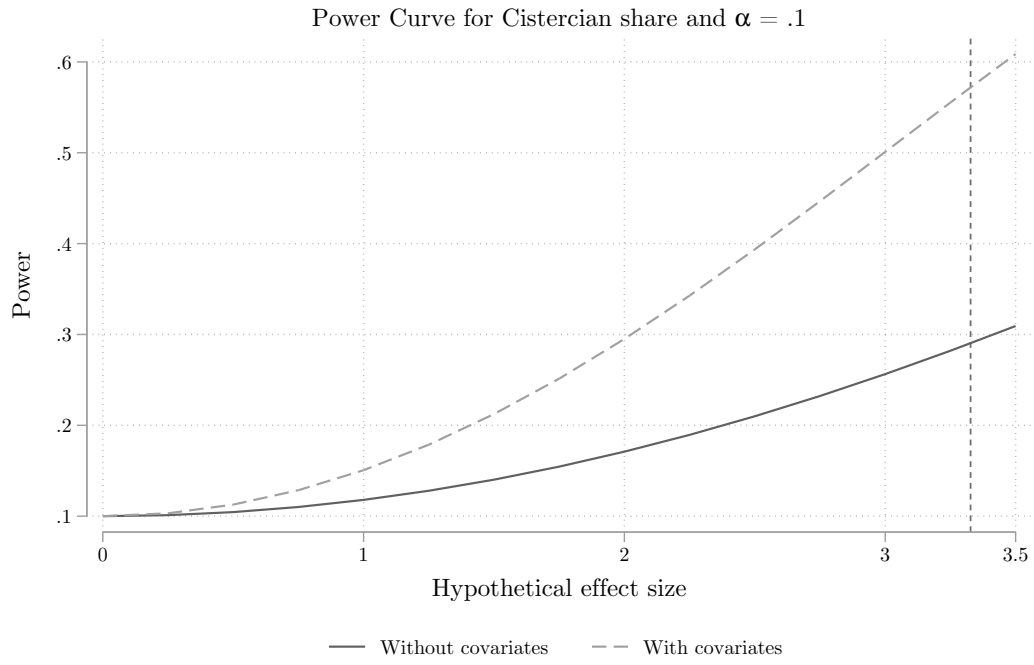


Figure 20: Power analysis for various hypothetical effect sizes of Cistercian share on the logarithm of population growth 1377-1801. The vertical line indicates the observed effect.

is the variance of the endogenous treatment and ρ is the correlation between instrument and endogenous treatment. If the model contains covariates $\mathbf{X}$, $\text{Var}(\pi)$ is replaced with the variance of the adjusted treatment, projecting out $\mathbf{X}$. Likewise, ρ is replaced with the correlation between the instrument and the treatment after both were adjusted for covariates $\mathbf{X}$. This is not documented in the paper but implemented in the same way in the R-command `IVpower` which is part of the R-package `ivmodel` (Kang et al., 2021).

Figures 19 and 20 show the results for $\alpha = 0.1$ and various hypothetical effect sizes. The power of the 2SLS estimator without including further covariates is well-below generally desired levels for all effect sizes. For example, an effect of 0.5 for the binary treatment variable (Cistercian presence) has to be considered large for it means that the presence of Cistercians before the dissolution of the monasteries increased subsequent population growth by more than 60 % compared to counties without Cistercian presence. The power improves when the calculation considers the actual control variables from the 2SLS models but stays below generally recommended levels of 80 %.

of covariates (Kang et al., 2021: 6).

F. Bias analysis plots for the exclusion restriction

This section briefly investigates the sensitivity of the original IV results to possible violations of the exclusion restriction. I focus on column 3 from Table 4 (Andersen et al., 2017: 1779), one of ABDS preferred model specifications. The effect of Cistercian share on the logarithm of population growth between 1377 and 1801 is estimated to be 3.327 using Rforest as an instrument.

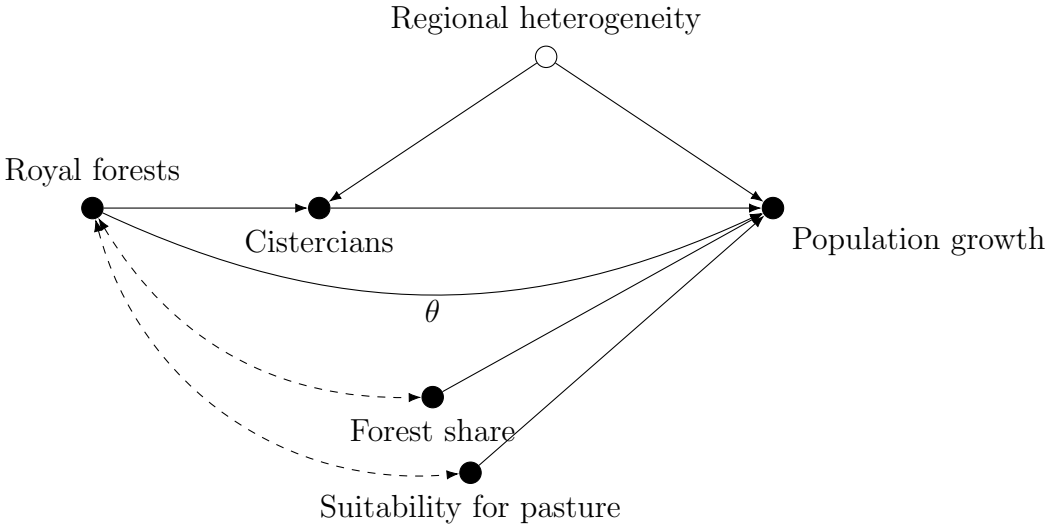


Figure 21: Directed acyclic graph with violated exclusion restriction. The instrument (*Rforest*) has a direct causal effect of size θ on the outcome.

The bias analysis follows the strategy outlined by Felton and Stewart (2022: 22): I specify a range of hypothetical values θ that represent the direct effect of the instrument Z (*Rforest*) on the outcome Y (population growth). The interval $[0; 0.4]$ includes the ITT effect of *Rforest* (.342) on population growth from a reduced form regression. Next, I calculate for every value of θ an adjusted outcome $Y_{adj} = Y - \theta Z$. Lastly, I estimate treatment effects with 2SLS based on the original specification using Y_{adj} instead of Y. The bias specification plots includes 90% confidence intervals based on the robust standard errors as well as the Anderson-Rubin weak instrument χ^2-test. The Anderson-Rubin confidence intervals are based on the version of the test implemented in the Stata ado `rivtest` (Finlay and Magnusson, 2009) using 1000 grid points for estimation. The bias analysis plot is shown in Figure 22. As can be seen, the estimated coefficient turns insignificant at the 10 %-level for fairly low values of θ and thus even small violations of the exclusion restriction. This finding again illustrates a general problem of the analysis of the English case: The small sample causes all estimates to be highly uncertain and sensitive to small violations of model assumptions.

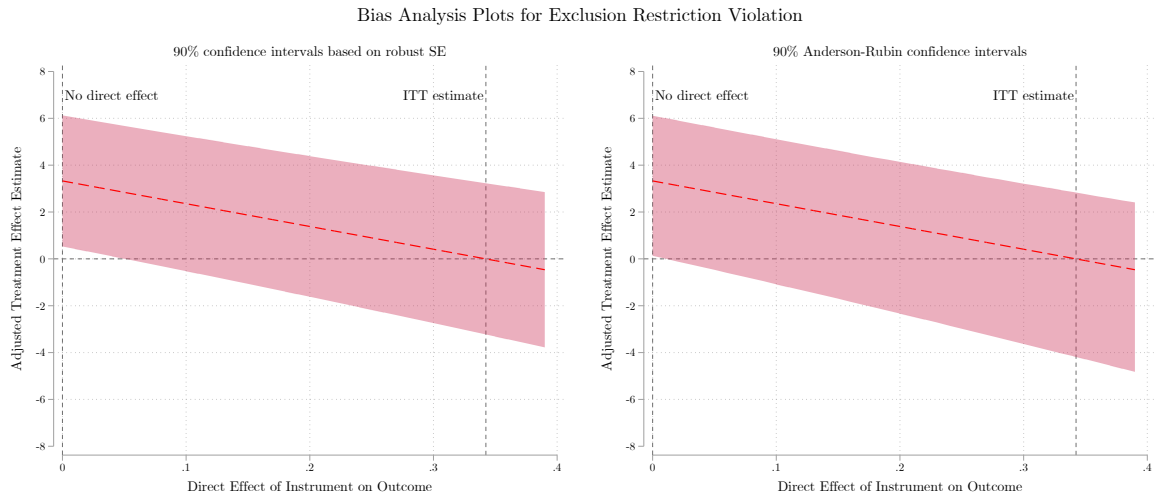


Figure 22: Adjusted effect of Cisterciansshare on the logarithm of population growth 1377–1801 for different values of θ.

G. Evidence of Cistercian legacies in the Netherlands

This paper has focussed the re-assessment of one particular study, although there are others that address similar questions about the long-term influence of religious movements on economic development. It is beyond the scope of this study to explore them in detail. I want to briefly consider evidence by Akçomak et al. (2016) for the role of Christian orders in the early economic development of the Netherlands. They argue that the religious movement of the *Brethren of the Common Life* (BCL) stimulated human capital accumulation during the early Renaissance. Dutch cities with BCL presence had higher rates of literacy and higher levels of book production. They also find evidence of an impact on city growth. Their OLS models show that population growth in the period 1400–1560 was about 35 % higher in cities with BCL communities. They use the distance to Deventer, the movement's city of origin, as an instrument and report an effect on population growth of more than 50 % in their 2SLS estimates.

Addressing Andersen's et al. research, they "find no effects of Cistercians or other religious orders on the economic development of the Netherlands" (Akçomak et al., 2016: 824). Even though they don't report analyses regressing population growth on Cistercian presence, the data provided in the supplementary materials allow some quick OLS calculations. Regressing city growth on a dummy variable of Cistercian presence, using the same control set as their main analyses, gives an effect size of about 24 % with $t = 1.98$. However, Akçomak et al. (2016) do not distinguish between male and female Cistercian monasteries. I recode the dummy variable based on the location of male Cistercian communities to make the results more comparable to Andersen et al. Male Cistercians were located in six of the 67 sampled cities. The coefficient remains nearly unchanged (26 %) but the estimates are less precise ($t = 1.56$). This is (very) weak

evidence for a relationship between Cistercian presence and Dutch economic development. The association is, however, stronger than associations between the other orders/religious movements[34] in the Akçomak et al. data and city growth. Overall, this is consistent with the broader picture: While we cannot reject the hypothesis that the presence of Cistercian monasteries stimulated local economic development in Europe, the evidence is far from being decisive. Whether one is persuaded by the evidence ultimately depends on the prior probability one assigns to the hypothesis.

H. Causal Mediation Analysis

In a causal mediation analysis, the total causal effect (TE) of a binary treatment $D_i \in \{0, 1\}$ and a continuous mediator M is defined as

$$\text{TE} = \text{E}[Y_i(1, M_i(1)) - Y_i(0, M_i(0))]$$

where $M_i(1)$ and $M_i(0)$ refer to the potential outcomes of i on the mediator variable under the two treatment conditions. The total effect can be decomposed into a natural indirect effect (NIE) and a natural direct effect (NDE)

$$\begin{aligned}\text{NIE} &= \text{E}[Y_i(1, M_i(1)) - Y_i(1, M_i(0))] \\ \text{NDE} &= \text{E}[Y_i(1, M_i(0)) - Y_i(0, M_i(0))]\end{aligned}$$

which correspond to non-parametric versions of the direct and indirect effects in traditional mediation analysis but with clarified assumptions regarding their causal interpretation. More precisely, four assumptions are required to identify all three effects: (A1) no unmeasured confounding of the treatment-outcome relationship; (A2) no unmeasured confounding of the mediator-outcome relationship; (A3) no unmeasured confounding of the treatment-mediator relationship; (A4) there is no mediator-outcome confounder that is affected by the treatment (VanderWeele, 2015: 24–26).[35]

In the case at hand, the mediator is the mean value orientation v (hard work) in sub-region (NUTS 2 level) s. The treatment is the dummy coded historical Cistercian presence C in sub-region s

$$v_s = \alpha_1 + \alpha_2 C_s + \mathbf{Z}_s'\zeta + c_r + \varepsilon_s.$$

In line with the models found in the Andersen's et al. appendix, the control set $\mathbf{Z}$ in

[34] Franciscans, Tertiarians, Beghards and Beguines, and Modern Devotion.

[35] There are several other controversial implicit assumptions and requirements. For example, the identification of the NIE as well as the NDE involves cross-world-counterfactuals, i. e. the term $\text{E}[Y_i(1, M_i(0))]$ assumes both the treatment condition $D = 1$ and, at the same time, the mediator value that would be realized under the different treatment condition $D = 0$. See discussion by VanderWeele (2015: 179–183).

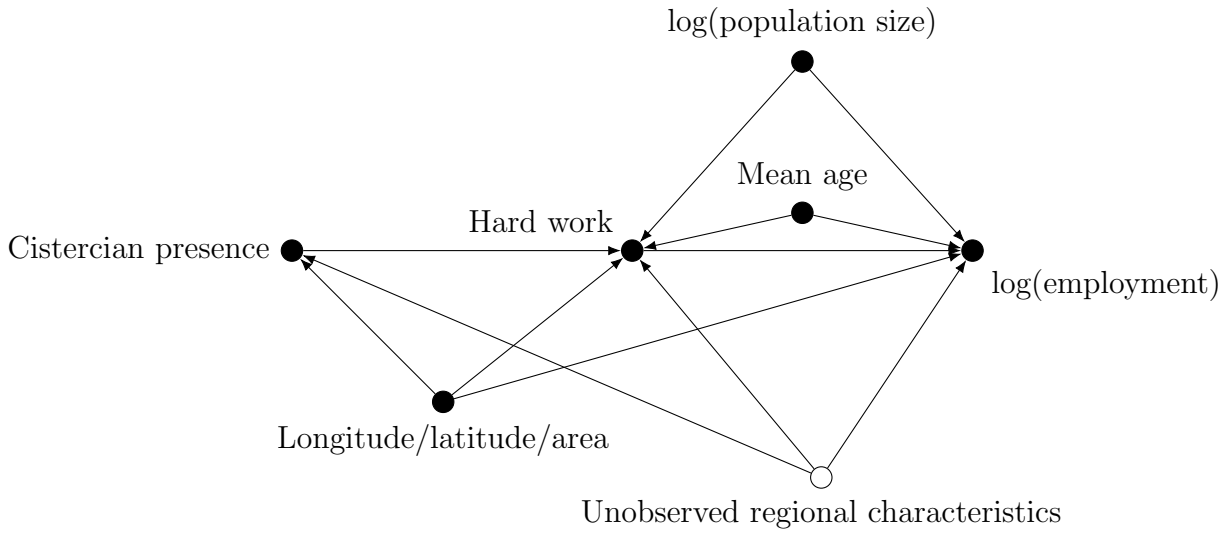


Figure 23: Directed acyclic graph summarizing the causal assumption of the mediation model. Unobserved regional characteristics are controlled for via NUTS 1-level fixed effects.

the mediator model includes only the region's area as well as the centroid latitude and longitude. Moreover, the model includes regional (NUTS 1 level) fixed effects c_r.

Because no parametric assumptions have been imposed, the outcome model includes interactions between the treatment and the mediator. Thus, the model of employment levels for the year 2007 in NUTS 1 regions is

$$\log(E_s) = \beta_1 + \beta_2 C_s + \beta_3 v_s + \beta_4 C_s v_s + \mathbf{W}'_s \gamma + c_r + \nu_s.$$

The set of covariates $\mathbf{W}$ includes the region's area, the centroid latitude and longitude, the mean age of respondents in region, and the natural logarithm of the population size in 2007 in addition to regional fixed effects.

Figure 23 summarizes the causal assumptions as a DAG. Even though they include control variables, ABDS are careful to avoid an explicit causal interpretation of their cross-sectional results. Hence, this mediation analysis makes stronger assumptions than ABDS. However, the idea is to show that – even under such strong assumptions – the analyses are not compatible with the theoretically expected mediating relationship. The mediation analysis was conducted using Stata's `mediate` command (StataCorp, 2023: 222).

References

Akçomak, I. S., D. Webbink, and B. ter Weel (2016). Why did the Netherlands develop so early? The legacy of the brethren of the common life. *The Economic Journal 126*(593), 821–860.

Andersen, T. B., J. Bentzen, C.-J. Dalgaard, and P. Sharp (2017). Pre-reformation roots of the Protestant ethic. *The Economic Journal 127*(604), 1756–1793.

Bazeley, M. L. (1921). The extent of the English forest in the thirteenth century. *Transactions of the Royal Historical Society 4*, 140–172.

Bengtson, H. and V. Milojčić (Eds.) (1995). *Großer historischer Weltatlas. Zweiter Teil: Mittelalter*. München: Bayerischer Schulbuchverlag.

Donkin, R. A. (1960). The Cistercian settlement and the English royal forests. *Cîteaux: Commentarii Cistercienses 11*, 39–50, 117–132.

Donkin, R. A. (Ed.) (1978). *The Cistercians: Studies in the Geography of Medieval England and Wales*. Toronto: Pontifical Institute of Medieval Studies.

Felton, C. and B. M. Stewart (2022). *Handle with Care: A Sociologist's Guide to Causal Inference with Instrumental Variables*. SocArXiv. osf.io/preprints/socarxiv/3ua7q.

Finlay, K. and L. M. Magnusson (2009). Implementing weak-instrument robust tests for a general class of instrumental-variables models. *The Stata Journal 9*(3), 398–421.

Jedin, H., K. S. Latourette, and J. Martin (Eds.) (1987). *Atlas zur Kirchengeschichte*. Freiburg im Breisgau: Herder.

Kang, H., Y. Jiang, Q. Zhao, and D. S. Small (2021). Ivmodel: An R package for inference and sensitivity analysis of instrumental variables models with one endogenous variable. *Observational Studies 7*(2), 1–24.

Langton, J. and G. Jones (Eds.) (2010). *Forests and Chases of Medieval England and Wales c 1000–c 1500: towards a survey and analysis.* Oxford: St John's College Research Centre.

StataCorp (2023). *Causal Inference and Treatment-Effects Estimation Reference Manual Release 18*. College Station, TX: Stata Press.

VanderWeele, T. (2015). *Explanation in Causal Inference: Methods for Mediation and Interaction*. New York, NY: Oxford University Press.