Online Appendix to "Reassessing the Link between Revolutionary Threats and Democratization"

A OLS and IV/2SLS Weights

Let riot intensity take on values in the finite set $r_i \in \{1, 2, 3, ..., \bar{r}\}$, and suppose that the true model has the form:

$$Whig\ Share\ 1831_i = \alpha_0 + \sum_{j=1}^{\bar{r}} \alpha_1^j D_{ij} + X_i'\theta + \upsilon_i, \tag{A1}$$

where $D_{ij} = 1[r_i \geq j]$ is a dummy variable equal to 1 if $r_i \geq j$ and 0 otherwise, and υ_i is a mean-zero error term. The α_1^j coefficient represents the marginal effect of a one unit increase from rioting level $j-1$ to j. The unrestricted model (A1) nests the linear model (1) in the main body of the paper, which restricts the riot-specific effects to be independent of riot intensity: $\alpha_1^j = \alpha_1$ for all $j > 0$.

Consider a multi-valued instrument, Z_i, and suppose that the standard IV assumptions hold. Mogstad and Wiswall (2010) show that the linear IV estimand for α_1 in (1), can be decomposed as:

$$\alpha_1^{2SLS} = \sum_{j=1}^{\bar{r}} w_j^{IV} \alpha_1^j,$$

where

$$w_j^{IV} = \frac{Cov(D_{ij}, Z_i)}{Cov(r_i, Z_i)}.$$

These two expressions have four important implications. First, estimating the misspecified linear model in equation (1) in the main body of the paper using IV would yield a consistent estimate of a weighted average of the marginal effects across the riot intensity distribution. Second, each weight $\hat{\omega}_j^{IV}$ attached to α_1^j will depend on the proportion of units that, because of the instrument, experience a change in treatment intensity. Third, α_1^{2SLS} will assign more weight to the marginal effects for the treatment levels of that are most affected by the instrument. Lastly, the weights can be computed using the sample analog of w_j^{IV}.[1]

As Lochner and Moretti (2015) note, OLS is a special case of IV estimation. Therefore, in the absence of endogeneity, the OLS estimator for model (1) in the main body of the paper also converges to a weighted average of the treatment-specific effects, α_1^j, where the weights are nonnegative and sum to 1. In this case, the OLS weight, w_j^{OLS} is defined as

$$w_j^{OLS} = \frac{Cov(D_{ij}, r_i)}{Var(r_i)}.$$

An examination of the last expression makes clear that the weights implied by OLS

[1] When $\sum_{j=1}^{\bar{r}} D_{ij} = r_i$, the w_j^{IV} will sum to one over $j \in \{1, ..., \bar{r}\}$, They will also be nonnegative as long as monotonicity in the effects of the instrument on r_i holds (Løken et al. 2012; Lochner and Moretti 2015).

estimation will not be necessarily equal to the weights implied by IV estimation. In fact, inappropriately assuming that all per unit effects are the same as in equation (1) in the main body of the paper, when per unit treatment effects vary across treatment levels as in equation (A1) will generally yield different OLS and IV/2SLS estimates even in the absence of endogeneity (Loken, Mogstad, and Wiswall 2012; Lochner and Moretti 2015). Therefore, the IV–OLS coefficient gap in AF could be an artifact of how the IV and OLS coefficients place different weights on different treatment margins.

In the case of AF's study, the weight for each rioting level j can be computed as the coefficient estimate on $Riots\ within\ 10\ km_i$ from a 2SLS regression of D_{ij} on the endogenous regressor, using $Distance\ to\ Sevenoaks_i$ as an instrument for the number of riots within a radius of 10 km from a constituency. In this case, a constituency's riot intensity falls as the distance to Sevenoaks increases. Therefore, the weight w_j^{IV} attached to α_1^j will be a function of the proportion of constituencies that, because of the distance to Sevenoaks, experience a change in rioting from more than j to j or less. The weights implied by OLS estimation, $\hat{\omega}_j^{OLS}$, in turn, correspond to the coefficient estimate on the endogenous regressor in an OLS regression of D_{ij} on $Riots\ within\ 10\ km_i$.

B Heterogeneous Treatment Effects

Table B1 presents 2SLS estimates of the impact of local Swing riots on the outcome of the 1831 election, both with and without controls, for three different sub-samples, based on agricultural production.

Table B1: Local Swing Riots and Reform Support by Agricultural Area

	Cereal	*Cereal*	*Dairy*	*Dairy*	*Farming*	*Farming*
			Second Stage			
Riots within 10 km	2.842	-0.224	10.812***	6.122***	2.034	0.205
(instrumented)	(1.979)	(1.080)	(3.031)	(2.366)	(1.425)	(6.664)
			First Stage			
Distance to Sevenoaks	-1.174**	-1.453**	-0.649***	-0.699***	-0.825**	-0.299
	(0.596)	(0.675)	(0.123)	(0.149)	(0.344)	(0.462)
Baseline controls	No	Yes	No	Yes	No	Yes
Observations	117	117	72	72	55	55

2SLS robust std. errors; *, **, and *** indicate statistical significance at the 10%, 5%, and 1% level, respectively. Following AF, the set of control variables includes: *Whig share 1826, Whig share 1826 Squared, Reform support in 1830, County constituency, Narrow franchise, Patronage index, Emp. fract. index, Agriculture (emp. share), Trade (emp. share), Professionals (emp. share), Population, Population density , Thriving economy, and Declining economy.*

Table B2 reports probit results (marginal effects evaluated at the mean) associating local Swing riots with the likelihood of a Whig being elected to a seat in 1831 by type of electoral race. Model (1) replicates the results in AF's Table II, Panel B, Column 2. Model (2) restricts the analysis to the 362 races where an MP serving in the 1830-31 parliament chose to run again for the same constituency in the 1831 election. Lastly, Model (3) examines the remaining 127 races where an open seat was available.

Table B2: Whig Elected 1831 by Electoral Race

	(1)	(2)	(3)
	All Races	*Incumbent Candidate*	*Open Seat*
Riots within 10 km	0.0056**	0.0032	0.0181***
	(0.0028)	(0.0027)	(0.0057)
Obs. (seats)	489	362	127
Mean Whig Elected	0.564	0.539	0.638
Mean Riots within 10 km	9.393	9.599	8.803
Bottom Quartile Riots	1	1	1
Top Quartile Riots	12	12	11

The standard errors (in parentheses) are clustered at the constituency level.; *, **, and *** indicate statistical significance at the 10%, 5%, and 1% level, respectively. The control variables include *Whig share 1826, Whig share 1826 Squared, and Reform support in 1830.*

C Unobserved Heterogeneity

Understanding the mechanisms that drive self-selection into treatment becomes crucial when interventions do not require mandatory treatment (Mogstad and Torgovitsky 2018). Bisin and Moro (2021) highlight the difficulty of representing the mechanisms driving selection into treatment directly, especially when dealing with group units of observation like countries, cities, or constituencies. However, they suggest that even a reduced-form model, devoid of explicit equilibrium micro-foundations, might aid in interpreting parameters defined in terms of choice behavior under actual or counterfactual manipulations of the instrument.

In the context of AF's study, the diverse constituencies in England may have faced unique costs or benefits associated with rioting. Based on the non-linear relationship between riots and reform support uncovered above, consider a generalized Roy model postulating treatment to be determined by a simple cutoff condition: each constituency $i = 1, 2, ...N$ selects into treatment if its benefit is greater than the cost:

$$T_i = \begin{cases} 1 & if \;\; b^i \geq c(Distance\ to\ Sevenoaks_i) \\ 0 & otherwise \end{cases}$$

where T_i is a dummy variable equal to 1 if $Riots\ within\ 10\ km_i \geq 6$ and 0 otherwise, b^i denotes the benefits of treatment for constituency i that is not observable to the econometrician, and $c(Distance\ to\ Sevenoaks_i)$ is the cost of treatment, which is increasing in each constituency's distance from Sevenoaks. According to this simple model of selection into treatment, for LATE effects to be distinct from ATE it is sufficient to hypothesize that the benefits of treatment are at least in part obtained through the effects of T_i on $Whig\ Share\ 1831_i$.

Vytlacil (2002) demonstrates that, assuming the instrument Z is exogenous, the monotonicity condition is comparable to the existence of a weakly separable selection (or choice) equation,

$$T = 1[v(X, Z) - U \geq 0],$$

where X are observable factors, v is an unknown function, and U is an unobserved continuous random variable. As Mogstad and Torgovitsky (2018) show, the distribution $U|X = x$ can be normalized to be uniformly distributed over $[0, 1]$ for every value of X. Under this normalization, $v(X, Z)$ can be expressed as the propensity score

$$p(X, Z) \equiv P[T = 1|X = x, Z = z].$$

Marginal Treatment Effects

The marginal treatment effect (MTE), developed by Heckman Vytlacil (1999; 2001; 2005) constitutes an important unifying concept for IV methods that maintain the weakly separable choice model delineated above. The MTE is defined as

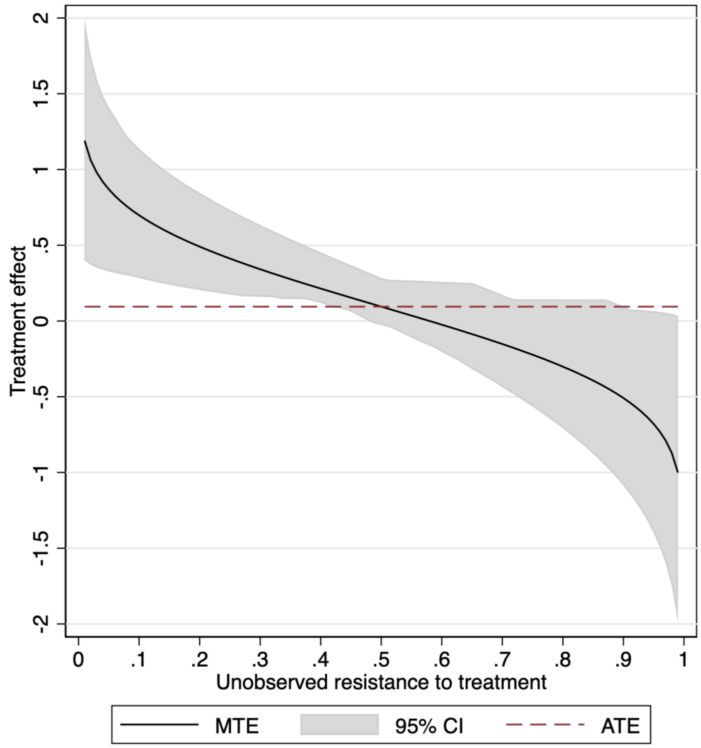
$$MTE(u,x) \equiv E[Y_1 - Y_0 | U = u, X = x].$$

In the context of AF's study, $MTE(u,x)$ is the average causal effect of experiencing six or more riots for constituencies with selection unonservable $U = u$ and observed characteristics $X = x$. The choice equation implies that, given X, constituencies with lower values of U are more likely to take treatment, regardless of their distance to Sevenoaks. Following the applied MTE literature, I assume separability between observed and unobserved heterogeneity in the treatment effects. Together with the assumption of an exogenous instrument that satisfies monotonicity, this restriction on the potential outcomes is sufficient to allow point identification of MTE over the unconditional support of the propensity score $p(X,Z)$.

I estimate the Marginal Treatment Effect (MTE) using the 489 individual seats contested in the 1831 election. Per the results presented above, showing the non-linearity in treatment effects, I group the constituencies in two groups. My binary treatment variable is thus a dummy that takes the value of 1 if a constituency experienced six riots or more, and zero otherwise. The dependent variable is also binary, taking the value of 1 if a Whig candidate was elected to seat j in constituency i, and 0 otherwise. The outcome equation uses the same specification as Model (1) in Table B2 in the main body of the paper. The choice model includes the same controls, alongside the instrument $Distance\ to\ Sevenoaks_i$, and is estimated using probit. Standard errors are computed using cluster bootstraping with 100 repetitions.

Figure C1 plots the MTE estimates by the unobserved resistance to treatment (i.e., the latent variable U) based on the local IV approach developed by Heckman and Vytlacil (1999, 2001, 2005) using a a joint normal error structure. Recall that U has been normalized to be unit uniform, so that tracing MTE over the values of u shows how the effect of rioting on reform support vary with different quantiles of the unobserved component of the willingness to take-up treatment. A simple test of whether the slope of the MTE is zero (i.e., constituencies do not select into treatment based on variability in α_1) is a test of whether the expected difference in the unobserved resistance to treatment for treated and non-treated units is zero. The results indicate that this difference is -0.471 with a standard error of 0.176, implying that the hypothesis can be rejected (p-value = 0.008).

Figure C1: Marginal Treatmement Effects



The graph illustrates that marginal treatment effects are positive among constituencies exhibiting low unobserved resistance to treatment. However, these effects diminish as the unobserved resistance to treatment increases. This finding indicates that the impact of experiencing six riots or more on reform support was more pronounced for constituencies with higher propensities to select into treatment, irrespective of their distance from Sevenoaks. In contrast, experiencing six or more riots had a negative impact on reform support for the constituencies with high unobserved resistance to treatment (although the estimated effects are statistically indistinguishable from zero). All standard treatment parameters can be expressed as distinct weighted averages by integrating the MTE over the propensity score for the relevant sample (Heckman and Vytlacil 2005; Mogstad and Torgovitsky 2018). These include the average treatment on the treated (ATT $= E(\alpha_1|T = 1)$), the average treatment effect (ATE $= E(\alpha_1)$), and the average causal effect for the subgroup of constituencies whose selection into treatment was influenced by their proximity to Sevenoaks (IV/LATE). The ATT estimates reveal that the effect of the Swing riots on the probability of electing a reform-friendly candidate is especially large for the treated: 0.358 (z-score 2.85). By comparison, the estimated LATE is 0.143 (z-score: 2.40), and the estimated ATE is 0.095 (z-score: 1.25).

Note that the MTE-based LATE was estimated using a binarised version of AF's originally multivalued treatment. It is thus possible that it may suffer from coarsening bias (Marshall 2016; Andersen and Huber 2021). The marginal effect of an additional riot on the probability of supporting a pro-reform candidate can be calculated by dividing the MTE-based LATE estimate by the difference in the average number of riots of constituencies that experienced at least five riots and those that experienced six riots or more. This per-riot effect amounts to 0.009 (0.143/15.54), which is larger that the IV estimate (0.006), but within the latter's 95% confidence interval (0.001- 0.013). Therefore, it does not seem that the LATE calculated using the MTE is overestimating the true marginal effect. As Heckman and Vytlacil (2005) note, the LATE is the integral of the MTE over a specific region of the domain of the unobserved resistance to treatment. In this case, as Figure C1 shows, the LATE of 0.143 corresponds to constituencies at the median value of the latent variable U.

D Quantifying Reformers' Gains Associated with Riots

The non-linear relationship between rioting levels and the outcome of the 1831 election complicates the estimation of the number of seats gained by reformers. Yet, it would still be possible to calculate the average value of $Whig\ Share\ 1831_i$ for each level of $Riots\ within\ 10\ km_i$ if AF's OLS estimates are consistent. Lochner and Moretti (2015) propose a test to assess the consistency of the OLS estimator for the model in equation (A1). It involves reweighting the OLS estimates of the $\alpha_1^j s$ using the estimated IV/2SLS weights. The reweighted sum of the treatment-specific OLS estimates can be then compared with the corresponding IV/2SLS estimator of α_1^{2SLS} in equation (1) in the main body of the paper.

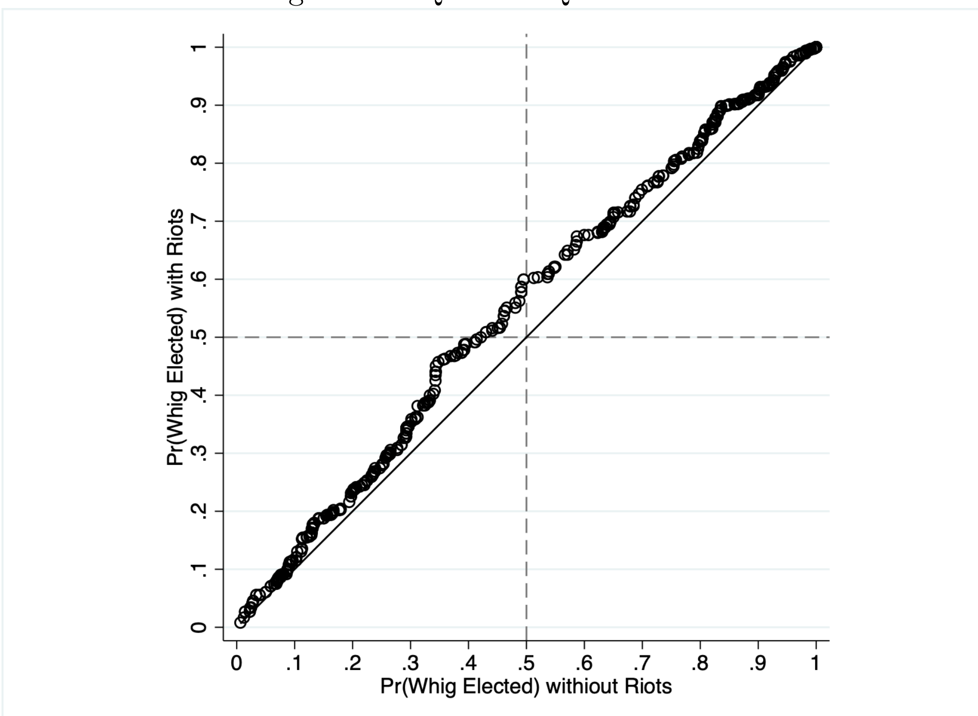
The test indicates that AF's estimates of the α_i^j are consistent, suggesting that the OLS estimator for the linear-in-riots model in equation (1) in the main body of the paper captures the *net effect* of riots within 10km of a constituency on mean electoral outcomes. According to AF's preferred specification (reported in column 5, panel A, Table II of their paper), exposure to one additional riot within a radius of 10 km from a constituency increased the share of Whigs elected in that constituency by 0.47 percentage points relative to past Whig support. This finding, as AF point out, implies that that reformers' parliamentary representation increased by approximately 5.2 percentage points – an additional 25.4 MPs–, because of the Swing riots.

The Lochner-Moretti test, however, is only valid if equation (A1) represents the true model. Therefore, mis-specification due to individual-level parameter heterogeneity would likely invalidate it. Yet, the MTE estimated in Appendix C can be used to estimate standard treatment parameters as well as their weights. An examination of the MTE curve for compliers presented in Figure C1 suggests that the impact of experiencing six or more riots on the support for pro-reform candidates becomes statistically indistinguishable from zero beyond the 43th percentile. Given the estimated LATE, it can thus be inferred that the riots contributed to approximately 30 seats obtained by the Whigs in the 1831 election.

While valid for a pure proportional system, these calculations may not square well with Britain's unreformed electoral system. Depending on the margin of victory in different districts these estimates may imply different conclusions on the effects of the riots on the Whig majority. To address this issue, I conduct the following counterfactual exercise. First, I calculate the predicted probability of a Whig victory for each of the 489 English seats contested in the 1831 election. I estimate these probabilities using AF's preferred specification (Table II, Panel B, Column 5). In the next step, the predicted probability of a Whig winning a seat is computed by setting the value of $Riots\ within\ 10\ km_i$ to 0 for all observations.

Figure D1 presents a plot comparing the quantiles of the predicted probabilities generated by the model where the values of $Riots\ within\ 10\ km_i$ are those calculated by AF with the quantiles of the estimated counterfactual probabilities (i.e. when $Riots\ within\ 10\ km_i =$ 0). If voters/patrons that experienced local riots were more likely to support pro-reform candidates, the former probabilities should be higher than the latter. As the graph shows, this is the case. The estimated slope coefficient of a regression of quantiles corresponding to the actual values of $Riots\ within\ 10\ km_i$ on the ones where they were set to zero is statistically different from one at conventional levels (0.97; z-score=143.14).

Figure D1: Quantile-Quantile Plot



Recall that MPs were chosen using the plurality rule. Therefore, the only cases where the outcomes of a given race would have changed are those where the counterfactual predicted probabilities are below 0.5, while the actual predicted probabilities are above that threshold. In all other cases, while the propensity to vote for a Whig would have been higher with, rather than without riots, this increased support for a pro-reform candidate would not have changed the election outcome. The top-left quadrant of Figure D1 shows the races where the occurrence of local riots could have altered the winning candidate's identity from anti-reform to pro-reform. There are a total of 23 observations falling into this category. Therefore, the outcomes of this counterfactual exercise closely align with the previously calculated results.

Notice that the calculations of the riots' influence on the Act's passage are based on the variable *Riots within* 10 km_i, which entail a comparison between areas more exposed to *local* riots to those that were less exposed, but potentially influenced. The untreated group is thus not entirely untreated, but simply untreated by a nearby riot. One way to address this concern is to use the number of riots within a radius of 50 km of each constituency to conduct the counterfactual exercise discussed above. With this alternative measure, 487 out of 489 seats were exposed to at least one riot, making it a reasonable proxy for overall riot exposure. The results suggest that in 46 of the contested seats, riots had the potential to shift the winning candidate's stance from anti-reform to pro-reform. By redistributing these 46 seats from the government to the opposition, the required margin of victory for the reform would increase to 92. This number is still below the Grey ministry's 96.5-vote average margin of victory in the 42 divisions related to parliamentary reform. Finally, the calculation of the 26 additional seats was based on the average impact of riots, which is susceptible to sampling error. According to AF's preferred specification (reported in column 5, panel A, Table II of their paper), the standard error associated with *Riots within* 10 km_i's coefficient estimate is 0.18. Given the average margin of victory for the Grey ministry at 96.5 votes, the likelihood that riot exposure contributed significantly to such a substantial majority is 0.01, or 1 in 100. Therefore, the Grey ministry would still have had sufficient support to pass the Reform Bill, with or without riots.

Grey Ministry's Margin of Victory

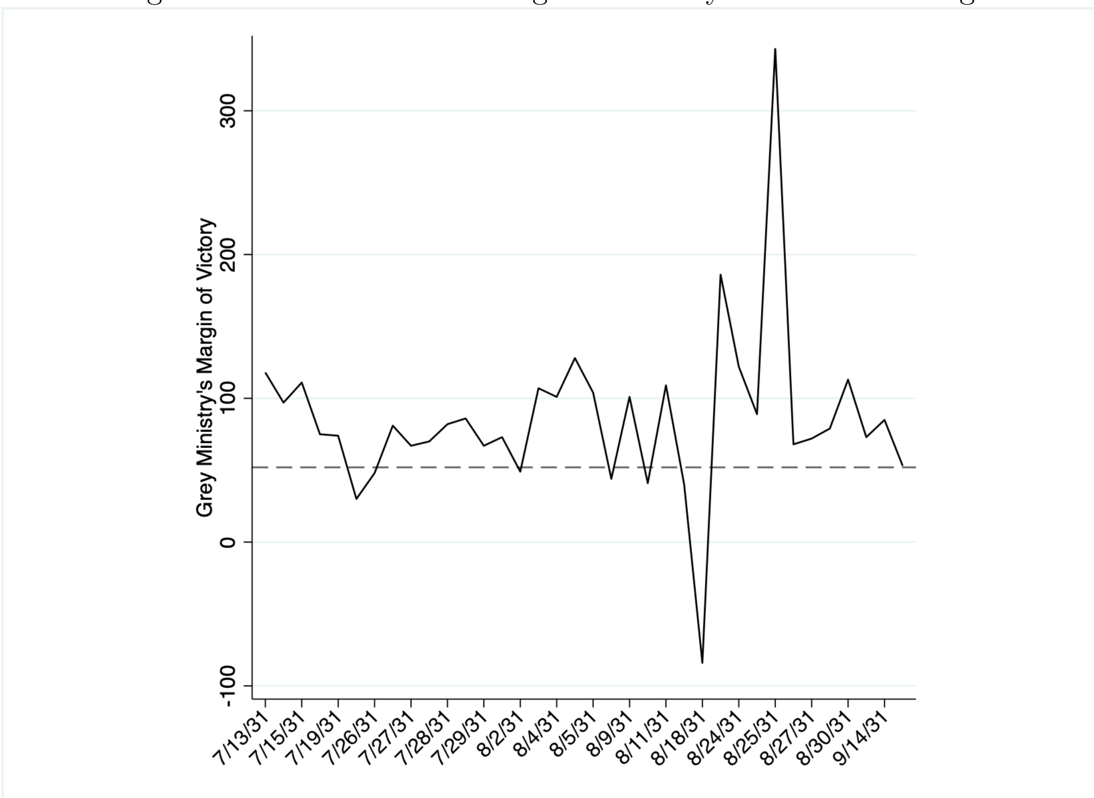
The support for reform can be also gauged by examining the 36 divisions that took place at the committee stage between its first meeting on July 12 and the last one on September 14, 1831.[2]. Figure D2 shows the Grey ministry's margin of victory in those divisions. The government prevailed in all but one of them, with an average margin of victory of 91 votes, with a minimum of 30 and a maximum of 343.[3] If one were to take the 26 seats away from the government and give them to the opposition, the margin of victory – depicted by the horizontal grey dashed line in Figure D2 – would be 52 votes. Excluding the government's single defeat, in only 6 of out of the 36 divisions reallocating the pro-reform votes could have

[2]The data are available at https://hansard.parliament.uk/Commons/

[3]The government's single committee defeat took place on 18 August 1831. An amendment to give substantial tenants-at-will county votes proposed by the Marquess of Chandos was carried by 232 to 148. According to Brock (1973), the cabinet decided that they would resist it in the Commons but accept defeat on it, hoping that its mildly democratic flavor would make it unacceptable to the Lords.

changed a vote's outcome. Five of them pertained to the placement of different boroughs into the proposed schedules reapportioning parliamentary seats. The other one was related to the the right of voting vested in non-resident freemen. Losses in these divisions, however, would have hardly jeopardized the Reform Bill's passage.[4]

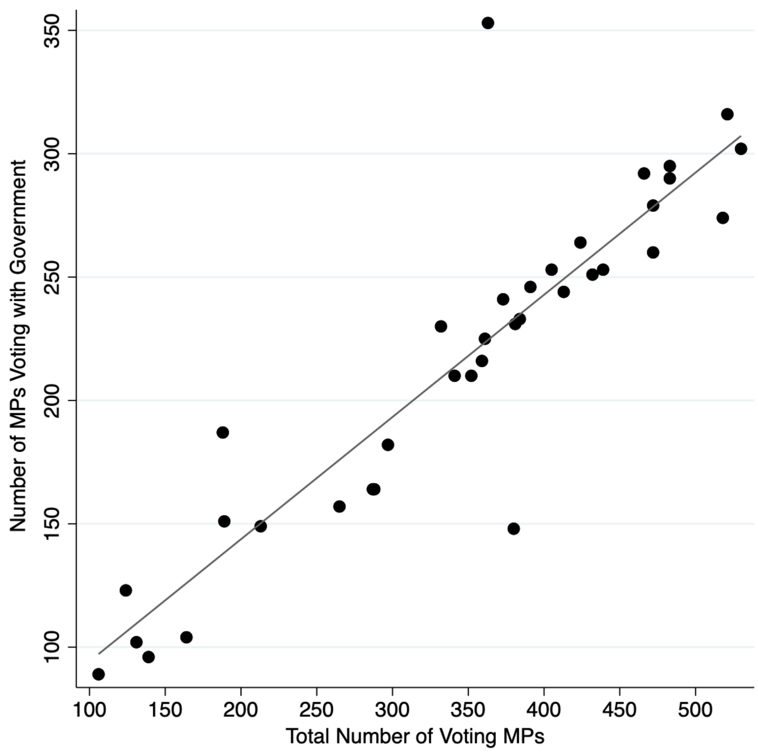
Figure D2: Government's Margin of Victory - Committee Stage



The variation in winning margins presented in Figure D2 raises questions about the Gray Ministry's ability to effectively manage its MPs. Figure D3 shows the number of MPs voting with the government as a function of the total number of voting MPs across the 36 divisions.

[4]In the words of Mr. Evelyn Denison, an MP who participated in the debate on freeholders' voting rights, " ... The [non-resident freemen] clause now before them was not of material importance, and therefore, whatever might be their decision respecting it, the Bill itself would not be materially affected. The only question was, whether town freeholders of a certain description should vote for the town where their freeholds were situated, or vote for the county ..." (cf. https://hansard.parliament.uk/Commons/1831-07-12/).

Figure D3: Share of MPs Voting with Government - Committee Stage



The results indicate that the government's voting coalition remained relatively stable regardless of the number of voting MPs. The estimated slope coefficient from the regression of the number of MPs voting with the government on the total number of voting MPs is statistically significant at conventional levels (0.49; z-score = 11.76), with an intercept of 44.68. These findings suggest that an increase in Whig MPs would likely have had a minimal impact on the Gray Ministry's majority and effectiveness.

E Riot Exposure and Reform Support

In this Appendix, I examine the relationship between local riots and the electoral support of the Whigs in the 1831 election across constituencies with varying numbers of non-resident voters. The point estimates and 95% confidence intervals associated with $Riots\ within\ 10\ km_i$ are calculated using AF's preferred least squares specification (column 5, panel A, Table II).

While precise data on the number of outvoters in each constituency do not exist, their relative presence can be inferred from the different franchise requirements. Both counties and freeman boroughs, should have been more prone to having non-resident voters. In contrast, scot and lot, as well as householder boroughs, which required voters to be residents, should have had fewer outvoters. Therefore, I use the voting qualifications in pre-reform England to create five different groups of constituencies based on their voters' degree of direct exposure to the local riots. The first group consists of the 48 constituencies where residency requirements existed, namely the scot and lot, as well as householder boroughs. The second and third groups include constituencies with less stringent residency requirements, specifically counties and freeman boroughs, respectively.

Additionally, changes in the franchise introduced by the 1832 Reform Act indicate that in some freeman boroughs, outvoters comprised a sizable proportion of the electorate. Therefore, the fourth and fifth groups account for the 29 constituencies identified by Salmon (2005) as having a significant number of non-resident voters. While the former group consists of freeman boroughs excluding these 29 constituencies, the latter includes only these 29 observations.[5] The first column in Table E1 identifies the 5 different groups of constituencies defined above, while the second one indicates the number of observations in each of these groups. Columns three and four show the average number of riots and voters, respectively, in each of these groups of constituencies.

The main results are presented in columns 5-8 of E1. The analysis shows that exposure to an additional riot within a 10 km radius of a constituency without a significant proportion of outvoters (i.e., Group 1) led to a 0.75 percentage point increase in the share of Whigs elected compared to previous Whig support. This effect is not significantly different in counties (Group 2) and in boroughs where outvoters dominated the freeman rolls (i.e., Group 5).

[5] These boroughs are Barnstaple, Beverley, Bridgnorth, Canterbury, Colchester, Coventry, Dover, Durham, Evesham, Gloucester, Grantham, Great Yarmouth, Hereford, Hertford, Lancaster, Leicester, Lichfield, Lincoln, Ludlow, Maldon, Newcastle-upon-Tyne, Rochester, St Albans, Sandwich, Southampton, Sudbury, Tewkesbury, Worcester and York. For estimation purposes, I calculate the coefficient and 95% confidence intervals associated with $Riots\ within\ 10\ km_i$ by interacting $Riots\ within\ 10\ km_i$ with *Outvoters* (a variable taking a value of 1 for these 29 constituencies, and 0 otherwise).

Notably, the 29 boroughs in Group 5 had a considerably larger number of voters in the pre-Reform era. All specifications include AF's set of covariates accounting for "selection on observables." Therefore, these findings do not appear to be driven by measurement error related to outvoters being only a small fraction of the constituency electorate or by the degree of control that patrons exercised over certain boroughs.

Table E1: Riot Exposure and Reform Support

Group	Obs.	Av. Riots	Av. Voters	Coefficient	Std. Error	95% Conf. Int.	
1	48	8.31	925.02	0.75	0.88	-1.04	2.55
2	40	9.35	6007.43	0.53	0.31	-0.10	1.16
3	90	8.44	992.34	0.57	0.28	0.02	1.12
4	61	8.00	687.80	0.79	0.47	-0.14	1.72
5	29	9.38	1632.93	0.45	0.29	-0.14	1.03

Notes: The size of the estimated electorates was obtained from https://www.historyofparliamentonline.org/volume/1820-1832/survey/appendix-v

F Instrument Validity

In this Appendix, I evaluate the validity of AF's instrument. Consider the MTE-based ATT discussed in Appendix C. According to this estimate, the effect of the Swing riots on the probability of electing a reform-friendly candidate is approximately 0.331. However, the ATT estimated using propensity score (PS) matching reported by AF suggests that local Swing riots increased the share of Whigs elected in 1831 by about 20 percentage points in the constituencies exposed to at least six riots within a 10 km radius (cf. Tables S7 and S10 in the Supplemental Material). The MTE-based estimate of the ATT is thus significantly larger than that from the PS-based method. The large discrepancy between the PS- and MTE-based ATT may be due to its underestimation by the former method and/or its overestimation by the latter one.

Zhou and Xie (2016) examine the identification assumptions and estimation performances of propensity score and marginal treatment effect methods. They show that when both the ignorability and exclusion restriction assumptions are met, both methods provide asymptotically unbiased estimates for ATE and ATT. However, if systematic baseline differences between treated and untreated units persist, even after accounting for observed covariates, PS-based methods may underestimate these parameters. Their findings also indicate that MTE-based methods might lead to a severe overestimation of ATT whenever the instrument is "weak," or the exclusion restriction is violated.

Relevance

AF use the first-stage partial F-statistic to summarize their instrument's strength. According to the results for the specification without controls (reported on Table VI, column (1)), the Kleibergen–Paap F-statistic is equal to 74.3. They also test the null hypothesis that $\hat{\alpha}_1 = 0$ in their second stage equation. The p-value associated with the Anderson–Rubin test is equal to 0.006, indicating that $\hat{\alpha}_1$ is statistically different than zero. Both tests, however, were computed using White robust standard errors. This is problematic because the constituencies are grouped within counties. If the the values of the regressor of interest do not vary much within groups, ignoring intra-class correlation can lead to under-estimated standard errors and consequent over-rejection using standard hypothesis tests (Moulton 1986; Young 2022; Lal et. al 2024).

In this case, the correlation of $\textit{Distance to Sevenoaks}_i$ within each county is 0.96, and the intra-class correlation coefficient for the residuals is 0.24. The estimated Moulton factor is roughly 3, indicating that the standard error of the coefficient associated with $Distance\ to\ Sevenoaks_i$ in the first-stage equation reported on Table VI, column (1) in AF should be multiplied by a factor of 1.74.

Table F1 compares the standard errors of the coefficient associated with the travel-time distance between each constituency and Sevenoaks in the first-stage equation using different approaches. I report six different estimates: (1) AF's White robust standard errors; (2) corrected standard errors using the Moulton formula; (3) random-effects standard errors; (4) clustered standard errors; (5) block-boostrapped standard errors; and (6) standard errors from weighted estimation at the group level.

Table F1: Standard Errors in First-Stage Regression

Variance Estimator	Standard Error
Robust	0.166
Moulton Correction	0.289
Random Effects	0.227
Clustered	0.281
Block Boostrap	0.285
Estimation using group means (weighted by group size)	0.246

Notes: The coefficient on $Distance\ to\ Sevenoaks_i$ is -1.433, with the exception of the random effects model (-1.246), and the estimation using group means (-0.943). The group level for clustering is the county. The number of observations is 244. The bootstrap estimates uses 1,000 replications. All specifications include $Distance\ to\ Sevenoaks_i$ as the single independent variable. However, the results are robust to the inclusion of AF's additional set of covariates accounting for "selection on observables".

The findings reveal that all adjustments deliver very similar results (a standard error of approximately 0.23-0.29), indicating that there are enough clusters for group-level asymptotics to work reasonably well.[6] Substantively, these larger standard errors imply a significantly lower first-stage F-statistic (26.01, rather than 74.3). The test of Olea and Pflueger (2013) is robust to heteroskedasticity, autocorrelation, and clustering. Therefore, it can be used to assess the instrument's strength in AF taking the group structure of the data into account. When the standard errors are clustered at the county level, the identification-robust Anderson-Rubin (AR) 5% confidence set is [-0.971, 3.05].[7] Therefore, these findings suggest

[6] If instead of using the sample with 244 observations aggregated at the constituency level one uses the one with the 489 individual seats, then $\hat{\pi}_i = -1.423$ with a robust standard error of 0.117, while the corrected standard error using the Moulton formula amounts to 0.302. This is another indication that making the data set larger by disaggregating the county-level one generates little to none new information.

[7] Another approach that is robust to clustering, the tF procedure proposed by Lee et al. (2021), yields

that ignoring intraclass correlation led AF to overstate the precision of the estimated effect of distance to Sevenoaks on riot intensity, and to reject the hypothesis $\hat{\alpha}_1 = 0$.

Validity

To assess the validity of their instrument, AF examine the relationship between the travel-time distance to Sevenoaks and the support for the Whigs in the 1830 election, which took place before the peak of the Swing riots. They argue that, if the exclusion restriction is satisfied, then the effect of the instrument on this pre-treatment outcome should be zero.[8] AF fail to reject the null of no effects of $Distance\ to\ Sevenoaks_i$ on the share of seats won by Whigs in the 1830 election (cf. panel B of Table V). Based on this zero-first-stage test, they conclude that their instrument satisfies the exclusion restriction.

While this test is a useful heuristic, it can never verify an instrument's validity. Nonetheless, we can still evaluate how the main effect of interest changes under moderate violations of the exclusion restriction. Consider equation (1) above, linking Swing riots to the share of seats won by reform-friendly candidates in the 1831 election, but with the addition of the term $\gamma \times Distance\ to\ Sevenoaks_i$. The exclusion restriction in AF amounts to the assumption $\gamma = 0$. Conley, Hansen and Rossi (2012) suggest that this assumption can be relaxed, and replaced with a plausible value, range or distribution of γ. They discuss different inference approaches for this parameter. The "local to zero" approximation, is obtained when the prior on γ follows a Normal distribution with mean μ_γ and variance Ω_γ, and the uncertainty about γ reduces with the sample size. This method provides a way to estimate the main parameter of interest when γ is different from zero, but it does not give any particular guidance on the plausible value, range or distribution of γ. In a sample where the instrument is uncorrelated with the treatment variable, the instrument's reduced-form coefficient is an estimator for γ. Therefore, as Kippersluis and Rietveld (2018) note, the estimator $\hat{\gamma}$ obtained from a zero-first-stage regression is a plausible estimate of the direct effect of the instrument on the outcome of interest. In this case, one can use the estimated coefficient associated with $Distance\ to\ Sevenoaks_i$ in the reduced-from regression where the dependent variable is the outcome of the 1830 election ($\hat{\gamma} = -0.84$) as the prior μ_γ, to observe how the effect of

very similar results. In this case, with county-level clustered standard errors, the First-Stage F-statistic is 26.76, and the AR 5% confidence set is [-.542, 3.174].

[8] Unlike a traditional "zero-first-stage" test, where the instrument is not expected to influence treatment assignment in a sub-sample of the data, the identification assumption here is that very few Swing riots took place before August of 1830, so the posited mechanism (linking geography to agricultural disturbances) has no traction in the case of the 1830 election.

interest α_1 changes upon a plausible violation of the exclusion restriction.

I estimate equation (1) with the addition of $Distance\ to\ Sevenoaks_i$ using the "plausibly exogenous" method (Conley, Hansen and Rossi 2012). To highlight the problem posed by an invalid, rather than a weak, instrument I first ignore the issue of how the intraclass correlation affects the precision of the estimated effect of distance to Sevenoaks on local rioting; namely, I rely on robust, rather than clustered, standard errors for inference.[9] The results indicate that when $\mu_\gamma = -.84$ and variance $\Omega_\gamma = 0$, the parameter $\hat{\alpha}_1^{2SLS}$ is 0.73, compared to $\hat{\alpha}_1^{2SLS} = 1.32$ under a perfect ($\gamma \equiv 0$) instrument. They also reveal that, once the estimated direct effect of $Distance\ to\ Sevenoaks_i$ is taken into account, the instrumented effect of Swing riots on the share of seats won by reform-friendly candidates in the 1831 election is much less precise.

With $\mu_\gamma = -.84$, the corresponding 95% confidence interval for α_1 is approximately [-.17, 1.63], compared to [.41, 2.22] under a perfect ($\gamma \equiv 0$) instrument. This latter finding suggests that, under a prior based on the estimator $\hat{\gamma}$ obtained from the zero-first-stage regression, the data are uninformative about the impact of riots on pro-reform electoral support.[10] As expected, the sensitivity of the 2SLS estimator to the violation of the exclusion restriction is greater when the effect of intraclass correlation on the instrument's strength is taken into consideration (Lal et. al 2024). With standard errors clustered at the county level, the 95% confidence interval for α_1 under $\gamma \equiv 0$ is s [-.19, 2.82]. In this case, even a very small violation of the exclusion restriction, for example $\gamma = -0.1$, significantly attenuates AF's 2SLS estimates.

[9] I also restrict my attention to a specification that does not include any additional controls.

[10] Following Kippersluis and Rietveld (2018), one could also incorporate uncertainty around $\hat{\gamma}$ assuming that the normalized difference in direct effects between the zero-first-stage group and the other group does not exceed one-quarter in 95% of the cases. Accounting for uncertainty about the direct effect of $Distance\ to\ Sevenoaks_i$, however, would move the p-value of $\hat{\alpha}_1^{2SLS}$ even further away from statistical significance.

G Further Threats to Inference

Table G1 examines the observable differences between constituencies in the cereal/non-cereal producing areas.

Table G1: Differences between constituencies in cereal/non-cereal region

	Cereal		Non-Cereal		Difference	
	N	Mean	N	Mean	Difference	p-value
	Constituency-Level Variation					
Distance to Sevenoaks (Days)	117	4.45	127	10.69	-6.23	0.0000
Riots within 10 km	117	15.81	127	3.57	12.24	0.0000
Reform support 1830	117	0.03	127	-0.08	0.12	0.0988
	County-Level Variation					
Threshing machines (1800–1829)	116	32.11	127	17.23	14.87	0.0000
Poor Law Expenses per Capita (1828-1830)	116	0.75	127	0.40	0.37	0.0000
Labor riots (1793–1822)	117	2.94	127	0.15	2.79	0.0000
Distance to Garrison (Logged)	114	10.46	126	10.83	-0.37	0.0000
Police within 10 km	114	0.03	126	0.02	0.01	0.0003

Note: The constituency-level data come from AF. County-level data on threshing machines, distance to . garrison, and police within 10km were obtained from Caprettini and Voth (2020). The data on labor riots at the county level come from AF. I collected the data on county-level law expenses from the House of Commons Parliamentary Papers Online (https://parlipapers.proquest.com/parlipapers)

References

Andersen, Martin E., and Martin Huber. 2021. "Instrument-based estimation with binarised treatments," *The Econometrics Journal*, 24: 536-558.

Bisin, Alberto, and Andrea Moro. 2021. "LATE for History," in Alberto Bisin and Giovanni Federico (Eds.). *Handbook of historical economics.* London: Elsevier.

Brock, Michael. 1973. *The Great Reform Act.* London: Hutchinson & Co.

Caprettini, Bruno, and Hans-Joachim Voth. 2020. "Rage against the machines," *American Economic Review: Insights*, 2: 305-20.

Conley, Timothy G., Christian B. Hansen, and Peter E. Rossi. 2012. "Plausibly exogenous," *Review of Economics and Statistics*, 94: 260-272.

Heckman, James J., Sergio Urzua, and Edward Vytlacil. 2006. "Understanding instrumental variables in models with essential heterogeneity," *Review of Economics and Statistics*, 88: 389-432.

Heckman, James J., and Edward J. Vytlacil. 1999. "Local instrumental variables and latent variable models for identifying and bounding treatment effects," *Proceedings of the national Academy of Sciences*, 96: 4730-4734.

Heckman, James J., and Edward Vytlacil. 2001. "Policy-relevant treatment effects," *American Economic Review*, 91: 107-111.

Heckman, James J., and Edward Vytlacil. 2005. "Structural equations, treatment effects, and econometric policy evaluation," *Econometrica*, 73: 669-738.

Kippersluis, Hans van, and Cornelius A. Rietveld. 2018. "Beyond plausibly exogenous," *The Econometrics Journal*, 21: 316-331.

Lal, Apoorva, Mac Lockhart, Yiqing Xu, and Ziwen Zu. 2024. "How Much Should We Trust Instrumental Variable Estimates in Political Science?" *Political Analysis*, available at https://doi.org/10.1017/pan.2024.2.

Lee, David S., Justin McCrary, Marcelo J. Moreira, and Jack R. Porter. 2022. "Valid t-ratio Inference for IV," *American Economic Review*, 112: 3260-3290.

Lochner, Lance, and Enrico Moretti. 2015. "Estimating and testing models with many treatment levels and limited instruments," *Review of Economics and Statistics*, 97: 387-397.

Løken, Katrine V., Magne Mogstad, and Matthew Wiswall. 2012. "What linear estimators miss: The effects of family income on child outcomes," *American Economic Journal: Applied Economics*, 4: 1-35.

Marshall, John. 2016. "Coarsening bias: How coarse treatment measurement upwardly biases instrumental variable estimates," *Political Analysis*, 24: 157-171.

Mogstad, Magne, and Alexander Torgovitsky. 2018. "Identification and extrapolation of causal effects with instrumental variables," *Annual Review of Economics*, 10: 577-613.

Mogstad, Magne, and Matthew Wiswall. 2010. "Linearity in Instrumental Variables Estimation: Problems and Solutions," *IZA Discussion Paper*, No. 5216. Available at SSRN: https://ssrn.com/abstract=1686527 or http://dx.doi.org/10.2139/ssrn.1686527

Moulton, Brent R. 1986. "Random group effects and the precision of regression estimates," *Journal of econometrics*, 32: 385-397.

Olea, José Luis Montiel, and Carolin Pflueger. 2013. "A robust test for weak instruments," *Journal of Business & Economic Statistics*, 31: 358-369.

Salmon, Philip. 2005. "Reform Should Begin at Home: English Municipal and Parliamentary Reform, 1818–32," *Parliamentary History*, 24: 93-113.

Young, Alwyn. 2022. "Consistency without Inference: Instrumental Variables in Practical Application," *European Economic Review*, doi: https://doi.org/10.1016/j.euroecorev.2022.104112

Zhou, Xiang, and Yu Xie. 2016. "Propensity-Score-Based Methods versus MTE-Based Methods in Causal Inference: Identification, Estimation, and Application," *Sociological Methods & Research*, 45:3-40.