CESifo WORKING PAPERS

10180
2022

December 2022

Response to Guinnane and Hoffman: Medieval Anti-Semitism, Weimar Social Capital, and the Rise of the Nazi Party: A Reconsideration

Nico Voigtländer, Hans-Joachim Voth

CESifo

Impressum:

CESifo Working Papers
ISSN 2364-1428 (electronic version)
Publisher and distributor: Munich Society for the Promotion of Economic Research - CESifo GmbH
The international platform of Ludwigs-Maximilians University's Center for Economic Studies and the ifo Institute
Poschingerstr. 5, 81679 Munich, Germany
Telephone +49 (0)89 2180-2740, Telefax +49 (0)89 2180-17845, email office@cesifo.de
Editor: Clemens Fuest
https://www.cesifo.org/en/wp
An electronic version of the paper may be downloaded
· from the SSRN website: www.SSRN.com
· from the RePEc website: www.RePEc.org
· from the CESifo website: https://www.cesifo.org/en/wp

CESifo *Working Paper No. 10180*

Response to Guinnane and Hoffman: Medieval Anti-Semitism, Weimar Social Capital, and the Rise of the Nazi Party: A Reconsideration

Abstract

Guinnane and Hoffman (subsequently GH) comment on two of our papers: Voigtländer and Voth: "Persecution Perpetuated" (2012, subsequently PP) and Satyanath, Voigtländer and Voth: Bowling for Fascism (2017, subsequently BF). They allege that our econometric results are fragile and depend on outliers in the state of Bavaria; that our results do not account for the role of institutional actors, and that we 'misinterpret' history. This brief response addresses these allegations and shows that i) GH's empirical criticisms are targeted at small subsets of our results; ii) use ad hoc, restrictive specifications – standard procedures to address GH's concerns about outliers actually confirm our results; iii) GH's conceptual critique is misguided and based on a misrepresentation of Weimar history, especially when it comes to the case of Bavaria. In sum, the empirical findings in PP and BF stand as in our original publications.

Nico Voigtländer
Anderson School of Management
University of California Los Angeles (UCLA)
110 Westwood Plaza
USA – Los Angeles, CA 90095-1481
nico.v@anderson.ucla.edu

Hans-Joachim Voth
UBS Foundation Economics Department
University of Zurich
Schönberggasse 1
Switzerland - 8001 Zurich
joachimvoth@gmail.com

December 21, 2022
We thank our co-author on the "Bowling for Fascism" paper, Shanker Satyanath, as well as Paola Giuliano and Sebastian Ottinger for helpful comments and suggestions.

Guinnane and Hoffman (subsequently GH) comment on two of our papers: Voigtländer and Voth: "Persecution Perpetuated" (QJE 2012, subsequently PP) and Satyanath, Voigtländer and Voth: Bowling for Fascism (JPE 2017, subsequently BF). They allege that our econometric results are fragile and depend on outliers in the state of Bavaria; that our results do not account for the role of institutional actors, and that we 'misinterpret' history. This brief response addresses these allegations and shows that our results are not fragile, but remarkably robust, and that our alleged 'misunderstandings' are actually well-founded in the historical record. The table below summarizes the main points of GH's critique and our responses. Section I provides our detailed reply regarding PP, and Section II discusses our responses regarding BF.

Summary of main points of critique by GH and our response

GH Comments	Summary of Response
Empirical results in PP are driven by outliers in the state of Bavaria	GH focus on only one of six outcome variables (NSDAP votes in 1928) and use a highly restrictive specification, completely dropping Bavaria. For the other five outcomes, the coefficient of interest barely changes, even when excluding Bavaria. When using a more efficient estimation (with a Bavaria dummy, and an algorithm that gives lower weight to 'outliers'), the results for 1928 NSDAP votes are also highly robust. (Section I.A)
The results in PP fail "basic placebo tests" because the vote for two selected other parties also correlates with medieval pogroms.	GH present a variety of specifications, using arbitrary subsamples of our data without justification as to why these are needed for the placebo tests. The basis of GH's placebo criticism disappears when we use the same specifications and samples as in our original paper – the placebo coefficients are small and almost always insignificant.
The empirics in PP do not appropriately account for the historical institutional context, there are problems with the coding of some variables, and placebo tests contradict our findings.	For institutions to bias our results, they would have to be correlated at the local level over six centuries. GH fail to recognize this point and do not provide any evidence why such a (historically unlikely) local pattern would exist. Instead, they focus on *state-level* institutions (in Bavaria). But these are absorbed by state fixed effects, which we already controlled for in PP. We also accounted for a possible role of local economic or political correlates in PP. Placebo tests with our main specification show no false positives. In general, GH's critique often builds up straw men, exaggerating the claims made in PP, while ignoring the evidence that we presented to address the respective concerns. We also show that GH's allegations about coding problems are misguided. (Section I.B – I.D).
In BF, the results that condition on government stability are not robust and depend on the coding of the political stability in Bavaria, which in turn hinges on the role of the BVP party.	This point focuses on a subset of results in a subsidiary analysis. Our main results are not questioned, and not even the principal result on political stability (Prussia vs. the rest of the Weimar Republic). Instead, GH focus on a subsample that drops Prussia and thus about half the observations. We show that even in this subsample, contrary to GH's claim, our results are robust. Regarding the historical context, GH argue that the BVP party was a source of stability, and that Bavaria – despite its tumultuous history – should therefore be classified as having an above-average political stability score. This is not merely doubtful; it is simply wrong. Leading historians who studied the BVP emphasize its conservative-authoritarian traditions and opposition to parliamentary democracy. (Sections II.A – II.B)

I. Responses to Comments on PP (2013)

The main specification in PP regresses six proxies for local anti-Semitism in the 1920s and 1930s on an indicator for whether a town had a pogrom against its Jewish community during the Black Death in 1348-50 (*POG1349*).

I.A. Robustness of the PP Evidence

While our main results in PP are based on six outcome variables, GH focus on just one of these in their paper – the share of Nazi Party votes in the 1928 election. They argue that "*the result is driven by outliers, many of which are in the federal state of Bavaria*" (GH, p. 4). GH assert that we omitted the role of political and religious authorities, and that this introduces a bias because "*Bavaria had more pogroms because of its splintered political authority […and…] The Nazi Party won a higher vote there too, because Bavaria was where Hitler got his start. This coincidence created the Bavarian outliers that biased the coefficient for POG1349.*" (GH p. 9). We note in passing that GH directly contradict their own computation, which shows that Bavaria "*did not have more Black Death pogroms than the rest of Germany*" (GH p. 7, emphasis added, see also GH footnote 12 for the underlying calculation).

We discuss GH's argument regarding political and religious authorities in more detail below. Here, we first show that our empirical results are robust. We note that GH's assertion about Bavaria being 'different' can be accounted for by a dummy variable for the state, allowing for a higher average level of Nazi Party votes.[1] In fact, our original publication already accounted for the (more general) possibility of differences across states or regions by presenting regressions with province fixed effects – one of which is Bavaria (PP Appendix Table A.6). Nevertheless, in what follows, we specifically address GH's point on Bavaria. Table 1 replicates our main regressions from PP Table VI (for the six outcome variables) and from PP Table VII (for the principal component of these outcomes). In Panel A1 of Table 1, we add a dummy for Bavaria in all specifications; in Panel A2, we completely drop Bavaria, as GH do; in Panels B and C we replicate the propensity score estimation from PP. Note that in the

[1] GH, instead, choose to drop Bavaria from the dataset, thus also dropping the variation *within* Bavaria. As we discuss in more detail below, this would only be warranted if there were concrete concerns about unobserved political or religious institutions at the local (municipality) level within Bavaria that are – additionally – correlated over six centuries between 1349 and the 1920s. GH do not make this point, let alone provide evidence along these lines. Nevertheless, we checked the results for the 1928 Nazi Party vote in a specification that also includes an interaction term between *POG1349* and Bavaria, thus allowing the coefficient of interest to be different in this state. We find a coefficient on *POG1349* of .0086 with a p-value of .062, confirming that our 1928 voting result also holds outside of Bavaria, albeit with a somewhat smaller coefficient.

propensity score estimations, the dummy for Bavaria means that the matching algorithm will preferentially match towns within Bavaria, thereby accounting for a higher average level of Nazi Party votes in this state. The results in Table 1, Panel A1, are remarkably close to those in PP. In fact, the OLS results for the 1924 election (when the Nazi Party was banned and the right-wing DVFP ran on an anti-Semitic platform) are even stronger than in our original OLS regression (Panel A1, col 3). Similarly, our results that use the first principal component of the six outcome variables also become stronger (col 7). Adding a Bavaria dummy, we obtain a coefficient of .333 (p-value .003) – as compared to a coefficient of .290 with a p-value of .029 in our original result (PP Table VII, col 1). The Bavaria dummy itself is statistically significant, and positive for 4 out of our 6 outcome variables. For the remaining two, it is *negative* (and significant for deportations). Thus, there is no coherent pattern for Bavaria across the six outcome variables in our paper.

Panel A2 in Table 1 presents the most restrictive specification, dropping Bavaria altogether. For 4 of the 6 outcome variables, the coefficient on *POG1349* remains statistically significant. The same is true for the principal component (col 7). It is also noteworthy that for the result that GH focus on (NSDAP votes in 1928 in col 2), the p-value of .117 is actually close to statistical significance at the 10% level. In addition, the coefficient estimates for *POG1349* remain similar for the five other outcomes when we drop Bavaria. That is, GH picked the one outcome for which dropping Bavaria leads to a smaller and statistically (barely) insignificant result. Furthermore, we note that none of the coefficients for *POG1349* when excluding Bavaria (Table 1, Panel A2) are statistically significantly different (at the 10% level) from our baseline results (PP Table VI and Table VII, col 1). In other words, GH's argument is subject to the fallacy that the "difference between 'significant' and 'not significant' is not itself statistically significant" (Gelman and Stern, 2006). In fact, GH do not even pass this bar for the one outcome that they emphasize to make their point (NSDAP votes in 1928).

Finally, the matching estimates in Panels B and C fully confirm the results in PP for all outcome variables and for the principal component.

GH also present quantile regressions for the 1928 Nazi votes, arguing that "*POG1349 has little effect on the conditional median*" (GH p. 4). There is no clear ex ante reason to favor conditional medians over conditional means. In Figure 1 we present the full quantile regression plot (for the 10^{th} to the 90^{th} percentile of the Nazi vote, not only for the median as GH). We follow GH's argument about Bavaria being 'different' and include a dummy for this state, as well as the other control variables used in Table 1, col 2. We find that the coefficient on *POG1349* is relatively stable, and it is generally statistically significant for the first six deciles. That is, our result is not driven by outliers with particularly high Nazi Party votes; if anything,

the contrary is true – at least in the specification that follows GH's argument and allows for Bavaria to be different by including a dummy for the state.

Similarly, we re-estimated the regression in Table 1, col 2 using the robust regression estimator. Instead of adjusting the sample "by hand," dropping observations deemed too influential, as GH do, the robust regression estimator (Huber 1964) works by i) dropping observations with a Cook's distance greater than 1, and ii) giving more weight to observations where the difference between predicted and actual values is small. In a final step, the regression is re-estimated using these weights. The robust regression returns a coefficient of .005 with a p-value of .006 (i.e., statistically significant at the 1% level).[2] While the coefficient on *POG1349* is somewhat smaller than in our baseline results, its sign and significance are robust to a rigorous examination of the effect of outliers when using a more efficient procedure than the ones by GH (quantile regression only at the median or fully dropping Bavaria).

I.B. GH's Placebo Exercise

GH claim that our results fail "basic placebo tests." Our original paper, we already performed placebo tests, examining election results for the Communist Party and the DNVP (Table XI). We discuss in PP why we believe these parties to be a natural choice for our tests. The DNVP was also a far-right party, but it lacked the radical anti-Semitism of the Nazi Party and its associated DVFP. We documented that "[v]otes lost by the DNVP are similar to votes gained by the DVFP in these cities. Because the two parties' programs were similarly right-wing overall, these findings point to anti-Semitism, not extreme political attitudes as the driver of voting behavior in cities with Black Death pogroms." [PP p. 1384] We also examined the relationship between medieval Pogroms and votes for the Communist Party (KPD) to check whether our results could be driven by political extremism more broadly. We find no results.

In their Table 2, GH perform placebo checks for two other parties: the DDP and the DVP in 1924 and 1928. We note in passing that there is nothing particularly compelling about using the DDP or DVP, which were bourgeois parties supporting the Weimar state, as a placebo – as the number of tests performed grows, a false positive becomes more likely. GH present a variety of specifications, using different subsamples of our data. The resulting coefficients on medieval pogroms range from negative to significantly positive, with the latter more common in specifications that use subsamples of our data (not in our original paper). In what follows,

---

[2] Weights are inversely related to absolute deviations of predicted from actual values. We implement this regression using the Stata command *rreg*. Observations in Bavaria receive a weight of 0.5 on average, and towns in the rest of the sample, an average weight of 0.877.

we show that GH's placebo criticism disappears when we use the specifications and samples that are consistent with our paper.

In Table 2, we implement the placebo tests for DDP and DVP in a way that is consistent with our original specifications (PP Table VI): OLS, propensity score matching, and matching by geographical proximity. We include all original controls from PP and add a dummy for Bavaria (as GH argue is important). For purposes of comparison, columns 1 and 2 in Table 2 show our main results for the DVFP and NSDAP votes in 1924 and 1928, respectively. For the Nazi party and its institutional stand-in, the coefficients on medieval pogroms are positive and statistically significant at the 1% or 5% levels throughout. Columns 3-6 present results for DVP and DDP. Across these specifications, we find no evidence of the strong positive relationship claimed by GH. Results for the DVP are always insignificant and sometimes even negative. For the DDP, the 1928 results are equally insignificant. Only for the DDP in 1924 do we find two positive, and marginally significant coefficients. In contrast to GH's weak and inconsistent placebo results for the DDP, and DVP, the results for the anti-Semitic parties in columns 1 and 2 show 6 out of 6 coefficients that are large, positive, and significant.

I.C. Shortcomings in GH's Argument about Historical Confounders

GH employ the standard econometric expression for the OLS estimator to illustrate potential biases in our estimation (p. 8). A bias would result if historical unobservables were correlated with both pogroms in 1349 and Nazi Party votes in 1928. GH elaborate on what they believe is a candidate for such a bias: political and religious authorities. They present a lengthy discussion that Bavaria had a divided authority after 1347, which made it difficult to protect Jews against pogroms. They also point out that Bavaria was a fertile ground for the Nazi Party and then allege that the covariances that would lead to biased estimates "*are unlikely to be zero, however, because PP omits a role for political and religious authorities*." (GH p. 9). We make several observations: 1) To the extent that Bavaria was systematically more anti-Semitic in both periods, this would be captured by our province fixed effects in the original paper, as well as by the Bavaria dummy that we added above.[3] 2) To bias our results, there would need to be local differences in political authorities that a) correlate with pogroms in 1349 and b) also correlate with Nazi Party votes in 1928. While GH fail to even discuss this precise condition for their assertion of a bias, we did acknowledge this possibility in PP Section IV.A. and provided some (admittedly imperfect) empirical checks to address this concern. 3) At the core of GH's argument is the presumption that more "splintered" political authority in Bavaria led

[3] However, we refer again to GH's own computation – based on our data – showing that Bavaria did not have more pogroms in 1349. See their footnote 12.

to more pogroms in 1349. However, GH's go-to example of Strasbourg actually undermines their own argument, showing that political authority can go either way: "*Strasbourg's political and religious authorities organized a pogrom in 1349; worked to stop one eleven years earlier; and then blocked anti-Semitic outrages in the 1920s and 1930s.*" (GH p. 7). For our argument (and econometrics) to be valid, we do not require that in each location, anti-Semitism was always at the highest level; we only require that places that murdered or attacked their Jewish communities in 1349/50 were more anti-Semitic, on average, than places that were not.[4] 4) GH's description of Strasbourg also illustrates another shortcoming of their argument: local institutions were highly volatile, even over relatively short horizons. Since the same political and religious institutions were not in place over 600 years, this cannot be driving our results. In addition, we note that anti-Semitism among *individuals* serving in local institutions is compatible with our argument. In fact, in PP we mention that one explanation for our results on deportations or the Night of Broken Glass is that local officials in the 1930s may have been more anti-Semitic in localities with Black Death pogroms.

I.D. Further Examples of GH's Misunderstandings

GH examine our results for the 1924 election, claiming that "*There are also deeper problems with the pogrom indicator. Placebo results (Appendix Section A.5) demonstrate that the POG1349 proxy is often positively correlated with votes for parties that were not anti-Semitic, a finding that casts doubt on the authors' interpretation of the proxy.*" (GH p. 5). We highlight this example because it illustrates how GH distort our results and apply double-standards.

In the 1924 election, we used a "political experiment to distinguish anti-Semitic from right-wing votes" (PP p. 1384). The right-wing party DNVP had expelled anti-Semites from its ranks in 1922, leading to a split of the party, with a second right-wing party emerging (the DVFP) that pursued a markedly more radical anti-Semitic agenda. In the 1924 election, the DVFP gained more seats in localities with medieval pogroms. We reasoned that "[i]f this is a reflection of anti-Semitism – and not more right-wing attitudes generally – then we should expect the closest (but less anti-Semitic) competitor DNVP to register fewer votes in towns and cities with an anti-Semitic past." (PP p. 1384). This is exactly what we documented in the data. GH assert that "*these results do not really address the other part of the claim in PP, which is that anti-Semitic voters preferred the DVFP or DNVP to other parties.*" (GH Appendix p. 17). We did not make such a claim. We merely proposed to compare the two parties, to distinguish far-right

[4] GH mention that "*Despite the deep roots of anti-Semitism in Strasbourg, in the 1920s and 1930s the city did not witness any of the anti-Semitic violence seen in other hotbeds of cultural hostility to Jews.*" This is disingenuous; Strasbourg was under French control in the 1920s, as GH admit in the same paragraph. Since we are not interested in cross-country differences in anti-Semitism, we did not include it in our sample.

voting and anti-Semitic attitudes. Nevertheless, GH compute the combined vote of the two parties: "*Rows (13) – (16) combine the votes for the two right-wing parties. If the pogrom dummy does pick up persistent anti-Semitism, then its coefficient in these regressions should be positive and significant.*" (GH App. p. 17). The results (GH Table A5.1) show that the coefficient is small, negative, and statistically insignificant in 3 out of 4 specifications. It is marginally significant (at the 10% level) in one specification that considers Prussia only, throwing out more than half of our sample. However, this is 'strong' enough for GH to build a case around it: "*This result undermines PP's core contention and demonstrates the importance of considering a wide range of electoral outcomes. The supposed indicator of persistent anti-Semitism reduces [emphasized also in GH] the vote share for anti-Semitic parties. This important outcome suggests that PP's results for the DNVP alone in 1924, and for the Nazis later, is weak evidence that areas with medieval pogroms were more likely to support anti-Semitic parties.*" We make two observations: 1) In fact, the non-result (which is how we read the small, insignificant coefficients) is in line with our original argument. As we state on PP p. 1384: "Votes lost by the DNVP are similar to votes gained by the DVFP in these cities [i.e., adding both parties should yield a zero result]. Because the two parties' programs were similarly right-wing overall, these findings point to anti-Semitism, not extreme political attitudes as the driver of voting behavior in cities with Black Death pogroms." 2) GH's argument entirely depends on one regression result that is only weakly marginally significant in one subsample, and insignificant in all other specifications – which is arguably not the kind of consistent statistical pattern that they themselves argue is essential to draw substantive conclusions.

Along similar lines, GH examine votes for the Communist Party, which we found to be unrelated to *POG1349*. GH find the same in 3 out of 4 specifications (Table A5.1, cols 17-20). In one subsample (when excluding Bavaria), the coefficient is marginally significant at the 10% level. Again, this is sufficient for GH to "*call into question the idea that POG1349 picks up enduring anti-Semitism reflected in Weimar elections.*" (GH App. p. 18).

I.E. Other Points by GH on PP

On page 11, GH discuss "*three troublesome patterns in the PP data.*" First, "*that many towns with a Black Death pogrom were close to places that escaped an attack on the Jews in 1348-50: sometimes only a few miles away.*" Rather than "troublesome," we believe this is an opportunity to further test our argument. GH only make vague claims why culture should not vary over short distances; as a matter of fact, the rich empirical literature on culture has unearthed many traits that vary over short distances or within small geographic units such as districts (Eugster et al. 2017; Suedekum 2018; Alesina, Giuliano, and Nunn 2011; Fernandez and Fogli 2009; Michalopoulos and Papaioannou 2013, Guiso, Sapienza and Zingales 2016,

Giuliano and Nunn 2021). Many confounding factors are arguably constant over short distances: economic incentives, geographical position and weather are all near-identical. If, nonetheless, two towns close to each other show significant differences in one dimension in 1349 (anti-Semitic attacks), and then display consistent, similar differences in six outcome measures 600 years later, the empirical basis becomes stronger. In our original publication, we implemented a matching exercise where we systematically conduct such comparisons, showing large, significant effects (see also Table 1, Panel C below). While hard evidence is not abundant, we reviewed the available information on historical migration patterns in PP; the existing historical studies suggest that the rate of *in-migration* into *small towns* (the vast majority of our sample) was probably very low over centuries, which would offer a ready explanation for why we find persistence of local differences even over short distances (PP, pg. 1381-82).

The second and third points by GH are simply assertions about what they *believe* should be true – why did Jews come back if places were anti-Semitic? And why did Catholic areas switch from being less anti-Semitic before 1933 to more hatred thereafter? It is hard to see the connection of these claims with our argument in PP. Many reasons drive location decisions by minorities, including economic opportunities; we never argued that places with a deep history of anti-Semitism had to be so toxic that no Jew ever settled there again. Similarly, GH's argument about Catholics is a straw man – in arguing that some of the geography of German anti-Semitism in the interwar period reflects deep historical roots, we are not required to claim that no other factor ever played a role.

In concluding their discussion about Bavaria, GH assert: "*The lesson is that just because a cause is in the past does not mean it is exogenous, particularly when political factors can affect observations across time and regions.*" (p. 9) While we agree with this statement per se, we are disappointed by GH's misrepresentation of our argument. We did <u>not</u> make any claim that medieval pogroms were 'exogenous.' To the contrary: In section IV.A., we asked: "Do medieval economic or political correlates of Jewish settlement and pogroms directly predict twentieth-century anti-Semitism? If so, then medieval pogroms might simply be proxying for geographical, economic, or political factors that have remained unchanged." (PP p. 1376). This is followed by an empirical analysis that addresses this concern. In sum, GH's critique often exaggerates and distorts the argument in PP, while ignoring the evidence that we presented to address the respective concerns.

II. Responses to Comments on BF

While GH claim that they take issue with our *main* findings in BF, their only empirical critique relates to a *sub-result* of an *extension* in our paper. Our main argument is that denser networks of associations and clubs at the city level are positively correlated with more entry into the Nazi Party. None of the claims in GH questions this conclusion. In addition, their various "conceptual points" and reflections on coding choices are mere speculation, as they do not show any data demonstrating potential biases.[5]

II.A. GH's Points on the Subsample that Examines the Role of Political Stability

The main argument in GH about BF relates to an extension of an auxiliary analysis, where we examine the question – was social capital always a conduit for Nazi recruitment, or did its effect depend on political context? Here, we mainly focused on the comparison between Prussia – known as a "bulwark of democracy" – and the rest of the country. This comparison is also not in doubt. We then go one step further and examine patterns in the non-Prussian part of the data – i.e., in a subsample that drops about half the observations. This is what Table 7 in our original paper seeks to address, splitting the remaining federal states into stable and unstable ones, depending on their record of government stability and parties in power. Here, GH take issues with our coding, arguing that we misclassify Bavaria.

In what follows, we show that a) our results are robust, independent of the coding chosen and b) GH wrongly allege that we misclassified Bavaria (and Prussia). Before we turn to these points, it is worth noting that all their analyses show that the coefficient of interest (associations and party entry) remains positive (albeit not always significant, which is not surprising given the small sample sizes).

GH argue that the Bavarian People's Party (BVP) should be treated like the Zentrum, a Catholic, middle-of-the road party in the rest of Germany. They write: "*The BVP split from the*

---

[5] For example, GH speculate about selection bias of towns into our sample: "*Better managed cities or towns that appealed to directory publishers might, for example, be more likely to land in the BF sample. So would communities that have long funded public goods such as local libraries and archives. Those characteristics could in turn be correlated with local political conditions and with social capital, both today and in the 1920s.*" (GH p. 17). None of these speculations is backed by citations or data. We also point the reader to Appendix B.1 in BF, where we discuss the sample construction and potential selection bias. In particular, Figure A.4 in the BF appendix shows that our main result is remarkably similar in two subsamples that were obtained in different data collection steps: the first one for cities whose archives are listed on nationwide directories and the second in the remaining towns that we had to contact individually. The first group reflects what GH would call "better managed cities." Since our results are identical for the two subsamples, it is unlikely that sample selection issues in the speculative spirit by GH are responsible for our result.

Zentrum just after World War I, yet it functioned much as a sister party to the Zentrum, almost like the Christian Social Union and the Christian Democrats today." (GH p. 13). On this basis, GH argue that Bavaria was more stable than, or as stable as, the rest of the Weimar Republic.

Such a claim stretches historical credulity. While both the BVP and the Zentrum were Catholic parties, their similarity largely ends there. There is no doubt that Bavaria was a hotbed of anti-democratic, anti-Semitic agitation in the period 1919-23. Bavaria is the federal state where leading politicians collaborated with Hitler in his first, violent bid for power in 1923 – the Beerhall Putsch. The BVP itself entered government in 1919 in "putsch-like" conditions (Schönhoven 1977). The analogy with CDU and CSU in post-1945 Germany could not be more wrong – the BVP continuously sought to overturn the established democratic constitution, even after its own, disastrous role in the Hitler Putsch: Benz (2010) describes how Gustav von Kahr (not a party member himself) was pushed by the BVP into the position of Prime Minister of Bavaria. In 1923, von Kahr planned for his own violent overthrow of democracy (pre-empted by Hitler), and told high-ranking officers that it was time for a showdown between the "international Marxist-Jewish [worldview] and the national Germanic one" (Deuerlein 1962).

Even after the putsch, according to the leading historian on the history of the BVP, politicians of the BVP "continued to be more committed to conservative-authoritarian traditions than to parliamentary democracy. The republican form of government was viewed by the BVP as a temporary solution…" (Schönhoven 1977, p. 341). When General von Hindenburg stood successfully for the presidency in 1925, he was supported by the NSDAP, the DNVP, the DVFP, the DVP, the Agricultural League – and the BVP, which thus sided against the parties of the "Weimar coalition," and in favor of a reactionary monarchist candidate.

To suggest that such a party's continued role in government was a source of stability, and that Bavaria – despite its tumultuous history – should therefore be classified as having an above-average political stability score, is not merely doubtful; it is simply wrong – and wrong on a scale of historical misjudgement that can only surprise from scholars who emphasize that getting history right is essential.

II.B. Further GH Comments on the Coding of Government Stability

In addition, GH allege: "*Instead of using this index in regressions, BF creates an indicator variable for more and less stable states. […] Creating the indicator variable throws away information, yet they do not defend this procedure.*" (GH p. 14). We do, in fact, use the continuous stability measure in BF Appendix D.3 (see in particular Figure A.7).

GH allege a coding error for Prussia, claiming that "*BF also erroneously assigns Prussia a value of .73 for this third component, even though every single Prussian government was*

headed by a Weimar coalition party." (GH p. 13). The coalition in 1921-25 included the DVP, which is not a Weimar coalition party.[6] Therefore, the value for Prussia is correct.

In their Table 2, GH argue: "*Column (2) corrects a subtle error: while the index is defined at the state level, BF computes the index using all of the city observations.*" (GH p. 14). We made this choice deliberately so that the principal-component-based index reflects the different frequencies of observations for different states in our sample. The GH procedure gives the same weight to Prussia (with 119 observations) as to Anhalt, Braunschweig, or Hamburg (each with 1 observation). The GH index is very similar to ours – the correlation coefficient is .9998, and Spearman's rho is .9956. Nevertheless, we checked our results for the GH state-level computation of the stability index (GH do not report these results in their Table 3; they only report in Table 2 that the index itself has different entries). We confirm our results in Table 3: For unstable states, we obtain a large and statistically significant coefficient (with a p-value of .025), and for stable states (with above-median stability index), a small and statistically insignificant coefficient. These two coefficients are also significantly different from each other.[7]

III. Conclusions

Replication is a key dimension of the scientific method; close scrutiny of published results is essential to scholarly integrity. It can advance our understanding by challenging accepted empirical findings, thereby stimulating new research. However, GH is not a contribution in this spirit. Instead, their comment is a broader attack on a style of research that combines granular

---

[6] Ditto for Bavaria, where every single coalition included also USPD, BBB, or DNVP – so there was never a coalition only made up of Weimar coalition parties (or a subset thereof). In our paper, we should have spelled out the coding more clearly and comprehensively. We say "(iii) a state was governed by at least one party from the Weimar coalition." We should have added "…and by no other party that was not a part of the Weimar coalition." Since GH did not contact us prior to circulating their critique, we did not have a chance to clarify this feature of our coding.

[7] In the same context, GH also state that "*Their description of the binary indicator is also inaccurate. The replication code shows the authors include the median values as part of the "unstable" group.*" Column 4 in Table 7 of BF includes a footnote that explains how we define the "stable" group: "$I_{Stable\ Govt}$ is a dummy variable for Weimar states with above-median government stability." Thus, the remainder (unstable) group includes the median itself. We emphasize that we compute this division using the standard Stata code to split our observations into two quantiles: xtile …, nq(2). That is, the choice to put the median itself into the "unstable" group is made by the Stata package, not by us (as GH seem to allege). It is also worth noting that the median – for both the city-level and the state-level index – is Bavaria. Our discussion above highlights that classifying Bavaria as "unstable" is coherent with the historical record.

data from the past to shed light on questions relevant to modern-day economics and politics. GH criticize our work as an "example of this genre" (GH p. 2), choosing two papers that are not closely connected in terms of topic – one is about persistence of cultural traits (PP), the other on the relationship between social capital and political extremism (BF).

In their comments on PP, GH narrowly focus on only one of the six indicators of anti-Semitic attitudes that we examined. While they allege that our statistical results are fragile for one of them, and driven by data from Bavaria, we demonstrate that a) our results are robust, and b) that Bavaria being different does not drive our results. Contrary to the placebo results shown by GH, we find no evidence of consistent false positives in 1924/28 electoral data. Where GH question the logic of our analysis, as in the case of the DNVP and DVFP, they misrepresent our argument and merely repeat a result that is in PP *because it strengthens our case.*

GH's critique of BF is mostly 'conceptual;' they make several suggestions of what social capital is and might be, and how this could relate to what we find. We believe there is free entry in the market for ideas, and the authors should feel free to publish a compelling contribution on mechanisms based on new data, if there is one to be written. GH levy no objection against our main finding, that areas with more social capital saw more rapid entry into the Nazi Party. Their empirical critique of our results focuses on a sub-analysis of an extension in the paper, where we examined interactions between social capital and political instability outside of Prussia. We have shown that this criticism is misguided and that our results hold when coding choices reflect the historical record. GH's critique of our alleged miscoding of Bavaria shows a remarkably poor understanding of historical context and the politics of the relevant parties. Far from overstating the instability of Bavaria by not coding the local Catholic party, the BVP, as equivalent to the middle-of-the-road Zentrum party, it is clear that Bavaria was highly unstable, and a hotbed of far-right activism. The party in question, the BVP, was unambiguously opposed to the democratic Weimar constitution and sought to overthrow the established order; it should not be counted as an innocent pillar of the republic, as GH argue.

In combination, GH's critiques of our papers reveal neither empirical fragility nor conceptual shortcomings; if anything, they have given us the opportunity to demonstrate both the strength and robustness of our results, as well as the conceptual reasoning behind our analysis.

References

Alesina, Alberto, Paola Giuliano, and Nathan Nunn. 2011. 'Fertility and the Plough'. *American Economic Review* 101 (3): 499–503.

Benz, Wolfgang. 2010. *Politik in Bayern 1919-1933: Berichte Des Württembergischen Gesandten Carl Moser von Filseck*. Vol. 22. Walter de Gruyter.

Deuerlein, Ernst. 1962. *Der Hitler-Putsch: Bayerische Dokumente Zum 8./9. November 1923*. Vol. 9. Deutsche Verlags-Anstalt.

Eugster, Beatrix, Rafael Lalive, Andreas Steinhauer, and Josef Zweimüller. 2017. 'Culture, Work Attitudes, and Job Search: Evidence from the Swiss Language Border'. *Journal of the European Economic Association* 15 (5): 1056–1100.

Fernandez, Raquel, and Alessandra Fogli. 2009. 'Culture: An Empirical Investigation of Beliefs, Work, and Fertility'. *American Economic Journal: Macroeconomics* 1 (1): 146–77.

Gelman, Andrew and Hal Stern. 2006. The Difference Between "Significant" and "Not Significant" is not Itself Statistically Significant'. *The American Statistician* 60(4).

Giuliano, Paola and Nathan Nunn. 2021. 'Understanding Cultural Persistence and Change'. *Review of Economic Studies* 88 (4): 1541-1581

Guiso, Luigi, Paola Sapienza and Luigi Zingales. 2016. 'Long-Term Persistence'. *Journal of the European Economic Association*, 14(6): 1401–1436.

Huber, Peter J. 1964. 'Robust Estimation of a Location Parameter'. *Annals Mathematics Statistics* 35.

Michalopoulos, Stelios, and Elias Papaioannou. 2013. 'Pre-colonial Ethnic Institutions and Contemporary African Development'. *Econometrica* 81 (1): 113–52.

Satyanath, Shanker, Nico Voigtländer and Hans-Joachim Voth. 2017. 'Bowling for Fascism: Social Capital and the Rise of the Nazi Party'. *Journal of Political Economy* 125(2): 478-526.

Schönhoven, Klaus. 1977. 'Zwischen Anpassung Und Ausschaltung'. *Historische Zeitschrift* 224 (1): 340–78.

Suedekum, Jens. 2018. 'Economic Effects of Differences in Dialect'. *IZA World of Labor*, Article #414.

Voigtländer, Nico and Hans-Joachim Voth. 2012. 'Persecution Perpetuated: The Medieval Origins of Anti-Semitic Violence in Nazi Germany'. *Quarterly Journal of Economics* 127(3): 1339-1392.

FIGURES

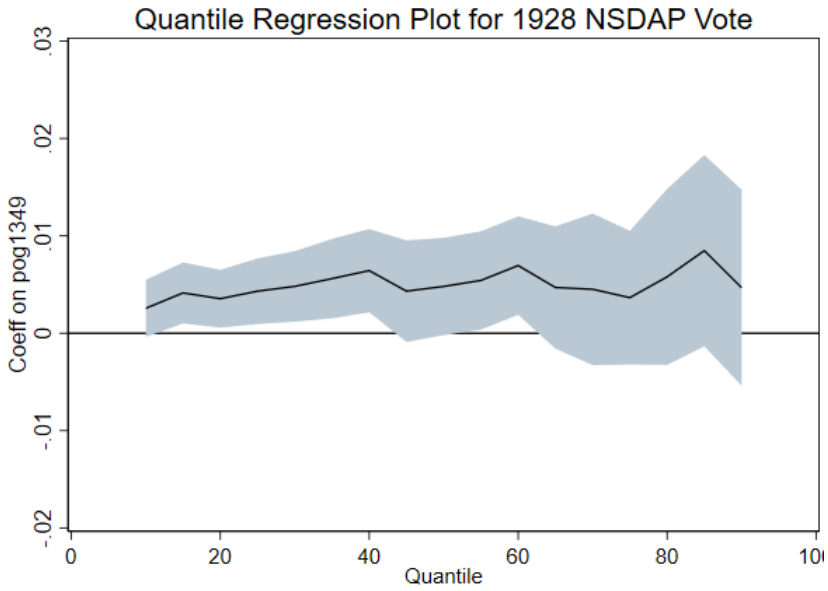

Figure I

Quantile Plot for Regression of NSDAP vote in 1928 on POG1349

The figure shows the quantile regression plot corresponding to the specification in Table 1, column 2. The solid line corresponds to the coefficient estimate of *POG1349* at the various quantiles, and the shaded area represents the 90% confidence interval.

Table 1

Replicating PP Main Results – Controlling for a Bavaria

Dep. variable:	(1) 1920s pogroms	(2) NSDAP 1928	(3) DVFP 1924	(4) Depor-tations	(5) *Stürmer* letters	(6) Synagogue attacks	(7) Principal Comp[a]
Panel A1: Baseline Regressions, Dummy for Bavaria							
	OLS	OLS	OLS	ML	ML	OLS	OLS
POG^{1349}	.0642***	.0164***	.0196**	.154**	.316**	.121**	0.333***
	(.0235)	(.00518)	(.00901)	(.0705)	(.138)	(.0529)	(0.110)
Bavaria	.0923**	.0701***	.154***	-.189***	.574***	-.0621	1.868***
	(.0418)	(.00834)	(.0174)	(.0586)	(.163)	(.0524)	(0.200)
Baseline Controls	✓	✓	✓	✓	✓	✓	✓
Observations	320	325	325	278	325	278	311
Adjusted R^2	.075	.366	.483			.101	0.495
Panel A2: Baseline Regressions, Dropping Bavaria							
POG^{1349}	.0461	.00685	.0126*	.162**	.373***	.135**	0.150*
	(.0291)	(.00435)	(.00686)	(.0771)	(.136)	(.0592)	(0.0799)
p-value	*[.115]*	*[.117]*	*[.068]*	*[.035]*	*[.006]*	*[.024]*	*[.061]*
Baseline Controls	✓	✓	✓	✓	✓	✓	✓
Observations	253	257	257	223	257	224	247
Adjusted R^2	.021	.057	.345			.089	0.228
Panel B: Matching Estimation (adding Bavaria as matching variable)[b]							
POG^{1349}	.0754***	.0129***	.0198**	164.6***	2.544***	.127**	0.264**
	(.0165)	(.00466)	(.00770)	(42.75)	(.587)	(.0612)	(0.127)
Observations	320	325	325	278	325	278	311
Panel C: Geographic Matching (adding Bavaria as matching variable)[c]							
POG^{1349}	.0819***	.0110**	.0250***	203.6***	2.843***	.168***	0.318***
	(.0168)	(.00430)	(.00608)	(32.07)	(.572)	(.0646)	(0.0819)
Median distance	21.6	21.5	21.5	23.8	23.8	25.7	20.5
Mean distance	25.3	25.0	25.0	30.4	34.1	29.4	24.1
Observations	320	325	325	278	325	278	311

Notes. The table replicates the main regressions from PP Table VI (for the six outcome variables – cols 1-6) and from PP Table VII (for the principal component of these outcomes – col 7). Panel A1, adds a dummy for Bavaria to all regressions, and Panel A2 drops Bavaria altogether. Controls are the same as in VV Table VI. All regressions run at the city level. Standard errors in parentheses, clustered at the county (*Kreis*) level. POG^{1349} takes the value 1 if a pogrom occurred in the years 1348-50, and 0 otherwise. City population is taken from the 1925 census in column 1 and from the election data for the respective year in columns 2 and 3; in columns 4–6, city population is from the 1933 census. %Jews is from the 1925 census for columns 1–3, and from 1933 census in columns 4–6. %Protestants is from the 1925 census. OLS = Ordinary least-squares estimation; ML = Poisson maximum likelihood estimation.

[a] First principal component (standardized) obtained from six proxies for twentieth-century anti-Semitism, as described in the notes to PP Table VII.

[b] Matching estimation based on the same set of control variables as used in Panel A. Treatment variable is POG^{1349}. Replicating PP Table VI, Panel B.

[c] Matching estimation based on geography; the matching characteristics are longitude and latitude. Replicating PP Table VI, Panel C.

* $p < .10$, ** $p < .05$, *** $p < .01$.

Table 2

Revisiting GH's Placebo Tests

	(1)	(2)	(3)	(4)	(5)	(6)
	Anti-Semitic Parties		**"Placebo" Parties Suggested by GH**			
Dep. variable:	DVFP 1924	NSDAP 1928	DVP 1924	DVP 1928	DDP 1924	DDP 1928[a]
Panel A: OLS Regressions						
POG^{1349}	.0198**	.0164***	.00670	.0116	.0103*	.00283
	(.00897)	(.00518)	(.00692)	(.00775)	(.00544)	(.00493)
Controls	✓	✓	✓	✓	✓	✓
Observations	325	325	325	325	325	325
Adjusted R^2	.483	.366	.486	.339	.269	.291
Panel B: Matching Estimation[b]						
POG^{1349}	.0219***	.0129***	.00346	.00416	.0110*	.00626
	(.00788)	(.00466)	(.00762)	(.00804)	(.00652)	(.00529)
Observations	325	325	325	325	325	325
Panel C: Geographic Matching[c]						
POG^{1349}	.0250***	.0110**	-.000342	.000706	.00850	.00391
	(.00608)	(.00430)	(.00829)	(.00938)	(.00741)	(.00712)
Observations	325	325	325	325	325	325

Notes: The table revisits GH's 'placebo checks' of PP's voting results in 1924 and 1928. Columns 1 and 2 table replicate the main regressions from PP Table VI (for DVFP votes in 1924 and NSDAP votes in 1928). Columns 3-4 consider votes for the DVP and DDP in 1924/28 (as in GH, Table 2). Controls are the same as in VV Table VI, adding a dummy for Bavaria, in line with GH's argument (see also Table 1 above). All regressions are run at the city level, for cities with documented Jewish communicates in 1348. Standard errors in parentheses, clustered at the county (*Kreis*) level. POG^{1349} takes the value 1 if a pogrom occurred in the years 1348-50, and 0 otherwise.

[a] For 1928, the DDP data are reported jointly with votes for the (marginal) Volksrechtpartei (VRP).

[b] Matching estimation based on the same set of control variables as used in Panel A. Treatment variable is POG^{1349}. Replicating PP Table VI, Panel B (adding Bavaria as a control).

[c] Matching estimation based on geography; the matching characteristics are longitude and latitude. Replicating PP Table VI, Panel C (adding Bavaria as a control).

* $p < .10$, ** $p < .05$, *** $p < .01$.

Table 3

Replicating BF for non-Prussia sample (BF Table 7, cols 3-4)

	(1)	(2)	(3)	(4)
Stability measure	City-Level		State-Level	
Gov't stability	Unstable	Stable	State	State
$ASSOC_{all}$	.349***	-.0116	.278**	.000914
	(.128)	(.0619)	(.121)	(.0594)
Beta-Coefficient	*.440*	*-.023*	*.357*	*.002*
Tests that beta coefficient are equal	col. 1 = col. 2: *p*-value: .017		col. 3 = col. 4: *p*-value: .075	
Baseline Controls	✓	✓	✓	✓
N	58	48	75	31

Notes. The table replicates BF Table 7 for the non-Prussia subsample. Col. 1+2 report the original results in Table 7 (there: col 3+4). Columns 3+4 here use the state-level data to the principal-component-based state stability measure and the corresponding below/above median assignment into politically unstable/stable states. Standardized beta coefficients; Standard errors in parentheses, * p < .10, ** p < .05, *** p < .01.
Dependent variable is the average (standardized) rate of Nazi Party entry (per 1,000 inhabitants) in each city over the period 1925–July 1932 (when the Prussian government was replaced by a *coup d'etat*). Robust standard errors are in parentheses. Standardized beta coefficients (in brackets) report by how many standard deviations the outcome variable changes as a result of a 1 SD increase in the explanatory variable. $ASSOC_{all}$ is the number of associations per 1,000 city inhabitants. Baseline controls are listed in BF Table 2.