Did Protestantism promote prosperity via higher human capital?

Jeremy Edwards
University of Cambridge and CESifo

March 2018

Abstract. This paper investigates the Becker-Woessmann (2009) interpretation of the Weber thesis - Protestants were more prosperous in nineteenth-century Prussia because they were more literate - and shows that it cannot be sustained. The econometric analysis on which Becker and Woessman based their argument is fundamentally flawed, because their instrumental variable does not satisfy the exclusion restriction. When an appropriate instrumental-variable specification is used, the evidence from nineteenth-century Prussia rejects the human-capital version of the Weber thesis put forward by Becker and Woessmann.

Acknowledgements: I thank Tim Guinnane, Erin Hengel, Alexander Klein and Sheilagh Ogilvie for very helpful comments. I also thank Sascha Becker and Ludger Woessmann for posting their data at http://www.cesifo-group.de/ifoHome/facts/iPEHD-Ifo-Prussian-Economic-History-Database/publications.html and thus enabling me to write this paper.

1. Introduction

Did Protestantism cause economic development? Max Weber's initial formulation of this hypothesis was that Protestantism did so by promoting the work ethic and wealth accumulation. Criticism of this specific argument, however, has led to the Weber thesis now being more commonly interpreted as a general claim that there is some causal connection between Protestantism and economic prosperity. Becker and Woessmann (2009) (henceforth BW) provide an interesting new interpretation of this general form of the Weber thesis. The authors argue, using data from late-nineteenth-century Prussia, that Protestantism did indeed lead to greater prosperity, the reason being that Protestants were better educated than Catholics and thus had more human capital. The human capital advantage of Protestants was an unintentional consequence of the emphasis Protestantism placed on education so that everyone could read the Bible.[1] BW do not rule out the possibility that other aspects of Protestantism, such as a work ethic, contributed to Protestants' greater prosperity, but they argue that higher human capital explains most of the difference, and possibly all.[2]

This paper shows that BW's interpretation of the Weber thesis is not supported by the Prussian data on which it is based. BW use an instrumental-variable (henceforth IV) analysis to show that there was a positive causal effect of Protestantism on literacy, their measure of human capital. They then use an IV analysis to show that there was a positive effect of Protestantism on economic prosperity in a regression model which does not include a measure of literacy. BW interpret these IV estimates as reflecting "the total causal effect of Protestantism on economic outcomes, including any indirect effect running through literacy".[3] What is required, they argue, is an estimate of the effect of Protestantism on prosperity once this indirect effect has been removed. In the absence of an IV for literacy, BW argue that this can be done by subtracting an assumed effect of literacy from observed economic prosperity and using IV to estimate the effect of Protestantism on prosperity after this adjustment for the effect of literacy. The resulting estimated effects of

[1] BW (2009), 539-41.
[2] BW (2009), 532.
[3] BW (2009), 569.

Protestantism vary with the assumed values for effect of literacy, but they are always considerably smaller than the estimated effect obtained without any adjustment for the effect of literacy, and never statistically significant. From this, BW conclude that most of the positive effect of Protestantism on economic prosperity is due to the greater literacy of Protestants.

There are two fundamental flaws in the BW argument. The first is that it is based on a misunderstanding of IV estimation. Suppose that there exists an IV for Protestantism that satisfies the exclusion restriction required to be a valid IV in a regression model that does not include literacy as a regressor. Then IV estimation will give a consistent estimate of the effect of Protestantism on economic prosperity which does not include any indirect effect due to literacy, the omitted variable. This is simply the basic argument for using IV estimation to deal with omitted-variable bias. The BW procedure of assuming values for the effect of literacy on economic prosperity and estimating the effect of Protestantism on prosperity adjusted for these assumed values is unnecessary if a valid IV for Protestantism exists. However, if the IV is valid, it will also give a consistent estimate of the effect of Protestantism on prosperity that is independent of any indirect effect via literacy in a regression model in which the dependent variable is prosperity adjusted for an assumed effect of literacy on prosperity, as section 3 of this paper shows. Thus if BW's IV for Protestantism is valid, there should be no significant difference between the estimated effect of Protestantism in the regression model in which literacy is an omitted regressor and that in a model in which prosperity is adjusted for an assumed effect of literacy. Furthermore, if the IV is valid, the estimated effect of Protestantism should not differ significantly according to the assumption made about the effect of literacy when prosperity is so adjusted, as section 3 also shows. However, BW's results do not satisfy these properties. This constitutes evidence that their IV for Protestantism is not a valid IV in the regression models they estimate.

The second fundamental flaw in the BW argument is that if, as they contend, Protestantism has a causal effect on literacy, then any variable that is correlated with Protestantism cannot be a valid IV for Protestantism in a regression model that omits literacy as a regressor. If Protestantism has a causal effect on literacy, any variable that is proposed as an IV for Protestantism in such a model must necessarily be

correlated with literacy, and so cannot satisfy the exclusion restriction. It is therefore unsurprising that, as noted above, the results BW obtain from their IV analysis provide evidence that their IV is not valid.

In order to obtain consistent estimates of the causal effects of Protestantism and literacy on economic prosperity which allow for a causal effect of Protestantism on literacy, it is essential to have a plausible IV for literacy. Becker et al. (2011) analyse the relationship between education and industrialisation in Prussia using a dataset for the 334 Prussian counties that existed in 1849, and argue that pre-industrial education, for which they have data, can be used as an IV for literacy. In this paper, I combine the dataset used in BW with the dataset used in Becker et al. (2011). This makes it possible to use a plausible IV for literacy and obtain consistent estimates of the causal effects of both Protestantism and literacy on economic prosperity.

The IV for Protestantism used by BW in their regression models is the distance of a county – the unit of observation in their dataset – from Wittenberg, the city where Luther's nailing of 95 theses to the door of the castle church in 1517 is conventionally taken to mark the beginning of the Protestant Reformation. In principle, distance to Wittenberg can serve as an IV for Protestantism, but care must be taken when so using it in regression models of county economic outcomes. A central theme in the literature on nineteenth-century German industrialisation is the importance of regional effects. BW's regression models of economic outcomes in Prussia do not include any variables that measure regional effects, and hence, because distance to Wittenberg is a geographical variable, it will be correlated with these omitted regional effects and is thus likely to be an invalid IV. The paper shows that in order for distance to Wittenberg to be a valid IV for Protestantism it is necessary for the regression model to include measures of regional effects within Prussia.

The paper proceeds as follows. Section 2 of the paper outlines the BW argument. Section 3 discusses the fundamental flaws in it. Section 4 considers the importance of regional effects in Prussian economic development and shows that distance to Wittenberg is correlated with measures of these effects. Section 5 describes the data used for the main analysis in this paper. Section 6 shows that the positive causal effect of Protestantism on literacy is still present in a regression model

in which, in order to allay concerns about the validity of distance to Wittenberg as an IV, measures of regional effects are included. This effect is, however, a very small one. Section 7 provides estimates of the causal effects of both Protestantism and literacy on the three measures of Prussian county economic outcomes used by BW. The results are not the same across all three measures, but they unambiguously reject BW's claim that Protestantism had a positive effect on economic prosperity which was largely due to the higher human capital of Protestants. Section 8 concludes.

2. An outline of the BW argument

BW's argument that Protestants were more prosperous than Catholics in late nineteenth-century Prussia because they had more human capital has three components. The first is the claim that, because of Luther's belief that everyone should be able to read the Bible, Protestantism promoted education, and hence Protestants had greater human capital than Catholics. To support this claim, BW use data for the 452 Prussian counties in the 1871 census to estimate regressions in which the share of literates aged 10 or over in a county's population is explained by the share of Protestants in a county's population, a number of county-level demographic variables, and a measure of the share of the county population for which literacy information is missing.[4] BW argue that the distance from a county to Wittenberg provides a way of obtaining variation in the share of Protestants in a county that is independent of economic and educational considerations.[5] Hence, they argue, the causal effect of Protestantism on literacy can be estimated by using distance to Wittenberg as an IV for the share of Protestants. BW's IV estimates show that Protestantism did have a positive causal effect on literacy, thus supporting the first component of their argument.

BW's initial arguments for the relevance of the distance to Wittenberg as an IV for the share of Protestants were strengthened by Cantoni (2012). That paper argued that distance to Wittenberg is a good predictor of the adoption of Protestantism by a territory of the Holy Roman Empire in the sixteenth and seventeenth centuries

[4] County is BW's translation of the German word *Kreis*. A *Kreis* is an administrative unit which is closer to the American than to the British sense of county.
[5] BW (2009), 557-63.

because of strategic neighbourhood interactions. The Catholic sympathies of the Holy Roman Emperors made it risky for territorial rulers to adopt the Reformation. If neighbouring rulers adopted the Reformation, the risk of a given ruler so doing was reduced, and hence this ruler was more likely to adopt Protestantism. Cantoni finds evidence that an increase in the proportion of neighbouring rulers who had adopted the Reformation led to an increase in the probability of a ruler subsequently adopting the Reformation. He argues that the early adoption of Protestantism by the Elector of Saxony, in which Wittenberg was located, is the key factor in explaining why distance to Wittenberg predicts Protestantism.

The second component of the BW argument is the claim that Protestantism led to economic prosperity. To support this claim, BW use three different measures of economic outcomes: income tax revenue per capita in each Prussian county in 1877 (their preferred measure), the natural logarithm of the average annual income of male elementary school teachers in each county in 1886, and the share of the labour force in manufacturing and service occupations in each county in 1882, which, following BW, will be referred to as the nonagricultural share in the remainder of the paper. IV regressions of these measures on the share of Protestants in each county (instrumented by distance to Wittenberg) and the demographic control variables show that Protestantism had a positive causal effect on all three economic outcomes. Since literacy is not included as an explanatory variable in these regressions, BW interpret the estimated coefficient of the share of Protestants as the total causal effect of Protestantism on economic outcomes, including any effect that operates because Protestantism promotes literacy.[6] These results seem to confirm the second component of the BW argument.

The third and final component of the BW argument is the claim that the positive causal effect of Protestantism on economic outcomes is largely due to the higher literacy of Protestants. To establish this claim, both literacy and Protestantism have to be included as explanatory variables in the regression analysis of economic outcomes. Literacy is an endogenous variable in the economic outcome regressions, since it is likely to depend on economic outcomes, and both literacy and economic

[6] BW (2009), 569.

outcomes are likely to depend on unobservables such as ability and work ethic. Thus an IV for literacy is required in order to estimate its causal effect on economic outcomes. However, no IV for literacy is available in BW's dataset. To circumvent this problem, BW assume a range of plausible values for the causal effect of literacy. Using these assumed values, they calculate a range of values for county economic outcomes adjusted for the assumed effect of literacy. IV regression of these adjusted economic outcomes on the share of Protestants (instrumented by distance to Wittenberg) and the demographic control variables produces estimates of the causal effect of Protestantism that control for the assumed effect of literacy. These estimates appear to show that literacy can account for most, or even all, of the positive effect of Protestantism on economic outcomes: once literacy is controlled for, the estimated effect of Protestantism is small and not statistically significant. Thus the third component of the BW argument also seems to be confirmed.

However, as the remainder of this paper will show, BW's argument is incorrect. The claim that Protestantism had positive effects on county economic outcomes that were largely due to the higher literacy of Protestants does not stand up to careful scrutiny.

3. Fundamental problems with the BW argument

BW argue that IV estimation of the effect of Protestantism on economic outcomes in a regression model that omits literacy produces estimates of the total causal effect of Protestantism, including the indirect effect operating via literacy. To obtain the causal effect of Protestantism alone, which I shall call the pure effect of Protestantism, BW adjust the measures of economic outcomes for assumed effects of literacy and then estimate the effect of Protestantism on these adjusted measures by IV.

The problem with this procedure is that it appears not to appreciate the standard case for IV estimation when there are omitted variables. If distance to Wittenberg is a valid IV for the share of Protestants, then a consistent estimate of the pure causal impact of Protestantism on economic outcomes can be obtained even if literacy is omitted from the regression model. Suppose that the population regression

model relating economic outcomes (Y_i), the share of Protestants (P_i) and the share of literates (L_i) is

$$Y_i = \alpha + \beta P_i + \gamma L_i + \varepsilon_i, i = 1,...,n \quad (1)$$

where ε_i is an error term that is uncorrelated with L_i.[7] Omitting L_i from the regression model results in

$$Y_i = \alpha + \beta P_i + e_i, i = 1,...,n \quad (2)$$

where

$$e_i = \gamma L_i + \varepsilon_i, i = 1,...,n.$$

Let Z_i be an IV which is relevant and satisfies the exclusion restriction, so that $\text{cov}(Z_i, P_i) \neq 0$, $\text{cov}(Z_i.L_i) = 0$ and $\text{cov}(Z_i.\varepsilon_i) = 0$, where these are all population covariances. From equation (2)

$$\begin{aligned}\text{cov}(Z_i, Y_i) &= \text{cov}(Z_i, \alpha + \beta P_i + e_i) \\ &= \beta \,\text{cov}(Z_i, P_i) + \gamma \,\text{cov}(Z_i, L_i) + \text{cov}(Z_i, \varepsilon_i).\end{aligned}$$

Since Z_i is assumed both to be relevant and to satisfy the exclusion restriction,

$$\beta = \frac{\text{cov}(Z_i, Y_i)}{\text{cov}(Z_i, P_i)} \quad (3)$$

so that the population coefficient β is the ratio of the population covariance between Z and Y to the population covariance between Z and P.

The IV estimator of β, denoted $\hat{\beta}$, replaces the population covariances in (3) with sample covariances between Z and Y (s_{ZY}) and between Z and P (s_{ZP}):

$$\hat{\beta} = \frac{s_{ZY}}{s_{ZP}}. \quad (4)$$

The sample covariance is a consistent estimator of the population covariance and hence the IV estimator $\hat{\beta}$ is consistent:

$$\hat{\beta} \xrightarrow{p} \frac{\text{cov}(Z_i, Y_i)}{\text{cov}(Z_i, P_i)} = \beta.$$

Thus, even if literacy is omitted from a regression model of county economic outcomes, the IV estimator will yield a consistent estimator of the pure causal effect of the share of Protestants provided that distance to Wittenberg is a valid IV for the

[7] If there are additional regressors that are uncorrelated with the error term, equation (1) can be thought of as the regression equation after these additional regressors have been partialled out.

share of Protestants. It is only when the OLS rather than the IV estimator is used that the omission of literacy will, if there is a positive correlation between Protestantism and literacy, result in the estimated pure effect of Protestantism being too high.

BW use IV to estimate a regression model which relates economic outcomes adjusted for an assumed effect of literacy to the share of Protestants. Does this approach also yield a consistent estimate of the pure causal effect of Protestantism on economic outcomes? Let $W_i = Y_i - \delta L_i, i = 1,...,n$ where δ is the assumed coefficient of literacy. Equation (1) can then be written as

$$W_i = \alpha + \beta P_i + (\gamma - \delta)L_i + \varepsilon_i, i = 1,...,n$$

and, since literacy is omitted from the regression model, BW estimate

$$W_i = \alpha + \beta P_i + u_i, i = 1,...,n \qquad (5)$$

where $u_i = (\gamma - \delta)L_i + \varepsilon_i$. Maintaining the assumptions that Z_i is a relevant IV and can be excluded from (5), the same argument that led to (3) shows that

$$\beta = \frac{\text{cov}(Z_i, W_i)}{\text{cov}(Z_i, P_i)}.$$

The IV estimator of β in (5), denoted $\tilde{\beta}$, is

$$\tilde{\beta} = \frac{s_{ZW}}{s_{ZP}}$$

where s_{ZW} denotes the sample covariance between Z and W. Hence the IV estimator $\tilde{\beta}$ is consistent:

$$\tilde{\beta} \xrightarrow{p} \frac{\text{cov}(Z_i, W_i)}{\text{cov}(Z_i, P_i)} = \beta.$$

This analysis shows that if, in the regression models estimated by BW, distance to Wittenberg is a relevant IV for Protestantism and satisfies the exclusion restriction, the following results should be expected. There should be no significant difference, at least in large samples, between the IV estimate of the effect of Protestantism on economic outcomes obtained from a model in which literacy is omitted and the IV estimate of this effect obtained from a model in which the effect of literacy is assumed to take a specific value. Furthermore, in the latter case, the IV

estimate of the effect of the share of Protestants should not differ significantly whatever assumption is made about δ, the effect of literacy on economic outcomes.

BW's estimates in Tables V and VI of their paper do not exhibit these features. Equation (4) in Table V shows that the IV point estimate of the effect of the share of Protestants on per capita income tax in their regression model which omits literacy is 0.586.[8] Table VI shows that the IV point estimates of the effect of the share of Protestants in their regression models which incorporate an assumed effect of literacy on economic outcomes vary from 0.309 to -0.061. It is not correct to interpret these findings as evidence that literacy can account for most of the positive effect of Protestantism on economic outcomes. Rather, these results are evidence that distance to Wittenberg is not a valid IV for Protestantism in BW's regression models, as I shall now show.

Suppose in the analysis above that, although the population covariance between Z_i and ε_i is zero, the population covariance between Z_i and L_i is non-zero: $\text{cov}(Z_i, L_i) \neq 0$. The IV Z_i does not therefore satisfy the exclusion restriction. Now

$$\frac{\text{cov}(Z_i, Y_i)}{\text{cov}(Z_i, P_i)} = \beta + \gamma \frac{\text{cov}(Z_i, L_i)}{\text{cov}(Z_i, P_i)} \tag{6}$$

and

$$\frac{\text{cov}(Z_i, W_i)}{\text{cov}(Z_i, P_i)} = \beta + (\gamma - \delta) \frac{\text{cov}(Z_i, L_i)}{\text{cov}(Z_i, P_i)}. \tag{7}$$

Neither the IV estimator $\hat{\beta}$ nor the IV estimator $\tilde{\beta}$ are consistent in this case. The former, obtained from the regression model that omits literacy, will converge in probability to the right-hand side of (6) while the latter, obtained from the regression model in which the effect of literacy is assumed to be δ, will converge in probability to the right-hand side of (7). The two estimators should yield different results, since

$$\hat{\beta} - \tilde{\beta} \xrightarrow{p} \delta \frac{\text{cov}(Z_i, L_i)}{\text{cov}(Z_i, P_i)}$$

and, from (7), the estimated effect of Protestantism in the regression model in which the effect of literacy is assumed to be δ will depend on the assumed value of δ.

[8] Per capita income tax is BW's preferred measure of county economic outcomes.

These are exactly the features of the results in Tables V and VI of the BW paper. The BW findings are what would be expected if their IV, distance to Wittenberg, is correlated with the omitted literacy variable, and thus is not a valid IV in BW's regression models relating economic outcomes to Protestantism.

There are two reasons why distance to Wittenberg is not a valid IV for the share of Protestants in BW's regression models of county economic outcomes. One is not specific to distance to Wittenberg but applies to any IV that might be used for Protestantism in models of economic outcomes that omit literacy as a regressor. The other applies specifically to distance to Wittenberg.

If Protestantism has a causal effect on literacy, then any variable that is sufficiently correlated with Protestantism to be an IV must also be correlated with literacy. Hence it is not possible to find an IV for the share of Protestants that will also satisfy the exclusion restriction in a regression model that omits literacy. The estimates obtained by BW using distance to Wittenberg as an IV for Protestantism in regression models of economic outcomes, irrespective of whether they are or are not based on assumed values for the effect of literacy, must be inconsistent if Protestantism influenced literacy. In such circumstances, distance to Wittenberg must be an invalid IV.

Even if this fundamental problem did not exist, distance to Wittenberg is unlikely to be a valid IV for the share of Protestants in BW's regression models of county economic outcomes. This is because those models do not include any regressors that reflect regional effects on economic outcomes. The importance of such effects is a central theme in the literature on German industrialisation in the nineteenth century, and any analysis of county economic outcomes must take them into account.[9] Regional effects will be correlated with distance to Wittenberg, since both are geographical in nature. The omission of regional effects from BW's regression models therefore means that distance to Wittenberg is likely to be an invalid IV, as is discussed in the next section.

[9] Tipton (1976), Ogilvie (1996a).

4. Regional effects in the Prussian economy and their implications

Nineteenth-century Prussia consisted of territories that had been part of the Prussian state for very different lengths of time. The Duchy of Prussia was created in 1525 and was unified with Brandenburg in 1618 to become the state of Brandenburg-Prussia, which also included some small territories in the Rhineland. During the seventeenth century this state acquired some territories in Westphalia and elsewhere. In 1701 Brandenburg-Prussia became the Kingdom of Prussia, and during the eighteenth century it expanded by acquiring, inter alia, Pomerania, Silesia, and additional parts of Poland. In 1815 Prussia acquired all the Rhineland, Westphalia, and various other territories, and in 1866 Prussia annexed Hannover, Hessen, and Schleswig-Holstein. Of the 452 counties in BW's database, 66 had been Prussian since 1525, 108 had become Prussian in the 1810s, and 86 had only been Prussian since 1866.

The institutional frameworks of different parts of Prussia varied a great deal, and this variation influenced economic development throughout the nineteenth century.[10] In the backward eastern parts of Prussia, the powers of feudal landlords remained strong even after the formal abolition of Prussian serfdom in 1806, and factory industrialisation here was delayed until the later nineteenth century. Silesia had very dense proto-industry in the early nineteenth century, but its industrialisation was hampered by the desire of feudal landlords to protect their proto-industrial feudal revenues through resistance to technological improvements in linen production, in which they were supported by the Prussian state. The most economically advanced part of Prussia in 1816 was the Rhineland, where the early decline in landlord power combined with extensive political fragmentation to enable proto-industries easily to cross territorial boundaries in order to locate where political and institutional conditions were least oppressive. The institutional framework here remained favourable to economic development throughout the nineteenth century.[11] Acemoglu et al. (2011) construct an index of German institutional reform based on the civil code, agrarian reform, the abolition of guilds, and the abolition of serfdom. The value of this index for the Rhineland was considerably higher, and the value for Saxony was

[10] Ogilvie (1996b), 265.
[11] Ogilvie (1996a), 124-5.

modestly higher, than for the provinces of Prussia, Brandenburg, Pomerania, Silesia and Westphalia in both 1850 and 1900.[12]

The omission of any variables measuring regional effects from BW's regression models of county economic outcomes might not result in inconsistent estimates of the effect of Protestantism if BW's IV for Protestantism could plausibly be argued to be uncorrelated with the omitted regional effects. But BW's IV is distance to Wittenberg, a geographical variable. As I shall now show, distance to Wittenberg is indeed correlated with measures of these regional effects, and hence it is unlikely to be a valid IV in regression models of county economic effects which omit them.

A natural way to allow for the effects of institutional frameworks that differ between Prussian provinces in regression models of county economic outcomes is to suppose that there are province fixed effects on these outcomes, which can be captured by including provincial dummy variables as regressors. The counties in BW's dataset can be categorised according to the provinces in which they were located in 1871 using the ifo Prussian Economic History Database.[13] There were 57 counties in the province of Prussia, 34 in Brandenburg, 29 in Pomerania, 27 in Posen, 63 in Silesia, 42 in Saxony, 20 in Schleswig-Holstein, 37 in Hannover, 35 in Westphalia, 35 in Hessen, 69 in the Rhineland, and 4 in Hohenzollern.

Province dummy variables on their own, however, are unlikely to capture all the effects of differing institutional frameworks on county economic outcomes, since within each province counties varied in the length of time they had been part of the Prussian state. In the Rhineland province, for example, 59 counties had become Prussian in 1815, but two had been Prussian since 1614, one since 1666, three since 1702 and one since 1713. I therefore used the year of a county's annexation by Prussia as a regressor in order to allow for possible effects of the length of time it had been Prussian on economic outcomes. In addition, I used interactions between the

[12] Acemoglu et al. (2011), Table 1, 3292.
[13] Becker et al. (2014).

year of annexation and the province dummy variables to allow for the possibility that the effects of the length of time a county had been Prussian varied by province.[14]

There is also evidence that three other geographical variables - the distance of a county from Berlin, from London, and from the nearest provincial capital - influenced the industrial development of Prussian counties in the latter part of the nineteenth century (Becker et al. 2011, Edwards 2018). The distance from London, in particular, had a strong negative effect on Prussian industrialisation, reflecting the importance for industrialisation of distance from the country that was the industrial leader for most of the nineteenth century. These distance variables are correlated with distance to Wittenberg, raising further questions about the validity of BW's IV in regression models that omit them as possible influences on county economic outcomes. BW's dataset does not contain measures of the distance of a county from London and the nearest provincial capital, but it does include the latitude and longitude of each county, from which, together with the latitude and longitude of London and the provincial capitals, I was able to calculate these two distance measures.

Table 1 reports the results of two regressions which show the association between distance to Wittenberg and the various geographical measures that are likely to affect county economic outcomes. The four observations for the province of Hohenzollern were not used in the estimation of these regressions, because consistent estimates of the fixed effect of this province cannot be obtained from such a small number of observations. Equation (1.1) omits, and equation (1.2) includes, the demographic control variables used by BW in their analysis of the relationship between economic outcomes and Protestantism. It is clear from Table 1 that distance from Wittenberg is correlated with several variables that were omitted from BW's regression analysis, irrespective of whether BW's demographic variables are included.

For both equations in Table 1, the null hypotheses that the coefficients of the ten province dummy variables were all zero and the coefficients of the eight terms

[14] The counties in the provinces of Schleswig-Holstein and Hessen all became Prussian in 1866 so that interaction terms for these provinces were not required.

Table 1: The relationship between distance to Wittenberg and possible omitted variables in regression models of county economic outcomes, Prussia 1871

	Dependent variable: Distance to Wittenberg	
Regressors	(1.1)	(1.2)
Brandenburg	-19.895	-6.080
	(15.367)	(17.617)
Pomerania	9.695**	10.644**
	(4.387)	(4.980)
Posen	-18.142***	-18.204***
	(2.818)	(3.557)
Silesia	-59.973***	-54.168***
	(3.122)	(4.103)
Saxony	-122.011***	-107.078***
	(15.043)	(16.985)
Schleswig-Holstein	-31.516***	-27.463**
	(11.358)	(10.703)
Hannover	-79.001***	-70.497***
	(15.793)	(16.383)
Westphalia	-95.062***	-83.464***
	(18.555)	(20.492)
Hessen	-122.715***	-108.912***
	(14.421)	(15.914)
Rhineland	-100.409***	-88.142***
	(19.650)	(20.978)
Year in which annexed by Prussia	-1.027***	-0.644**
	(0.195)	(0.259)
Distance to Berlin	0.968***	0.979***
	(0.017)	(0.018)
Distance to London	0.064**	0.064**
	(0.027)	(0.027)
Distance to nearest provincial capital	-0.002	0.009
	(0.047)	(0.049)
Constant	-25.070	-222.783**
	(83.450)	(105.180)
Adjusted R^2	0.990	0.990

Notes: Number of observations for both equations is 448. Figures in parentheses are heteroscedasticity-robust standard errors. ** and *** denote significance at the 0.05 and 0.01 levels respectively. Equation (1.2) also includes the following regressors, the coefficients of which are not reported: percentage aged below 10, percentage Jewish, percentage female, percentage native residents, percentage Prussian nationality, average household size, ln (population size), population growth 1867-71, percentage blind, percentage deaf-mute, and percentage insane. See text for interpretation of estimated effects of year of annexation and province dummy variables.

that interacted province dummies with the year of annexation were all zero were strongly rejected. Thus there is clear evidence that these variables were associated with the distance of a county from Wittenberg. The results for year of annexation and province reported in Table 1 are the marginal effects for each variable. The marginal effect of year of annexation is evaluated at the mean values of the province dummies for the entire sample, while the marginal effects of the provinces are evaluated at the

mean values of year of annexation for the province in question. The omitted province dummy is that for the province of Prussia, so the province marginal effects show the difference compared to that province.

It is clear from Table 1 that distance to Wittenberg is statistically significantly associated with the variables that measure regional effects on Prussian economic development as well as the measures of the distance of a county from both Berlin and London. In many cases these associations are large. The mean value of the distance to Wittenberg for the 448 observations used in Table 1 is 324.848 kilometres. Several of the province marginal effects are more than 25 per cent of this value. The marginal effect of the year of annexation in equation (1.2) corresponds to an elasticity of -3.47 while that in equation (1.1) corresponds to one of -5.53. Here, and throughout the paper, all elasticities are calculated at sample mean values. In equation (1.1) the partial correlation between the distance to Wittenberg and the distance to Berlin is 0.955 while in equation (1.2) this value is 0.954, so there is a particularly strong association between distance to Wittenberg and distance to Berlin. If the regional effect and distance variables in Table 1 did have an effect on county economic outcomes, then their omission from BW's regression models of economic outcomes would mean that distance to Wittenberg was an invalid IV for Protestantism in the BW models. The analysis which follows shows that this is indeed the case.

5. Data

If, as BW argue, Protestantism has a causal influence on literacy, then, as explained in section 3, a consistent estimate of the effect of Protestantism on economic development requires that the regression model used for this purpose does not omit literacy. Since both Protestantism and economic outcomes potentially have causal effects on literacy, IV estimation of the effect of literacy in such a regression model is needed. BW were unable to obtain such an estimate of the effect of literacy because their dataset did not contain a suitable IV. However, Becker et al. (2011) used a dataset for the 334 Prussian counties that existed in 1849 to analyse the relationship between education and industrialisation. This dataset provides a possible solution to the problem of the missing IV for literacy, because it has observations for the level of education in 1816. Becker et al. argue that this variable can serve as an IV for the

1871 literacy rate. In addition, this dataset allows a number of other potentially important variables to be included in the analysis of county economic outcomes.

I therefore used the ifo Prussian Economic History Database to supplement the database of 334 counties in 1849 by adding to it those variables used by BW in their analysis that were not in the Becker et al. dataset.[15] Several counties that existed in 1849 were subdivided between then and 1871, so in order to do this it was necessary to aggregate the data from 1871 to reconstruct the county structure of 1849, as recommended by Becker et al. (2014). Using a database for the 334 counties that existed in 1849 means, of course, that the analysis which follows excludes all the counties that became Prussian after 1849: thus there are no observations for the provinces of Hannover, Hessen, Hohenzollern and Schleswig-Holstein.

Table 2 presents summary statistics for the main variables from the sample of 334 counties used in the following analysis. Comparison of Table 2 with BW's Table I shows that the economic outcome measures and the literacy rate are similar in the two samples, although teacher income is rather more dispersed in the BW sample.[16] There is a somewhat greater difference in the share of Protestants between the two samples, with the mean value being 4.2 percentage points lower in the smaller sample, but this difference is not statistically significant. The main features of BW's dataset are preserved in the smaller sample.

To check that BW's analysis was replicated using only the 334 Prussian counties that existed in 1849, I estimated their IV regression models on this smaller sample. As section A1 of the Appendix shows, restricting the analysis to the smaller sample of 334 counties for which there are observations of the share of Protestants in 1816 does not alter the BW findings to any significant extent.

[15] Becker et al. (2014).
[16] In Table I of BW (2009), income tax per capita is measured in hundreds of marks, rather than in marks as in Table 2 of this paper. The apparent large difference is merely a matter of different units.

Table 2: Summary statistics for the main variables in the sample of 334 Prussian counties

	Mean	Standard deviation	Min	Max
Economic outcome variables				
Income tax revenue per capita	190.41	68.78	56.34	562.62
Income of male elementary school teachers	965.91	181.44	726.76	1913.21
Ln (income of male elementary school teachers)	6.86	0.17	6.59	7.56
Nonagricultural share of labour force	32.95	15.20	7.93	79.17
Explanatory variables of primary interest				
Share of Protestants in population 1871	59.98	38.98	0.26	99.80
Literacy rate 1871	85.10	14.18	9.51	99.33
Geographical variables				
Distance to London	940.34	327.74	416.00	1534.00
Distance to Berlin	329.27	160.59	1.00	650.04
Distance to nearest provincial capital	85.02	43.41	0.00	280.00
Distance to Wittenberg	333.20	164.74	0.00	731.46
County area (in thousand square kilometres)	0.81	0.45	0.00	2.54
Pre-industrial variables				
Paved streets 1815 (dummy variable)	0.22	0.42	0.00	1.00
Public buildings per capita 1821	0.004	0.003	0.0003	0.021
Sheep per capita 1816	0.55	0.44	0.00	2.58
Tonnage of ships per capita 1819	0.01	0.05	0.00	0.55
Looms per capita 1819	0.01	0.02	0.00	0.23
Share of farm labourers in population 1819	9.52	3.95	0.00	24.13
Share of population living in cities 1816	24.83	18.74	0.00	100.00
Year in which annexed by Prussia	1720.50	109.01	1525.00	1816.00

Notes: Number of observations for all variables except income tax per capita is 334. Income tax data for the city counties was unavailable so number of observations for this variable is 321.

6. Protestantism and literacy

The first component of the BW thesis is the argument that Protestantism had a positive causal effect on literacy. They support this claim by IV estimation of the effect of the share of Protestants on the literacy rate, using distance to Wittenberg as an IV, in a regression which also includes several demographic variables. The results, in Table III of their paper, suggest that there is a causal effect of Protestantism on literacy, but of a rather modest size – the point estimate corresponds to an elasticity of literacy with respect to Protestantism is 0.138.

As section 4 showed, the distance of a county to Wittenberg is correlated with measures of regional effects and the county's distance from Berlin and London. If these variables influenced the literacy rate, distance to Wittenberg would not be a valid IV in BW's literacy regression model and the resulting estimate of the effect of Protestantism on literacy would be inconsistent. In order to investigate whether this problem arises, I estimated the following basic regression model:

$$L_i = \alpha_0 + \alpha_1 P_i + \alpha_2' PR_i + \alpha_3 A_i + \alpha_4' PR_i * A_i + \alpha_5' D_i + \eta_i .$$

Here *i* denotes county, L_i denotes the literacy rate, P_i denotes the share of Protestants, PR_i is a vector of dummy variables indicating the province in which a county is located, A_i denotes the year in which a county was annexed by Prussia, D_i is a vector of variables measuring the distance of the county to, respectively, Berlin, the provincial capital, and London, and η_i is an error term.

It is possible that literacy is influenced by earlier economic conditions, so I also estimated a second regression model in which a number of pre-industrial characteristics of counties were added to the basic model. The share of the population living in cities in 1816 and the number of looms per capita in 1819 are included as indicators of pre-industrial development. Since textiles was an important industrial sector, the number of sheep in 1816 is used as a proxy for the availability of wool for the textile industry. The share of farm labourers in the population in 1819 is included as an indicator of the agricultural orientation of a county. Various measures of pre-industrial public infrastructure which might have influenced subsequent economic development are also included as regressors: the number of public buildings per capita in 1821, a dummy variable registering the presence of paved inter-regional roads in 1815, and a measure of the capacity of river transport ships in 1819. I also include the geographical area of each county as a regressor.

BW included both the share of the population for which literacy information was missing and a number of demographic variables as regressors in their analysis of literacy. I omitted these from the basic regression model. The demographic variables – the proportions of the population in 1871 that were, respectively, aged below 10, female, Jewish, native-born, Prussian, missing educational information, blind, deaf-mute, and insane, together with the average household size, the log of population size,

and population growth from 1867 to 1871 – are, in many cases, bad controls: "variables that are themselves outcome variables in the notional experiment at hand".[17] However, in order to check whether the estimated effect of Protestantism on literacy was affected by the omission of these variables, I also estimated a third regression model in which they, together with the pre-industrial variables, were added to the basic model as regressors.

As pointed out in section 4, the partial correlation between a county's distance to Wittenberg and its distance to Berlin is extremely high. This meant that, when distance to Berlin was included as a regressor in the three literacy regression models, distance to Wittenberg was an extremely weak IV for the share of Protestants. I therefore did not use distance to Wittenberg as an IV for the share of Protestants in county population in 1871, but instead estimated the regression models for literacy using the share of Protestants in county population in 1816 as the only IV for the share of Protestants in 1871. I shall discuss the justification for so doing shortly, but for the moment I proceed on the assumption that the former is a valid IV for the latter. I follow Becker et al. (2011) in clustering standard errors at the level of the 280 independent units of observation in 1816. This is because the data reported in the 1816 census had to be adjusted in order to construct a consistent dataset for the 334 Prussian counties as they existed in 1849, and hence the data for 1816 were not based on 334 independent observations.[18] For each regression model, I tested the null hypothesis that the coefficient of distance to Berlin was zero. This null hypothesis was not rejected in any of the three models (the p values were all greater than 0.4 and the elasticities corresponding to the point estimates ranged from -0.023 to -0.001). I therefore removed distance to Berlin as a regressor and estimated the models using both distance to Wittenberg and the share of Protestants in 1816 as IVs for the share of Protestants in 1871. Table 3 shows the results.

Both the Kleibergen-Paap and Montiel Olea-Pflueger first-stage F statistics in equations (3.1), (3.3) and (3.5) are extremely large, reflecting the high correlation between the share of Protestants in 1816 and the share of Protestants in 1871

[17] Angrist and Pischke 2009, 64.
[18] Becker et al. (2011), 105 n.12, and online Appendix A1 (https://assets.aeaweb.org/assets/production/articles-attachments/aej/mac/app/2010-0021_app.pdf).

Table 3: The relationship between the literacy rate and Protestantism in Prussia, 1871

	Dependent variable: literacy rate in 1871					
Regressors	IV	OLS	IV	OLS	IV	OLS
	(3.1)	(3.2)	(3.3)	(3.4)	(3.5)	(3.6)
Share of Protestants in population 1871	0.091***	0.103***	0.089***	0.100***	0.084***	0.094***
	(0.017)	(0.017)	(0.019)	(0.018)	(0.017)	(0.016)
Brandenburg	10.553***	10.667***	8.840***	8.959***	6.597**	6.673**
	(3.022)	(3.022)	(3.348)	(3.347)	(3.315)	(2.759)
Pomerania	11.141***	10.783***	10.312***	9.956***	11.747***	11.388***
	(3.055)	(2.970)	(2.976)	(2.912)	(2.908)	(2.811)
Posen	-7.367**	-7.175*	-6.177*	-6.056*	-1.101	-1.062
	(3.663)	(3.684)	(3.512)	(3.353)	(3.311)	(3.280)
Silesia	16.299***	16.336***	16.715***	16.797***	11.390***	11.532***
	(2.913)	(2.909)	(3.031)	(3.040)	(2.663)	(2.666)
Saxony	13.879***	13.785***	10.869***	10.903***	9.504**	9.149**
	(3.840)	(3.792)	(3.962)	(3.927)	(3.955)	(3.909)
Westphalia	12.186**	12.803***	11.258**	11.861**	15.637***	15.696***
	(4.914)	(4.993)	(4.968)	(5.038)	(5.638)	(5.647)
Rhineland	11.063**	11.774**	9.074	9.787*	11.046*	11.262*
	(5.272)	(5.374)	(5.671)	(5.770)	(6.340)	(6.391)
Year in which annexed by Prussia	0.364	0.378	0.358	0.373	0.171	0.176
	(0.297)	(0.290)	(0.285)	(0.279)	(0.208)	(0.203)
Distance to London	-0.020***	-0.020***	-0.022***	-0.022***	-0.005	-0.004
	(0.007)	(0.007)	(0.007)	(0.007)	(0.006)	(0.006)
Distance to nearest provincial capital	-0.064***	-0.063***	-0.051***	-0.050***	-0.036***	-0.036***
	(0.012)	(0.012)	(0.012)	(0.012)	(0.009)	(0.009)
Pre-industrial variables	No	No	Yes	Yes	Yes	Yes
BW demographic controls	No	No	No	No	Yes	Yes
Share of population missing literacy information	No	No	No	No	Yes	Yes

Constant	128.882*** (27.789)	122.595*** (27.674)	136.385*** (26.692)	130.818*** (26.866)	217.095** (100.773)	211.639** (101.093)
Adjusted R^2	0.735	0.735	0.749	0.749	0.796	0.797
Kleibergen-Paap *F* statistic	1467.58	-	1865.77	-	1101.76	-
Montiel Olea-Pflueger *F* statistic	1146.86	-	1082.00	-	768.04	-
Standard 95 per cent confidence interval for share of Protestants	[0.058, 0.124]	[0.069, 0.137]	[0.052, 0.126]	[0.064,0.136]	[0.050, 0.118]	[0.063, 0.125]
Weak-IV-robust 95 per cent confidence interval for share of Protestants	[0.063, 0.129]	-	[0.556, 0.130]	-	[0.054, 0.122]	-
J test *p* value	0.343	-	0.738	-	0.349	-
C test *p* value	0.097	-	0.107	-	0.194	-

Notes: Number of observations for all equations is 334. Figures in parentheses are standard errors clustered at the level of the 280 independent units of observation in 1816. *, ** and *** denote significance at the 0.10, 0.05 and 0.01 levels respectively. The results reported for year of annexation and province are the marginal effects for each variable, calculated in the same way as those in Table 1. The omitted province dummy variable is that for the province of Prussia.

(0.981).[19] Such a high correlation between an IV and the endogenous variable may raise the concern that the IV is essentially the same as the endogenous variable. However, provided that the share of Protestants in 1816 is uncorrelated with the error in the regression equation being estimated, the IV estimate of the effect of Protestantism on literacy will be consistent however high the correlation between the IV and the endogenous regressor. Just-identified estimates of these three equations using only distance to Wittenberg as an IV are reported in section A2 of the Appendix, and these yield similar point estimates of the effect of Protestantism on literacy. This suggests that, conditional on distance to Wittenberg being a valid IV, so too is the share of Protestants in 1816. Distance to Wittenberg is not a redundant IV: the null hypothesis that it is was rejected for each of equations (3.1), (3.3) and (3.5) with *p* values of 0.007, 0.010 and 0.020 respectively, so the asymptotic efficiency of the estimation is improved by using distance to Wittenberg as an IV. Thus, if the regression model does not omit relevant variables that are correlated with distance to Wittenberg, and BW's arguments that distance to Wittenberg is in principle a valid IV for Protestantism are accepted, the *J* test of overidentifying restrictions is a test of the validity of the share of Protestants in 1816 as an IV for the share of Protestants in 1871. The *p* values of this test reported in Table 3 suggest that it is justifiable to regard the share of Protestants in 1816 as a valid IV for the share of Protestants in 1871.

The estimated causal effect of Protestantism on literacy in equations (3.1), (3.3) and (3.5) hardly differs. In all three cases it is precisely estimated, as shown by both the standard and weak-IV-robust 95 per cent confidence intervals.[20] However, the size of this effect is very small: the largest point estimate of the coefficient of the share of Protestants, in equation (3.1), corresponds to an elasticity of only 0.064. Thus

[19] The Stata command *ivreg2* (Baum et al. 2010), which was used for all the estimates in this paper, reports the Kleibergen-Paap (2006) *F* statistic in the case of non-iid errors. As Baum et al. (2007) emphasise, this statistic does not provide a formal test for weak instruments in such circumstances. Montiel Olea and Pflueger (2013) provide such a formal test in the case of a single endogenous regressor: their *F* statistic enables a test of the null hypothesis that the approximate asymptotic bias of the IV estimator exceeds a specified fraction of the worst-case bias. The critical values at which it is possible to reject at the 0.05 level the null hypothesis that this bias is respectively 5 and 10 per cent of the worst-case bias are 37.42 and 23.11 in the case of one IV, and 17.35 and 11.34 in the case of two IVs. The Montiel Olea and Pflueger *F* statistics reported in this paper were calculated using the Stata command *weakivtest* (Pflueger and Wang 2015)

[20] The weak-IV-robust confidence intervals were obtained using the conditional likelihood ratio procedure, as recommended by Andrews et al. (2006) for situations in which the regression model is overidentified, and was computed by the Stata *weakiv* command of Finlay et al. (2013).

there is evidence in favour of the first component of the BW thesis, but the positive causal effect of Protestantism on literacy is a very small one. All three IV estimates in Table 3 correspond to an elasticity approximately half the size of the one corresponding to BW's estimate in equation (2) of their Table III.

Table 3 also reports the *p* values of the *C* statistic which tests the null hypothesis that the education variable being treated as endogenous in IV estimation is actually an exogenous regressor.[21] Especially for equations (3.1) and (3.3), the *p* values do not conclusively fail to reject the null hypothesis. However, the OLS estimates in equations (3.2) and (3.4) are very similar to the IV estimates in equations (3.1) and (3.3), which suggests that IV estimation is not required and the education variable can be treated as exogenous. The point estimate of the coefficient of the share of Protestants in equation (3.2) corresponds to an elasticity of 0.072.

The reason that the estimates of the effect of Protestantism on literacy in Table 3 differ from the BW estimate is that the latter is inconsistent. Distance to Wittenberg is not a valid IV in the BW regression model because it is correlated with omitted variables. There is clear evidence that the province dummy variables and their interactions with the year of annexation were determinants of the literacy rate and hence should not have been omitted from the BW regression model. The null hypothesis that, in the literacy regression models, the coefficients of both the seven province dummies and the seven interaction terms were all zero was always rejected. The results for year of annexation and province reported in Table 3 are the marginal effects for each variable. These are calculated in the same way as in Table 1. The omitted province dummy is that for the province of Prussia, so the province marginal effects show the difference compared to that province. It is also clear from Table 3 that distance to London and distance to the nearest provincial capital influenced the literacy rate and should not have been omitted from the BW regression model. Protestantism did have a causal effect on literacy in Prussia, but it was a very small one: the upper bound of the 95 per cent confidence interval for the elasticity of literacy with respect to Protestantism is a little less than 0.1.

[21] Hayashi (2000), pp. 218-21; Baum et al. (2003).

7. Protestantism, literacy and economic outcomes

I turn now to BW's claim that Protestantism had a positive effect on county economic outcomes due largely to the higher literacy of Protestants. In principle, the positive causal influence of Protestantism on literacy shown in the previous section means that, in order to obtain a consistent estimate of the effect of Protestantism on economic outcomes, the literacy rate in 1871 must also be included as a regressor. In practice, however, the causal effect of Protestantism on literacy is so very small that there is a case for avoiding the additional complications created by having two potentially endogenous variables as regressors. I therefore begin by estimating models of economic outcomes which omit literacy. A further reason for doing this is that BW reported IV estimates of the effect of Protestantism in regressions that excluded literacy, so omitting the literacy rate as a regressor enables a clear comparison with BW's findings.

I therefore began by estimating the following regression model of the determinants of county economic outcomes:

$$Y_i = \beta_0 + \beta_1 P_i + \beta_2' PR_i + \beta_3 A_i + \beta_4' PR_i * A_i + \beta_5' D_i + \beta_6' X_i + v_i$$

where Y_i denotes a measure of county economic outcome, X_i is a vector of pre-industrial and other characteristics as described in section 6, v_i is an error term, and other variables are as previously defined. I used BW's three different measures of economic outcomes – per capita income tax, the log of teacher income, and the nonagricultural share of the labour force – as dependent variables. I did not include BW's demographic control variables as regressors. Economic conditions are likely to influence demographic and migration behaviour, and hence population size, population growth, average household size as well as the shares of the population that were below 10, native-born, and Prussian are all potentially endogenous variables in a regression model of county economic outcomes. These variables are likely to be bad controls. Section A3 of the Appendix shows that the broad conclusions about the effect of Protestantism on economic outcomes do not depend on the omission of these demographic variables. It also provides a more detailed justification for the specification of the regression model used in Table 4.

For the same reason as explained in section 6, I first estimated a model that included distance to Berlin as a regressor and used only the share of Protestants in 1816 as an IV for the share of Protestants in 1871. I then tested the null hypothesis that the coefficient of the distance to Berlin variable was zero. This was not rejected at the 0.05 level when the dependent variables were per capita income tax and log teacher income. Furthermore, the point estimates of the effect of distance to Berlin on these two variables were economically insignificant, corresponding to elasticities of -0.004 and -0.023 respectively. For these two dependent variables, I therefore estimated regression models that excluded distance to Berlin as a regressor and in which both the share of Protestants in 1816 and distance to Wittenberg were used as IVs. However, when the dependent variable was the nonagricultural share of the labour force, this null hypothesis was rejected, and the point estimate corresponded to an elasticity of 0.293. It was not therefore justifiable to omit distance to Berlin as a regressor, and so distance to Wittenberg could not be used as an IV.

The estimated effects of Protestantism and the other regressors on the three economic outcome measures are shown in Table 4. The enormous Kleibergen-Paap and Montiel Olea-Pflueger *F* statistics reported for the three IV regressions in Table 4 reflect the very high correlation between the shares of Protestants in 1816 and 1871 that has already been noted. As in the previous section, doubts about the validity of the former as an IV for the latter can be allayed, at least in equations (4.1) and (4.3), by the *J* test of overidentification restrictions. In both these equations, the null hypothesis that distance to Wittenberg is a redundant IV was rejected (with *p* values of 0.005 and 0.013 respectively). Thus, provided that distance to Wittenberg is a valid IV conditional on the regressors, the *p* values of the *J* test reported for equations (4.1) and (4.3) suggest that the IV estimates obtained using the share of Protestants in 1816 are consistent. As a further reinforcement of this point, section A3 of the Appendix reports the just-identified estimates of the effect of Protestantism on per capita income tax and log teacher income obtained using the share of Protestants in 1816 and distance to Wittenberg as IVs: the estimates obtained using only the former IV are similar to those obtained using solely the latter.

Table 4 reports both standard 95 per cent confidence intervals for the estimated effects of Protestantism, based on the asymptotic distribution of the IV

Table 4: The effects of Protestantism on county economic outcomes in Prussia

	Dependent variable: per capita income tax		Dependent variable: ln(teacher income)		Dependent variable: nonagricultural share of labour force	
Regressors	IV (4.1)	OLS (4.2)	IV (4.3)	OLS (4.4)	IV (4.5)	OLS (4.6)
Share of Protestants in population 1871	-0.083	-0.082	0.0012***	0.0011***	0.105***	0.094***
	(0.136)	(0.122)	(0.0002)	(0.0002)	(0.037)	(0.031)
Brandenburg	29.874	29.886	0.150***	0.148***	9.205**	8.608**
	(21.786)	(21.827)	(0.034)	(0.034)	(4.390)	(4.307)
Pomerania	24.354	24.320	0.021	0.025	0.118	0.306
	(17.201)	(16.933)	(0.026)	(0.026)	(2.931)	(2.896)
Posen	-40.531***	-40.518***	0.016	0.015	-1.415	-1.673
	(11.359)	(11.399)	(0.020)	(0.020)	(2.017)	(1.945)
Silesia	5.205	5.214	0.120***	0.119***	6.837**	6.647**
	(12.922)	(12.928)	(0.021)	(0.021)	(2.740)	(2.721)
Saxony	-10.324	-10.315	0.126***	0.125***	3.412	3.399
	(25.741)	(25.776)	(0.041)	(0.040)	(5.067)	(5.061)
Westphalia	-70.837*	-70.773*	0.171***	0.163***	-12.641*	-12.417*
	(37.318)	(37.344)	(0.054)	(0.053)	(7.588)	(7.540)
Rhineland	-78.269**	-78.192**	0.222***	0.213***	-17.320*	-16.807*
	(39.793)	(39.945)	(0.060)	(0.060)	(9.610)	(9.550)
Year in which annexed by Prussia	-2.453***	-2.451***	-0.0014	-0.0016	-0.266**	-0.284**
	(0.856)	(0.892)	(0.0019)	(0.0019)	(0.119)	(0.113)
Distance to London	-0.156***	-0.156***	-0.0001	-0.0001	-0.041***	-0.041***
	(0.045)	(0.045)	(0.0001)	(0.0001)	(0.010)	(0.010)
Distance to nearest provincial capital	-0.128	-0.128	0.0000	0.0000	0.021*	0.022*
	(0.085)	(0.085)	(0.0001)	(0.0001)	(0.012)	(0.012)
Distance to Berlin	-	-	-	-	0.029**	0.027**
					(0.014)	(0.013)
County area	-30.048***	-30.039***	-0.038***	-0.038***	-1.649	-1.752
	(9.492)	(9.467)	(0.014)	(0.014)	(1.506)	(1.488)

Paved streets 1815	24.758***	24.757***	0.054***	0.054***	3.606**	3.596**
	(8.700)	(8.689)	(0.014)	(0.014)	(1.709)	(1.714)
Public buildings per capita 1821	-1,719.162	-1,720.750	-6.373***	-6.158***	-821.143***	-810.319***
	(1,421.260)	(1,422.459)	(2.266)	(2.226)	(195.608)	(197.360)
Sheep per capita 1816	-9.309	-9.303	-0.027**	-0.027**	-5.570***	-5.637***
	(8.696)	(8.668)	(0.012)	(0.012)	(1.553)	(1.554)
Tonnage of ships per capita 1819	207.789	207.756	0.206***	0.208***	15.174	15.334
	(168.739)	(168.695)	(0.078)	(0.078)	(13.241)	(13.345)
Looms per capita 1819	-4.365	-4.450	0.863***	0.883***	183.091***	184.867***
	(114.497)	(115.048)	(0.226)	(0.228)	(47.281)	(47.339)
Share of farm labourers in population 1819	117.612*	117.647*	0.182	0.181	-7.839	-7.835
	(69.202)	(69.093)	(0.127)	(0.127)	(14.508)	(14.429)
Share of population living in cities 1816	101.935***	101.924***	0.316***	0.316***	17.690***	17.607***
	(31.599)	(31.479)	(0.041)	(0.041)	(3.491)	(3.495)
Constant	395.868***	395.290***	6.509***	6.578***	24.417	31.277
	(139.037)	(136.894)	(0.222)	(0.209)	(28.700)	(25.966)
Observations	321	321	334	334	334	334
Adjusted R^2	0.488	0.488	0.762	0.762	0.632	0.632
Kleibergen-Paap *F* statistic	1700.53	-	1865.77	-	1097.55	-
Montiel Olea-Pflueger *F* statistic	980.43	-	1082.00	-	1098.18	-
Standard 95 per cent confidence interval for share of Protestants	[-0.349, 0.183]	[-0.321, 0.157]	[0.00077, 0.00168]	[0.00069, 0.00150]	[0.033, 0.178]	[0.033, 0.155]
Weak-IV-robust 95 per cent confidence interval for share of Protestants	[-0.368, 0.180]	-	[0.00085, 0.00171]	-	[0.031, 0.176]	-
p value for test of redundancy of distance to Wittenberg	0.005	-	0.010	-		-
J test *p* value	0.972	-	0.539	-	-	-
C test *p* value	0.979	-	0.013	-	0.370	-

Notes: Figures in parentheses are standard errors clustered at the level of the 280 independent units of observation in 1816. *, ** and *** denote significance at the 0.10, 0.05 and 0.01 levels respectively. The results reported for year of annexation and province are the marginal effects for each variable, calculated in the same way as those in Tables 1 and 3. The omitted province dummy variable is that for the province of Prussia.

estimator, and weak-IV robust ones, computed using the Stata *weakiv* command of Finlay et al. (2013). For equations (4.1) and (4.3), the weak-IV-robust confidence intervals were obtained using the conditional likelihood ratio procedure, while for the just-identified equation (4.5), the Anderson-Rubin (1949) test was used to obtain the weak-IV-robust confidence interval. Unsurprisingly, given the large first-stage *F* statistics, there is little difference between the inferences based on the two types of confidence interval.

The results for year of annexation and province reported in Table 4 are the marginal effects for each variable, calculated in the same way as in Tables 1 and 3. The omitted province dummy variable is that for the province of Prussia. Before discussing the estimated effects of Protestantism on economic outcomes, I note that the results reported in Table 4 provide evidence that regional effects and pre-industrial variables influence all three measures of such outcomes. The year of annexation by Prussia and the distance variables influence per capita income tax and the nonagricultural share of the labour force, but not the log of teacher income. If distance to Wittenberg is to be a valid IV in regression models of county economic outcomes, therefore, regional effects and distance variables cannot be omitted as regressors.

Table 4 shows a clear difference between the estimated effect of Protestantism when per capita income tax is the measure of county economic outcomes and that when the other two measures are used. The point estimate of the effect of Protestantism on per capita income tax in equation (4.1) is negative, though poorly determined. The *p* value of the *C* statistic for this equation is so large as to suggest that there is no need for IV estimation in this case. Of course, when the IV estimate is poorly-determined, the *C* test may have low power, but the OLS point estimate in equation (4.2) is so similar to the IV one in (4.1) that there is very little danger of making a type II error in concluding that IV estimation of the regression model with per capita income tax as dependent variable is not required. The OLS point estimate of the effect of Protestantism on this measure of county economic outcomes corresponds to an elasticity of -0.029: it is therefore of essentially no economic significance.

However, when county economic outcomes are measured by either the log of teacher income or the nonagricultural share of the labour force, the estimated effect of Protestantism is positive and statistically significant. For the former measure, the *p* value of the *C* statistic in equation (4.3) suggests that IV estimation is required, although there is very little difference between the point estimates in equations (4.3) and (4.4). The economic significance of the effect of Protestantism on teacher income is very small: the point estimate in equation (4.1) corresponds to an elasticity of only 0.073. For the latter measure, the *C* test suggests that IV estimation is not necessary. The OLS point estimate of the effect of Protestantism on the nonagricultural share corresponds to an elasticity of 0.171, so its economic significance is rather modest but not trivial.

How different are the IV estimates of the effect of Protestantism on economic outcomes in Table 4 from those reported by BW? The comparisons are most easily made in terms of elasticities, because BW's results were obtained using a different sample. The elasticities corresponding to the point estimates of the effect of Protestantism in equations (4) - (6) of BW's Table V are 0.190, 0.067 and 0.155 for economic outcomes measured respectively by per capita income tax, teacher income, and nonagricultural share of the labour force. Section A1 of the Appendix reports the results of estimating the BW regression model on the sample of 334 counties used in Table 4: in this case the elasticities are 0.295, 0.070 and 0.128. The elasticities corresponding to the point estimates in equations (4.1), (4.3) and (4.5) are -0.026, 0.073 and 0.293 respectively. Thus there is a major difference between the IV point estimate of the effect of Protestantism on per capita income tax in Table 4 and the BW point estimate: the latter is clearly positive, while the former is negative though very close to zero. There is also a difference between the estimated effects on the nonagricultural share: although both are positive, the Table 4 point estimate is roughly twice the size of the BW one. For these two economic outcome measures, the different regression specification used in Table 4 leads to substantial changes from the BW estimates. However, there is essentially no difference between the Table 4 and BW estimates of the point effect of Protestantism on teacher income.

The estimated effects of Protestantism on economic outcomes reported in Table 4 will be consistent estimates of the pure effect of Protestantism only if

Protestantism has no causal effect on literacy. The results in Table 3 show that the causal effect of Protestantism on literacy is very small, but it is not zero. Thus, although it is unlikely that the estimates reported in Table 4 are seriously inconsistent, it remains a possibility. In order to be sure that the estimates in Table 4 are not seriously inconsistent, it is necessary to include both the share of Protestants and the literacy rate in 1871 as regressors in models of county economic outcomes. Since the literacy rate in 1871 may be influenced by economic conditions, it must be regarded as a potentially endogenous variable when it is added as a regressor, and thus an IV for it is required.

I follow Becker et al. (2011) by using education in 1816, measured by enrolment in elementary and middle schools as a share of the population aged from six to 14, as an IV for the literacy rate in 1871. Becker et al. use this variable as an IV for education at later dates in their analysis of the role of education in Prussian industrialisation. They argue that differences in education levels among Prussian counties in 1816 reflected exogenous historical idiosyncracies and therefore had no direct effect on subsequent industrialisation. Edwards (2018) shows that, in order for education in 1816 to be a valid IV for this purpose, it is necessary for the regression model of industrialisation to include measures of regional effects and the distances to Berlin and London. These variables should, of course, also be included in regression models of county economic outcomes if distance to Wittenberg is to be a valid IV.

I added the literacy rate in 1871 to the regressors in the general specification above and estimated this model for each of the three economic outcome measures. Since the general specification included distance to Berlin, distance to Wittenberg could not be used as an IV, and the share of Protestants in 1871 and the literacy rate in 1871 were instrumented by the share of Protestants in 1816 and education in 1816 only. The null hypothesis that the coefficient of the distance to Berlin variable was zero was not rejected at the 0.05 level when the dependent variables were per capita income tax and log teacher income. The elasticities at sample mean values corresponding to the point estimates of the effect of distance to Berlin were -0.003 and -0.024 respectively. For these two dependent variables, I therefore estimated regression models that omitted distance to Berlin as a regressor and in which the share of Protestants in 1816, education in 1816 and distance to Wittenberg were used as

IVs. However, omission of distance to Berlin was not justifiable when the dependent variable was the nonagricultural share of the labour force and so in this case only the share of Protestants and education in 1816 were used as IVs. The resulting estimates of the effects of Protestantism and literacy on economic outcomes are shown in Table 5.

The presence of two endogenous variables – the share of Protestants in 1871 and the literacy rate in 1871 – means that the conventional first-stage *F* statistic cannot be used to test the weakness of the IVs. Table 5 therefore reports the Sanderson-Windmeijer (2016) conditional first-stage *F* statistics for the share of Protestants and the literacy rate, which suggest that weak IV problems have to be taken seriously, especially for the literacy rate in equations (5.1) and (5.3). I used an IV estimator that has good properties in such circumstances - Fuller's modified limited information maximum likelihood procedure with parameter one (Fuller 1977) - but the estimates hardly differed from those obtained by two-stage least squares, so Table 5 reports the latter results, since the Sanderson-Windmeijer conditional *F* statistics apply to this estimator.

The 95 per cent weak-IV-robust confidence intervals for the estimated effects of the share of Protestants and the literacy rate in Table 5 were calculated using the Stata *weakiv* command. For equations (5.1) and (5.3), these confidence intervals were obtained from a grid search using Kleibergen's (2005) GMM extension of the conditional likelihood ratio test to find the set of values of the coefficients of the two endogenous variables such that the null hypothesis that both these coefficients were zero could not be rejected.[22] The confidence interval for equation (4.5) was obtained from a grid search using the Anderson-Rubin test. The confidence intervals reported in Table 5 are the ranges of values of the relevant coefficients that were included in the sets obtained by grid search.

The null hypothesis that distance to Wittenberg is a redundant IV is rejected with *p* value 0.025 in equation (5.1) and 0.041 in equation (5.3). Thus there is a good

[22] I thank Mark Schaffer for clarifying this point.

Table 5: The effects of Protestantism and literacy on county economic outcomes in Prussia

Regressors	Dependent variable: per capita income tax		Dependent variable: ln(teacher income)		Dependent variable: nonagricultural share of labour force	
	IV (5.1)	OLS (5.2)	IV (5.3)	OLS (5.4)	IV (5.5)	OLS (5.6)
Share of Protestants in population 1871	-0.129	-0.136	0.0016***	0.0011***	0.164***	0.086***
	(0.184)	(0.121)	(0.0004)	(0.0002)	(0.059)	(0.032)
Literacy rate 1871	0.580	0.584***	-0.0038	-0.0002	-0.667	0.074
	(1.854)	(0.222)	(0.0032)	(0.0006)	(0.464)	(0.076)
Year of annexation, province dummy variables, and interactions	Yes	Yes	Yes	Yes	Yes	Yes
Distance to London and distance to nearest provincial capital	Yes	Yes	Yes	Yes	Yes	Yes
Distance to Berlin	No	No	No	No	Yes	Yes
Pre-industrial variables	Yes	Yes	Yes	Yes	Yes	Yes
Observations	321	321	334	334	334	334
Adjusted R^2	0.490	0.490	0.738	0.762	0.512	0.632
SW conditional F statistic for share of Protestants	14.68	-	16.55	-	32.83	-
SW conditional F statistic for literacy rate	7.29	-	9.81	-	19.69	-
Standard 95 per cent confidence interval for share of Protestants	[-0.546, 0.287]	[-0.373, 0.101]	[0.0007, 0.0025]	[0.0007, 0.0015]	[0.030, 0.298]	[0.024, 0.149]
Weak-IV-robust 95 per cent confidence interval for share of Protestants	[-0.546, 0.343]	-	[0.0009, 0.0027]	-	[0.030, 0.369]	-
Standard 95 per cent confidence interval for literacy rate	[-3.612, 4.773]	[0.149, 1.020]	[-0.0110, 0.0034]	[-0.0013, 0.0010]	[-1.717, 0.383]	[-0.074, 0.222]
Weak-IV-robust 95 per cent confidence interval for literacy rate	[-5.289, 5.891]	-	[-0.0149, 0.0024]	-	[-2.697, 0.243]	-
J test p value	0.997	-	0.478	-	-	-
C test p value	0.985	-	0.036	-	0.129	-

Notes: as for Table 4.

case for using distance to Wittenberg as an IV to estimate these two equations in order to improve the asymptotic efficiency of the estimation. Doing so also means that the *J* test of overidentifying restrictions can be used to test whether, conditional on accepting BW's argument that distance to Wittenberg is a valid IV for the share of Protestants in 1871 and Becker et al.'s argument that education in 1816 is a valid IV for the literacy rate in 1871, the share of Protestants in 1816 is also a valid IV. The *p* values for the *J* test reported for equations (5.1) and (5.3) in Table 5 show that it is, subject to this proviso. A weak-IV-robust version of the *J* test yields the same conclusion for all values of the coefficients of Protestantism and literacy in the weak-IV-robust 95 per cent confidence intervals.

As in Table 4, the estimates when per capita income tax is the dependent variable are different from those for the other two economic outcome measures. The *p* value of the *C* statistic for (5.1) is so large as to suggest that there is no need for IV estimation in this case. The OLS point estimates in equation (5.2) are so similar to the IV ones in (5.1) that there is very little danger of making a type II error in rejecting the null hypothesis that the potentially endogenous variables are exogenous. The OLS point estimate of the effect of Protestantism on per capita income tax corresponds to an elasticity of -0.043: it is therefore of essentially no economic significance as well as being poorly determined. In contrast, the OLS point estimate of the effect of literacy is well-determined and of modest, but non-trivial, economic significance, corresponding to an elasticity of 0.260.

For equation (5.3), the *C* test rejects the null hypothesis that the potentially endogenous variables are exogenous, so IV estimation is required. The 95 per cent weak-IV-robust confidence interval for the estimated effect of Protestantism shows that this is clearly positive, though its economic significance is small: the point estimate in (5.3) corresponds to an elasticity of teacher income with respect to Protestantism of 0.094. The point estimate of the effect of literacy corresponds to an elasticity of -0.325, but this effect is very poorly determined.

The *p* value of the *C* statistic for equation (5.5), in which the dependent variable is the nonagricultural share of the labour force, is 0.129, so the null hypothesis of no difference between IV and OLS estimates cannot be rejected at

conventional levels. But in this case one must take seriously concerns about the low power of the *C* test when IV estimates are not well-determined. The IV and OLS point estimates of the effects of Protestantism and literacy differ substantially, as do many of the other coefficient estimates in equations (5.5) and (5.6). The possibility of making a type II error by concluding that IV estimation is not required is sufficiently great that I focus on the results of equation (5.5). The 95 per cent weak-IV-robust confidence interval for the estimated effect of Protestantism in this equation shows that it is clearly positive. The point estimate corresponds to an elasticity of 0.299, so the economic significance of Protestantism, though modest, is non-trivial. The point estimate of the effect of literacy is much larger (in absolute value), corresponding to an elasticity of -1.723, but it is very poorly determined.

Although literacy is estimated to have a modest positive effect on per capita income tax, the estimates of its effect on the other two economic outcome measures are very imprecise, and the point estimates are negative. Is a negative effect of literacy on teacher income and the nonagricultural share of the labour force plausible? A possible explanation of the negative causal effect of literacy on the latter measure is that greater literacy in 1871 reflected increased education in the decades before that date, which reduced the supply of child labour to factories, thus increasing the cost of labour and lowering the profitability of industrial activity, so that the share of the labour force in manufacturing was lower in 1871. Some support for this explanation comes from the debates preceding the enactment of the Prussian child labour law in 1839: many opponents of this new legislation were concerned that removing children from their jobs in order to send them to school would be damaging to industry (Anderson 2013). A negative effect of literacy on the share of manufacturing occupations is therefore consistent with contemporary evidence on the Prussian economy. Whether this explanation can also apply to teacher income is unclear. On the one hand, a negative effect of literacy on industrialisation might lower the general level of incomes in a county, including teacher income. On the other hand, increased education might be expected to raise the relative income of teachers due to greater demand for them. In any event, a negative effect of literacy on this measure of county economic outcomes is not impossible.

The IV point estimates of the effects of Protestantism on economic outcomes in Table 5 are somewhat different from the corresponding estimates in Table 4, although the null hypothesis that there was no difference between the two estimates was not rejected for any of the three dependent variables. It is therefore not possible to say definitely that it is necessary to include literacy as a regressor in order to obtain consistent estimates of the effect of Protestantism on economic outcomes. Even if it is, the effect of so doing is very small, as would be expected given the very small causal effect of Protestantism on literacy.

However, BW's hypothesis that there is a positive effect of Protestantism on economic outcomes that largely reflects the higher human capital of Protestants is clearly rejected by the evidence in Tables 4 and 5. Literacy does have a positive effect on per capita income tax, but Protestantism has a negative, though poorly-determined, effect on this economic outcome measure irrespective of whether literacy is included or omitted as a regressor and whether IV or OLS is used as the estimation method. For the other two economic outcome measures, literacy has a negative, though poorly-determined, effect, while Protestantism has a well-determined positive effect, irrespective of the method of estimation and whether literacy is included or omitted. BW's hypothesis requires that the effect of Protestantism on county economic outcomes should be positive when literacy is omitted from the regression model, but not when literacy is included. The results in Tables 4 and 5 do not exhibit these features.

Although the results in Tables 4 and 5 clearly reject BW's human capital version of the Weber thesis, they do not provide a clear answer to the question of whether, in late nineteenth-century Prussia, there was any evidence of a positive pure causal effect of the share of Protestants in the population on county economic outcomes. For one measure – per capita income tax – there is no evidence of any such effect, but Protestantism did have a positive, though very small, effect on teacher income, and a positive and somewhat larger effect on the nonagricultural share of the labour force. The Prussian evidence is inconclusive, but it leaves open the possibility that there is a grain of truth in the idea that Protestantism had a positive causal effect on economic prosperity.

8. Conclusion

This paper has shown that it is not possible to sustain the argument that, in late nineteenth-century Prussia, Protestantism had a positive effect on economic prosperity which can largely be explained by the higher human capital of Protestants. BW's claim that this is the case is rendered untenable by fundamental problems with their empirical strategy. If Protestantism does have a causal effect on literacy, it is simply not possible to obtain consistent estimates of the respective causal effects of Protestantism and literacy in the absence of an IV for literacy. This paper obtained such an IV by combining the BW dataset with another dataset, used by Becker et al. to study education and industrialisation in Prussia. The combined dataset was used to estimate the causal effects of both Protestantism and literacy for a subsample of the Prussian counties considered by BW. In addition to providing IV estimates of the effects of literacy as well as Protestantism, this paper also addressed another problem with the BW empirical strategy: the absence of any measures of regional effects on Prussian economic development in their regression models. The omission of regional effects makes distance to Wittenberg (BW's IV for Protestantism) an invalid IV even if Protestantism has no causal influence on literacy..

The results in this paper show that Protestantism had a very small positive causal effect on literacy in nineteenth-century Prussia. This effect is so small as to make it unlikely that any positive effect of Protestantism on economic prosperity can be mainly due to the greater literacy of Protestants, and this is conformed by the analysis in this paper. The estimated effects of Protestantism and literacy on county economic outcomes cannot be reconciled with the BW view. Literacy had a positive effect on per capita income tax, but the pure effect of Protestantism on this economic outcome measure was negative (though poorly determined) irrespective of whether literacy was or was not included as a regressor. For the other two measures of economic outcomes, the pure effect of Protestantism was positive, irrespective of whether literacy was included as a regressor, while that of literacy was negative, though poorly determined.

Although the empirical analysis in this paper clearly rejects the BW human capital interpretation of a link between Protestantism and economic prosperity, it

leaves open the question of whether either Protestantism or literacy had a positive causal effect on county economic outcomes in nineteenth-century Prussia. The results vary according to the economic outcome measure used. There is no evidence of a positive effect of Protestantism on per capita income tax, but literacy had a modest positive effect on this measure. The effect of Protestantism on teacher income was positive, but small, while the effect on the nonagricultural share in the labour force was also positive and large enough to be of some economic significance. However, for both these measures, there is no evidence of a positive effect of literacy. The evidence for nineteenth-century Prussia provides no clear support for the hypotheses that either Protestantism or literacy had positive effects on economic prosperity, but neither does it unambiguously reject them.

References

Acemoglu, D., D. Cantoni, S.H. Johnson and J.A. Robinson (2011). 'The Consequences of Radical Reform: The French Revolution', American Economic Review, 101, 3286-3307.

Anderson, E. (2013). 'Ideas in Action: The Politics of Prussian Child Labor Reform, 1817-1839', Theory and Society, 42, 81-119.

Anderson, T.W. and H. Rubin (1949). 'Estimation of the Parameters of a Single Equation in a Complete System of Stochastic Equations', Annals of Mathematical Statistics, 20, 46-63.

Andrews, D.W.K., M.J. Moreira and J.H. Stock (2006). 'Optimal Two-sided Invariant Similar Tests for Instrumental Variables Regression', Econometrica, 74, 715-52.

Angrist, J.D. and J-S. Pischke (2009). Mostly Harmless Econometrics. Princeton: Princeton University Press.

Baum, C.F., M.E. Schaffer and S. Stillman (2003). 'Instrumental Variables and GMM: Estimation and Testing', The Stata Journal, 3, 1-31.

Baum, C.F., M.E. Schaffer and S. Stillman (2007). 'Enhanced Routines for Instrumental Variables/Generalized Method of Moments Estimation and Testing', The Stata Journal, 7, 465-506.

Baum, C.F., M.E. Schaffer and S. Stillman (2010). ivreg2: Stata module for extended instrumental variables/2SLS, GMM and AC/HAC, LIML and k-class regression. http://ideas.repec.org/c/boc/bocode/s425401.html

Becker, S.O. and L. Woessmann (2009). 'Was Weber Wrong? A Human Capital Theory of Protestant Economic History', Quarterly Journal of Economics, 124, 531-596.

Becker, S.O., E. Hornung and L. Woessmann (2011). 'Education and Catch-Up in the Industrial Revolution', American Economic Journal: Macroeconomics, 3, 92–126.

Becker, S.O., F. Cinnerella, E. Hornung and L. Woessmann (2014). 'iPEHD - The ifo Prussian Economic History Database', Historical Methods: A Journal of Quantitative and Interdisciplinary History, 47, 57-66.

Cantoni, D. (2012). 'Adopting a New Religion: The Case of Protestantism in 16th Century Germany', Economic Journal, 122, 502-31.

Edwards, J.S.S. (2018). 'A replication of "Education and Catch-Up in the Industrial Revolution" (American Economic Journal: Macroeconomics, 2011)', Economics: The Open-Access, Open-Assessment E-Journal, 12 (2018-3): 1–33. http://dx.doi.org/10.5018/economics-ejournal.ja.2018-3

Finlay, K., L.M. Magnusson, and M.E. Schaffer (2013). weakiv: Weak-instrument-robust Tests and Confidence Intervals for Instrumental-Variable (IV) Estimation of Linear, Probit and Tobit models. http://ideas.repec.org/c/boc/bocode/s457684.html

Fuller, W.A. (1977), 'Some Properties of a Modification of the Limited Information Estimator', Econometrica, 45, 939-53.

Hayashi, F. (2000). Econometrics. Princeton: Princeton University Press.

Kleibergen, F. (2005). 'Testing Parameters in GMM Without Assuming That They Are Identified', Econometrica, 73, 1103-23.

Kleibergen, F. and R. Paap (2006). 'Generalized Reduced Rank Tests Using the Singular Value Decomposition', Journal of Econometrics, 133, 97-126.

Montiel Olea, J.L. and C. Pflueger (2013). 'A Robust Test for Weak Instruments', Journal of Business and Economic Statistics, 31, 358-69.

Ogilvie, S.C. (1996a). 'Proto-Industrialization in Germany', in S.C. Ogilvie and M. Cerman (eds.), European Proto-Industrialization, Cambridge: Cambridge University Press.

Ogilvie, S.C. (1996b). 'The Beginnings of Industrialization', in S.C. Ogilvie (ed.), Germany: a New Social and Economic History, Vol. II: 1630-1800, London, Edward Arnold.

Pflueger, C. and S. Wang (2015). 'A Robust Test for Weak Instruments in Stata', The Stata Journal, 15, 216-25.

Sanderson, E . and F. Windmeijer (2016). 'A Weak Instrument F-Test in Linear IV Models with Multiple Endogenous Variables', Journal of Econometrics, 190, 212-221.

Tipton, F.B. (1976). Regional Variations in the Economic Development of Germany During the Nineteenth Century. Middletown: Wesleyan University Press.

Appendix

A1. The BW analysis on the sample of 334 counties that existed in 1849

BW used a sample of 452 counties for their analysis while I used a smaller sample of 334 counties. Table A1 shows the effects of Protestantism on the literacy rate and county economic outcomes that were obtained when the BW regression models were estimated on the sample of 334 counties. In addition to the share of Protestants in the population in 1871, BW used a number of demographic controls as regressors. These are described in the main text and were included in the Table A1 regressions, although their coefficient estimates are not reported. BW also included the share of county population with missing information on literacy in their regression model for the literacy rate in 1871, and this was included in equation (A1.1). Following BW, the estimates in Table A1 were obtained using distance to Wittenberg as an IV for the share of Protestants. Equation (A1.1) of Table A1 should be compared to equation (2) of BW's Table III, and equations (A1.2) – (A1.4) to equations (4), (5) and (6) of BW's Table V. The one difference between the estimation procedures used by BW and in Table A1 is that the standard errors in Table A1 are clustered at the level of the 280 independent units of observation in 1816.

The estimated effect of Protestantism on literacy reported in Table A1 is similar to that in equation (2) of BW's Table III. The point estimate in equation (A1.1) corresponds to an elasticity of 0.152, while the elasticity corresponding to the BW point estimate is 0.138.

The estimated effects of Protestantism on economic outcomes in Table A1 are broadly similar to those in equations (4), (5) and (6) of BW's Table V. The elasticity corresponding to the point estimate in equation (A1.2) is 0.295, compared to 0.190 in BW's equation (4). For equation (A1.3) it is 0.070, compared to 0.067 in BW's equation (5), and for equation (A1.4) it is 0.128 compared to 0.155 in BW's equation (6). As in BW's Table V, the equations for log teacher income and the share of the nonagricultural labour force in Table A1 fit much better than the equation for per capita income tax. The only difference worth noting is that the point estimate of the

Table A1: IV estimates of BW literacy and economic outcome regression models on sample of 334 counties

	Dependent variable			
	Literacy rate in 1871	Economic outcome measures		
		Per capita income tax	ln(teacher income)	Nonagricult-ural share of labour force
Regressors	(A1.1)	(A1.2)	(A1.3)	(A1.4)
Share of Protestants in population 1871	0.216***	0.938***	0.0012**	0.070
	(0.056)	(0.354)	(0.0006)	(0.046)
BW demographic controls	Yes	Yes	Yes	Yes
Share missing literacy information	Yes	No	No	No
Observations	334	321	334	334
Adjusted R^2	0.629	0.257	0.653	0.670

Notes: Figures in parentheses are standard errors clustered at the level of the 280 independent units of observation in 1816. ** and *** denote significance at the 0.05 and 0.01 levels respectively.

effect of Protestantism on the nonagricultural share in equation (A1.4) is less well-determined than in BW's equation (6), having a p value of 0.129. This is not due to the clustering of the standard errors: even if, following BW, no adjustment is made to the standard errors for heteroscedasticity or clustering, the p value is 0.117. Overall, the main features of BW's estimates from their sample of 452 counties are preserved in the smaller sample of 334.

A2. Just-identified IV estimates of the effect of Protestantism on literacy

The very high correlation between the share of Protestants in the 1816 and 1871 county populations raises the question of whether the former is a valid IV for the latter or merely the endogenous variable measured with error. To answer this question, Table A2 reports just-identified IV estimates of the equations in Table 3 of the main text. The J test of overidentification restrictions reported in Table 3 suggests that there is no statistically significant difference between the estimated effect of Protestantism on literacy obtained using distance to Wittenberg and the share of Protestants in 1816 separately as IVs. However, this may be because the point estimate using distance to Wittenberg, though very different from that using the share

Table A2: Just-identified IV estimates of the relationship between Protestantism and literacy

	IV: Distance to Wittenberg		
Regressors	(A2.1)	(A2.2)	(A2.3)
Share of Protestants in population 1871	0.121***	0.101***	0.119***
	(0.034)	(0.036)	(0.042)
Year of annexation, province dummy variables, and interactions	Yes	Yes	Yes
Distance to London and distance to nearest provincial capital	Yes	Yes	Yes
Pre-industrial variables	No	Yes	Yes
BW controls	No	No	Yes
Adjusted R^2	0.734	0.749	0.795
Kleibergen-Paap F statistic	35.62	35.85	26.32
Montiel Olea-Pflueger F statistic	35.64	35.87	26.34
Standard 95 per cent confidence interval for share of Protestants	[0.053, 0.188]	[0.031, 0.171]	[0.036, 0.202]
Weak-IV-robust 95 per cent confidence interval for share of Protestants	[0.057, 0.201]	[0.032, 0.179]	[0.044, 0.228]
	IV: Share of Protestants in population 1816		
Regressors	(A2.4)	(A2.5)	(A2.6)
Share of Protestants in population 1871	0.090***	0.089***	0.083***
	(0.017)	(0.019)	(0.018)
Year of annexation, province dummy variables, and interactions	Yes	Yes	Yes
Distance to London and distance to nearest provincial capital	Yes	Yes	Yes
Pre-industrial variables	No	Yes	Yes
BW controls	No	No	Yes
Adjusted R^2	0.735	0.749	0.796
Kleibergen-Paap F statistic	1829.55	2130.87	1333.72
Montiel Olea-Pflueger F statistic	1830.61	2132.10	1334.49
Standard 95 per cent confidence interval for share of Protestants	[0.057, 0.123]	[0.052, 0.127]	[0.049, 0.118]
Weak-IV-robust 95 per cent confidence interval for share of Protestants	[0.059, 0.126]	[0.053, 0.129]	[0.051, 0.120]

Notes: Figures in parentheses are standard errors clustered at the level of the 280 independent units of observation in 1816. *** denotes significance at the 0.01 level.

of Protestants in 1816, is also very poorly determined. Table A2 shows that this is not the case.

For all three regression models, the point estimate of the effect of Protestantism on literacy is somewhat higher when distance to Wittenberg is the IV (the top panel) than when it is the share of Protestants in 1816 (the bottom panel).

However, the point estimates in equations (A2.1) - (A2.3) are all well-determined, so it is reasonable to conclude from the failure of the *J* test to reject the overidentification restrictions that these differences are not statistically significant. Furthermore, the economic significance of the difference between the two sets of just-identified IV estimates is small: for example, in equation (A2.1) the point estimate corresponds to an elasticity of 0.085 while in (A2.4) it corresponds to an elasticity of 0.064. Thus there is no evidence in Table A2 of significant differences between the just-identified IV estimates obtained using distance to Wittenberg and the share of Protestants in 1816. Conditional on the former being a valid IV, so too is the latter.

A3. The specification of the regressions reported in Table 4

In this section I discuss several alternative specifications of regressions used to estimate the effect of Protestantism on county economic outcomes in the main text of the paper. I also present estimates which use different IVs. Doing this provides justification for the specification of the regressions reported in Table 4 of the main text.

There are nine columns and three panels in Table A3. The three panels show the results obtained using the three different measures of county economic outcomes. The nine columns show the estimates obtained with different sets of regressors and IVs.

For all nine columns and all three panels, the null hypothesis that the coefficients of the province dummy variables, the year of annexation by Prussia, and interactions between the dummies and the year of interaction are all zero is strongly rejected. There is overwhelming evidence that these variables should be included in a satisfactory regression model of Prussian county economic outcomes. If they are omitted, their correlation with distance to Wittenberg, documented in Table 1 of the paper, will make it an invalid IV. Columns (A3.1) – (A3.3) show the results from a regression that, as well as the share of Protestants in county population in 1871, includes these variables. The estimates in the three columns are obtained using, in column (A3.1), distance to Wittenberg as the sole IV, the share of Protestants in county population in 1816 as the sole IV in column (A3.2), and the two IVs together

Table A3: Alternative IV estimates of county economic outcome regression models

	Dependent variable: per capita income tax								
Regressors	(A3.1)	(A3.2)	(A3.3)	(A3.4)	(A3.5)	(A3.6)	(A3.7)	(A3.8)	(A3.9)
Share of Protestants in 1871	0.819	0.170	0.185	-0.169	0.015	0.010	-0.074	-0.083	-0.222*
	(0.304)	(0.134)	(0.134)	(0.306)	(0.134)	(0.133)	(0.284)	(0.137)	(0.128)
Year of annexation, province dummy variables, and interactions	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
p value for year of annexation, province dummy variables and interactions	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Distance to London and distance to nearest provincial capital	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes
p value for distance variables	-	-	-	0.000	0.000	0.000	0.001	0.000	0.006
Pre-industrial variables	No	No	No	No	No	No	Yes	Yes	Yes
p value for pre-industrial variables	-	-	-	-	-	-	0.000	0.000	0.002
BW demographic controls	No	No	No	No	No	No	No	No	Yes
Number of observations	321	321	321	321	321	321	321	321	321
Adjusted R^2	0.319	0.365	0.365	0.402	0.407	0.407	0.488	0.488	0.559
Distance to Wittenberg used as IV	Yes	No	Yes	Yes	No	Yes	Yes	No	Yes
p value for test of redundancy of distance to Wittenberg	-	-	0.067	-	-	0.004	-	-	0.001
Share of Protestants in 1816 used as IV	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Kleibergen-Paap *F* statistic	46.88	1842.23	1120.43	33.94	1674.03	1389.00	34.48	1864.79	1137.90
Montiel Olea-Pflueger *F* statistic	46.91	1843.37	882.00	33.96	1675.06	1065.00	34.50	1865.94	731.11
J test *p* value	-	-	0.016	-	-	0.530	-	-	0.686
Weak-IV-robust 95 per cent confidence interval for share of Protestants	[0.254, 1.504]	[-0.090, 0.441]	[-0.086, 0.498]	[-0.860, 0.425]	[-0.256, 0.275]	[-0.270, 0.257]	[-0.670, 0.500]	[-0.370, 0.182]	[-0.502, 0.027]

	Dependent variable: ln(teacher income)								
	(A3.1)	(A3.2)	(A3.3)	(A3.4)	(A3.5)	(A3.6)	(A3.7)	(A3.8)	(A3.9)
Share of Protestants in 1871	0.0023***	0.0017***	0.0017***	0.0012*	0.0015***	0.0015***	0.0015***	0.0012***	0.0009***
	(0.0006)	(0.0003)	(0.0003)	(0.0006)	(0.0004)	(0.0004)	(0.0005)	(0.0002)	(0.0002)
Year of annexation, province dummy variables, and interactions	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
p value for year of annexation, province dummy variables and interactions	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Distance to London and distance to nearest provincial capital	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes
p value for distance variables	-	-	-	0.018	0.012	0.011	0.559	0.351	0.108
Pre-industrial variables	No	No	No	No	No	No	Yes	Yes	Yes
p value for pre-industrial variables	-	-	-	-	-	-	0.000	0.000	0.000
BW demographic controls	No	No	No	No	No	No	No	No	Yes
Number of observations	334	334	334	334	334	334	334	334	334
Adjusted R^2	0.512	0.524	0.524	0.531	0.531	0.531	0.760	0.762	0.849
Distance to Wittenberg used as IV	Yes	No	Yes	Yes	No	Yes	Yes	No	Yes
p value for test of redundancy of distance to Wittenberg	-	-	0.095	-	-	0.007	-	-	0.020
Share of Protestants in 1816 used as IV	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Kleibergen-Paap *F* statistic	47.42	1989.80	1198.52	35.62	1829.55	1467.58	35.85	2130.87	1127.24
Montiel Olea-Pflueger *F* statistic	47.45	1990.95	954.13	35.64	1830.61	1146.86	35.87	2132.10	796.24
J test *p* value	-	-	0.202	-	-	0.652	-	-	0.036
Weak-IV-robust 95 per cent confidence interval for share of Protestants	[0.0013, 0.0036]	[0.0010, 0.0024]	[0.0011, 0.0025]	[-0.00003, 0.0026]	[0.0008, 0.0022]	[0.0008, 0.0022]	[0.006, 0.0026]	[0.0008, 0.0017]	[0.0006, 0.0013]
	Dependent variable: nonagricultural share of labour force								
Share of Protestants in 1871	0.094	0.122***	0.122***	0.636	0.159***	-	-	0.105***	0.033
	(0.061)	(0.041)	(0.040)	(1.089)	(0.045)	-	-	(0.037)	(0.025)

	(A3.1)	(A3.2)	(A3.3)	(A3.4)	(A3.5)	(A3.6)	(A3.7)	(A3.8)	(A3.9)
Year of annexation, province dummy variables, and interactions	Yes	Yes	Yes	Yes	Yes	-	-	Yes	Yes
p value for year of annexation, province dummy variables and interactions	0.000	0.000	0.000	0.005	0.000	-	-	0.000	0.017
Distance to London and distance to nearest provincial capital	No	No	No	Yes	Yes	-	-	Yes	Yes
p value for distance variables	-	-	-	0.306	0.000	-	-	0.000	0.000
Distance to Berlin	-	-	-	0.158	0.053***	-	-	0.029**	0.005
				(0.238)	(0.016)	-	-	(0.014)	(0.010)
Pre-industrial variables	No	No	No	No	No	-	-	Yes	Yes
p value for pre-industrial variables	-	-	-	-	-	-	-	0.000	0.000
BW demographic controls	No	No	No	No	No	-	-	No	Yes
Number of observations	334	334	334	334	334	-	-	334	334
Adjusted R^2	0.379	0.379	0.379	-0.002	0.413	-	-	0.632	0.809
Distance to Wittenberg used as IV	Yes	No	Yes	Yes	No	-	-	No	Yes
p value for test of redundancy of distance to Wittenberg	-	-	0.075	-	-	-	-	-	-
Share of Protestants in 1816 used as IV	No	Yes	Yes	No	Yes	-	-	Yes	Yes
Kleibergen-Paap *F* statistic	47.42	1989.80	1198.52	0.841	1117.35	-	-	1097.55	827.02
Montiel Olea-Pflueger *F* statistic	47.45	1990.95	954.13	0.841	1118.00	-	-	1098.18	827.50
J test *p* value	-	-	0.619	-	-	-	-		
Weak-IV-robust 95 per cent confidence interval for share of Protestants	[-0.033, 0.217]	[0.043, 0.205]	[0.040, 0.200]	(-∞, ∞)	[0.069, 0.246]	-	-	[0.031, 0.176]	[-0.018, 0.083]

Notes: Figures in parentheses are standard errors clustered at the level of the 280 independent units of observation in 1816. *, ** and *** denote significance at the 0.10, 0.05 and 0.01 levels respectively.

in (A3.3). When per capita income tax is the dependent variable, the just-identified IV point estimate of the effect of Protestantism using distance to Wittenberg is very different from that using the share of Protestants in 1816, and unsurprisingly the *J* test in column (A3.3) rejects the over-identification restriction. However, for each of the other two dependent variables, the two just-identified IV point estimates of the effect of Protestantism are much more similar, and the *J* test in column (A3.3) does not reject the over-identification restriction. In both these cases, the *p* values for the test of whether distance to Wittenberg is a redundant IV conditional on the share of Protestants in 1816 being an IV show that distance to Wittenberg is rather a weak IV, so not too much weight should be placed on the results of the *J* tests. However, for both log teacher income and nonagricultural share, a weak-IV-robust version of the *J* test also fails to reject the over-identification restriction for all values of the coefficient of Protestantism in the weak-IV-robust 95 per cent confidence intervals.

Columns (A3.4) – (A3.6) show the results when the distances of a county from London, the nearest provincial capital, and - in the case when the dependent variable is the nonagricultural share of the labour force - Berlin are added to the set of regressors. As discussed in the main text, the very high partial correlation between distance to Berlin and distance to Wittenberg makes the latter an unusably weak IV if the former is included as a regressor. It is justifiable to omit distance to Berlin from the regressions when the dependent variable is per capita income tax or log teacher income, but not when it is the nonagricultural share of the labour force. The weakness of distance to Wittenberg as an IV when distance to Berlin is a regressor is driven home by the results reported in column (A3.4) for the case in which the dependent variable is the nonagricultural share. These show that when distance to Wittenberg is used as the sole IV, the Kleibergen-Paap and Montiel Olea-Pflueger *F* statistics are tiny and the weak-IV-robust 95 per cent confidence interval for the coefficient of Protestantism is infinite.[23] Because of this, no other regressions involving the use of distance to Wittenberg are reported in Table A3 when the dependent variable is the nonagricultural share and distance to Berlin is a regressor.

[23] Since this confidence interval was calculated using a grid search, it was not, of course, calculated for an infinite range of values: strictly, the confidence interval includes all values from -1000 to 1000.

When per capita income tax is the dependent variable, the two just-identified IV point estimates of the effect of Protestantism in columns (A3.4) and (A3.5) are still rather different, though they are so poorly determined that the *J* test in column (A3.6) does not reject the over-identification restriction. In this case the null hypothesis that distance to Wittenberg is a redundant IV is clearly rejected. For log teacher income, the two just-identified IV point estimates are reasonably similar and well-determined, distance to Wittenberg is not a redundant IV, and the over-identification restriction is not rejected. Thus the addition of distance to London and distance to the nearest provincial capital as regressors leads to the two just-identified IV estimates of the effect of Protestantism on both per capita income tax and log teacher income being not statistically significantly different. There is, however, still some doubt about whether there is no economically significant difference between the two just-identified IV point estimates in the case of per capita income tax.

Columns (A3.7) and (A3.8) show how the just-identified IV estimates of the effect of Protestantism on per capita income tax and log teacher income are changed when the pre-industrial variables are added to the list of regressors. For the nonagricultural share of the labour force, only the just-identified IV estimate using the share of Protestants is reported. This is, of course, identical to equation (4.5) in Table 4 of the main text.

When the pre-industrial variables are added, the two just-identified IV point estimates of the effect of Protestantism on per capita income are very similar indeed. Equation (4.1) in Table 4 reports the results of estimating this regression using both distance to Wittenberg and the share of Protestants in 1816 as IVs. The null hypothesis that the former is a redundant IV is strongly rejected, and the *J* test does not reject the over-identification restriction. For this economic outcome measure, when the regression equation is as specified in columns (A3.7) and (A3.8), the share of Protestants in 1816 is clearly a valid IV, conditional on distance to Wittenberg being valid.

The addition of the pre-industrial variables reverses the two just-identified IV point estimates of the effect of Protestantism on log teacher income by comparison with the case when these regressors were omitted, but they remain reasonably similar.

The over-identified estimates of this regression are shown as equation (4.3) in Table 4. Distance to Wittenberg is not a redundant IV and the *J* test does not reject the over-identification restriction. When log teacher income is the dependent variable and the regression equation is as specified in columns (A3.7) and (A3.8), if distance to Wittenberg is a valid IV then so too is the share of Protestants in 1816.

All the regressions reported in columns (A3.1) – (A3.8) of Table A3 omit the demographic control regressors used by BW in their analysis. As pointed out in the main text, many of these variables are likely to be endogenous in a regression model of economic outcomes: population growth and migration, for example, are influenced by economic conditions in a county. These demographic variables are therefore likely to be bad controls, and for this reason they are omitted from the regressions reported in Table 4 of the paper. In order to show that the conclusions drawn from Table 4 do not depend on omitting these variables, column (A3.9) of Table A3 shows the estimated effect of Protestantism on the three economic outcome measures when the demographic controls are added to equations (4.1), (4.3) and (4.5) of Table 4.

For all three economic outcome measures, the addition of these variables lowers the point estimate of the effect of Protestantism by comparison with the IV point estimates reported in Table 4. The reduction is least pronounced when the dependent variable is log teacher income and most pronounced when it is per capita income tax. The signs of the point estimates are, however, unchanged by the addition of the demographic controls. Thus the conclusions from Table 4 about the effect of Protestantism on economic outcomes would not be altered in substance by the inclusion of BW's demographic controls as regressors, though the magnitudes of the point estimates would all be somewhat lower.