Appendix to Jeremy Edwards "Did Protestantism Promote Prosperity via Higher Human Capital? Replicating the Becker-Woessmann (2009) Results"

This Appendix provides, in Sections A.1 – A.6, greater detail on a number of points that are touched upon in the main text. Sections A.7 and A.8 discuss the implications of the analyses of Cantoni (2015), Dittmar (2011), and Rubin (2014) for the results reported in this paper.

A.1. Regional Effects on Prussian Economic Outcomes

The importance of regional effects is a central theme in the literature on nineteenth-century German economic development.[1] Nineteenth-century Prussia consisted of territories that had been part of the Prussian state for greatly differing 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 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 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.[2] Of the 452 counties in BW's database, 66 had been Prussian since 1525, 108 had become Prussian in the 1810s, and 86 had been Prussian only since 1866. The remaining counties had become Prussian at various dates after 1525 but before 1866.

[1] Tipton (1976); Ogilvie (1996b), 265; Tilly and Kopsidis (2020), 4, Ch. 2.
[2] The Peace of Tilsit in 1807 involved Prussia ceding many counties to France and states associated with France: nearly half of Prussian territory was lost. These counties were returned to Prussia in 1815, and in this paper are treated as having been Prussian since their original date of acquisition.

Contrary to BW's claim that institutional heterogeneity is not a problem for their empirical analysis,[3] the institutional frameworks of different parts of Prussia differed substantially, and this variation influenced economic development throughout the nineteenth century.[4] In the eastern parts of Prussia, manorialism survived until the early nineteenth century, with consequent harmful effects on agricultural productivity and entrepreneurship in general.[5] The powers of feudal landlords remained strong even after the formal abolition of Prussian serfdom in 1806, and factory industrialisation in eastern Prussia was delayed until the later nineteenth century. Silesia had very dense proto-industry in the early nineteenth century, but its factory industrialisation was hampered by the desire of feudal landlords to protect their proto-industrial feudal revenues by resisting technological improvements in linen production, in which they were supported by the Prussian state. Westphalia too had some linen-based industrial regions at the beginning of the nineteenth century, but landlords and village communities had considerable institutional control over rural society. Furthermore, towns were strong enough to compel rural proto-industrial producers to sell through urban staple markets, and the Prussian state supported them in so doing well into the nineteenth century. The result was that many Westphalian linen regions de-industrialised, and mechanised linen production did not begin until the 1850s.[6] The most economically advanced part of Prussia in 1815 was the Rhineland. Here there was a decline in landlord power by the sixteenth century, which allowed flexible land use, livelier commerce, more open rural markets, and labour markets to operate free of the constraints of serfdom.[7] The Rhineland was also characterised by extensive political fragmentation, which enabled proto-industries easily to cross territorial boundaries in order to locate where political and institutional

[3] BW (2009), 533.
[4] Ogilvie (1996b), 265.
[5] Ogilvie (2014).
[6] Ogilvie (1996a), 125, Ogilvie (1996b), 289.
[7] Kisch (1989); Ogilvie (1996b), 283.

conditions were least oppressive. The Rhineland institutional framework remained favourable to economic growth throughout the nineteenth century.[8]

The regional variation in Prussian economic development was already evident at the beginning of the nineteenth century. Hardach (1991) dates the beginning of the German industrial revolution to 1784, when a mechanised spinning plant was opened in the Rhineland town of Ratingen. Kaufhold (1986) identifies 39 industrial regions, defined as having an above-average density of industrial employment and a large proportion of output sold beyond the region, in Germany around 1800. Of these, 16 were in territory that was part of Prussia by 1815: nine in the Rhineland, five in Westphalia, and two in Silesia. In contrast, there were no such industrial regions in the central and eastern parts of Prussia around 1800, which reflects the institutional obstacles to economic development that existed in these parts of Prussia in the eighteenth century.

It might be argued that these regional variations in the Prussian institutional framework were actually determined by whether regions were predominantly Protestant or Catholic. If this were the case, including regional variables as regressors when estimating the effect of Protestantism on economic outcomes would be inappropriate: the regional variables would be bad controls (Angrist and Pischke 2009, 64) and the estimated effect of Protestantism would be biased. However, there is no reason to think that the regional variations were determined by Protestantism. In the eastern parts of Prussia, irrespective of their religious composition, manorialism and the powers of feudal landlords persisted until well into the nineteenth century. The provinces of Prussia and Posen were the two most easterly ones in the entirety of Prussia.[9] In 1871, the mean value of the percentage of Protestants for all counties in Prussia was 64.2 and the median was 83.1. For the counties in the province of Prussia, the mean was 72.4 and the median was 92.5; for Posen, the

[8] Kisch (1989), Ogilvie (1996a), 124-5.
[9] From 1829 to 1878 there was a province of Prussia in the Kingdom of Prussia.

corresponding values were 31.7 and 28.4. Both provinces had a long history of strong feudal landlords (in the case of Posen much of this was as part of the Polish-Lithuanian Commonwealth) which did not depend on whether they were predominantly Protestant or Catholic.

The institutional features which allowed the Rhineland to be the most economically advanced part of Prussia in the early nineteenth century were a weak manorial system and political fragmentation. These features dated from the medieval period (Kisch 1989 15-18, Tilly and Kopsidis 37). They cannot therefore have been a consequence of the religious changes brought about by the Reformation in the sixteenth century.

Differences in institutions such as landlord and town power were an important and deep-rooted influence on Prussian economic development, which did not depend on Protestantism. It is essential to take account of them in any analysis of the determinants of prosperity in nineteenth-century Prussia.

A.2. Identification of the Effect of Protestantism on Prosperity

As Section A.1 showed, there are strong reasons to expect prosperity in nineteenth-century Prussia to be influenced by long-standing regional differences in institutions such as landlord and town power. In the Rhineland in particular, but also to some extent in Westphalia and Silesia, the institutional framework was more favourable to economic development than in Pomerania, Posen, and the province of Prussia. Thus an analysis of the effects of Protestantism on prosperity in Prussia must either specify a regression model that allows for possible effects that differ by location in Prussia, or omit any such location variables and use an IV for Protestantism that can plausibly be argued to be independent of these omitted locational effects on prosperity.

BW omit location variables from their regression models and use distance from Wittenberg as an IV for Protestantism. BW (2009, 557-63) and Cantoni (2012) provide

strong arguments that distance from Wittenberg is related to the spread of Protestantism and is not related to some possible omitted influences on prosperity in Prussia. But, by its nature as a spatial variable, distance from Wittenberg is very unlikely to be independent of the omitted locational influences on prosperity, a point that neither BW nor Cantoni address.

The regions of Prussia that had institutional frameworks better or worse suited to economic development were not uniformly either nearer to or further from Wittenberg. The province that was furthest from Wittenberg – the province of Prussia – had one of the institutional frameworks least favourable to economic development in Prussia. The Rhineland province was geographically nearer to Wittenberg than the province of Prussia, but still further from it than most other provinces. It had the institutional framework that was probably most conducive to prosperity in all Prussia. The provinces of Posen, Silesia, and Westphalia were all about the same distance from Wittenberg, and closer to it than the Rhineland. However, their institutional frameworks differed, with that of Posen being similar to that of the province of Prussia, while those of Silesia and Westphalia were relatively favourable to economic development. The province of Pomerania was geographically closer still to Wittenberg, but had an institutional framework similar to that of Posen and the province of Prussia. The relationship between distance from Wittenberg and the parts of Prussia with more or less favourable institutions for economic development was therefore not a straightforward one, but this does not mean that distance from Wittenberg satisfies the exclusion restriction required for it to be a valid IV for Protestantism in regressions that omit any variables capturing regional effects. Even a complicated relationship between omitted regional effects and distance from Wittenberg will make it an invalid IV. It is essential to investigate whether BW's conclusions about the effects of Protestantism on prosperity in Prussia are robust to modifications of their regression specification which allow for regional effects on Prussian economic development.

A natural way to do this is to suppose that there are unobserved district effects on prosperity. Late-nineteenth-century Prussia was divided into 13 provinces, which in turn were further subdivided into a total of 35 administrative districts (*Regierungsbezirke*). Modelling the unobserved regional effects as district-specific components of the overall error term in the cross-section regressions used by BW means that standard panel-data techniques can be employed to estimate the effect of the share of Protestants in a county in 1871 on prosperity while allowing for district effects.

District effects alone may not capture all the effects of differing institutional frameworks on county economic outcomes, since in 18 of the 35 districts there was some within-district variation in the length of time that counties had been part of the Prussian state, and it is possible that the institutional framework operating in a county depended partly on the length of time it had been Prussian. Thus there are two ways in which regional effects on prosperity across Prussia can be incorporated into regression specifications. One is simply to add unobserved district effects to the BW regression model. The other is to include, alongside unobserved district effects, a linear spline of the year in which a county became Prussian, in order to allow for any effects of the length of time a county had been Prussian on its institutional framework. The spline begins in 1525 and has knots at 1742, when the major Prussian acquisition of territories in the eighteenth century began, and 1815, when Prussia expanded substantially at the end of the Napoleonic Wars.

The doubts expressed above about whether distance from Wittenberg is a valid IV for Protestantism in regression models of prosperity were due to the omission of regional effects from the BW regression model. Provided that regional effects are incorporated into the regression specification, the arguments of Becker and Woessmann and Cantoni that distance from Wittenberg is a valid IV for Protestantism are convincing. The analysis which follows maintains the hypothesis that it satisfies the exclusion restriction when unobserved district effects are taken into account.

The appropriate panel-data estimation technique to use depends on whether the district-specific component of the overall error term is correlated with the regressors. If it is not, then both between- and within-district variation can be used for estimation. Table A1 reports the estimated total effect of Protestantism on the three prosperity measures obtained when this random-effects framework is used. The regressions in Table A1 include the demographic controls and the linear spline in the year of becoming Prussian, and their errors are assumed to be clustered at the district level. The confidence intervals are therefore based on cluster-robust estimates of the variance matrix using Stata's finite-sample adjustments.[10] The regressions in Table A1 are thus comparable to the regressions reported in the third and fourth columns of Table 1 in the main text.

The use of both between- and within-district variation for the estimates in Table A1 means that first-stage *F* statistics for the IV estimates in it are larger than those for the estimates in the third column of Table 1. However, they are not large enough to exceed the Montiel Olea-Pflueger critical value of 23.109 at which it is possible to reject at the 0.05 level the null hypothesis that the approximate asymptotic bias of the IV estimator is 10 per cent of a worst-case benchmark, so weak IV problems are still present.[11] Table A1 shows that taking account of unobserved district effects in a random-effects framework produces estimates of the total effect of Protestantism on income tax per capita and teacher income that are somewhat different from the BW estimates. For the former, both IV and OLS estimates of the effect are lower, though still imprecise, while for the latter the IV estimate is larger and clearly positive. The estimated total effects on Protestantism on the non-agricultural share in Table A1 are, however, not very different from the BW estimates, and the Moran test shows that spatial correlation of the idiosyncratic errors is present in equation (A1.6).

[10] These involve multiplying the cluster-robust variance matrix by (N-1)/(N-K)*G/(G-1), where N is the number of observations, K is the number of parameters estimated, and G is the number of clusters, and using critical values for hypothesis tests from the *t* distribution with G-1 degrees of freedom.
[11] This worst-case benchmark is when the IV is completely uninformative and the first- and second-stage errors are perfectly correlated.

Table A1. Random-Effects Estimates of the Total Effect of Protestantism on County Prosperity in Prussia

	Dependent variable: Income tax per capita	
Estimation method	IV	OLS
	A1.1	A1.2
Share of Protestants	0.173	-0.009
	[-0.58, 1.52]	[-0.23, 0.21]
F statistic	23.468	-
p value of Moran test	0.276	0.595
p value of Mundlak test	0.000	0.000
Number of observations	426	426
	Dependent variable: Log teacher income	
Estimation method	IV	OLS
	A1.3	A1.4
Share of Protestants	0.196	0.086
	[0.05, 0.41]	[0.04, 0.13]
F statistic	20.808	-
p value of Moran test	0.252	0.468
p value of Mundlak test	0.000	0.000
Number of observations	452	452
	Dependent variable: Nonagricultural share	
Estimation method	IV	OLS
	A1.5	A1.6
Share of Protestants	0.092	0.055
	[-0.02, 0.21]	[0.01, 0.10]
F statistic	20.821	-
p value of Moran test	0.366	0.001
p value of Mundlak test	0.000	0.000
Number of observations	452	452

Notes. All equations include the following demographic control variables: the proportions of the population of each county in 1871 that were, respectively, aged below 10, female, Jewish, native-born, Prussian, blind, deaf-mute, and insane, and average household size, log of population size, and population growth from 1867 to 1871 in each county. They also include a linear spline in the year a county became Prussian with knots at 1742 and 1815. Figures in brackets are 95 per cent confidence intervals based on cluster-robust estimates of the variance matrix. The confidence intervals for the IV estimates are weak-IV-robust ones. The null hypothesis of the Mundlak test is that there is no correlation between the unobserved district effects and the regressors.

However, for all six regressions in Table A1, the *p* values of the Mundlak tests (Mundlak 1978; Wooldridge 2010 pp. 331-3) reject the null that the district-specific error component is uncorrelated with the regressors. Therefore these estimates are not consistent. In order to obtain consistent estimates when the unobserved district effects are correlated with the regressors, it is necessary to use a fixed-effects framework, in which only within-district variation can be used for estimation of the regression models.

Is there evidence that regressions which incorporate unobserved district effects in a fixed-effects framework are preferable to the BW specification? Mundlak tests can also be

used to test the BW regression model against both the fixed-effects model that incorporates unobserved district effects and the fixed-effects model that allows for the year of becoming Prussiam as well as unobserved district effects. For all three prosperity measures, the results of these tests provide very strong support for the two models which allow for district effects. The *p* values for the test of the null that there are no district effects are all 0.000.

Thus there is compelling evidence that district effects should be taken into account in a satisfactory regression model of late nineteenth-century Prussian county prosperity, and that this has to be implemented in a fixed-effects framework. If district effects are not taken into account, then distance from Wittenberg is likely to be an invalid IV for Protestantism, and the effect of Protestantism on prosperity will not be identified. However, dealing with this identification difficulty may create another identification problem. Using only within-district variation to estimate the regression models is likely to make distance from Wittenberg a weak IV, in which case the estimated effect of Protestantism may be so imprecise that very little can be said about it.

Does the use of within-district variation only change distance from Wittenberg from a strong but invalid IV for Protestantism to a weak but valid IV? Table A2 shows that the answer to this question is no. This table reports, for the sample of 452 observations, the estimated effect of distance from Wittenberg on the share of Protestants in the first-stage regressions used for the IV estimates reported in Table 1 of the main text and Table A2 in this Appendix.[12] The additional regression specification in Table A2 that is not also in Table 1 of the main text is the one which allows for unobserved district effects, but not the year of becoming Prussian.

Table A2 reports three different 95 per cent confidence intervals for the effect of distance from Wittenberg on the share of Protestants, obtained from three different

[12] The estimates using the smaller sample of 426 observations are very similar.

assumptions about the variance matrix of the first-stage regression errors. The first is that these errors are independently and identically distributed (IID), which is what BW assume. The second is that the errors are heteroscedastic, in which case the confidence intervals are based on heteroscedasticity-robust estimates of the variance matrix. The third is that the errors are clustered by district, in which case the confidence intervals are based on a cluster-robust estimate of the variance matrix using Stata's finite-sample adjustments. Table A2 also reports *F* statistics corresponding to each of these three assumptions about the variance-covariance matrix of the first-stage regression errors.

The point estimate of the effect of distance from Wittenberg on the share of Protestants hardly differs between the BW specification and the specification which includes unobserved district effects and year of becoming Prussian. This point estimate is somewhat lower for the specification which allows only for district effects. However, as the elasticities reported in Table A2 show, none of these differences are of more than very modest economic significance. Furthermore, the differences are not statistically significant whatever assumption is made about the error variance matrix.

The most striking feature of the results in Table A2 is the difference in the precision of the estimated effects of distance from Wittenberg according to whether the regression errors are assumed to be clustered by district. As discussed in the main text, there is a strong case for clustering the errors of the BW regression models. The omission from these models of any variables reflecting regional effects means that the regression errors in particular locations are likely to have the same sign. The Moran test discussed in the main text shows that there is evidence of positive spatial correlation of the errors in BW's regression models of prosperity. This is consistent with the existence of unobserved district effects that result in clustered regression errors. Table A2 also reports *p* values for Moran tests of spatial correlation in the first-stage regression errors. The null hypothesis of no spatial correlation is rejected for the first-stage regressions that do not take account of district effects. The Moran

Table A2. The First-Stage Relationship between Protestantism and Distance from Wittenberg in Prussia in 1871

	Dependent variable: Share of Protestants in County in 1871		
	A2.1	A2.2	A2.3
Distance from Wittenberg	-0.095	-0.079	-0.093
IID 95% conf. int.	[-0.12, -0.07]	[-0.13, -0.03]	[-0.14, -0.04]
Heteroscedasticity-robust 95% conf. int.	[-0.12, -0.07]	[-0.13, -0.03]	[-0.14, -0.04]
Cluster-robust 95% conf. int.	[-0.16, -0.03]	[-0.14, -0.02]	[-0.15, -0.04]
IID *F* statistic	75.071	8.825	13.071
Heteroscedasticity-robust *F* statistic	70.493	10.430	14.165
Cluster-robust *F* statistic	9.555	7.677	10.580
p value of Moran test	0.000	0.351	0.389
Elasticity	-0.484	-0.400	-0.474
District effects	No	Yes	Yes
Year Prussian variables	No	No	Yes
Adjusted R^2	0.401	0.720	0.740

Notes. The number of observations is 452 for all equations. All equations include the following demographic control variables: the proportions of the population of each county in 1871 that were, respectively, aged below 10, female, Jewish, native-born, Prussian, blind, deaf-mute, and insane, and average household size, log of population size, and population growth from 1867 to 1871 in each county. The elasticities of the share of Protestants in 1871 with respect to the distance from Wittenberg are calculated at sample mean values.

statistic in this case indicates positive spatial correlation of the errors, so nearby counties tend to have similar error values, which again is consistent with clustering due to unobserved district effects.

In the regression in Table A2 which does not take account of district effects, and so is estimated using both between- and within-district variation, the first-stage *F* statistics are 70 or higher when the errors are assumed to be either IID or heteroscedastic. However, if the errors in this regression are assumed to be clustered by district, the first-stage *F* statistic is less than 10. Both regressions which allow for district effects, irrespective of the assumption made about the error variance matrix, have first-stage *F* statistics of about 14 or less. But, in the case when the errors are assumed to be clustered by district, the first-stage *F* statistic is a little larger for the regression that allows for district effects and the year of becoming Prussian (A2.3) than it is for the regression that takes no account of district effects (A2.1). Whether distance from Wittenberg is a strong or a weak IV for the share of Protestants therefore depends primarily on whether estimation allows for errors that are clustered by district, not on whether unobserved district effects are taken into account. The apparent

strength of the IV in BW's regressions derives from their implicit assumption that the errors in their regression models of county prosperity are not clustered by district.

Table A2 shows that distance from Wittenberg becomes a weak IV for the share of Protestants if the regression errors are assumed to be clustered by district, irrespective of whether estimation uses between- and within-district variation, or solely the latter. Taking account of district effects deals with the threat to identification of the effect of Protestantism on prosperity posed by the probable invalidity of distance from Wittenberg as an IV when regional effects are omitted, and only creates a new identification problem if one is willing to argue that the errors in BW's regressions should not be clustered by district. The evidence of spatial correlation in these errors makes this an argument that is very difficult to maintain.

A.3. Alternative Regression Models Incorporating District Effects

Table A3 reports the results of estimating the two regression specifications which incorporate unobserved district effects on prosperity, but differ according to whether a linear spline in the year of becoming Prussian is included as a regressor. The appropriate assumption to make about the errors in the regressions reported in Table A3 is not clear. The overall error comprises a district-specific component and an idiosyncratic observation-specific component. The district-specific component of the overall error may be sufficient to control for clustering, in which case it would not be necessary to cluster the idiosyncratic error by district. Table A3 therefore reports two 95 per cent confidence intervals for each estimated effect of a regressor. The upper confidence interval is obtained from heteroscedasticity-robust estimates of the variance matrix, while the lower one is obtained from estimates of this matrix that are robust to both heteroscedasticity and clustering by the 35 districts, which, for simplicity, will be described as cluster-robust. The cluster-robust estimates reported in Table A3 are obtained using Stata's finite-sample adjustments to standard errors and hypothesis tests.

Table A3. Alternative Fixed-Effects Estimates of the Total Effect of Protestantism and the Year of Becoming Prussian on County Prosperity in Prussia

	A. Dependent variable: Income tax per capita			
Estimation method	IV	OLS	IV	OLS
	A3.1	A3.2	A3.3	A3.4
Share of Protestants	-0.572	-0.003	-0.351	0.009
	[-2.70, 0.98]	[-0.19, 0.18]	[-1.81, 0.92]	[-0.19, 0.20]
	[-3.06, 1.16]	[-0.19, 0.19]	[-2.10, 1.02]	[-0.19, 0.21]
Elasticity	-0.185	-0.001	-0.114	0.003
Year Prussian to 1742	-	-	0.089	0.146
	-	-	[-0.16, 0.34]	[-0.01, 0.31]
			[-0.28, 0.46]	[-0.14, 0.43]
Year Prussian 1742-1815	-	-	-0.455	-0.501
	-	-	[-0.80, -0.11]	[-0.80, -0.13]
			[-1.03, 0.12]	[-1.01, 0.01]
Year Prussian from 1815	-	-	0.895	0.943
	-	-	[0.25, 1.54]	[0.18, 1.58]
			[0.39, 1.40]	[0.44, 1.45]
District effects	Yes	Yes	Yes	Yes
Adjusted R^2	0.570	0.598	0.594	0.604
F statistic	10.349	-	14.651	-
	10.204		13.469	
p value of Moran test	0.319	0.308	0.311	0.302
p value of test of exogeneity	0.424	-	0.541	-
	0.402		0.519	
	B. Dependent variable: Log teacher income			
Estimation method	IV	OLS	IV	OLS
	A3.5	A3.6	A3.7	A3.8
Share of Protestants	0.359	0.096	0.309	0.101
	[0.04, 0.95]	[0.04, 0.16]	[0.04, 0.73]	[0.04, 0.16]
	[0.04, 0.87]	[0.04, 0.15]	[0.04, 0.65]	[0.04, 0.16]
Elasticity	0.231	0.062	0.199	0.065
Year Prussian to 1742	-	-	0.048	0.014
	-	-	[-0.01, 0.11]	[-0.02, 0.04]
			[-0.03, 0.13]	[-0.03, 0.06]
Year Prussian 1742-1815	-	-	-0.031	-0.018
	-	-	[-0.10, 0.04]	[-0.08, 0.04]
			[-0.14, 0.08]	[-0.10, 0.07]
Year Prussian from 1815	-	-	0.183	0.192
	-	-	[-0.10, 0.47]	[-0.06, 0.44]
			[0.01, 0.36]	[0.04, 0.35]
District effects	Yes	Yes	Yes	Yes
Adjusted R^2	0.642	0.728	0.677	0.728
F statistic	10.430	-	14.165	-
	7.677		10.580	
p value of Moran test	0.260	0.273	0.258	0.268
p value of test of exogeneity	0.073	-	0.092	-
	0.055		0.072	
	C. Dependent variable: Non-agricultural share			
Estimation method	IV	OLS	IV	OLS
	A3.9	A3.10	A3.11	A3.12
Share of Protestants	0.211	0.054	0.182	0.058
	[-0.03, 0.62]	[0.01, 0.09]	[-0.02, 0.47]	[0.01, 0.10]
	[-0.20, 0.71]	[-0.01, 0.12]	[-0.14, 0.51]	[-0.01, 0.13]
Elasticity	0.399	0.101	0.344	0.110
Year Prussian to 1742	-	-	0.035	0.015
	-	-	[-0.01, 0.08]	[-0.01, 0.04]
			[-0.01, 0.08]	[-0.02, 0.05]
Year Prussian 1742-1815	-	-	-0.030	-0.022

	-	-	[-0.09, 0.03]	[-0.08, 0.03]
			[-0.10, 0.04]	[-0.09, 0.04]
Year Prussian from 1815	-	-	0.105	0.111
	-	-	[-0.05, 0.26]	[-0.02, 0.24]
			[0.01, 0.20]	[0.04, 0.18]
District effects	Yes	Yes	Yes	Yes
Adjusted R^2	0.707	0.750	0.725	0.750
F statistic	10.430	-	14.165	-
	7.677		10.580	
p value of Moran test	0.402	0.421	0.403	0.418
p value of test of exogeneity	0.170	-	0.211	-
	0.294		0.353	

Notes. The number of observations is 426 for Panel A and 452 for Panels B and C. All equations include the following demographic control variables: the proportions of the population of each county in 1871 that were, respectively, aged below 10, female, Jewish, native-born, Prussian, blind, deaf-mute, and insane, and average household size, log of population size, and population growth from 1867 to 1871 in each county. The year of becoming Prussian variables are a linear spline in the year a county became Prussian with knots at 1742 and 1815. The IV estimates were obtained using the Stata commands *ivreg2* and *xtivreg2* (Baum et al. 2010, Schaffer 2010). Figures in brackets are 95 per cent confidence intervals. In each pair of confidence intervals, the upper one is based on heteroscedasticity-robust estimates of the variance matrix, and the lower one is based on cluster-robust estimates of this matrix. The confidence intervals for the IV estimates are weak-IV-robust ones obtained using the Stata command *weakiv* (Finlay et al. 2013). In each pair of *F* statistics, the upper one is based on heteroscedasticity-robust estimates of the variance matrix, and the lower one is based on cluster-robust estimates of this matrix. In each pair of *p* values for the test of exogeneity, the upper one is based on heteroscedasticity-robust estimates of the variance matrix, and the lower one is based on cluster-robust estimates of this matrix. The elasticities with respect to the share of Protestants are calculated at sample mean values. In Panel B, the elasticity is of teacher income with respect to Protestantism.

Table A3 also reports two first-stage *F* statistics for each IV regression. These are Kleibergen-Paap *F* statistics: in each pair, the upper one is obtained from the heteroscedasticity-robust estimates and the lower one from the cluster-robust estimates. The Montiel Olea-Pflueger critical values at which it is possible to reject at the 0.05 level the null hypothesis that the approximate asymptotic bias of the IV estimator is respectively 20 or 30 per cent of a worst-case benchmark are 15.062 and 12.039. The IV regressions in Table A3 therefore suffer from serious weak-IV problems, and the point estimates of the total effect of Protestantism in them are unreliable. The 95 per cent confidence intervals reported for the share of Protestants in the IV regressions are therefore weak-IV-robust ones.

The Moran tests show no evidence of spatially correlated idiosyncratic errors for any of the regressions reported in Table A3. This provides support for the view that it is not necessary to cluster the idiosyncratic errors by district. However, there are no great differences between the heteroscedasticity- and cluster-robust confidence intervals reported

in Table A3, which suggests that whether the idiosyncratic errors are assumed to be clustered does not matter very much. In almost all cases, the clustered confidence intervals are the larger ones, but this is not invariably so. For the third part of the spline in year of becoming Prussian (the years from 1815 onwards), the cluster-robust confidence intervals are always smaller than the heteroscedasticity robust ones. The same is true for the share of Protestants when log teacher income is the measure of prosperity. In the following discussion I focus on the cluster-robust estimates.

In all the regressions which include the year of becoming Prussian from 1815, the cluster-robust 95 per cent confidence intervals for the effect of this variable contain only positive values. The size of this effect is economically significant. Other things equal, a county that became Prussian in 1866 rather than 1815 is estimated to have higher prosperity by an amount which ranges from about 10 per cent of the sample mean value when teacher income is the measure of prosperity to about 23 per cent when the measure is income tax per capita. The estimated effects of becoming Prussian at dates between 1525 and 1814, however, are in most cases small and always poorly determined.

Thus there is clear evidence that the year of becoming Prussian after 1815 had a positive effect on prosperity. This means that the preferred regression specification is the one in the third and fourth columns of Table A3. The equations in the third column also have the advantage over those in the first column that distance from Wittenberg is a less weak IV for Protestantism when the year of becoming Prussian is included as a regressor. Although the first-stage F statistics for (A3.3), (A3.7), and (A3.11) still indicate weak-IV problems, they are larger than those for (A3.1), (A3.5), and (A3.9), and the 95 per cent confidence intervals for the estimated effects of Protestantism are narrower.

There are some differences between the IV estimates of the total effect of Protestantism in the first and third columns of Table A3. The 95 per cent weak-IV-robust confidence intervals are wider in the first column, and their upper bounds are higher than the

corresponding intervals in the third column. However, the similarities outweigh the differences. The total effect of Protestantism on income tax per capita is estimated very imprecisely in both (A3.1) and (A3.3), and most of both confidence intervals consist of negative values. The total effect on the non-agricultural share is also imprecisely estimated in both (A3.9) and (A3.11), but in this case positive values comprise the majority of both confidence intervals. The total effect on log teacher income is more precisely estimated, and is unambiguously positive in both (A3.5) and (A3.7). The overall conclusion from the IV estimates of the total effect of Protestantism in Table A3 does not depend on the particular regression specification: for two of the three prosperity measures there is no clear evidence of a positive total effect, but for log teacher income there is such an effect.

Table A3 reports the *p* values of a control function test of whether the share of Protestants is an exogenous regressor in the IV regressions. In each pair of *p* values, the upper one is obtained from the heteroscedasticity-robust estimates and the lower one from the cluster-robust estimates. If the null that the share of Protestants is exogenous is not rejected, then there is no need to use IV rather than OLS as the estimation method. A problem with this test when the IV is weak is that the IV estimates are biased towards the OLS ones, so the power of the test is low. Thus the test results in Table A3 have to be interpreted carefully. The *p* values for the exogeneity tests in equations (A3.5) and (A3.7) certainly cannot justify any claim that IV estimation is unnecessary in these cases.

In the main text, the discussion of the estimated total effects of Protestantism which allow for unobserved district effects is based on the specification that includes year of becoming Prussian as a regressor, since the results in Table A3 show that this is the preferred specification. However, the conclusions in the main text about the total effects of Protestantism when district effects are taken into account do not depend on this choice of regression specification.

A.4. Comparing different estimates of the total effect of Protestantism

A natural question to ask is whether there are statistically significant differences between the estimated total effects of Protestantism on prosperity from the regressions that do and do not take account of unobserved district effects. Answering this question is not straightforward, because of the weakness of distance from Wittenberg as an IV for the share of Protestants. This creates a number of problems. First, the IV point estimates of the total effects of Protestantism in the regressions reported in Table 1 of the main text are biased towards the OLS ones. Second, standard tests of the difference between point estimates cannot be used, as these are based on asymptotic approximations to the distribution of IV estimators which are unreliable when the IV is weak. Third, an IV that is weak as well as invalid can easily produce estimates that are more inconsistent than OLS.

This third problem has particularly serious implications for the IV estimates of the total effect of Protestantism on prosperity in the three BW regression equations in Table 1 of the main text. In contrast to the results reported by BW, which are based on the assumption that the regression errors are IID, distance from Wittenberg is a weak IV in these regressions when the errors are assumed to be clustered by district. As discussed in the main text, such clustering of the errors in the BW regressions is to be expected if regional effects influence prosperity, and as section A.1 of this Appendix points out, the historical evidence provides strong reasons to expect such effects to be important. The weak-IV-robust 95 per cent confidence intervals for the total effect of Protestantism based on the clustered errors in equations (1.1), (1.5), and (1.9) contain both positive and negative values, so no unambiguous statement about the sign of these effects is possible. At best, there is no clear evidence of positive total effects of Protestantism from the BW regression specification when the errors are clustered. But there are further problems with the BW estimates. The evidence of spatially correlated errors in the BW regressions suggests that they are misspecified because they do not take account of regional effects on prosperity. The omission of any variables reflecting

such regional effects from the BW regressions means that it is likely that distance from Wittenberg is an invalid IV in them as well as a weak one. It is therefore very unclear what interpretation can be given to the IV estimates of the total effect of Protestantism from the BW regressions in Table 1 of the main text.

Inference about the differences between the two sets of IV estimates of the total effect of Protestantism on prosperity in Table 1 has to be based on the 95 per cent weak-IV-robust confidence intervals. These confidence intervals are so wide that there are no clear differences between the IV estimates from the BW specification of the total effect of Protestantism, in the first column of Table 1, and the corresponding IV estimates from the alternative specification in the third column of Table 1. For income tax per capita, the 95 per cent confidence intervals for the total effect of Protestantism in equations (1.1) and (1.3) differ in that the former contains mostly positive, and the latter mostly negative, values, but there is a substantial overlap between the two. There is more of a difference for the log of teacher income, where, although there is considerable overlap between the confidence intervals in (1.5) and (1.7), the former includes some negative values and has an upper bound markedly smaller than the latter, which consists entirely of positive values. For the non-agricultural share, the confidence interval in (1.9) is wholly contained in that for (1.11).

The fact that the 95 per cent weak-IV-robust confidence intervals for the total effect of Protestantism in Table 1 of the main text do not show clear differences between the estimates from the two different specifications does not mean that there is no basis for preferring one set of estimates to the other. As has been pointed out, the BW specification is unsatisfactory. The spatial correlation of the errors in the BW specification, and the problems created by the combination of weakness and invalidity of distance from Wittenberg as an IV in the BW regression model, favour the IV estimates from the specification incorporating district effects and the year of becoming Prussian over those from the BW one.

In contrast to the IV estimates, where the weakness of the IV prevented the use of standard tests of the difference between the estimates, it is possible to test whether there is a difference between the OLS estimates of the total effect of Protestantism from the two specifications in Table 1. There are no statistically significant differences between these estimates for the three prosperity measures. In two cases this is because the corresponding estimates are fairly well determined but similar, but in the case of income tax per capita the two estimates are rather different but poorly determined. Once again, this does not mean that there is no basis for preferring the specification that allows for district effects and the year of becoming Prussian to the BW specification. The evidence of spatial correlation in the BW regression errors suggests that regional effects have been omitted, and hence that the alternative specification is preferable.

A.5. OLS Point Estimates of the Association between Literacy and Prosperity

As explained in the main text, in order to estimate the direct effect of Protestantism on prosperity, BW adjust their measures of prosperity by a range of values for the effect of literacy on prosperity. These values are calculated by multiplying the share of literates in a county in 1871 by a range of values based on the point estimates of the association between prosperity and literacy from OLS regressions of prosperity measures on the share of Protestants in a county in 1871, the share of literates, the demographic controls, and the proportion of the county population for which literacy information is missing (BW 2009, p. 571 and equations (7)-(9) in Table V). However, these regressions do not take account of regional effects on prosperity, and hence the estimated associations between literacy and prosperity are likely to be biased.

Table A4 reports OLS estimates of the association between literacy and the two prosperity measures for which there is evidence of a positive total effect of Protestantism. For each measure, two sets of estimates are shown: the BW ones (equations A4.1 and A4.3)

Table A4. OLS Estimates of the Association between Prosperity and Literacy in Prussia

	Dependent variable			
	Ln teacher income		Non-agricultural share	
	A4.1	A4.2	A4.3	A4.4
Share of literates	0.636	0.156	0.490	0.260
	[0.48, 0.79]	[0.003, 0.31]	[0.34, 0.64]	[0.08, 0.43]
Elasticity	0.56	0.14	1.26	0.67
District effects	No	Yes	No	Yes
Year Prussian variables	No	Yes	No	Yes
Adjusted R^2	0.572	0.728	0.643	0.755
p value of Moran test	0.000	0.263	0.000	0.391
No. of observations	452	452	452	452

Notes. The number of observations is 452 for all equations. These estimates are obtained from regressions all of which include the following regressors: the share of Protestants in county population in 1871, the proportions of the population of each county in 1871 that were, respectively, aged below 10, female, Jewish, native-born, Prussian, blind, deaf-mute, insane, and missing literacy information; and the average household size, the log of population size, and population growth from 1867 to 1871 in each county. Figures in brackets are 95 per cent confidence intervals based on cluster-robust estimates of the variance matrix. The elasticities with respect to the share of literates are calculated at sample mean values. For equations (A4.1) and (A4.2), the elasticity is of teacher income with respect to Protestantism.

which do not allow for district effects, and the alternative ones (equations A4.2 and A4.4) which take account of district effects and the year of becoming Prussian. The point estimates of the association between literacy and the prosperity measures are much smaller when allowance is made for these additional influences on prosperity. The differences between the point estimates which do and do not take account of the additional influences are significant both economically, as can be seen from the elasticities shown in the table, and statistically. Table A4 also reports the *p* values of the Moran tests for spatial correlation in the regression errors. There is strong evidence of such correlation in the errors of the regressions BW use to estimate the associations between literacy and prosperity, which suggests that they omit spatial influences. The Moran statistics for equations (A4.1) and (A4.3) are all positive, which is consistent with the presence of unobserved district effects on prosperity. In contrast, there is no evidence of spatial correlation in the errors of equations (A4.2) and (A4.4). BW's failure to allow for district effects and other influences on prosperity means that their measures of prosperity net of the effect of literacy are based on point estimates of the association of prosperity with literacy that are too large.

A.6. Is There an Urban-Rural Difference in the Total Effect Of Protestantism?

Cantoni (2015) analysed the effect of Protestantism on the economic development of 272 cities in the German Lands over the period 1300-1900, using city population as a measure of economic growth, and found no evidence of such an effect. In an attempt to reconcile his findings with those of BW, Cantoni suggested that Protestantism had a positive effect on prosperity in rural counties of Prussia in the latter part of the nineteenth century, which constitute the majority of BW's sample, but no effect on prosperity in urban counties. Cantoni's argument was that in the late nineteenth century differences in literacy mattered for economic development in rural counties, so that BW's version of the Weber thesis generated a link between Protestantism and prosperity. In the urban counties, however, literacy was high by the beginning of the nineteenth century and did not differ between Protestant and Catholic cities, so that there was no relationship between Protestantism and prosperity. The cities analysed by Cantoni were drawn from the Bairoch et al. (1988) data set. Using BW's OLS regression specification, Cantoni shows that there is no association between Protestantism and BW's prosperity measures in a subsample consisting of those counties in the full BW sample that contained one or more cities from this data set (Cantoni 2015, Appendix B.7). In Cantoni's view, the difference between his and BW's results can be explained by distinguishing urban and rural parts of Prussia, and restricting BW's version of the Weber thesis to apply only to the latter.

If district effects and the year of becoming Prussian are taken into account, does the association between Protestantism and prosperity differ between urban and rural parts of Prussia? To answer this question, Table A5 reports the results of OLS regressions analysing the association between Protestantism and prosperity in two subsamples of the full BW sample of Prussian counties. Following Cantoni, those counties that contained one or more cities from the Bairoch et al. data set are taken to be urban, and the remainder are assumed to

Table A5. The Total Effect of Protestantism on Prosperity in Urban and Rural Counties in Prussia

	Dependent variable: Income tax per capita			
	BW specification		District effects and other regressors	
	Counties with cities in Bairoch et al.	Counties without cities in Bairoch et al.	Counties with cities in Bairoch et al.	Counties without cities in Bairoch et al.
Share of Protestants	-0.037	0.210	0.203	0.083
	[-0.48, 0.40]	[-0.07, 0.49]	[-0.33, 0.74]	[-0.13, 0.30]
District effects	No	No	Yes	Yes
Year Prussian variables	No	No	Yes	Yes
Adjusted R^2	0.078	0.411	0.514	0.668
p value of Moran test	0.268	0.000	0.217	0.347
Number of observations	108	318	101	317
	Dependent variable: Log teacher income			
	BW specification		District effects and other regressors	
	Counties with cities in Bairoch et al.	Counties without cities in Bairoch et al.	Counties with cities in Bairoch et al.	Counties without cities in Bairoch et al.
Share of Protestants	0.011	0.053	0.111	0.087
	[-0.09, 0.11]	[-0.004, 0.11]	[-0.01, 0.24]	[0.03, 0.14]
District effects	No	No	Yes	Yes
Year Prussian variables	No	No	Yes	Yes
Adjusted R^2	0.519	0.394	0.784	0.689
p value of Moran test	0.000	0.000	0.276	0.232
Number of observations	134	318	131	317
	Dependent variable: Non-agricultural share			
	BW specification		District effects and other regressors	
	Counties with cities in Bairoch et al.	Counties without cities in Bairoch et al.	Counties with cities in Bairoch et al.	Counties without cities in Bairoch et al.
Share of Protestants	-0.032	0.042	0.131	0.029
	[-0.10, 0.04]	[-0.002, 0.09]	[0.05, 0.22]	[-0.05, 0.11]
District effects	No	No	Yes	Yes
Year Prussian variables	No	No	Yes	Yes
Adjusted R^2	0.609	0.519	0.793	0.675
p value of Moran test	0.001	0.005	0.269	0.440
Number of observations	134	318	131	317

Notes. All equations include the following demographic control variables: the proportions of the population of each county in 1871 that were, respectively, aged below 10, female, Jewish, native-born, Prussian, blind, deaf-mute, and insane, and average household size, log of population size, and population growth from 1867 to 1871 in each county. Figures in brackets are 95 per cent confidence intervals based on cluster-robust estimates of the variance matrix. The number of observations for the regressions that allow for district effects is smaller than for those that do not because, for some districts, there was only one observation, which could not therefore generate within-district variation and so had to be dropped.

be rural. Table A5 follows Cantoni in not reporting IV regressions, the reason being that distance from Wittenberg is such a weak IV for Protestantism in the subsample of urban counties that in many cases the estimates were completely uninformative. One complication is that I found 134 counties that contained cities from the Bairoch et al. data set, while

Cantoni uses only 115 such counties. However, the main feature of Cantoni's findings is also present in this larger subsample of urban counties.

The point estimates of the association between Protestantism and prosperity in the regressions in Table A5 which use the BW specification are all lower for the subsample of urban counties than for that of rural counties. This is consistent with the results reported in Table B.5 of the Appendix to Cantoni (2015), although the values of the point estimates differ because of the different subsample sizes. The null hypothesis that the association in urban and rural counties differs is, however, rejected only when prosperity is measured by the non-agricultural share. Furthermore, in all but one of these six regressions, the Moran test rejects the null of no spatial correlation in the errors, so this specification is not satisfactory.

When the regression specification includes unobserved district effects and the year of becoming Prussian, the Moran tests show no evidence of spatial correlation in the idiosyncratic errors. For this specification, the point estimates of the association between Protestantism and prosperity are all larger in the urban county subsample, though again it is only when prosperity is measured by the non-agricultural share that the difference is statistically significant. Cantoni's suggested reconciliation of the difference between his and BW's findings does not apply, therefore, when regional effects and other sources of heterogeneity in different parts of Prussia are taken into account: there is no evidence that the association between Protestantism and prosperity was present only in rural counties. The question why there is no evidence of a positive total effect of Protestantism on city growth in the German Lands over a very long period, but evidence of a positive total effect on some measures of prosperity in Prussian counties in the later nineteenth century, remains to be resolved.

A.7. Prosperity, Protestantism and the Spread of Printing

The effects of the spread of the printing press in the second half of the fifteenth century have been analysed by Dittmar (2011) and Rubin (2014). Using distance from Mainz, the city in which Gutenberg established the first printing press in Europe, as an IV for the establishment of a printing press in a city, Dittmar shows that cities which had a press by 1500 grew substantially more rapidly in the sixteenth century than those which did not. Also using distance from Mainz as an IV, Rubin shows that cities which had a press by 1500 were substantially more likely to adopt Protestantism by 1600. These findings mean that a relationship between city growth and the adoption of Protestantism in sixteenth-century Europe obtained from an analysis which fails to take account of the establishment of printing presses might be spurious. If these effects of the spread of the printing press persisted over centuries, the estimated effects of Protestantism on prosperity in nineteenth-century Prussia obtained from regressions which do not take account of the spread of printing on both prosperity and Protestantism may also be biased.

To address this possibility, Table A6 reports regressions in which the distance of the counties in the BW dataset from Mainz is added as an exogenous regressor to the specifications in Table 1 of the main text. Dittmar (2011) argues that printing technology diffused in concentric circles from Mainz in the second half of the fifteenth century, and that the distance of a city from Mainz was not correlated with other influences on city growth. Hence the distance of a county from Mainz can be taken as a exogenous measure of the likelihood that printing was established at a relatively early date in that county. Including distance from Mainz as a regressor ensures that the estimated total effect of Protestantism is not biased because of the possible correlation of Protestantism with an omitted variable that influenced prosperity: the early establishment of a printing press.

The IV estimates of the BW regression specification with distance to Mainz added as a regressor all show a statistically and economically significant effect of this variable on the

Table A6. The Total Effect of Protestantism on County Prosperity in Prussia When Distance from Mainz is Included

	A. Dependent variable: Income tax per capita			
Estimation method	IV	OLS	IV	OLS
	A6.1	A6.2	A6.3	A6.4
Share of Protestants	0.077	0.110	-1.254	-0.003
	[-1.13, 1.36]	[-0.18, 0.40]	[-7.43, 0.21]	[-0.20, 0.20]
Distance from Mainz	-0.110	-0.110	-0.149	-0.117
	[-0.18, -0.04]	[-0.19, -0.03]	[-0.30, 0.003]	[-0.24, 0.002]
District effects	No	No	Yes	Yes
Year Prussian variables	No	No	Yes	Yes
Adjusted R^2	0.347	0.347	0.483	0.607
F statistic	18.784	-	7.651	-
p value of Moran test	0.000	0.000	0.314	0.278
p value of test of exogeneity	0.943	-	0.092	-
	B. Dependent variable: Log teacher income			
Estimation method	IV	OLS	IV	OLS
	A6.5	A6.6	A6.7	A6.8
Share of Protestants	-0.071	0.053	0.271	0.100
	[-0.37, 0.12]	[0.004, 0.10]	[-0.16, 0.61]	[0.04, 0.16]
Distance from Mainz	-0.037	-0.035	-0.006	-0.009
	[-0.05, -0.02]	[-0.05, -0.02]	[-0.03, 0.02]	[-0.03, 0.02]
District effects	No	No	Yes	Yes
Year Prussian variables	No	No	Yes	Yes
Adjusted R^2	0.540	0.587	0.694	0.727
F statistic	17.704	-	7.474	-
p value of Moran test	0.055	0.093	0.254	0.264
p value of test of exogeneity	0.139	-	0.154	-
	C. Dependent variable: Non-agricultural share			
Estimation method	IV	OLS	IV	OLS
	A6.9	A6.10	A6.11	A6.12
Share of Protestants	-0.040	0.028	0.061	0.057
	[-0.31, 0.11]	[-0.01, 0.06]	[-0.92, 0.39]	[-0.02, 0.13]
Distance from Mainz	-0.025	-0.024	-0.019	-0.019
	[-0.04, -0.01]	[-0.03, -0.01]	[-0.05, 0.02]	[-0.05, 0.01]
District effects	No	No	Yes	Yes
Year Prussian variables	No	No	Yes	Yes
Adjusted R^2	0.627	0.646	0.751	0.751
F statistic	17.704	-	7.474	-
p value of Moran test	0.303	0.241	0.374	0.374
p value of test of exogeneity	0.330	-	0.981	-

Notes. The number of observations is 426 for Panel A and 452 for Panels B and C. All equations include the following demographic control variables: the proportions of the population of each county in 1871 that were, respectively, aged below 10, female, Jewish, native-born, Prussian, blind, deaf-mute, and insane, and average household size, log of population size, and population growth from 1867 to 1871 in each county. The IV estimates were obtained using the Stata commands *ivreg2* and *xtivreg2* (Baum et al. 2010, Schaffer 2010). Figures in brackets are 95 per cent confidence intervals based on cluster-robust estimates of the variance matrix. The confidence intervals for the IV estimates are weak-IV-robust ones obtained using the Stata command *weakiv* (Finlay et al. 2013).

three prosperity measures (the point estimates correspond to elasticities at sample mean values of -0.25, -0.16, and -0.33 for income tax per capita, teacher income, and the non-agricultural share respectively). Adding distance to Mainz also somewhat increases the

strength of distance from Wittenberg as an IV for Protestantism compared to the corresponding regressions in Table 1. The first-stage *F* statistics in equations (A6.1), (A6.5), and (A6.9) are all greater than the Montiel Olea-Pflueger (2013) critical value at which it is possible to reject the null hypothesis that the approximate asymptotic bias of the IV estimator is 20 per cent of a worst-case benchmark.

The point estimates of the total effect of Protestantism in these equations are all much smaller than those in the corresponding equations of Table 1, though imprecisely estimated. The results of the exogeneity tests for these regressions do not suggest that IV estimation is necessary, and there is less reason to be concerned about the IV estimates being biased towards the OLS ones here, since distance from Wittenberg is less weak an IV. Although distance from Mainz continues to be economically and statistically significant in the OLS regressions that do not take account of district effects, there is little difference between the estimated total effects of Protestantism on prosperity in (A6.2), (A6.6), and (A6.10) and those in the corresponding equations in Table 1. However, the Moran test shows that, for two of the three prosperity measures, the errors in the regressions which include distance from Mainz are spatially correlated, so that the BW specification continues to be unsatisfactory even when this spatial variable is added as a regressor.

The regressions in the third and fourth columns of Table A6 take account of unobserved district effects and include the spline in year of becoming Prussian. The Moran tests for these regressions show no evidence of spatially correlated errors. The distance of a county from Mainz was, of course, correlated with the district in which it was located, so it is possible that any effects of the former on prosperity will be absorbed into the unobserved district effects. However, the point estimates for the effect of distance from Mainz on the non-agricultural share are a little smaller than in the corresponding equations in the first two columns of Table A6, while those for its effect on income tax per capita are a little larger. For these two prosperity measures, the point estimates correspond to elasticities (at sample mean

values) of about -0.25 to -0.35, although they are poorly determined. In contrast, the point estimates of the effect of distance from Mainz on log teacher income are substantially reduced by making allowance for district effects and the year of becoming Prussian. For whatever reason, distance from Mainz has an ambiguous effect on prosperity in these regressions.

In the IV regressions in the third column of Table A6, distance from Wittenberg is an even weaker IV for Protestantism than it is in Table 1 of the main text, and the estimated total effects of Protestantism in equations (A6.3), (A6.7), and (A6.11) are therefore more imprecise. The 95 per cent weak-IV-robust confidence interval for the total effect of Protestantism on income tax per capita in (A6.3) almost entirely comprises negative values, and the *p* value of the exogeneity test is 0.092, so the evidence that the total effect of Protestantism on income tax per capita is negative is stronger when distance from Mainz is included. However, for the other two prosperity measures, the OLS confidence interval for the total effect of Protestantism is wholly contained in the weak-IV-robust one, and the exogeneity tests do not suggest that IV estimation is necessary, although once again it must be remembered that weak IV estimates are biased towards OLS ones. The exogeneity test for (A6.7) does not, therefore, provide strong evidence that IV estimation for log teacher income is unnecessary. There is hardly any difference between the OLS estimates of the total effect of Protestantism on log teacher income and the non-agricultural share in the final column of Table 1 of the main text and those in the final column of Table A6. A comparison of the relevant weak-IV-robust confidence intervals in the third columns of Table 1 and Table A6 also suggests that any differences are quite small.

The overall conclusion is that adding distance from Mainz as a regressor to take account of the possible long-term effects of the spread of the printing press on prosperity in nineteenth-century Prussia does not lead to very different estimates of the total effects of Protestantism. BW's IV estimates are indeed changed, but they were not satisfactory in any

case. In the regressions which make allowance for unobserved district effects and include the spline in year of becoming Prussian as well as distance from Mainz, there are no real differences in the estimated total effects of Protestantism on log teacher income and the non-agricultural share compared to those from the regressions in the last two columns of Table 1 in the main text. However, for income tax per capita the inclusion of distance from Mainz changes the estimated total effect from being possibly negative to almost certainly negative. As for distance from Mainz itself, its point estimate in the regressions which include year of becoming Prussian and take account of district effects is negative and economically significant in the case of income tax per capita and the non-agricultural share, but not well-determined.

Additional references for Appendix

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

Bairoch, P., J. Batou and P. Chèvre (1988). La Population des Villes Européenes, 800-1850: Banque de Données et Analyse Sommaire des Résultats. Geneva: Droz.

Cantoni, D. (2015). 'The Economic Effects of the Protestant Reformation: Testing the Weber Hypothesis in the German Lands', Journal of the European Economic Association, 13, 561-598.

Dittmar, J.E. (2011). 'Information Technology and Economic Change: The Impact of the Printing Press', Quarterly Journal of Economics, 126, 1133-1172.

Kaufhold, K.-H. (1986). 'Gewerbelandschaften in der frühen Neuzeit', in H. Pohl (ed.), Gewerbe- and Industrielandschaften vom Spätmittelalter bis ins 20. Jahrhundert. Stuttgart: Steiner.

Kisch, H. (1989). From Domestic Manufacture to Industrial Revolution. New York and Oxford: Oxford University Press.

Mundlak, Y. (1978). 'On the Pooling of Time Series and Cross Section Data', Econometrica, 46, 69-85.

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.

Rubin, J. (2014). 'Printing and Protestants: An Empirical Test of the Role of Printing in the Reformation', Review of Economics and Statistics, 96, 270-286.

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

Wooldridge, J.M. (2010). Econometric Analysis of Cross-Section and Panel Data (second edition). Cambridge: MIT Press.